A COMMERCIAL OUTCOME PREDICTION SYSTEM FOR UNIVERSITY TECHNOLOGY TRANSFER USING NEURAL NETWORKS

A thesis submitted for the degree of Doctor of Philosophy

by

Ling Chu

School of Design & Systems Engineering, Brunel University

October 2006

ABSTRACT

This thesis presents a commercial outcome prediction system (CPS) capable of predicting the likely future monetary return that would be generated by an invention. The CPS is designed to be used by university technology transfer offices for invention assessment purposes, and is based on the data from their historical invention cases. It is aimed at improving technology transfer offices' invention assessment performance.

Using qualitative critical factors suggested by literature, a prototype CPS based on decision tree induction was developed. The prediction performance achieved by the prototype CPS was unreliable. Three surveys with various technology transfer offices were then performed, and the findings were incorporated into a final version of the CPS, which was based on neural networks.

Subject to information obtained in the surveys, a number of potentially predictive attributes were proposed to form part of the predictor variables for the CPS. The CPS starts with a number of data reduction operations (based on principal component analysis and decision tree techniques), which identify the critical predictor variables. The CPS then uses five neural-network training algorithms to generate candidate classifiers, upon which the final classification is based.

The prediction results achieved by the CPS were good and reliable. Additionally, the data reduction operations successfully captured the most discriminative invention attributes. The research demonstrated the potential of using the CPS for invention assessment. However, it requires sufficient historical data from the technology transfer office using it to provide accurate assessments.

TABLE OF CONTENTS

NOMENCLATURE

SYMBOLS

b_{neuron}	bias value of a neuron
$\mathbf{c}$	class vector
c	class label
cc	correlation coefficient
$\mathbf{CC}$	correlation coefficient matrix
cov	covariance
D	data set
$\mathbf{e}$	original-attribute coefficient vector
$\mathbf{E}$	original-attribute coefficients matrix
$e(D)$	entropy value required to classify data set D
$\mathbf{f}$	transfer function
g	gradient
Δg	gradient change
$\mathbf{H}$	Hessian matrix
$\mathbf{I}$	identity matrix
$i(N)$	the impurity of node N
$\mathbf{J}$	Jacobian matrix
l	learning rate
L	attribute list
$\mathbf{P}$	randomised matrix
$\hat{P}(c)$	the fraction of cases that belong to class c

$\mathbf{r}$	network error vector
r_order	random permutation vector
R	squared error
$\mathbf{v}$	a vector containing the variance represented by principle components
$\mathbf{w}$	weight vector
$\Delta\mathbf{w}$	weight change
x	value of an attribute
X	attribute
$\mathbf{X}$	attribute matrix
$\mathbf{X}^{\mathrm{T}}$	transposed attribute matrix
$\mathbf{X}_{pca}$	principal component matrix of $\mathbf{X}$
$\mathbf{y}$	output vector
μ	arithmetic mean
σ	standard deviation

LIST OF EQUATIONS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

The author would like to express her indebted thanks to the following people and organisations, without which this research would not have been possible:

- Dr. J. Shackleton and Dr. Q.P. Yang, for their encouragement, supervision and advice throughout the duration of this research.
- Brunel University for sponsoring this research and providing the necessary resources.
- The anonymous university technology transfer office for providing the data.
- Dr. S. Jones, for his valuable comments regarding the final draft of this research.
- My family who have patiently awaited the completion of this work, providing encouragement throughout.

CHAPTER ONE

INTRODUCTION

University Technology Transfer is the process of transferring university-developed inventions to the private sector. The most common transfer mechanisms are the licensing of inventions to companies, and the establishment of spin-out companies based on inventions. Famous world-changing inventions that brought to the market through licensing include the Seat Belt and the LCD (Liquid Crystal Display). Successful university spin-outs include the internet search engine Lycos and the biotechnology firm Genentech. Due to the significant contributions brought by University Technology Transfer, large amount of funding and resources have been committed to this area by both universities and Governments. As more universities demonstrate how lucrative a source of income can be from merely one successful invention commercialisation, universities all over the world are increasingly focusing on University Technology Transfer. However, only a small percentage of universities manage to gain, and the majority are struggling to break even.

The process of University Technology Transfer is risky. Each invention usually involves a significant upfront investment, and the return can only be realised in the medium or long term. Generally speaking, the process of University Technology Transfer can be interpreted as a series of three steps: 1) invention disclosure, 2) invention assessment, and 3) invention commercialisation actions. Invention disclosure is the process of disclosing both verbally, and in written documents to a university the confidential technical details of an invention. Invention assessment refers to the procedure of evaluating and grading an invention's potential to succeed, prior to investment by a university technology transfer office. Invention commercialisation actions encompass a range of activities required to commercialise an invention, after a university technology transfer office decides to invest in the invention. These actions can include the marketing of the invention, negotiation with potential licensees and assisting the formation of a spinout based on the invention. Among the great number of empirical studies regarding University Technology Transfer, little has been done on the second step of the process: invention assessment. This represents a serious gap in the knowledge required for both effective invention disclosure and invention commercialisation. With insufficient knowledge of invention assessment, critical invention information may be missed during the invention disclosure step. Moreover, an inaccurate assessment can give rise to the wrong invention commercialisation actions, including mistaken investment and the loss of opportunities and resources.

Therefore, this research has endeavoured to investigate the area of invention assessment for the purpose of University Technology Transfer. Several issues require clarification before the aims and scope of this research can be stated. These issues are discussed below.

The current research concentrates on patentable inventions, as they form the majority of the projects handled by technology transfer offices in universities. Additionally, it should be noted that the term invention assessment can be interpreted as either the project selection process, or the performance evaluation process. Though the two processes are closely related, they require different approaches for research investigation. One of the obvious differences would be the perspective emphasis of the former, and the retrospective emphasis of the latter. Under the increasing demand for public accountability and performance indicators, it is especially important to prevent the potential misunderstanding by pointing out that the current research defines invention assessment as the project selection process.

Each country has its own unique development of University Technology Transfer. For instance, the United States, being the world leader in University Technology Transfer, has a relatively long history of practising University Technology Transfer. In the United Kingdom, however, University Technology Transfer has only been wide spread in the late 1990s. The University Technology Transfer industry in the United Kingdom is therefore still catching up in some areas. By limiting the scope of the current research to only universities in the United Kingdom, the present research focuses on their particular characteristics, and hence formulates a more suitable solution.

University Technology Transfer is a fairly controversial area. There is a school of thought that objects to the concept of University Technology Transfer for various reasons. For instance, some people are suspicious about the ability of universities, as public institutions, to conduct commercial activities. In terms of ethical issues, it is suggested that University Technology Transfer activities are in effect limiting the societal benefits, by privatising the transfer of knowledge. While acknowledging these opinions, this research takes the stance that research regarding University Technology Transfer is worthwhile, since it can contribute to over 80 percent of the universities in the United Kingdom, which have established technology transfer offices, promoting University Technology Transfer.

Depending on the chosen performance indicator, there are a variety of aspects on which an invention can possibly be assessed. An invention can score high in terms of its contribution to the society, but score low in regards of monetary return generated for the university. This research focuses on the assessment of an invention's potential to generate monetary return. This is due to the fact that monetary return is clearly one of the most widely used performance indicators, as well as being one of the most dominant objectives for universities conducting technology transfer. This is evidenced by the observation that almost all the relevant University Technology Transfer literature or surveys report the level of monetary return as an indicator of success. Wealth creation is also one of the most frequently stated objectives by university technology transfer offices.

While limiting the focus of the current research to monetary return, this is not suggesting that monetary return is the only or most important consideration for invention selection.

The size of a technology transfer office can vary from tens of employees down to a one man unit. In larger technology transfer offices, assessment decisions may be jointly considered by a number of managers, whereas the decisions made in smaller offices are likely to involve fewer managers. In order to conduct a piece of research that benefits technology transfer offices of all sizes, this research endeavours to concentrate on methods that minimise the input of human resources. In this regard, the research uses a quantitative approach. While it is acknowledged that managers might stop short of relying on computer generated solutions based on higher mathematics, this is not an issue when the solution forms merely one of many screening methods used to better assess the decision. This is supported by wide documentation on the contribution and application of quantitative tools towards imperative investment decisions in the financial sector.

Finally, quantitative applications come in many different forms, such as information visualisation, information organisation, prediction, and so on. The current research attends to the implementation of prediction. This is because the result of such an application could directly solve the problem of which invention is more likely to give the better monetary return. Because of this, the performance of such application is also measurable.

Subject to the issues discussed above, the precise aim of the present research is as follows: to develop a commercial outcome prediction system, which can be used by individual technology transfer offices in the United Kingdom for the purpose of estimating an individual invention's likely future monetary return for invention assessment purposes. For convenience, such a system will from now on be called the CPS (the Commercial Outcome Prediction System).

To achieve the above aim, the remainder of this thesis is organised as follows. Chapter Two reviewed relevant literature from the University Technology

Transfer sector, as well as other sectors relevant to the development of the CPS. Based on the findings from the literature review, Chapter Three developed a prototype version of the CPS, and the results it generated demonstrated the need for further investigation. Chapter Four conducted several surveys regarding the invention assessment methods adopted by various university technology transfer offices. Combining the insights derived from Chapter Two, Three and Four, Chapter Five developed the final version of the CPS. The results generated using the CPS are then presented and evaluated in Chapter Six. Lastly, important implications derived from the current research are discussed, and conclusions are drawn in Chapter Seven.

CHAPTER TWO

LITERATURE SURVEY

2.1 INTRODUCTION

This chapter provides the necessary background materials and reviews the relevant literature in order to achieve the research aim of developing the CPS (Commercial Outcome Prediction System) for university technology transfer offices in the United Kingdom. First of all in Section 2.2, a number of background materials are presented to locate the task of invention assessment for University Technology Transfer sector in space, time and culture. It then moves on to review literature from the University Technology Transfer domain in relevance to the invention assessment process. Based on the findings generated from Section 2.2, additional inputs of assessment methods from other relevant sectors are found necessary. Section 2.3 therefore examines literature covering evaluative bibliometrics from the domain of Science and Research Assessment, which are also applicable to the invention assessment task to a certain extent. Furthermore, Section 2.4 reviews literature regarding the use of quantitative classification for option selection/outcome prediction for investment purposes in the financial sector. Based on these discussions, Section 2.5 concludes this chapter by pointing out the applicability of evaluative bibliometrics and quantitative classification to the invention assessment problem of the University Technology Transfer domain as well as its relevance towards the development of the CPS.

2.2 UNIVERSITY TECHNOLOGY TRANSFER

This section starts by clarifying necessary definitions of the terms and processes involved in University Technology Transfer, followed by an account of the rise of the sector with a particular focus on the situation in the United Kingdom. It then reviews the difficulties involved in invention assessment as well as relevant studies contributing to the development of assessment solutions. This section ends with the establishment of three assumptions derived from these studies.

Definitions

An invention is defined in this research as a patentable novel entity with commercial potential. University Technology Transfer is defined in this thesis as the process of transferring university-developed inventions to the private sector. Despite the definition adopted by the current research, there are other variations of definitions proposed. Gopalakrishnan and Santoro (2004) provide a thorough review regarding these definitions. In particular, they attempted to clarify the differences between knowledge-transfer and technology-transfer. Licensing and spinout formations are the most common mechanisms to transfer university-developed inventions to the private sectors. Licensing refers to the process of letting the inventions to the private sectors, either exclusively or non-exclusively, in return for royalty incomes. Spinout formation refers to the creation of a spinout company selling products or services based on the inventions. A general model of University Technology Transfer reflecting the conventional wisdom of the process among University Technology Transfer practitioners (AUTM 2000) is presented in Figure 2-1.

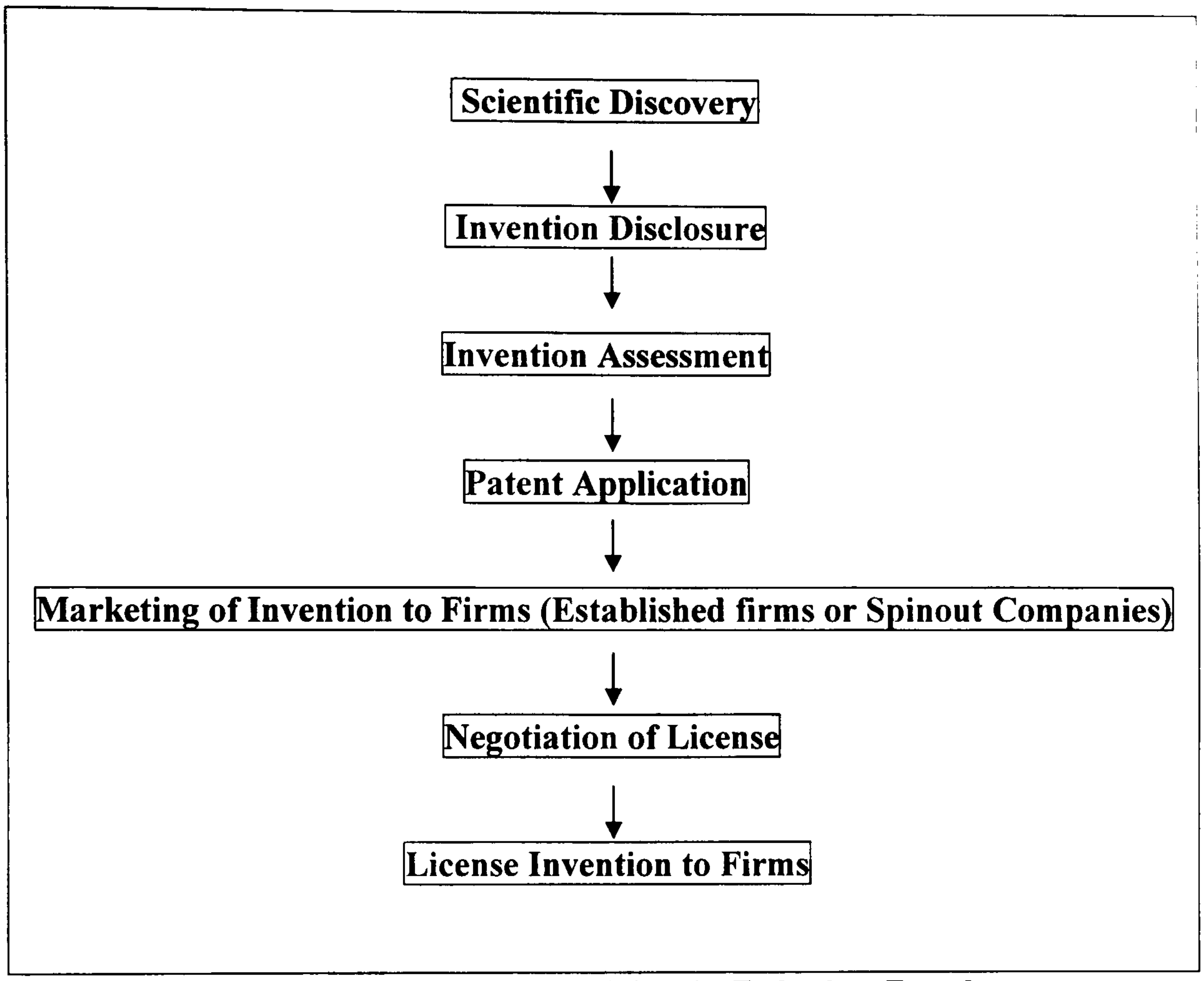


Figure 2-1: General Model of University Technology Transfer

The transfer process starts with a scientific discovery by an academic inventor. The detail of the invention derived from the discovery is then reported to the university, often in an invention disclosure form. The university then assesses the invention and decide whether resources will be allocated to this invention. If the invention assessment result is positive, a patent application is then filed for the invention and marketing of the invention to potential licensees or preparation for spinout formation starts. This is often followed by lengthy negotiation process regarding details of contracts. Finally, contracts are signed and the invention is licensed either to established firms or to a spinout. Though the University Technology Transfer process in reality deviates to a certain extent from this model, it captures the essence of the process. To assist the management of these steps, universities nowadays are increasingly establishing their own technology-transfer offices. In particular, during invention assessment (the second step of the model), an invention's potential to generate future income for the university will be evaluated, given the prerequisites that the invention is beneficial to humankind and patentable. Hence, invention assessment is defined in this thesis as the procedure to evaluate and grade an invention's potential to generate monetary return for a university.

The rise of University Technology Transfer

University Technology Transfer is not merely a recent phenomenon. Its history can be dated back to the history of modern universities since the early 19th century, when universities emerged as a combination of teaching and research in Germany. While it is still not clear, some believed that the Germans were the pioneer in University Technology Transfer and the first documented spinout company was from a German university (Gustin 1975). Nevertheless, University Technology Transfer was not widely popularised until the passing of several regulations including the 1980 Bayh-Dole Act, the 1980 Stevenson-Wydler Act, and the 1985 Federal Technology Transfer Act in the United States (Markman et al. 2005). Particularly, the Bayh-Dole Act simplified the University Technology Transfer process by granting universities the ownership of the patents arising from federal research grants. Since the passing of this law, the number of university technology transfer offices and the number of patents granted to universities in the United States have increased rapidly, so did the level of licensing income and the number of spinout companies established (AUTM 1997). Similar story is happening increasingly across Europe and beyond (Wright et al. 2004).

In the United Kingdom, University Technology Transfer became a wide spread phenomenon in the late1990s when there was a sudden increase in the numbers of university technology transfer office and spinout company established. This trend is further accelerated by the introduction of the Third Stream Funding introduced by the Government to support University Technology Transfer activities in England in 1999 (Crown 2003). The Third Stream Funding covers initiatives including the Higher Education Innovation Fund, Science Enterprise

Challenge and University Challenge Funds (DTI et al. 2002). Currently, over 80% of universities in the United Kingdom have established their own technology transfer offices and staff numbers are still increasing by about 25% per year, licensing rates and income has also increased (UNICO et al. 2003). Nonetheless, many technology transfer offices and spinouts in the United Kingdom were only recently established as a result of Government funding (Crown 2003). The quality of university technology transfer offices in the United Kingdom is highly skewed and most are in the red (UNICO et al. 2003). Deficit university technology transfer offices are often inexperienced and are believed to have filed for patents indiscriminately, depending on Government funding to create spinouts while failing to attract any license deals or external funding to support the spinouts. Experienced technology transfer offices, on the other hand, focus on licensing rather than spinouts (Markman et al. 2005). In 2002, the United Kingdom created three times more spinouts than the United States, its licence income, however, is merely a third of that of the United States and is predominantly achieved by the top five universities (UNICO et al. 2003). Very few spinouts in the United Kingdom have been sold or floated on a stock market (Mike et al. 2004). The difficulties encountered by the University Technology Transfer field do not merely exist in the United Kingdom but throughout the world.

Invention Assessment for University Technology Transfer

Invention assessment is very difficult largely due to the complex nature of University Technology Transfer, which involves early-stage inventions from a wide range of disciplines, upfront investment, multiple stakeholders and long developmental time. Unlike inventions developed in the private sectors, university-developed inventions are often embryonic technologies with a high level of technical and market uncertainty (Jensen and Thursby 2001; Owen-Smith and Powell 2003). As illustrated by the comment from a senior technology transfer officer (Owen-Smith and Powell 2003): "In most cases you don't even have a prototype, let alone an established market." Furthermore, inventor cooperation is normally necessary, which leads to the issues of moral hazard (Jensen and Thursby 2001) and contracting for tacit knowledge (Arora 1996). Being academic experts of particular fields, inventors tend to be optimistic about the resources required to develop the invention and relatively naïve about efforts required in marketing (McAdam et al. 2005). A study investigating the cause of failure in University Technology Transfer projects reported that half of the failure is due to a technical failure, while 18% are due to inventors failing to cooperate or deliver know-how during further development (Thursby and Thursby 2003). Prior to the location of licensees or establishment of spinouts, upfront investment is required for the intellectual property protection for these early-stage inventions. The application process for a patent is, however, lengthy and expensive, especially when the number of designated countries increases and/or infringement issues arise. In addition, the involvement of multiple stakeholders, including inventors, technology transfer office personnel, private partners and investors is problematic. A considerable amount of evidence showed that these parties possess contrasting values, standards and organisational cultures (Roberts 1988; Nelson 2001; Friedman and Silberman 2003; Crown 2003). Siegel et al. (2004) explained that their different cultures may constitute distinct "thought worlds" with distinct languages and organizational routines that impede technology transfer. For instance, Siegel et al. (2004) asked inventors, technology transfer office personnel and companies what the output of University Technology Transfer is. Companies largely answered *the number of licences* (75%) and *informal transfer of know-how* (70%); technology transfer office personnel mostly replied *the number of licences* (86.7%) and *licensing income* (66.7%). Inventors, on the other hand, considered a wide range of entities evenly, including *product development* (35%), *the number of licences* (25%) and *the number of patents* (20%). Lastly, university-developed invention projects often require five to ten years of further development before any return can be realised (AURIL 2003b). During this long period of lag time, multiple stakeholders come and go while each contributes to the final outcome (success or failure) of the collaboration project. User values and market properties are also likely to have changed after such a long period of time. Not surprisingly, a significant number of studies continues to point out the large proportion of failure cases of University Technology Transfer (Carlsson and Fridh 2002; Crown 2003; UNICO et al. 2003), and universities are commonly urged by Governments and researchers to improve their assessment methods before rushing to spend resources on patent protection and creating spinouts (Bozeman 2000; Shane 2002; Crown 2003; Siegel et al. 2003c; McAdam et al. 2005; Markman et al. 2005).

Despite the typical emphases on the importance of invention assessment, this process is largely omitted or avoided both in practice and among the research community. Evidence from interviews with practitioners and archival materials from technology transfer offices often shows that where blockbusters have been realised, few received appropriate assessment at their early stage (Owen-Smith and Powell 2003). Evidence from case studies showed that spinouts were formed with technology assessment decisions still outstanding (McAdam et al. 2005). Some researchers also avoided the topic of invention assessment by concentrating on the university technology transfer office's role as a facilitator and neglecting the role of critical assessor (Colyvas et al. 2002; Friedman and Silberman 2003; Thursby and Thursby 2003). Even research studies regarding the definition of the whole University Technology Transfer process do not cover this area. Based on the general model presented in Figure 2-1 several researchers carried out surveys attempting to map the process in greater details, including the works of Siegel et al. (2004) and McAdam et al. (2005). These studies added considerable details to every step of the general model but except the step of *Invention Assessment*. Such avoidance coexists with abundant complaints by technology transfer practitioners regarding the difficulty in invention assessment and the literature's usual recommendation to practise robust invention assessment (Bozeman 2000; Shane 2002; Crown 2003; Siegel et al. 2003c; McAdam et al. 2005; Markman et al. 2005).

While the lack of invention assessment and its importance have been widely recorded, its understanding remains highly unclear and studies regarding invention assessment methods or processes are rarity. Invention assessment is generally understood as the assessment of an invention's technical and market potential (AURIL 2003a), which is presented as follows. Technical potential basically refers to the novelty and advantage of the invention against patents or products existing in the field. Assessment for technical potential is generally done by conducting prior art searches (using internet search engines and/or search facilities on national patent office websites) and looking at competitors' activity. Market potential is generally assessed using a range of factors, including market size, inventor experience and contribution, market impact, time to market and technology transfer office experience in the field. The assessment process for these factors is intuitive, where an assessor (a technology transfer officer) would give a score (such as on a scale of 1 to 10) to each factor after talking to inventors and/or looking at information contained in invention disclosure forms. Such intuitive invention assessment process is reflected in the sample invention evaluation form developed by AURIL (Association for University Research and Industry Links), as shown in Appendix A1.1. Assessors often possess basic understanding regarding intellectual property issues and a doctorate degree in a scientific discipline. Depending on the resources of the technology transfer office, there is often one assessor who is responsible for the assessment for inventions coming from all disciplines with the university.

A study conducted by Meseri and Maital (2001) is one of the rare studies regarding invention assessment. They compared the criteria used for invention assessment by six Israeli universities against those employed by venture capitalists and MIT (Massachusetts Institute of Technology[1]) and found that they are very similar. Their research methods included asking technology transfer offices personnel to score (on a scale of 1 to 5) a list of predetermined success factors regarding their importance during invention assessment. The factors with highest scores are then compared with the selection criteria derived from the Israel Venture Association. The six factors scored highest were "1) Market need; 2) Market size; 3) Existence of patent; 4) Success chances for R&D stage; 5) Level of innovativeness; 6) Degree of maturity of the idea". The findings from this study, however, are not compatible with those from other studies. According

[1] One of the most successful universities in terms of University Technology Transfer in the United States.

to Owen-Smith and Powell (2003), the market is rarely identified at the time of invention assessment. Factors like 'Market need' and 'Market size' are therefore not applicable. Also, invention assessment is viewed by many as a step prior to patent application (McAdam et al. 2005). Nonetheless, such disputes among the studies do not present invalidity of their findings. Rather, they provide further evidence that varying policies exist in different university technology transfer offices, as each study is based on different samples of university technology transfer offices. Furthermore, the use of the research method of asking technology transfer personnel to score a list of factors showed that Meseri and Maital (2001) agree with AURIL (2003a) that practitioners generally adopt an intuitive approach. Due to the intuitive nature of assessment process, the focus of Meseri and Maital (2001) is to find out which assessment criteria/factors are relatively more commonly adopted. In addition, the fact that Meseri and Maital (2001) compared criteria adopted by university technology transfer offices and those used by venture capitalists demonstrated the perceived similarity by the researchers regarding the project assessment process between the two parties. Literature regarding the project assessment process of venture capitalists (Zacharakis and Meyer 2000; Zacharakis and Shepherd 2005) also reported that venture capitalists' assessment process are highly intuitive and the criteria found are also similar to those located by Meseri and Maital (2001).

Informal venture capitalists, so-called business angels, represent another important area related to University Technology Transfer. Business Angels are largely researched among the Small Business and Entrepreneurial literature, due to their key role in promoting emergent businesses (Mason and Harrison 1999). In particular, their focus on early-stage ventures based on high technologies (Freear et al. 2002) means that they are potentially an important source of finance for commercialising university-developed inventions. This in turn makes the invention assessment criteria adopted by the business angels relevant to the topic of university invention assessment. An understanding regarding the angels' investment behaviour and criteria is therefore imperative.

Though the topic of business angels has been well established in the United States, it basically started in the United Kingdom in the early 1990s (Harrison and Mason 1996). Since then, the supply of risk capital from business angels has been increasing significantly (TSBS 2001). Being private investors, angels are different from formal venture capitalists (VCs) in a number of ways. First of all, VCs are more visible than angels, and VCs often invest bigger amount per investment (Freear et al. 2002). While both VCs and angels aim for capital gain, angels also emphasise on enjoyment and fun derived from the investment process. Since VCs report to their shareholders, they regard involvement as a cost to their investment (Southon and West 2006). Business angels, on the other hand, report to themselves, and regard involvement as a way of increasing their control over their investment (Van Osnabrugge and Robinson 2000; Mason and Stark 2002). Business angels, hence, are more concern about agency risk (i.e. risk caused by conflicts between investors and entrepreneurs), whilst VCs are more concern about market risk (i.e. risk caused by unforeseen market condition (Fiet 1995). Business angels invest both money and time to new ventures, and add values in various dimensions including technical, managerial and networking issues (Aernoudt 1999; Ehrlich et al. 1994). It is therefore suggested that strong business angels also reduce the 'liability of newness' of start-ups (Sørheim 2000). Not surprisingly, it is widely found that business angels concentrate on early-stage ventures where past experience and business know-how from angels are crucial, whilst VCs focus on later-stage ventures to avoid the cost of involvement (Van Osnabrugge 1999; Freear et al. 2002; Aernoudt 2005).

The angels' stress on business involvement probably explains the focus of their investment criteria on the quality of the management team. While the findings regarding angel investment criteria vary, it is unanimously agreed that the quality of the entrepreneur or the management team is the angels' primary criterion (Mason and Harrison 1996; Van Osnabrugge 1999; Gracie 1999; TSBS 2001; Mason and Stark 2002; Jungman et al. 2005; Southon and West 2006; Beer 2006). A management team's quality has been translated as the team's understanding of the business and their potential ability to create profit, pitching ability and rapport between the team and the angels, and track record of the team

(Van Osnabrugge and Robinson 2000; Beer 2006). The findings regarding other investment criteria adopted by business angels have been variable among the relevant literature. They generally include the potential satisfaction and enjoyment from involvement (Van Osnabrugge and Robinson 2000), possibility of exit (Beer 2006), 'investment ready' business plan[2] (Beer 2006; Mason and Stark 2002), industry and geographical proximity[3] (Jungman et al. 2005), growth potential and product readiness such as the status of the intellectual property rights (Van Osnabrugge 1999). Nonetheless, the heterogeneity exhibited by the findings is largely due to the personalised investment approach adopted by the business angels (Mason and Stark 2002).

The emphasis on the choice of factors/criteria and the neglect of the methods or processes involved in the invention assessment task are widely exhibited in other University Technology Transfer literature. A large number of literature adopt the notion that University Technology Transfer is generally a process that can only be learned through experience and cannot be taught (Crown 2003). Owen-Smith and Powell (2003) quoted a comment from a senior licensing associate: 'There is no curriculum for training someone. We try to send people to the AUTM (Association of University Technology Managers) seminars but they are really going to learn more by being here on the job,......This business is very much learn as you go and the more deals you are involved with, the more quickly you learn.' This notion of 'University Technology Transfer cannot be learned' together with the intuitive approach to invention assessment underlined the ad hoc management style of university technology transfer offices.

A group of literature that can be utilised to develop invention assessment methods is the so-called critical-factor literature. Critical-factor literature proposed positive or negative factors influencing the success of University Technology Transfer. Successful University Technology Transfer is largely accepted by these studies as inventions leading to monetary return for the

[2] Investment ready business plans refer to business plans which cover adequate information on market and financial issues, with limited technical jargon, and clear and probable assumptions.
[3] Industry and geographical proximity refer to business opportunities which are compatible with the business angels' expertises, and are close to where the angels are geographically based.

university. The level of success is commonly quantified as the level of output of a university technology transfer office, which is further quantified as the number of signed licensing contracts with established companies and spinouts created. Some added that non-sustainable spinouts should not be included (Crown 2003). The common approach of these studies is to locate the critical factors affecting the output of a University technology transfer office through empirical data generated from interviews, questionnaires, case studies and patent records, either at university-level or invention-level. Studies conducted at university-level looked at factors affecting the aggregate output of individual technology transfer offices, whereas invention-level studies examined factors influencing the outcome of individual inventions. In light of the aim of the current research to develop the Commercial Outcome Prediction System (as explained in Chapter One), factors not applicable to discriminate among inventions within a university, often suggested by certain university-level studies, would be outside the scope of the current research. For instance, Gregorio and Shane (2003) proposed that universities geographically locating near to venture capitalist are more likely to attract funding for spinout establishments. Such a factor, if valid, would be applicable to all inventions within these universities and therefore does not serve to discriminate among inventions.

However, though university-level studies do not generate critical factors that directly discriminate among inventions, important insights can be learned from some of the university-level literature. The most frequently found critical factors among the university-level literature include *experienced university technology transfer office* (Rogers et al. 2000; Siegel et al. 2003b; Carlsson and Fridh 2002; Mowery et al. 2002; Owen-Smith and Powell 2003; Friedman and Silberman 2003), *active networking with the private sectors* (Zucker et al. 1997; Shane and Stuart 2002; Friedman and Silberman 2003; Owen-Smith and Powell 2003; Lindel鰂 and L鰂sten 2004), and *strong scientific base of the university* (Foltz et al. 2000; Thursby and Kemp 2002; Owen-Smith and Powell 2003; Gregorio and Shane 2003; Siegel et al. 2004). Other less frequently stated critical factors from the university-level studies include high *university prestige* (Sine et al. 2002;

Gregorio and Shane 2003), *university patent impact* (Zacks 2000; Mowery et al. 2002) and *great rewards for inventors* (Friedman and Silberman 2003).

Among the invention-level studies, the most frequently found critical factors include *significant inventor's scientific capacity and reputation* (Foltz et al. 2000; Rogers et al. 2000; Owen-Smith and Powell 2003; Zucker et al. 1997; Jensen and Thursby 2001), *certain invention discipline* (Shane 2002; Thursby and Kemp 2002; Crown 2003; Owen-Smith and Powell 2003), *personal relationship with firms or investors* (Hsu and Bernstein 1997; Shane and Stuart 2002; Thursby and Thursby 2003; Owen-Smith and Powell 2003, Siegel et al. 2004), *developmental stage* (Jensen and Thursby 2001; Mike et al. 2004; Markman et al. 2005), and *secured external funding* (Hsu and Bernstein 1997; Crown 2003; Mike et al. 2004). Other less frequently stated critical factors located by invention-level studies include *inventor's business experience* (Shane and Khurana 2000), *types of research: basic or applied* (Bozeman 2000), effective patents (Shane 2002), and *lack of inventor's input in supplying contacts of potential licensees* (Hsu and Bernstein 1997).

When examining the details of the critical-factor literature, it is noted that they generally fall into one or both of the following two categories, namely qualitative survey and latent variable model. A large proportion of the critical-factor studies belong to the qualitative model category, such as the works of Hsu and Bernstein (1997), Thursby and Thursby (2003), Owen-Smith and Powell (2003) and McAdam et al. (2005). They are generally based on inductive methods like interview survey, case studies, and content analysis. These studies contributed to the sector by revealing the perceptions, inside stories and management strategies adopted by the various technology transfer stakeholders. The contributions from the literature are vital as they added new findings to the general understanding of the sector, which may have assisted the research design of subsequent studies. Moreover, the dominance of studies in this category is mainly because of the embryonic nature of the University Technology Transfer sector as well as the paucity of online databases. Owing to the newness of the field, researchers are still developing conceptual frameworks in attempts to establish unifying theories.

Also, University Technology Transfer is confidential in nature, since the inventions represent potential commercial opportunities. Readily available data for quantitative analysis is therefore limited. These studies generally involve extensive painstaking collection of qualitative verbal or written field data which the researchers have access to. However, qualitative data is highly vulnerable to ambiguity. Factors are often concluded as critical factors because they are repeatedly quoted by interviewees during the survey, such as the works of Owen-Smith and Powell (2003) and Thursby and Thursby (2003). However, the interviewee's interpretation of a factor may well be different from that of others'. For example, Hsu and Bernstein (1997) concluded from their interview findings that the value of an invention is one of the most important factors affecting the success of University-Technology Transfer. The value of an invention, however, has been interpreted or operationalised in terms of licensing income (UNICO et al. 2003), contribution to humankind (Crown 2003), or even patent citation (Shane 2001). Despite the wide variety of interpretations offered by previous researchers to many generic terms, subsequent qualitative survey studies continue to use the terms in the absence of any form of definitions. Such ambiguity nature of qualitative data, when appropriately controlled with more explicit definitions, serves to synthesize and summarise the properties of more quantifiable variables. Without necessary definitions, it can lead to unnecessary inconsistency and confusion as interviewees might have been attaching opposing meanings to the same term.

A subset of the critical-factor literature belongs to the second category, latent variable models, including the works of Thursby and Thursby (2002), Gregorio and Shane (2003), Friedman and Silberman (2003), Siegel et al. (2004), and Markman et al. (2005). The latent variable model refers to the construction of latent/hypothetical variables to represent various concepts, and using observable and measurable variables to represent the latent variables. Due to the quantitative nature or more quantitative nature of the observable variables, statistical data analysis methods are often applied to these models. A number of potential factors were often treated as independent variables and the outcome of technology transfer was treated as the dependent variables. The common goal of these studies was to identify the critical factors among the potential factors. Since the choices of observed variables used to represent the latent variables are up to the researchers, inconclusive findings were generated among the studies. For example, it appears that both Shane and Stuart (2002) and Lindelöf and Löfsten (2004) share the same conclusion in which they proposed that *connection with the private* sectors is an important critical factor. Yet, Shane and Stuart (2002) adopted counts of personal contacts with firms as the observed variable as a measure of the level of connection whereas the observed variable used by Lindelöf and Löfsten (2004) is the count of the number of firms located within a university science park. Though one can argue that the ambiguous nature of latent variable allows a wider margin for further investigations, it became a problem when it spread to the common dependent variable, the outcome of technology transfer. A closer look at these studies showed that a variety of observed variables have been assigned to represent the level of University-Technology Transfer success. They include the number of signed license contract (Siegel et al. 2004), the number of spinouts formed (Gregorio and Shane 2003), the number of invention disclosed (UNICO et al. 2003), the amount of licensing income (Markman et al. 2005), and a combination of the above (Friedman and Silberman 2003). The use of different observed variables for the common dependent variable is dangerous. Researchers build upon each other's work based on the assumption that they are solving the same dependent variable when they are not. Particularly, there is a lack of consensus regarding the validity of using the status of 'started a spinout' to represent 'successful University-Technology Transfer'. Some argue that most of the university spinouts in the United Kingdom created after the late 1990s are merely possible because of the Government's Third Stream Funding and they lack the ability to attract any external capital (Crown 2003). These spinouts therefore do not represent successful cases of University-Technology Transfer. Under such argument, factors closely correlated with spinout formation thus are not necessarily critical factors for successful University-Technology Transfer. However, Shane (2001) argued that spinouts are often resulted from important inventions and therefore regarded spinout formation as a valid measure of University-Technology Transfer success. Besides, the number of signed licence contracts can also be a problematic representation for success. Some consider signed licence contracts to cover licensing to established companies only (Thursby and Thursby 2003). others consider it to include both licensing to established companies and spinouts (Shane 2002). This ambiguity problem is also an industry-wide problem where technology transfer personnel in practice as well as other stakeholders involved, such as investors and companies, also hold different opinions for the definition of a successful case of University-Technology Transfer (Siegel et al. 2004). These problems are, nonetheless, due to the embryonic nature of the field where there are insufficient standards and established norms. In particular, patent data has been increasingly used to represent latent variables. This is largely due to the fact that most inventions involved in University-Technology Transfer activities are patented or patentable entities. For instance, Shane (2001) used patent citation, the number of different patent subclasses of the previous patents cited by a given patent, the number of patent classes assigned to a given patent and the number of patents owned by the invention in the past to represent the invention value, the invention's radicalness, the scope of protection for the invention and the inventor's firm founding experience respectively. Furthermore, the TR's University Research Scorecard used a patent related measure (including patent number and patent citation data) to represent the technological strength of research universities (Zacks 2000). The number of patents has also been widely used as a measure of a university's productivity in technology transfer (Thursby and Thursby 2002). In fact, patent data is one of the common bibliometric entities which have been heavily investigated by the Science and Research Assessment literature. Other bibliometric entities such as faculty research grant have also been used to represent critical factors (Friedman and Silberman 2003). Furthermore, the analysis methods employed by those latent variable model studies include descriptive statistics (Meseri and Maital 2001; Siegel et al. 2003), correlation and regression analyses (Friedman and Silberman 2003; Shane 2001b; Markman et al. 2005). These studies provide valuable evidence showing the close relationship and co-occurrences of certain factors and the technology transfer outcome. However, the old aphorism of 'correlation does not imply causation' seems to have been forgotten by some of these studies.

Moreover, several drawbacks were applicable to both categories of studies. While the studies concluded that the critical factors are likely to be predictive factors for the final technology-transfer outcome, they fail to provide evidence as they are retrospective in nature. In addition, due to the lack of publicly available invention data, most of the studies depended on privileged access to a certain or a small number of university technology transfer offices. Such privileged access may have given rise to inappropriate generalisation and biased findings caused by non-representative samples. Most studies, such as Shane (2001) and Shane and Stuart (2002), are using the data from non-representative institutions such as MIT (Massachusetts institute of Technology) where extremely favourable conditions for University-Technology Transfer may be available. Inconclusive and inconsistent findings are therefore generated from these studies, where critical factors proposed by one study as success factors can be rejected by another. For instance, evidence from Markman et al. (2005) disagreed with the success factor *great rewards for inventors* proposed by Friedman and Silberman (2003) by showing that the level of rewards for inventors is actually negatively related to a technology transfer office's performance. Another concern is that most of the literature regarding University-Technology Transfer is from the United States. This is mainly due to their more developed state of the sector, which provides more readily available data. Among these United States studies, many obtained valuable data from AUTM (The Association of University Technology Managers), which is a central body providing various resources to the sector. While similar organisations, such as the AURIL (Association for University Research and Industry Links), exists in the United Kingdom, the scale and resources offered are relatively limited when compared with AUTM. While the critical factor studies often refer to and build upon the findings from one another, it is important to note that critical factors derived from the United States may not be applicable to the United Kingdom. This is also supported by the findings by Schmiemann and Durvy (2003), who concluded that there are substantial cultural, legal and regulatory differences between the United States and Europe, and especially within Europe.

Despite the inconsistency resulting from the incomplete pictures presented by individual critical-factor studies, their findings provoke healthy controversy and together they have pushed the understanding of the field forward. Three assumptions have been developed based on the findings from the literature as a whole.

Firstly, while the studies typically use their own unique data source, some factors are widely agreed whereas others have stirred disputes. It is therefore highly probable that certain critical factors are transferable across universities (transferable factors) and others are not (non-transferable factors). While individual university technology transfer offices may be able to learn about the transferable factors from the findings based on other universities, non-transferable factors can only be learned from the office's own data. Additionally, it is previously stated that only invention-level critical factors apply to the discrimination of inventions within a university. The transferable critical factors important to invention assessment are therefore the most frequently found critical factors among the invention-level literature.

Assumption one: There are two types of critical factors, one is transferable and the other is not. The former can be learned from others whereas the latter can be learned from a university technology transfer office's own data.

Secondly, inter-relationships may exist among the critical factors. According to the explanations suggested by the critical-factor studies, many of the factors are interrelated. For instance, inventions from certain scientific disciplines, such as biotechnology, are more likely to succeed in University Technology Transfer than others (factor *invention discipline*). This is largely due to the industry's origin from universities. Most researchers from the industry therefore have close relationship with academic researchers (factor *networking between universities and the private sectors*). They understand the development of each other's work. Academics with strong scientific capacity (factor *inventor's scientific capacity and reputation*) are well known to both the academic and industrial communities, and often play a central role in commercialising the invention (Zucker et al.

1997). Furthermore, technology transfer is affected by the nature of an industry (factor *invention discipline*). Process based industries are more prone to moral hazards. For instance, technology transfer is more difficult for inventions from the manufacturing discipline, as they also require the transfer of tacit knowledge (Bozeman 2000). Pharmaceutical and semiconductor sectors have strong links to basic science (Mike et al. 2004), inventor's scientific capacity (factor *inventor's scientific capacity*) therefore often have a greater impact on these sectors. In addition, inventions from certain industries (factor *invention discipline*), such as the biotechnology sector, tend to be more radical and embryonic than those from other industries (factor *invention discipline* and *developmental stage*). Although embryonic or 'proof of concept' inventions are found to be more difficult to license than 'prototype' inventions (Jensen and Thursby 2001); early-stage or 'proof of concept' inventions tend to be more innovative and are more likely to raise venture capital funding (factor *secured external funding*) to establish spinouts (Mike et al. 2004). Having venture capital funding is found to be the largest contributor to the likelihood that a spinout undergoes Initial Public Offering (Shane and Stuart 2002). These inter-relationships also appear to be mainly triggered by the factor *invention discipline*. In other words, the inter-relationships are hierarchical in nature where the factor *invention discipline* may be located close to the top of the pyramid.

Assumption Two: Inter-related hierarchical relationships exist among critical factors.

Lastly, recursive learning is highly likely to be the major mechanism to enhance the ability to identify the appropriate inventions. In other words, the assessment accuracy is likely to be increased as a result of recursive learning. For instance, the factor of *university patent impact* is generally measured as the number of citation by subsequent patents received by the university's patents in a given year (Mowery et al. 2002; Owen-Smith and Powell 2003). Mowery et al. (2002) suggested that the effect of this factor is a result of universities learning over time to identify more valuable patents. In addition, the most frequently supported critical factor is *experienced technology transfer offices*, and experiences are accumulated by recursive learning.

Assumption Three: Invention assessment accuracy can be increased through recursive learning.

This section has reviewed literature from the University Technology Transfer domain that is relevant for the process of invention assessment. While studies investigating critical factors for intuitive invention assessment abound, very little has been done regarding non-intuitive methods to make use of these factors. The next two sections therefore review literature from other domains regarding non-intuitive assessment methods that are potentially applicable for the University Technology Transfer sector, namely evaluative bibliometrics applied in the Science and Research Assessment domain and quantitative classification models for the financial sector. Though the projects involved in these sectors do not share exact features as university-developed inventions, they are analogous to university-developed inventions in the respect that they all involve a high level of uncertainty and compete against each other for limited resources.

2.3 EVALUATIVE BIBLIOMETRICS

This section examines evaluative bibliometrics literature from the domain of Science and Research Assessment. Though peer review is the dominant assessment methods used in this domain, it will not be covered in this section as it is not a non-intuitive assessment method. The domain of Science and Research Assessment refers to the evaluation of scientific research projects funded by public organisations. The relevant method located in this domain is evaluative bibliometrics. This method is relevant to the invention assessment task in the University-Technology Transfer domain as evaluative bibliometrics are often proposed as valid measures regarding different aspects of research projects.

which in return are closely related to the critical factors proposed by the University-Technology Transfer sector.

Evaluative bibliometrics include simple counts or more advanced analyses of various bibliometrics indicators. Entities widely perceived to be appropriate bibliometrics indicators include refereed journal papers, patents, paper and/or patent citations, patent data (such as patent classes), peer reviewed books, keynote addresses, conference proceedings, competitive grants and so on. Among these entities, refereed journal papers are the most widely agreed bibliometrics indicator across different technical disciplines. Moreover, most bibliometrics literature focuses on refereed journal paper, patents, their citation and patent data (Pakes 1986; Lerner 1994; Podolny and Stuart 1995, 1996; Agrawal and Henderson 2002). Most evaluative bibliometrics involve counts of these entities, more advanced versions include simple arithmetic of these entities such as weighted counts and normalised counts (Schubert and Braun 1996; Kostoff 1997). These analyses are not trivial given the large and increasing number of technical disciplines and the enormous volume of bibliometrics entities produced. It is estimated that for every working day worldwide, about 5000 new papers are published in refereed journals and about 1000 new patents are issued (Narin et al. 1994). The literature found in this area generally interprets evaluative bibliometrics as one or more of the following four types of measurement. Firstly, counts of papers and patents are often suggested to be a valid measurement of the level of research and development activity (Butler 2004). Secondly, the number of times those papers and patents are cited by subsequent papers or patents (citations) are commonly proposed to be a valid measurement of the level of impact or importance of the cited papers and patents. Thirdly, the citations from papers to papers, from patents to patents and from patents to papers are usually advocated as a linkage/network/spillage measurement of intellectual linkages between the authors/organisations and the knowledge linkages between the subject areas (Jaffe and Trajtenberg 1996; Narin et al. 1997). Fourthly, a variety of patent data, including patent citation, have been considered to be measurements for various aspects of technical disciplines. For instance, counts of patents in a patent class and counts of patent classes assigned to a patent by a patent office have been used to measure the crowdedness of a technical area (Podolny and Stuart 1995) and the radicalness of a technology respectively (Lerner 1994).

Among these four types of measurement, patent data received the most attention. This is largely because patent citation, especially the citation of papers by patents, has been widely seen as a potentially effective indictor for the conversion of science to technology (Narin et al. 1984; Schmoch 1993; Narin et al. 1997; Meyer 1999). Large scale analysis of a variety of patent data have also been commercialised as electronic databases for various research purposes, such as the TECH-LINE for financial application of technology indicators.

As reviewed in Section 2.2, patent data has also been increasingly researched by the University-Technology Transfer literature. In light of this, various aspects regarding patent data are reviewed in more details below. Their relevance to University-Technology Transfer is also pointed out when necessary.

Patents provide a legal right to prevent others from imitating a particular technological development in areas delineated by the patent claims. The scope of the patent is important because "the broader the scope, the larger number of competing products and processes that will infringe the patent" (Merges and Nelson 1990). When a patent is narrow in scope, the holder of the patent will have less incentive to develop the technology through the creation of a new firm as it will have a smaller space of technology that is protected against imitation by other firms (Merges and Nelson 1990). In University-Technology Transfer, inventors revealed that they often ask patent attorneys for a judgement as to the scope of patent protection before they decide to establish spinouts to exploit their inventions (Shane 2001). Moreover, it is reported that investors are concerned with the breadth of the patents and prefer broader patents in the decision of whether or not to fund a new venture (Shane 2001; Lerner 1994). However, the potential economic values created by patented technologies are highly varied (Trajtenberg 1990). Empirical evidence has showed the skewness of the distribution of patent value (Pakes 1986) and indicated that most patents have no commercial value and only a few have a large value (Trajtenberg 1990; Trajtenberg et al. 1997; Scherer and Harhoff 2000). Besides, various measures have been proposed for patent value, among which patent citation is the most widely accepted measure. Evidence confirmed that patent citations are significantly correlated with the economic value of invention. Furthermore, Harhoff et al. (2002) found that the higher the estimated private value of an invention, the more the patent was cited by later patents. Hall and Ham Ziedonis (2001) also found that companies with highly cited patents have higher stock market values, all other things being equal. Citations influence the legal boundaries of the property rights to an invention (Jaffe et al. 1993). According to the United States Patent and Trademark Office (USPTO), patents are divided into approximately 100,000 nine-digit patent classes, which aggregate to approximately 600 three-digit classes, and which represent distinct technical areas. The assignment of a patent to a particular patent class represents the USPTO's assessment that the patent belongs in a particular technical field. Patent examiners also determine what previous inventions must be cited in a patent by searching prior patents. Because patents belong to technical classes and because they cite previous patents, citations to patents in particular technical fields represent the USPTO's assessment that a particular invention builds upon (cites) knowledge in that technical field. When a patent cites previous patents in classes other than the ones it is in, that pattern suggests that the invention builds upon different technical paradigms from the one in which it is applied. While citation inflation may exist in journal citation, this problem is less likely for patent citation. Jaffe et al. (1993) explained that gratuitous citations in patent applications are costly to inventors and yet may be common in journal publications. Firstly, patent citations determine the scope of the inventor's monopoly and unnecessary citations limit what the inventor can claim as his or her monopoly right. Secondly, it is the patent examiner's job to correct citation of previous patents and to remove gratuitous citations from the patent before it is issued. Other measures related to patent value include the use of patent lifetime, patent breadth/scope and the number of claims. One of the earliest attempts to estimate patent value by patent-lifetime was by Nordhaus (1967), who assumed that the accumulated profit flows of the patent obeyed a function similar to that of discounted values. However, following findings from technology cycles, Matutes et al. (1996) disagreed with the 'constant return per period' assumption. Moreover, patents breadth is the degree of protection granted upon the invention realisation, where the patent owner may be granted to realise the invention in a certain number of specific ways. It is assumed that the patent breadth correlates positively with the innovator's profit (Klemperer 1990; Gilbert and Shapiro 1990). Lerner (1994) operationalised patent breadth as the number of four-digit classes assigned by the International Patent Classification system. Lastly, the number of claims is also found to be a value determinant of a patent. The number of claims is found to be correlated positively with national research performance (Tong and Frame 1992).

In terms of research proposal selection, these evaluative bibliometrics, in theory, can be used as valid measures for the past performance of the researchers or similar previous research projects, and thus reflecting the potential of future research proposals (Lanjouw 1998). In practice, peer review continues to remain as the dominant project/program assessment method used by Government agencies and private companies. However, evaluative bibliometrics often form a part of or a minor part of national research assessment programs, such as the RAE (Research Assessment Exercise) in the United Kingdom and the REPP (Research Evaluation and Policy Project) in Australia (Butler 2004). The limited use of evaluative bibliometrics in practice is largely due to the criticism that their validities vary significantly across authors, technical disciplines and types of organisations. Empirical studies showed that Nobel Prize winning papers were often initially rejected by one or more journals (Campanario 1995). Very few of the active researchers, such as Alfred Lotka, William Shockley, and Derek de Solla Price, are found to be producing the heavily cited papers (Narin 1976). Some claimed that publication and citation habits differ considerably across fields and between the science and social science streams (Korevaar 1996). For instance, paper citations are only informative about the value of pharmaceutical and chemical patents but not in other technical fields (Harhoff 2002). The impact factor, the bibliometrics indicators extensively used to assess scientific production, is also found to be biased when different subfields are compared together (Schwartz and Lopez Hellin 1996). Other typical drawbacks include: counts of papers may be biased towards prestige journals and English language journals; counts of papers and patents merely measure quantity but not quality (Chan 1976); a subset of technologies may be kept secret and not publicly disclosed (Cordes et al. 1999); influences are under-cited (MacRoberts and MacRoberts 1996); citation decisions are often arbitrary judgements by authors (Liu 1993) and artificially inflated through self-citation (Weingart 2005) and ethnic bias (Greenwald and Schuh 1994; Egghe et al. 1999).

2.4 QUANTITATIVE CLASSIFICATION IN FINANCE

This section starts by introducing the basic concepts of quantitative classification and the rise of its applications in the financial sector. Relevant literature regarding the use of classification systems for different financial applications is then reviewed.

Basic Concepts

Quantitative classification is defined in this thesis as a mathematical procedure in which individual cases are placed into classes based on quantitative information on one or more attributes inherent in the cases, based on a training set of previously classified cases. Quantitative classification is closely related to other quantitative fields including statistics, knowledge discovery in databases or data mining, pattern recognition or artificial intelligence. Classification algorithms commonly used by these fields include Bayesian belief networks, maximum-likelihood estimation, k-nearest neighbour, linear classifiers, neural networks, decision tree induction, principal component analysis, and so on.

The basic components involved in a quantitative classification system are presented in Figure 2-2.

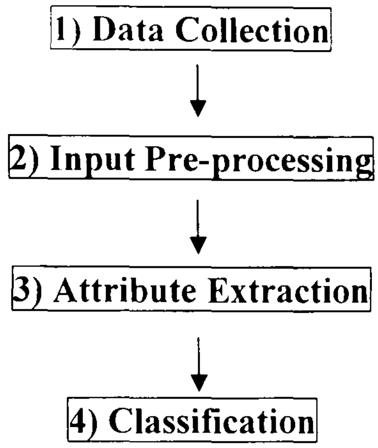


Figure 2-2: Basic components of a quantitative classification system

These steps differ according to the application domain. Generally speaking, data collection refers to the collection and storage of raw input data from one or more data sources, which can be achieved manually or through input devices such as a camera or recorder. Input pre-processing covers operations applied to the raw input data. This can include segmentation operations to group data into individual cases, cleaning operations to alleviate the effect of missing data, and normalisation operations to standardise data derived from different sources. Attribute extraction is the process of identifying discriminative attributes that are invariant to input pre-processing operations. Discriminative attributes refer to measurements applicable to every case, whose values are very similar among cases of the same class and very different for cases in different classes. The conceptual difference between the attribute extraction step and classification step is highly arbitrary. An effective attribute extractor can make the job of the classifier trivial, whereas a powerful classifier may not need any attribute extractor. Classification, the final step of the process, refers to the task of using the information provided by the attribute extractor to assign cases into appropriate classes. Since perfect classification for non-trivial problems are often impossible, the common approach is to determine the probability for each of the possible classes for individual cases.

The use of quantitative classification is widely researched and practised in the financial sector (Zhang and Zhou 2004). This is largely driven by three factors, substantial amount of financial data, availability of theoretical models, and the non-linearity of financial data. As a result of increased economic globalisation and improved information technology, the rate of data generation in the financial sector has excluded the practicality of manual analysis. Moreover, there are many well-developed finance and accounting models, such as the capital asset pricing model and the Kareken-Wallance model, identifying important attributes as well as their inter-relationship. In addition, a large proportion of financial data are times series which are noisy, non-linear, non-stationary and deterministically chaotic. A time series is a sequence of real numbers representing values of a real variable, such as a stock price or an exchange rate, measured at equal time intervals. Conventional statistical analysis and tests indicated that financial time series has non-random behaviour (Taylor 1986). These factors have contributed in the widespread use of classification algorithms for financial prediction applications. To date, quantitative classification systems have been used in various financial areas, including stock selection, loan risk assessment, fraud detection, bankruptcy prediction, and so on. These classification systems have supported proactive and knowledge-driven decisions, achieved increased revenue, reduced cost, and improved market responsiveness and awareness.

Existing Classification Systems in Finance

Classification algorithms are widely used in the financial sectors for a variety of problems. The five most popular application areas are stock selection, asset selection for portfolio management, foreign exchange market prediction, bankruptcy/business failure prediction and credit risk evaluation. These application areas all involve option selection decisions where classification systems are employed to assist the decision making process by predicting future outcomes for individual options. Literature regarding classification applications in these five areas are grouped into two groups, namely option selection and outcome prediction applications, and are reviewed below.

Option Selection

This group of applications include the areas of stock selection, asset selection for portfolio management, and foreign exchange market prediction. Stock market investors maximise their returns by choosing the appropriate stocks and appropriate times to trade (to buy or sell stock). Portfolio management concerns about the selection of various securities and assets to form a portfolio under an acceptable level of risk, and the ongoing monitoring of these investments to meet specific investment goals for the benefit of the investors. Foreign exchange is the simultaneous act of buying one currency and selling another. The market for foreign exchange is the largest financial market in the world, generating an average daily turnover of over US$1 trillion. Investors therefore maximise return by selecting the appropriate currencies to buy and sell. The common aim of these three areas is to select the right options for investment optimisation.

Macro economic variables, factor models, technical indicators, and critical factors based on economic fundamentals are widely used in real life financial forecasting applications. Popular factor models include the Capital Asset Pricing Model and Arbitrage Pricing Theory (Ross 1976). Examples of technical indicators for stock market prediction are Exponential Moving Average, Relative Strength Index and Bollinger Band (Ince and Trafalis 2004). For instance, critical factors developed to assist the prediction of the future returns of individual stocks include the growth rates of the followings: revenues, earnings per share, capital investment, debt, and market share (Walczak 1999). Financial database containing extensive financial data such as the Russell Indexes and stock prices on NASDAQ are often used for research and benchmark purposes (Sorensen et al. 2000). Traditionally, parametric pricing methods such as linear regression models have been used to develop applications for these three financial areas. However, due to those models' inability to cope with the highly non-linear and time-variant nature of financial data (Roman and Jameel 1996), classification algorithms especially neural networks are increasingly becoming the dominant techniques for financial prediction over the past decades largely due to its universal function approximating ability (Hutchinson et al. 1994; Wang and Leu 1996; Walczak 1999; Leigh et al. 2002; Chen et al. 2003; Enke and Thawornwong 2005).

Despite the wide spread support for neural networks (as explained in Chapter Five), there is a lack of consensus on the superiority of neural networks over other techniques such as support vector regression. While some confirmed that neural networks performed better (Refenes et al. 1994; Ince and Trafalis 2004), others found otherwise (Tay and Cao 2002; Trafalis and Ince 2002). Furthermore, a study demonstrated that the addition of relevant external market indicators improved the neural networks' prediction performance (Walczak 1999). For this, principal component analysis has often been used to pre-process the input prior to the classification by neural networks (Walczak 1999; Ince and Trafalis 2004). Additionally, decision tree induction, statistical analysis, genetic algorithm have also been used in financial market prediction (John et al. 1996; Saad et al. 1998; Sorensen et al. 2000). For instance, the Recon system, a stock selection tool which induces classification rules to model the given data, is reported to have achieved more than double of the total return generated by the benchmark over four years (John et al. 1996). Neural networks are also often used in conjunction with genetic algorithm for portfolio management, where neural networks are used to predict the future returns of individual assets and genetic algorithm is used to determine the optimal weights for each asset (Xu and Cheung 1997; Lazo et al. 2000). For instance, LBS Capital Management Inc. is one of the examples which use neural networks and genetic algorithm to manage their portfolio worth US$600 million. Moreover, it is increasingly accepted that the identification of relevant technical trading rules increases the returns achieved from the foreign exchange market. Genetic algorithm and neural networks have also been used to identify these rules or simulate the volatility of exchange rates (Neely et al. 1997; Walczak 2001). CART (classification and regression tree) is another popular method widely adopted for asset selection (Sorensen et al. 1994, 1996).

Outcome Prediction

This group of applications include the areas of corporate bankruptcy prediction and credit risk evaluation. Corporate bankruptcy is the legal process in which a firm declares inability to pay debts. It is the final state of corporate failure. To date, corporate bankruptcy has reached an unprecedented level, causing huge economic losses to industrial units, financial institutions, government, and considerable national social and economical damage. Corporate bankruptcy prediction systems therefore act as early warnings to relevant stakeholders regarding the firms' potential industrial failure. Credit risk refers to the risk that a borrower, a corporate or an individual, will not repay all or a portion of a loan on time. Credit risk evaluation aims at identifying non-deserving clients through evaluation of various factors that can lead to the non-payment of obligations. Corporate bankruptcy prediction is therefore essentially a form of credit risk evaluation, in which bankrupt firms represent non-deserving clients. While financial credit is a lucrative global industry with annual credit card transactions, consumer debt and high interest credit card loans amounting to trillions of pounds, both personal and corporate bankruptcy are on the rise. In the United States alone, over a million bankruptcies were filed between 2002 and 2003 (West et al. 2005). It is thus important to distinguish deserving clients from non-deserving clients by careful credit risk evaluation.

Both corporate bankruptcy prediction and credit risk evaluation are often implemented as binary classification applications aiming at distinguishing non-bankrupt firms/deserving clients from bankrupt firms/non-deserving clients. The breakthrough in bankruptcy prediction was the Z-score model, a highly accurate diagnostic tool forecasting the probability of corporate bankruptcy within a 2 year period, developed by Edward I. Altman in 1968 using multiple discriminant analysis and five key financial ratios (Altman 1968). The five key financial ratios are: working capital/total assets, retained earnings/total assets, earnings before interest and taxes/total assets, market value of equity/book value of total liability, and sales/total assets. As illustrated, a financial ratio is a quotient of two numbers, in which both number consist of financial statement items computed from a company's balance sheet. While these financial ratios are also relevant for corporate credit risk evaluation, the traditional approach for consumer credit risk evaluation is to decide intuitively through inspection of the application form details of the applicant including the socio-demographic status, economic conditions and intentions. Depending on the type of loan, whether it involves corporate or whether it is a long term loan, other possibly relevant information for evaluation includes the character of the borrower, collateral, sources of repayment, interest rate and economic condition (Altman and Haldeman 1995). As the advancement in information technology facilitated the electronic storage of all information regarding the characteristics and repayment behaviour of credit applicants, statistical and machine learning algorithms are increasingly applied in the area of credit risk evaluation. Real life credit data, such as the German credit data set[4] and the Bene data sets[5], is often publicly available for research and benchmark purposes, further motivating the use of automatic classification methods for credit risk evaluation.

Since the advent of the Z-score model (Altman 1968), discriminant analysis and financial ratios have been widely adopted for bankruptcy prediction (Sung et al. 1999). More recent models adopted for bankruptcy prediction are genetic algorithms (Sung et al. 2005), decision trees and CART (classification and regression trees) (Foster and Stine 2004), and neural networks (Odom and Sharda 1990; Tam and Kiang 1992; Altman et al. 1994; Wilson and Sharda 1994; Alici 1995; Boritz and Kennedy 1995; Atiya 2001). While many corporate bankruptcy prediction models choose financial ratios according to a choice based criteria, principal component analysis has been used to allow any number of financial ratios as input (Rekha Pai and Annapoorani 2002; Rekha Pai et al. 2004). In the area of credit risk evaluation, neural networks also received a lot of attention (Srivastava 1992; Jensen 1992; Davis et al. 1992; Salchenberger et al. 1992; Altman et al. 1994; Lacher et al. 1995; Desai et al. 1996; West 2000; Baesens et

[4] The German credit data set is publicly available at the UCI repository: http://www.ics.uci.edu/~mlearn/mlrepository.html.
[5] The Bene data sets can be obtained from major Benelux financial institutions.

al. 2003). Other classification models adopted for credit risk evaluation include k-nearest neighbour (Henley and Hand 1997), classification trees (Davis et al. 1992) and genetic algorithm (Desai et al. 1997). While many comparison studies regarding the performance of neural networks and other methods have been carried out, the results were contradictory and inconclusive (Altman et al. 1994; Desai et al. 1996; Sung et al. 1999; Rekha Pai et al. 2004; Tang and Chi 2005).

2.5 CONCLUSION

Quantitative classification has been widely adopted in the financial sector for option selection and outcome prediction applications. Both applications are basically realised by applying classification algorithms to a combination of two types of entities, namely financial indicators/financial ratios and economic knowledge/critical factors. While the financial sector is making use of the two types of entities, the current research proposed that University-Technology Transfer can adopt a similar approach as these two types of entities also exist in the University-Technology Transfer sector. Financial indicators/financial ratios from the financial sector is analogous to evaluative bibliometrics from the Science and Research Assessment domain, in which they are both quantitative indices proposed to be predictive of investment outcomes to enhance project selection. Furthermore, critical factors have been widely researched in the University-Technology Transfer literature.

Technology transfer practitioners often complian about the difficulties involved in invention assessment due to the complex nature of University-Technology Transfer, long time lag of invention development and the uniqueness of individual inventions. Such complexity, time lag and uniqueness, indeed, also apply to the financial sector. Financial time series are well known to be complex, non-stationary and deterministically chaotic. A large number of stakeholders are also involved in financial decision making. Areas such as credit risk evaluation involve long time lag while upfront investment decisions are required. Stock and asset price behaviours also vary significantly across industrial sectors. Financial indicators useful for one stock can be ineffective for another stock.

Among the University-Technology Transfer literature, though quantitative methods such as linear regression have been applied to data representing critical factors and technology transfer outcome to demonstrate the impact of the factors, only a few critical factors were investigated at a time. Studies covering a large number of factors are rare. Moreover, these studies are retrospective and very little has been done actually using or combining the findings to generate predictive classification systems.

The University-Technology Transfer literature generally focus on the identification of critical factors and adopt an intuitive approach to invention assessment, with very little attention paid to developing quantitative methods for invention assessment. As illustrated in Section 2.4, the financial sector indeed went through a stage similar to the current situation of University-Technology Transfer, where literature concentrated on investigating critical factors with predictive power and practitioners adopted an intuitive approach to investment assessment. The major difference between the financial sector and the University-Technology Transfer sector is that project/option data is publicly available for the former but not the latter. This is largely due to the confidential nature of invention data as well as the embryonic nature of the University-Technology Transfer sector. Nonetheless, though invention data from other universities is unlikely to be obtainable, publicly assessable patent and publication databases as well as the internal databases of individual technology transfer offices represent useful sources of input to construct predictive classification systems. The three hypotheses developed in Section 2.2 are also relevant to the application of quantitative classification for invention assessment. Firstly, publicly available bibliometrics entities may be representatives of transferable critical factors and data contained in university internal databases may be representatives of non-transferable critical factors. Secondly, the application of classification algorithms to invention data set may identify the hierarchies among critical factors. Lastly, the accuracy of invention assessment may be increased through recursive learning of the classification algorithms.

Today, intuitive decision making continues to dominate the financial sector while the sector simultaneously uses quantitative methods as some of the many tools to conduct thorough assessments. The current review therefore argues that quantitative classification is potentially a way forward as one of the tools applicable for the invention assessment task in the University-Technology Transfer sector.

CHAPTER THREE

PROTOTYPE COMMERCIAL OUTCOME PREDICTION SYSTEM

3.1 INTRODUCTION

This chapter presents the prototype Commercial Outcome Prediction System (prototype CPS), and the classification results achieved by the prototype CPS. Because the prototype CPS is based on the decision tree method, this chapter starts by describing the general classification mechanism of the decision tree method in Section 3.2. The prototype CPS is then presented in Section 3.3. This is followed by the presentation and discussion of the resultant classifier, and its classification results achieved in Section 3.4. Lastly, this chapter ends with Section 3.5, which presents the conclusions derived from the findings.

3.2 DECISION TREE CLASSIFICATION METHOD

Basic mechanism

Given the availability of a set of input data, a tree classifier can be generated by means of Decision Tree Induction. The basic mechanism of generating and testing a tree classifier can be explained using the following three sequential steps:

1. Division of the input data set into the training data and testing data.
2. Generate a tree classifier from the training data.
3. Evaluate the tree classifier using the testing data.

Step One: Data Division

The input data set consists of data about a number of cases, where each case is made up of its attribute data and class data. The first step is to divide the cases of the input data set into the training data set and testing data set using a certain ratio. For instance, if the ratio is 7:3, 70% of the cases will be used as the training data set, and the remaining 30% will be used as the testing data set. The ratio used depends on the size of the input data set. The generation of a tree classifier requires a decent amount of training data. If the input data set consists of a small number of cases, a higher ratio should be used.

Step Two: Classifier Generation using Training Data

The second step is to generate a tree classifier through the application of the Decision Tree Induction algorithm to the training data set. The tree classifier produced would have a flow-chart-like structure, like an inverted tree. The single node at the top is called the root node, and the lines connecting different nodes are called branches. Each node continues to split into two or more nodes until they reach bottom nodes, where the splitting stops. These bottom nodes are called the leaves. Each node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf displays the resultant class label. Figure 3-1 displays a typical tree classifier.

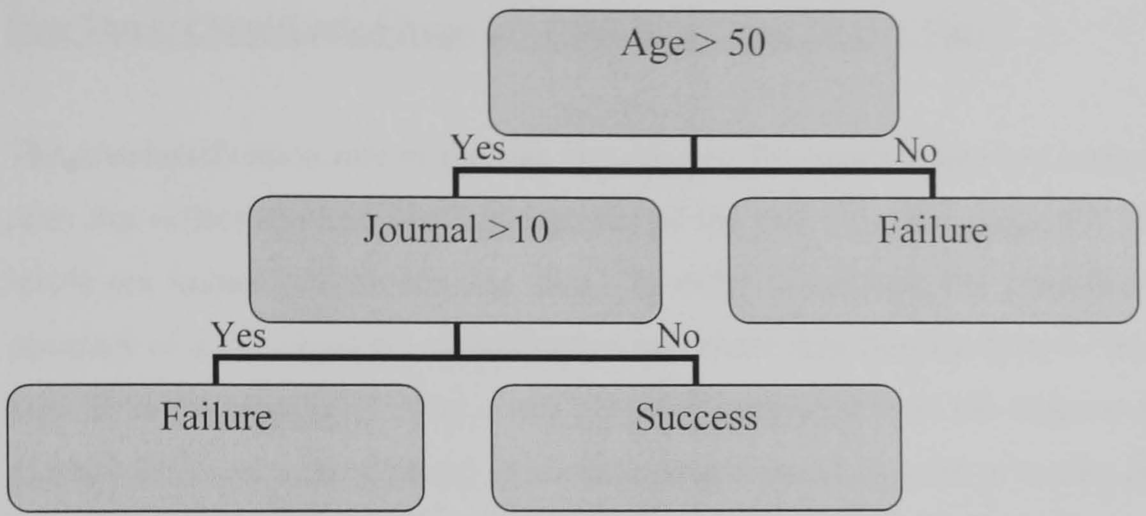


Figure 3-1: A typical decision tree

In order to classify a case with an unknown class label, the attribute values of the case are tested according to the tree classifier. A path is traced from the root to a leaf which holds the class predicted by the tree classifier. Figure 3-1 shows an example of a tree classifier, in which each left descending branch denotes a positive answer to the test, and each right descending branch represents a negative answer. According to this tree classifier, an invention would belong to the class 'success' if the inventor is aged over 50 and had not published more than ten journal papers. Nevertheless, though the tree classifier is generated from the training data set with known class, misclassification can occur even when the tree is used to classify the cases from the training data. This is due to the use of decision tree parameters that determine the resultant appearance of a tree classifier. One of the common parameters is pruning, which is a technique commonly employed to avoid over-fitting. Depending on the level of pruning, the resultant tree classifier can be a full tree or a pruned tree. Generally, a pruned tree contains a smaller number of nodes than the full tree. The pruned tree therefore depicts a more generic classification, where the full tree follows the training data more closely. However, even a full tree may still misclassify the training data, since the training data can contain cases with exactly the same attribute data but different class labels. This often happens when the training data consists of real life data, rather than computer generated data.

Step Three: Classification Accuracy Evaluation Using Testing Data

The misclassification rate of the tree classifier on the training data set, however, does not reflect the true predictive power of the tree classifier, since the class labels are known for the training data. In order to estimate the classification accuracy of a tree classifier, classification on unseen data (testing data) is carried out. It is the process of using a tree classifier generated from the training data (cases with known class labels) to conduct classification on a set of testing data, where the class labels of the testing data is not provided to the tree classifier. The testing data set generated in Step One is used for this step.

The testing data set consists of both the attribute data and class data of a number of cases. In this step, the tree classifier generated from the last step is used to compute the predicted class labels of the cases from the testing data set using only their attribute data. The predicted class label for each case of the testing data set is computed by tracing a path through the tree classifier. The classification error of the tree classifier is then calculated by comparing the predicted class labels and the class data of the testing data set, which contains the actual class labels.

Information Gain Algorithm

Starting from the root node, a decision tree splits the whole data set into subsets from the nodes, where each node in turn further splits its subset into more subsets at lower nodes. This process goes on until a leaf is reached. Splitting only ends at each leaf node, because this is where every case of the current subset belongs to the same class. The node can therefore be represented by one class label, hence a leaf node. Leaf nodes are said to be 'pure' since all the cases within the node share the same class label, whereas non-leaf nodes are 'impure' as they contain cases of a mixture of classes. In other words, a node is 'purer' when it would lead to less splitting. The principle of tree creation is to reach leaf nodes with the minimum of splitting, the attribute with the highest 'purity' is thus chosen as the current node. Instead of formalising 'purity', it is more convenient to formalise 'impurity'. Different trees can be generated using different impurity measures.

The Information Gain algorithm, one of the most common impurity measures, works by choosing the attribute with the greatest impurity reduction (or the highest information gain) as the test attribute for the current node. The calculations of the information gain of every attribute are accomplished by first computing the information (also called the entropy value) required to classify the data set, *e(D)*, and then the entropy values of every attribute. Where *e(X)* denotes the entropy value of a particular attribute, *X*, the encoding information gained using *A* is therefore *e(D) – e(X)*. This algorithm is presented below.

Where *D* consists of *d* cases and the number of classes is *n*. Let d_j be the number of cases of *D* in class c_j. Hence $e(D) = e(d_1, d_2, \ldots, d_n)$. Based on the definition of Information Gain, the entropy of the data set *D*, *e(D)*, is defined in Equation 3-1.

Equation 3-1: Entropy of the data set *D*

$$e(D) = \sum_{j=1}^{n} \hat{P}(c_j) \log_2 \hat{P}(c_j)$$

Where $\hat{P}(c_j)$ is the fraction of cases at node *N* that belong to class c_j. It is actually the expected error rate at node *N* if the class label is selected randomly from the class distribution presented at *N*. *n* denotes the number of different classes. A log function to the base 2 is used since the information is encoded in bits.

Let *X* have *m* values as in $\{x_1, x_2, \cdots, x_m\}$, where *X* can partition *D* into *m* subsets as in $\{D_1, D_2, \cdots, D_m\}$, and D_i contains those cases in *D* that have the value x_h of *X*. Let d_{jh} be the number of cases of class c_j in the subset D_h. The information required to partition the data set into subsets using *X*, *e(X)*, is defined in Equation 3-2.

Equation 3-2: Entropy of the attribute *X*

$$e(X) = \sum_{h=1}^{m} \frac{d_{1h} + \cdots + d_{nh}}{d} e(D)$$

Where $\frac{d_{1h} + \cdots + d_{nh}}{d}$ is the fraction of cases in the subset having value x_h.

The encoding information gained or the impurity drop by using *X* is therefore *e(D) − e(X)*.

Though in principle the leaf nodes possess zero impurity, there is no assurance in practice. As it is possible for real life data to contain two cases having the same value for each attribute, but having different class labels.

3.3 PROTOTYPE INVENTION PREDICTION SYSTEM

This section presents the prototype Commercial Outcome Prediction System (prototype CPS). The prototype CPS is designed to predict the likely class label (measured in future monetary return) of an invention, and is designed to be used by university technology transfer offices for invention assessment purposes.

Briefly speaking, the prototype CPS works by specifying the input data set based on the literature from the University Technology Transfer domain, and then generating a tree classifier using several procedures based on the decision tree method. The procedures involved in the prototype CPS are listed as follows:

1. Specification of the input data set
2. Division of the original input data set into smaller input data sets
3. Leave-one-out sampling for dividing data into the training and testing data
4. Classifier generation based on a decision tree method with adaptive boosting

Each of these procedures is explained sequentially below, followed by an overview description of the prototype CPS, explaining how the 4 procedures are used together to conduct classification.

3.3.1 Specification of the input data set

As previously explained in Section 3.2, an input data set is divided into the training data and testing data, upon which a tree classifier is generated and evaluated respectively.

The first procedure of the prototype CPS is thus to prepare the input data set, which is comprised of an attribute matrix ($\mathbf{X}_1$) and a class vector ($\mathbf{c}_1$), as illustrated in Equation 3-3. The attribute matrix contains the values of a number of (n) invention attributes for a number of (k) historical invention cases, in which the matrix consists of rows of cases and columns of attributes. The class vector contains the class labels of these invention cases.

Equation 3-3: Definitions of the input data set for the prototype CPS

$$\mathbf{X}_1 = \begin{pmatrix} x_{11} & x_{1j} & \cdots & x_{1n} \\ x_{i1} & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{kj} & \cdots & x_{kn} \end{pmatrix} \qquad \mathbf{c}_1 = \left[c_1 \ldots c_i \ldots c_k\right]$$

Where x_{ij} represents the value of the j-th attribute for the i-th case, n is the number of attributes, and k is the number of cases.

The attributes to be recorded in the attribute matrix is comprised of the suggested critical factors derived from relevant literature from the University Technology Transfer domain, and other data available from the source. The reason for using these critical factors as attributes is that these factors are suggested as predictive factors based on the findings from relevant surveys, such as interviews with technology transfer personnel and case studies of invention projects.

Critical factors suggested by the University Technology Transfer literature together with their respective references are listed in Table 3-1. All factors listed in Table 3-1 are suggested to be positive indicators for technology transfer outcome, except for those denoted with '(N)', which are negative indicators.

Table 3-1: List of Critical Factors

1. The discipline of the invention being either biotechnology, informatics or engineering (Mowery et al. 2002; Shane 2001)
2. The experts of the invention's discipline are mostly found in universities (Zucker et al. 1997).
3. The customers of the market for the invention's discipline are willing to trade off cost for efficacy (Shane 2001)
4. Strong patent protection for the invention's discipline (Hsu and Bernstein 1997; Shane 2002)
5. Inventions not needing complementary technologies or assets to be effective (Shane 2001)
6. Young age of the invention's technical field (Shane 2001)
7. The market segments of the invention are not the attention of established firms (Shane 2001)
8. Small average firm size for the industry where the invention belongs (Shane 2001)

9. The industry where the invention belongs does not involve tacit knowledge (Shane 2001)
10. Universities hold more patents than the industry for the invention's disciplines (Hsu and Bernstein 1997; Shane 2001)
11. Historical research leading to the invention was funded by the private sector (Gregorio and Shane 2003)[6]
12. Industrial partners involve in technology transfer activities for the inventions (Owen-Smith and Powell 2003; Lindelöf and Löfsten 2004)
13. Non-embryonic inventions – late stage inventions with low level of unknown regarding their technical feasibility and market applications (Bond and Houston 2003)
14. Inventions derived from basic research, as opposed to applied research (Bond and Houston 2003)
15. Radical inventions (Shane 2001)
16. Inventions with balanced objectives of various stakeholders (Siegel et al. 2004)
17. Inventions accredited by technology assessors (McAdam et al. 2005)
18. Inventions with future funding secured (Gregorio and Shane 2003)
19. Personal relationship exists among the stakeholders (inventors, technology-transfer office administrator, investor) of the invention (Siegel et al. 2004; Shane and Stuart 2002; Hsu and Bernstein 1997)
20. The stakeholders (inventors, technology-transfer office administrator, investor / market) are well connected (Siegel et al. 2004; Bond and Houston 2003)
21. Inventors being the 'entrepreneurial type': people who have been wanting to start their own businesses (Shane and Khurana 2003)
22. The team of inventors includes industrial inventors (Zucker et al. 1997)
23. Inventors with previous firm creation experience (Shane 2004)
24. Inventors with previous experiences in technology-transfer activities, such as collaborative research projects or consultancy with industrial partners (Owen-Smith and Powell 2003)

[6] This factor is only suggested at the university level.

25. Inventions of high values (Hsu and Bernstein 1997)
26. Inventors not having an overly simplified view of business and management issues (McAdam et al. 2005)
27. Inventors not underestimating the time and resources required to create proof of concept (Hsu and Bernstein 1997. McAdam et al. 2005)
28. Inventors not underestimating marketing efforts (McAdam et al. 2005; Shane and Stuart 2002)
29. Inventors of high status (Shane and Khurana 2003)
30. Inventors being star researchers who are experts in their field and have a high volume of publication and some highly significant publications (Zucker et al. 1997; Owen-Smith and Powell 2003)
31. Inventions with higher expenditures on external lawyers (Siegel et al. 2003)
32. Inventions having not too tight connections with firms (Owen-Smith, J., and Powell, W.W., 2003)
33. (N) Inventions lacking inventor-supplied contacts for potential licensees (Hsu, D.H., and Bernstein, T., 1997)
34. (N) Radical inventions being underappreciated by investors (Hsu and Bernstein 1997)
35. (N) Inventions with misunderstandings between inventors and universities on intellectual property rights issues (Siegel et al. 2003, 2004)

Although these critical factors are abstract and subjective, they are used as the guidelines for data collection regarding invention attributes. Lastly, the class label is either 'success' or 'failure', which is measured in terms of the level of monetary return generated by an invention.

Furthermore, the value of each attribute is to be collected at three points in time, which are related to the patent application process, as illustrated in Figure 3-2. Attribute data is to be collected at these three points and class data is to be collected at the end of an invention's life. The three points in time are denoted with reference to the number of month from the first patent filing date. For instance, 'Month 0' is equivalent to the first filing date, as it represents zero month away from the filing date.

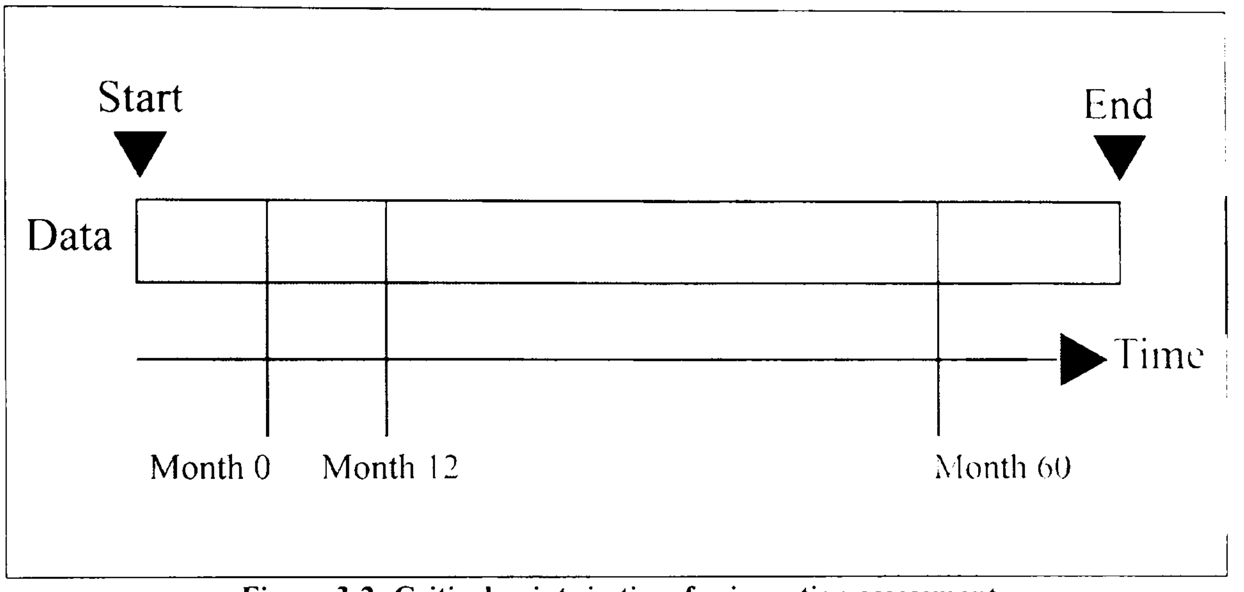


Figure 3-2: Critical points in time for invention assessment

The value of an attribute data collected at 'Month 0' is the value accounting from the start point of an invention to 'Month 0'. The values of data collected at 'Month 12' and 'Month 60' are the marginal values accounting from the previous critical points to the current critical points. For instance, given an attribute 'The amount of expenditure incurred', the attribute value collected at 'Month 0' for a particular invention project would be the expenditure amount during the period from the start of the project to 'Month 0'. Similarly, the attribute value collected at 'Month 12' is the marginal expenditure value during the period from 'Month 0' to 'Month 12'.

Data is collected at the three points in time as they are the times when the potential of an invention project is often assessed to decide whether or not to end the project. The respective decisions are explained below, which are in relation to the typical issues considered during the patent application process.

Month 0

Data collected at 'Month 0' is based on the period between the start of an invention project till its first patent filing day. This is the period just before any patent application has taken place. During this period, inventors normally provide the details of the invention to the technology-transfer office, generally through filling out an invention disclosure form. Necessary documents regarding technical specification of the invention and due diligence are also submitted. This is normally followed by an assessment based on the information provided to see if it is worth filing a United Kingdom patent application for the invention. This decision at 'Month 0' is crucial as it is the first gatekeeper that determines whether or not the invention would enter the whole technology-transfer process at all.

Month 12

'Month 12' is one year away from the first patent filing date. If an informal patent application had been submitted at 'Month 0', 'Month 12' is then the deadline to submit a formalised United Kingdom patent application. The difference between an informal and formal patent application is that the specification of the latter is written according to the legal format required by the patent offices. There is, however, no specific format for the specification for an informal application, which can be merely a general description of the invention. Thus, professional consultation is often sought for a formal patent application at 'Month 12'. Moreover, 'Month 12' is also the deadline to submit the PCT (Patent Cooperation Treaty) application, which is a preliminary international examination for the convenience of qualifying an invention for up to 108 countries. PCT is usually as costly as three times or more of the expenses required in a formal United Kingdom patent application. Yet, inventions seeking international patent protection often apply for PCT, for various reasons such as delaying patent applications for individual countries. Due to the high cost of PCT, it is usually only adopted when foreign patent applications are pursued.

Similarly, foreign patent applications are often expensive, especially when the number of designated country increases. The decision at 'Month 12' therefore involves high costs as well as crucial international strategies for technology transfer.

Month 60

There is no guarantee as to how long it may take to have a patent granted. It could be five or more years for a foreign patent application, or less for a United Kingdom patent application. For simplicity, the final data collection point is allocated at 'Month 60', which is five years away from the first filing date. At this point, usually at least one of the designated countries would have granted a patent. After the grants, annual renewal fees would need to be paid to each country in order to keep the patents alive. Fees will only be paid to maintain the patents if there is commercial interests. 'Month 60' is therefore another important point in time to decide whether or not to terminate the invention.

3.3.2 Partition of the original input data set into smaller input data sets

This procedure divides the original input data set into three smaller input data sets, based on attribute data collected at different points in time. This is explained below.

While the attribute matrix ($\mathbf{X}_1$) contains attributes collected at three points in time, attributes collected at a later point (such as at 'Month 60') are so-called 'older attributes', and attributes collected at an earlier point are so-called 'younger attributes'. For instance, given the attribute 'the number of publications created by the inventor' collected at the three points in time, three attributes are developed. The attribute collected at 'Month 0' is so-called a 'zero-year old' attribute, and is 'younger' than the attributes collected at 'Month 12' or 'Month 60'.

The decision tree method generates a tree classifier consisting of m attributes (m nodes), from a training data with n attributes, where m is a subset of n. If the n attributes are all 'one-year old' attributes, the m attributes would also be 'one-year old' attributes. The tree classifier generated from these n attributes can then perform classification once the future cases are at least one-year old. However, if those n attributes are 'two-year old' attributes, then the tree classifier can only be used until the future cases are at least two years old. Since the training data is a subset of cases of the input data, the training data and input data share the same n attributes. In other words, an input data set consisting of younger attributes enables earlier classification which requires younger attributes. Since the amount of resources consumed by inventions increases with time, it is better to perform classification (predict the class labels of inventions) earlier. In short, younger attributes are preferred.

Furthermore, when the number of attributes required by the input data set is smaller, data collection is likely to be easier for each case. When data collection is easier for each case, the data of a larger number of cases may be collected. An input data set with a larger number of cases enables a training data set with a larger number of cases, which in turn gives rise to a more accurate classifier. Since a classifier generated from a larger training data set is generally more robust. In short, it is important to minimise the number of attributes used for an input data set.

In order to minimise the age and the number of the attributes used for the input data set, this procedure divides the attribute matrix of the original input data set into various smaller attribute matrices according to data collection points in time. Each of the smaller matrices is then teamed up with the class vector of the original input data set to form a smaller input data set. These smaller input data sets would then be used to generate various classifiers. However, it is possible that a classifier based on the original input data set (which contains the attribute matrix covering attributes from all three data collection points) gives the highest classification accuracy. Therefore, the original input data set and the three reduced input data sets are all used to generate classifiers.

Let n be the number of attributes contained in the original attribute matrix, and p, q, r be the numbers of attribute collected at the three data collection points respectively. The three smaller attribute matrices therefore contain p, q, and r attributes respectively, where they contain the same number of cases. This division process of the original attribute matrix ($\mathbf{X}_1$) is illustrated in Equation 3-4.

Equation 3-4: Partition of the input data set

$$\mathbf{X}_1 = \begin{pmatrix} x_{11} & x_{1j} & \cdots & x_{1n} \\ x_{i1} & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{kj} & \cdots & x_{kn} \end{pmatrix}$$

$$\mathbf{X}_1 = \left(\begin{pmatrix} x_{11} & \cdots & x_{1b} \\ x_{i1} & \cdots & x_{ib} \\ \vdots & & \vdots \\ x_{k1} & \cdots & x_{kb} \end{pmatrix} \begin{pmatrix} x_{1c} & \cdots & x_{1d} \\ x_{ic} & \cdots & x_{id} \\ \vdots & & \vdots \\ x_{kc} & \cdots & x_{kd} \end{pmatrix} \begin{pmatrix} x_{1e} & \cdots & x_{1n} \\ x_{ie} & \cdots & x_{in} \\ \vdots & & \vdots \\ x_{ke} & \cdots & x_{kn} \end{pmatrix} \right)$$

$$\mathbf{X}_{1_{month0}} = \begin{pmatrix} x_{11} & \cdots & x_{1b} \\ x_{i1} & \cdots & x_{ib} \\ \vdots & & \vdots \\ x_{k1} & \cdots & x_{kb} \end{pmatrix} \qquad \mathbf{X}_{1_{month12}} = \begin{pmatrix} x_{1c} & \cdots & x_{1d} \\ x_{ic} & \cdots & x_{id} \\ \vdots & & \vdots \\ x_{kc} & \cdots & x_{kd} \end{pmatrix}$$

$$\mathbf{X}_{1_{month60}} = \begin{pmatrix} x_{1e} & \cdots & x_{1n} \\ x_{ie} & \cdots & x_{in} \\ \vdots & & \vdots \\ x_{ke} & \cdots & x_{kn} \end{pmatrix}$$

Where $\mathbf{X}_1$ denotes the attribute matrix of the original input data set. $\mathbf{X}_{1_{month0}}$, $\mathbf{X}_{1_{month12}}$, and $\mathbf{X}_{1_{month60}}$ denote the reduced attribute matrices extracted from $\mathbf{X}_1$. Where $b = p$, $c = p + 1$, $d = q + p$, $e = q + 1$, $n = p + q + r$.

Equation 3-4 illustrates the division of the attribute matrix $\mathbf{X}_1$ into three smaller matrices. The first p attributes of every case in $\mathbf{X}_1$ form $\mathbf{X}_{1_{month0}}$. The c-th to the d-th attributes of every case in $\mathbf{X}_1$ form $\mathbf{X}_{1_{month12}}$. Lastly, the e-th to the n-th attributes of every case in $\mathbf{X}_1$ form $\mathbf{X}_{1_{month60}}$. The four input data sets generated in this procedure are therefore $\mathbf{X}_1$ and $\mathbf{c}_1$, $\mathbf{X}_{1_{month0}}$ and $\mathbf{c}_1$, $\mathbf{X}_{1_{month12}}$ and $\mathbf{c}_1$, and $\mathbf{X}_{1_{month60}}$ and $\mathbf{c}_1$.

3.3.3 Leave-one-out sampling

As previously described in Section 3.2, the cases of the input data set have to be divided into the training and testing data, upon which a classifier tree is generated and evaluated respectively. Such division of cases is usually done according to a certain ratio, such as 7:3, where 70% of the cases are used for training and the remainder is used for testing. In this procedure, no fixed ratio is adopted unless the input data set contains a very limited number of cases, with less than 50 cases. When the input data set contains less than 50 cases, leave-one-out sampling technique is employed to divide the input data set into the training and testing data. The reason for not using a fixed ratio is because the ratio depends on the size (i.e. the number of cases) of the input data set. A ratio is valid as long as there are sufficient cases for training (classifier generation). A large input data set (with thousands of cases) therefore can afford to adopt a more even ratio such as 5:5. A medium input data set (with hundreds of cases) probably uses the prevalent ratios where a higher proportion is reserved for the training data set, such as 8:2 or 7:3.

By leave-one-out sampling, one case from the input data set with *n* cases is used as the testing data, while the remaining *n*-1 cases are used as the training data. A tree classifier generated using this training data set is then used to conduct classification on the testing data set. The classification accuracy is then recorded. This process is iterated *n* times so that each case of the input data set takes turn to be the testing data set. *n* tree classifiers and *n* records of classification accuracy are therefore resulted. The average of the *n* classification accuracies then represents an estimate of the classification accuracy of a tree classifier generated using all *n* cases of the input data set. The derivation of the classification accuracy using the leave-one-out sampling technique is therefore different from the usual way. While the usual way is to derive the accuracy using a subset of the input data as the testing data, the leave-one-out method generates *n* accuracies of *n* tree classifiers, and uses their average to estimate the accuracy of a tree classifier that was not generated. Nonetheless, this method preserves maximum number of cases for the training data set, which is especially important when the size (i.e. number of cases) of the input data set is very limited.

3.3.4 Classifier generation using decision tree with adaptive boosting

Due to the qualitative nature of most of the critical factors, the resultant attributes collected are likely to be comprised of categorical values. Most classification algorithms, such as the nearest-neighbour classifier and regression analysis, are based on measurements of distance between sample cases. These methods work when variables are real-valued and can be defined by metric. However, qualitative attributes such as 'source of funding leading to the invention', are comprised of categorical values like 'Government' or 'Charity'. Such categorical data is discrete, and lacks any natural notion of similarity or ordering. Therefore, the decision tree method is adopted to generate classifiers, as it is capable to analyse both numeric and categorical data.

Based on different settings of several parameters, different tree classifiers can result from the same training data. Five decision tree parameters are employed for classifier generation, they are listed as follows:

1. The branch factor – two
2. The number of attributes represented by each node – monothetic trees
3. The impurity measure employed – Information Gain
4. Over-fitting measure: Minimum Case Threshold
5. The number of trial adopted for the adaptive boosting technique

The values for the first three parameters have been pre-determined, which are the branching factor of 2, monothetic trees, and the Information Gain impurity measure. For the remaining two parameters, guided ranges of values are used. Each of the five parameters is presented below.

Branching Factor of Two

The number of branches descending from each node of a tree classifier is called the node's *branching factor* or *branching ratio*. In general, this can be specified and can vary throughout the tree, though every node can always be represented with a branching factor of 2 (binary tree). For instance, Figure 3-3 and 3-4 show how the same information can be expressed in a non-binary and binary tree respectively. Due to the fact that binary trees are expressive, and are comparatively simpler to train, binary trees are adopted.

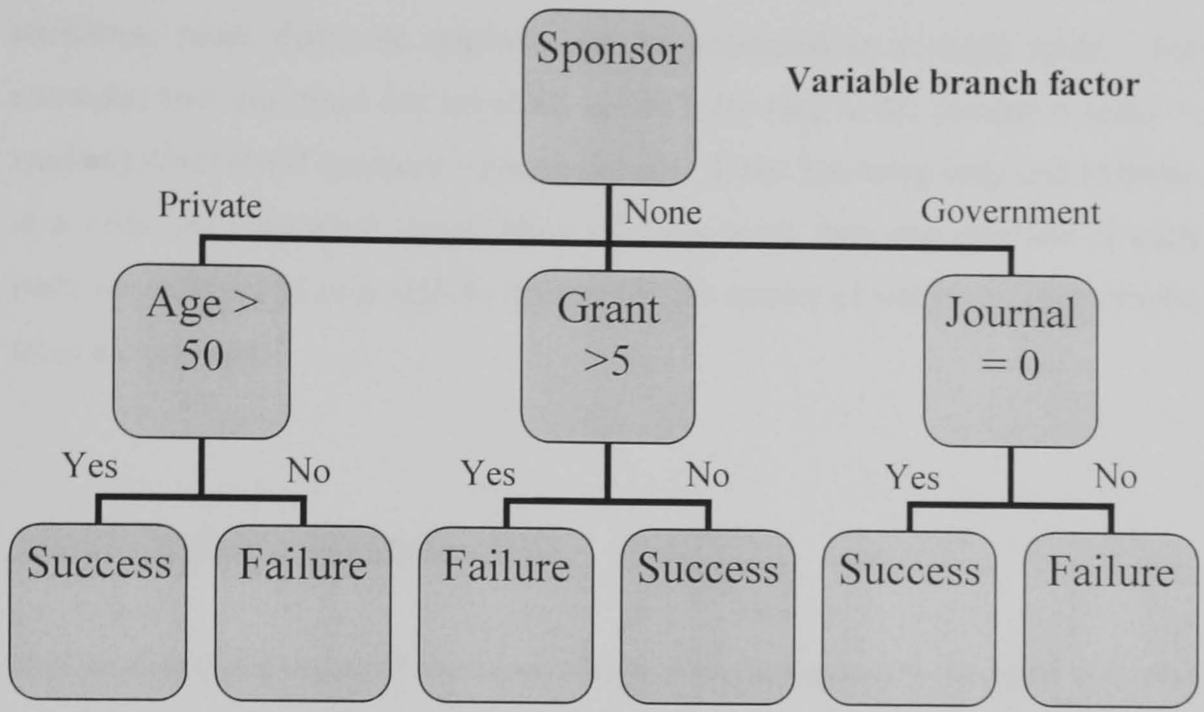


Figure 3-3: A decision tree with variable branch factor

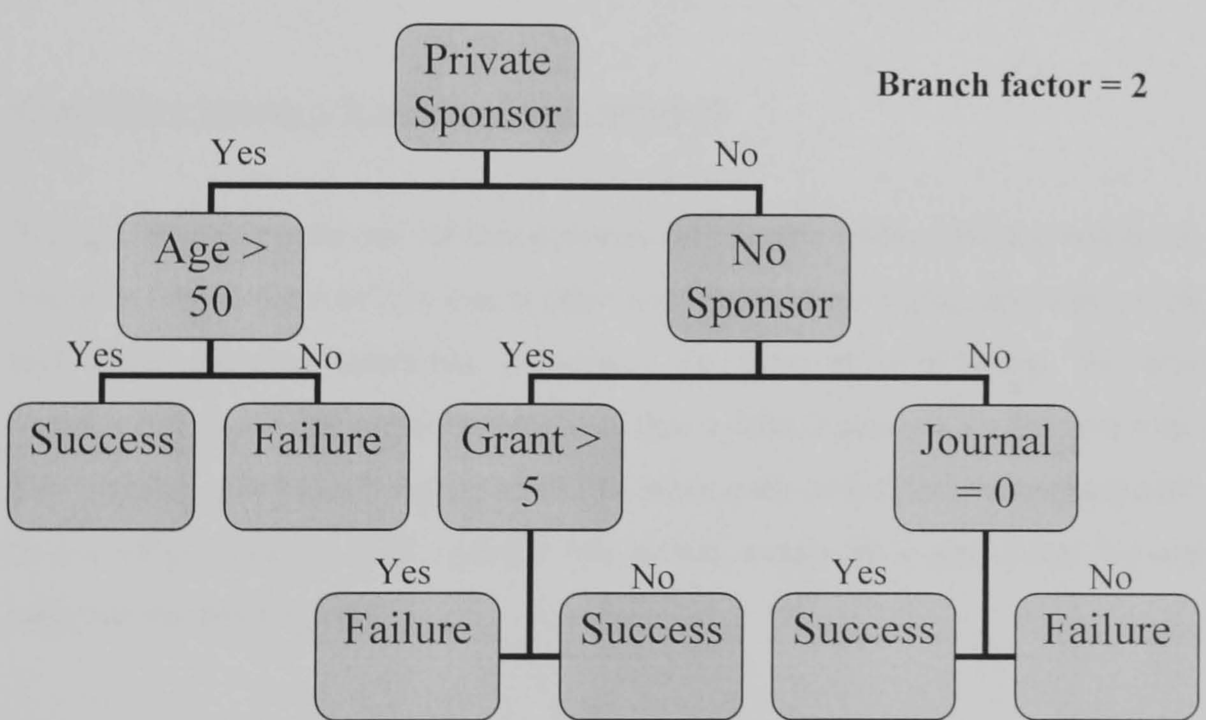


Figure 3-4: A decision tree with branch factor = 2

Number of attributes represented by each node – Monothetic Trees

While it seems a norm for each node to represent one attribute, it is possible to involve more than one attribute at each node. By using logical combinations of attributes, more than one attribute can be presented at a single node. For example, two attributes are involved in the following node: (*inventor status* = *student*) AND (NOT (*subject* = *engineering*)). Trees involving only one attribute at a node are monothetic trees, trees allowing more than one attribute at each node are referred to as polythetic trees. For the reason of simplicity, monothetic trees are adopted.

Impurity measure – Information Gain

Information Gain is one of the most widely accepted impurity measure, it is also a natural default in most classification problems. It is therefore adopted here for the impurity measure.

Over-fitting Measure: Minimum Case Threshold

Though the basic principal for tree generation is to split nodes until the minimum impurity is met, the resultant tree is often over-fitting the training data when each leaf node contains minimum impurity. In case of over-fitting, the tree corresponds so closely to the training data that it cannot generalise new data well. The extreme case of over-fitting would be when each leaf of the tree corresponds to one single training case. Such a tree would merely be a convenient lookup table for the training data.

To avoid over-fitting, nodes have to stop splitting before reaching the minimum impurity. In other words, the splitting has to be stopped before the tree over-fits. However, if the splitting is stopped too early, the tree may under-fit the training data. This is when the classification error on the training data is not sufficiently low, leading to poor generalisation performance.

A parameter used to identify levels of node splitting is called Minimum Case Threshold (MCT), by means of which splitting stops when a node represents fewer than the threshold number of cases. This has a benefit analogous to that in a *k*-nearest neighbour classifier, that the sizes of partitions are smaller in regions where data is denser, but larger where the data is sparser.

If a tree splits until the minimum impurity is met, it is called a full tree. When the splitting is stopped before the minimum impurity, the resultant tree is a sub-tree of the full tree. A full tree will be generated when the MCT value is set to be 2. Changing the MCT may or may not lead to a different sub-tree. It is possible that the tree generated by setting MCT to 4 is equivalent to that generated by setting MCT to ten. Generally, there would be more sub-trees if the full tree contains a larger number of nodes. When the MCT value is too high, the tree does not split and hence only consists of one node (a leaf node), suggesting that all cases should be classified to the same class. In such case, the one node tree is definitely under-fitting the training data and offering no classification insights.

In order to prevent over-fitting, instead of generating only one tree classifier from the training data set, this procedure requires a number of tree classifiers to be generated using a range of MCT values, which starts from an MCT value of 2 (i.e. a full tree) until a single-node tree is resulted.

Adaptive boosting technique

Adaptive boosting is a technique adopted to minimise classification error, where the final classification is based on generating several trees classifiers rather than one tree classifier from the input data. The process of classification system generation using the adaptive boosting technique is described below.

First of all, the number of trials of tree generation is pre-determined. A tree classifier is then generated form the training data set, where each of the subsequent tree classifiers is generated by focusing on the mistakes made by the previous tree. Tree classifier creation goes on until the pre-determined number of trials is completed. This process also stops when the current tree classifier generated is either extremely accurate or inaccurate. For instance, when it is specified to conduct 3 trials of tree generation, the procedure to produce a classification system is as follows. A tree classifier is first generated from the training data set. Next, a second tree classifier is constructed by paying more attention to those cases misclassified by the first tree classifier. Again, the misclassification error of the second decision tree will then become the focus when generating the third decision tree. Lastly, a classification system is resulted, whose computation of the predicted class of a case is determined by the votes from the three tree classifiers.

It is found for many empirical problems that the classification error often decreases when the number of boosting trial is increased up to the optimal level. While each classification problem often has its unique optimal level, the level of about ten trials is commonly found to be working reasonably well for many large and small classification problems.

In order locate the optimal trial number for boosting; this procedure requires the experimentation with a range of five trial numbers: 3, 6, 9, 12, and 15. These numbers are evenly distributed around the number of ten. They are adopted since the level of ten is empirically proven to be approximately the optimal trial number. Though it is possible that the optimal level exist between 0 and n (where n is the number of cases contained by the input data set), a huge number of experimentation would be resulted if every number between 0 and n is to be tested and when the value of n is large. Therefore, only five trial numbers are experimented.

3.3.5 Overview description of the prototype CPS

This section presents an overview description to explain how the four procedures presented above work together to form the prototype CPS to perform classification. The prototype CPS is explained using the following six steps.

1. The first step is to generate an input data set according to the specifications presented in Section 3.3.1,

2. The second step is to generate four input data sets using the procedure described in Section 3.3.2. The third and fourth steps are then applied to each of the four input data sets.

3. The third step is to state the parameter values for *m* candidate classifiers. The parameter variables are namely, the MCT number, and the number of trials used for adaptive boosting.

4. The fourth step is to compute the estimated classification accuracy of each of the *m* candidate classifiers, using the procedures described in Section 3.3.3 and 3.3.4.

5. The fifth step is to choose the best performing candidate classifier. Among the candidate classifiers generated, the candidate classifier achieving the highest estimated accuracy is regarded as the best performing candidate classifier. However, it is possible to have candidate classifiers generated based on different input data sets sharing the same level of estimated accuracy. Under such a situation, the candidate classifier generated using the input data set with the youngest attributes are preferred. The best performing candidate classifier is therefore the classifier with the highest estimated accuracy and generated using the youngest attributes. Candidate classifier based on younger attributes is preferred as it enables prediction based on merely attributes collected at earlier points in time. For instance, if 2 candidate classifiers generated using attributes collected at 'Month 0' and 'Month 12' achieved the same level of estimated accuracy, the candidate classifier based on 'Month 0' is preferred. This is because the 'Month 0' candidate classifier can predict the class label of an invention once the invention is one year old, while the 'Month 12' candidate classifier can only perform predictions until inventions are two years old. Furthermore, when several candidate

classifiers using the same input data set obtained the same highest estimated accuracies, the best performing candidate classifier is therefore the one generated using the smallest number of trials for boosting and/ MCT value.

6. The final step is to generate a classifier using the parameter values of the best performing candidate classifier identified in the last step, using every case of the input data set. This resultant classifier is then used to perform classification.

3.4 RESULTS AND DISCUSSION

Data Collection

One source for data collection is secured, after sending data collection requests to 7 university technology transfer offices. The source provided limited access to the information base for 23 of their invention cases, based on a specific confidentiality agreement.

The information base provided consists of hard copies of patent related documents of 23 invention projects during their life time. The record for each project starts when it was considered to have potential, which usually began from the preparation of the invention's specification for the first patent application. The record ends as soon as the invention was considered to have no potential, which can be before or after the patent grant of the invention.

The data available from the information base is very different from the data required by the prototype CPS. Based on the available data, only a subset of the critical factors listed in Table 3-1 has found analogous invention attributes for data collection. These invention attributes are presented Table 3-2, where the second column shows the data type of the attribute. Categorical values for nominal and ordinal data are also shown in the second column, and the values are presented in descending order for ordinal data. Finally, the last column shows the index number of the respective critical factors suggested by Table 3-1.

Table 3-2: Invention Attributes Collected

Invention Attributes extracted	Coding categories	Factors
1. The first source of funding leading to the current development of the invention	Nominal: Firm (F), Government (G), Charity (C), Spinout (S), Other university (O), Department (D), Technology-transfer office (T), Not applicable (N)	11
2. The second source of funding leading to the current development of the invention	Same as the first attribute	11
3. The third source of funding leading to the current development of the invention	Same as the first attribute	11
4. The forth source of funding leading to the current development of the invention	Same as the first attribute	11
5. The area of discipline of the invention	Nominal: Life science (L), Engineering (E), Informatics (I)	1
6. The number of collaboration partner	Scale	-
7. Whether any prior art searches has been conducted	Ordinal: Professional (P), Non-professional (NP), None (N)	17

8. Whether funding for next stage has been secured	Ordinal: Yes (Y), Partial (P), No (N)	18
9. The number of personal contacts for potential source of investment	Scale	33, 19
10. The number of related publication published by the inventors	Scale	29, 30
11. The number of significant publication published by the inventors	Scale	29, 30
12. The developmental stage of the invention	Ordinal: Proven design made using manufacturing tooling (A), Proven final design (B), Optimisation by prototype variation (C), Prototype (D), Proof of concept (E), Early stage (F)	13
13. Previous business experiences of the inventors	Ordinal: Setting up own business (B), Substantial collaboration (SC), Little collaboration (LC), None (N)	23
14. Whether there are industrial partners that were involved in decision making processes	Ordinal: Yes (Y), No (N)	12, 16
15. The level of communication between the technology transfer office (TTO) and the inventors	Scale	20

16. The level of communication between the TTO/inventors and the market/potential investors	Scale	20
17. Whether an estimation for future costing has been done	Ordinal: Thorough (T), Brief (B), None (N)	26
18. Whether a market research has been done	Ordinal: Professional (A), Non-profession thorough (B), Non-professional brief (C), None (N)	28
19. The number of firms that have shown a positive interest of the invention	Scale	-
20. The chosen type of the first patent application	Nominal: Formal (F), Informal (I)	-
21. The number of other patent applications owned by the inventors	Scale	24
22. Whether there are industrial inventors	Ordinal: Yes (Y), No (N)	22
23. Whether an invention disclosure form has been filled in	Nominal: Yes (Y), No (N)	-
24. The amount of expenditure incurred	Scale	31
25. The number of agreements signed	Scale	-
26. The number of licensing contracts signed	Scale	-
27. Whether a spinout company has been formed	Nominal: Yes (Y), No (N)	-
28. The year of the first filing date	Scale	-
29. The number of months took to have the first patent granted	Scale	-

30. The rate of patent grant	Scale	-
Income	**Ordinal:** **Success (S), Failure (F)**	-

As shown in Table 3-2, 30 invention attributes are extracted from the information base, where the last column indicates that most of them are generated based on 18 out of the 35 critical factors listed in Table 3-1. The data required by the remaining 17 critical factors cannot be collected as it is not available from the information base provided. The attributes not representing any critical factors are other attribute data available from the information base.

Whether an attribute is represented using ordinal or nominal data type depends on the suggestion from the critical factor. For instance, the 8th attribute in Table 3-2 is based on the 18th critical factor in Table 3-1. Since the 18th critical factor suggests that inventions with future funding secured are more likely to have successful technology-transfer outcome, the 8th attribute is therefore coded using ordinal data. Another example is the 5th attribute that is based on the 1st critical factor. This critical factor suggests that inventions from the three disciplines are all likely to have more successful technology-transfer outcomes than inventions from other disciplines. Yet the critical factor does not suggest what is the superiority order of the three disciplines. The 5th attribute is therefore coded using nominal data. Lastly, attributes like the 23rd attribute is not based on any critical factors. Such attributes are therefore coded using nominal data as there is no suggestion regarding the superiority order of the categorical values.

The last row of Table 3-2 contains the data 'Income', which is used to compute the class labels. Due to the issue of confidentiality, the values of this data cannot be disclosed. The values of this data are represented using two class labels, namely 'success' and 'failure'. The range of income values for each class is determined by a staff member from the data source.

Data Collection at Critical Points

The attribute data is collected at the three points in time namely 'Month 0', 'Month 12', and 'Month 60'. However, not every attribute has data collected at all three data collection points. Attributes such as the 20th and 23rd attributes in Table 3-2 only require data collection at 'Month 0'. The attribute with data collected in a particular point in time is marked 'X', as displayed in Table 3-3, where the first and the fifth columns denote the attribute number.

Table 3-3: Results of data collection at the three points in time

Attribute	Month 0	Month 12	Month 60	Attribute	Month 0	Month 12	Month 60
1	X	X	X	16	X	X	X
2	X	X	X	17	X	X	X
3	X	X	X	18	X	X	X
4	X	X	X	19	X	X	X
5	X	-	-	20	X	-	-
6	X	X	X	21	X	X	X
7	X	-	-	22	X	X	X
8	X	X	X	23	X	-	-
9	X	X	X	24	X	X	X
10	X	X	X	25	X	X	X
11	X	X	X	26	X	X	X
12	X	X	X	27	-	X	X
13	X	-	-	28	X	-	-
14	X	X	X	29	-	-	X
15	X	X	X	30	-	-	X

Moreover, Table 3-2 only denotes brief descriptions for each of the thirty invention attributes. Further explanations are needed for some of the attributes, which are presented successively below.

First to fourth attributes: While explicit statements of funding sources are usually missing, sources of funding are often traced from the receipts of paid bills. The four attributes are defined in terms of nominal data rather than ordinal data because the superiority of different funding source is not clear from the relevant (13^{th}) critical factor. Also, the data is categorical rather than numeric because the total amount of expenses consumed is not provided by the data source.

For example using imaginary invention project A, the data collected for these four attributes at the three critical time points is displayed in Table 3-4. As shown, the sources of funding have evolved over the period from the start of the project till 'Month 60'. There are two sources of funding (Government and Charity) at 'Month 0', and there is one source of funding, Government and technology-transfer office respectively, for 'Month 12' and 'Month 60'.

Table 3-4 also serves as a good example showing how attribute data is collected at the three data collection points.

Table 3-4: Data collection example for project A

Critical Points	Month 0				Month 12				Month 60			
Attributes	1	2	3	4	1	2	3	4	1	2	3	4
Data Value	G	C	N	N	G	N	N	N	T	N	N	N

Where 'G', 'C', 'N' and 'T' represent 'Government', 'Charity', 'Not applicable' and 'Technology-transfer office' respectively.

Fifth attribute: There is no definition of disciplines from the data source. Constant debates regarding disciplinary definitions go on even among the literature of research communities. Due to the interdisciplinary nature of most academic departments as well as their resultant invention projects, the same invention project can often be addressed using different disciplinary titles.

Furthermore, consultation with experts regarding the disciplinary issue is restricted due to the confidentiality of the data. The disciplinary labels given to the invention projects are thus based on a subjective judgement of the author.

Sixth attribute: The number of collaboration partners refers to the number of organisations where any relevant stakeholders belong to. The relevant stakeholders have rights to be involved in the technology-transfer decision making at varying degrees. They include inventors, material sponsors, investors, licensors, and spinout company management personnel. For instance, when the relevant stakeholders only involves one inventor and one investor who both come from the same organisation, the data for this attribute is therefore one.

Seventh attribute: These prior art searches refer to those conducted before the filing of patent application. The value of the attribute is 'professional' when these searches are done by professional intellectual property related personnel such as patent attorneys. The value is 'non-professional' when the searches are done by others such as administrators from technology-transfer offices or inventors. When both types of searches have been conducted, the value of this attribute is 'professional'.

Eighth attribute: Information about the total funding required for the next stage, so does other future planning information, is often missing from the data source. Unless a complete future funding is explicitly stated, any evidence of securing future funding sources merely represents the secure of partial future funding. In addition, this attribute refers to the secure of funding source in advance of the next stage. Evidence from email correspondence showed that internal funds from the university, such as from academic departments or the technology-transfer office, sometimes act like urgent sources of funding to settle bills at the last minute. Under such circumstances, the internal funding is not regarded as an advance funding source.

Ninth attribute: A contact for potential source of investment is accounted as a personal contact when it did not previously exist in the information base and is introduced by inventors or administrators from the technology-transfer office. This assumption has been made as there is no record regarding whether contacts are derived from personal sources or not.

Tenth attribute: A publication's relation to a specific invention is obvious to the inventors but not to others. Also, the author would need to familiarise with the technologies involved in the invention in order to judge each publication's relation to the invention, which is a very time-consuming process. Based on these considerations, only publications submitted by the inventors as relevant to the inventions are accounted as related publications.

Eleventh attribute: The significance of a publication is a subjective judgement since one definition may be accepted by one but not the others. For instance, publication citation is commonly viewed as a significance measure among the academics. This is however not agreed by the private sectors. Under the circumstance of lacking a commonly agreed measure, only publications receiving a third party approval are accounted as significant publications, such as awarded publications and granted patent.

Twelfth attribute: The definition for the developmental stages of an invention depends on its industry standard. The definition adopted for this attribute, as reflected in the ordinal data definition, is based on the industry standard from the engineering industry. This is due to the fact that most invention projects from the data source are engineering based.

Thirteenth attribute: 'Substantial collaboration' and 'little collaboration' refer to the situations when the numbers of past collaboration projects are three or more, and below three, respectively.

Fifteenth and sixteenth attributes: The level of communication is measured in terms of the number of pages of written correspondence such as emails and letters.

Seventeenth attribute: Estimation for future costing can be written in various degrees of details. For example, it can be represented in one numeric amount or through sections of pricing for individual items. The attribute value is based on the subjective judgement by the author.

Eighteenth attribute: Whether a research report is thorough or brief is determined by the author's judgement, based on the same argument stated for the 17^{th} attribute. Moreover, a professional market research refers to that completed by professional external organisations such as a marketing consultancy, whereas a non-professional one is that completed by inventors or technology-transfer administrators.

Twentieth attribute: this attribute refers to the priority patent application. The difference between a formal and an informal patent application is that the specification of a formal application is written in the format required by the UK patent office for examination purpose.

Thirtieth attribute: this rate is calculated by the number of patents granted divided by the number of years between the filing date of the first patent application and the grant date of the last application.

Resultant Input Data set

By collecting attribute data at the three data collection points, and the class data at the end of each case, an input data set is generated. This input data set contains the data for 23 cases, in which 7 and 16 cases are labelled as 'Success' and 'Failure' respectively. In addition, each 'X' in Table 3-3 denotes one attribute, the total number of attribute contained in the attribute matrix is therefore 73. The attribute data is stored in matrix $\mathbf{X}_1$, and the class data is stored in vector $\mathbf{c}_1$, as illustrated in Equation 3-5.

Equation 3-5: The resultant input data set stored in matrix $\mathbf{X}$ and vector $\mathbf{c}_1$

$$\mathbf{X}_1 = \begin{pmatrix} x_{11} & x_{1j} & \cdots & x_{1n} \\ x_{i1} & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{kj} & \cdots & x_{kn} \end{pmatrix} \qquad \mathbf{c}_1 = [c_1 \ldots c_i \ldots c_k]$$

Where x_{ij} represents the value of the j-th attribute for the i-th case. n is the number of attributes, and k is the number of cases. For this input data set, n is 73 and k is 23.

Generation of four input data sets

Four input data sets are then generated from the original input data set according to the procedure described in Section 3.3.2. Out of the 73 attributes in matrix $\mathbf{X}_1$, the first 27 attributes were collected at 'Month 0', the following 22 attributes were collected at 'Month 12', and the last 24 attributes were collected at 'Month 60'. The three resultant smaller attributes matrices are denoted as $\mathbf{X}_{1_{month0}}$, $\mathbf{X}_{1_{month12}}$, and $\mathbf{X}_{1_{month60}}$, which they consist of 27, 22, and 24 attributes respectively for 23 cases. The four attribute matrices are therefore $\mathbf{X}_1$, $\mathbf{X}_{1_{month0}}$, $\mathbf{X}_{1_{month12}}$, and $\mathbf{X}_{1_{month60}}$.

By combining each attribute matrix with the class vector ($\mathbf{c}_1$), four input data sets are generated, as shown in Table 3-5.

Table 3-5: Four input data sets generated using the original input data set

Input Data set	Attribute matrix	Class vector
D_1	$\mathbf{X}_1$	$\mathbf{c}_1$
D_2	$\mathbf{X}_{1_{month0}}$	$\mathbf{c}_1$
D_3	$\mathbf{X}_{1_{month12}}$	$\mathbf{c}_1$
D_4	$\mathbf{X}_{1_{month60}}$	$\mathbf{c}_1$

The first column of Table 3-5 showed the notations for each input data set, where the second and third columns showed their respective attribute matrices and class vectors.

Classification accuracies of the resultant candidate classifiers

The third step of the prototype CPS is to define each candidate classifier in terms of the two parameters: MCT (minimum case threshold) and number of trials for adaptive boosting. The fourth step of the prototype CPS is to calculate the estimated classification accuracy of each candidate classifier. Due to the limited number of cases contained by the input data set (23 cases), leave-one-out sampling technique is employed. The results generated from step three and four, which are the definitions of each candidate classifier as well as their estimated classification accuracy (in terms of the percentage of cases correctly classified), are displayed in the following four tables. Each table shows the results for a particular input data set.

Table 3-6: Results generated from the first input data set D_1

Classifier	Boost trial	MCT	Accuracy	Classifier	Boost trial	MCT	Accuracy
1	3	2	65.2	9	12	3	60.9
2	6	2	47.8	10	15	3	56.5
3	9	2	60.9	11	3	4	73.9
4	12	2	52.2	12	6	4	69.6
5	15	2	52.2	13	9	4	73.9
6	3	3	60.9	14	12	4	73.9
7	6	3	60.9	15	15	4	73.9
8	9	3	60.9	16	3	5	-

Table 3-7: Results generated from the second input data set D_2

Classifier	Boost trial	MCT	Accuracy	Classifier	Boost trial	MCT	Accuracy
1	3	2	60.9	9	12	3	73.9
2	6	2	60.9	10	15	3	78.3
3	9	2	60.9	11	3	4	60.9
4	12	2	65.2	12	6	4	60.9
5	15	2	65.2	13	9	4	60.9
6	3	3	73.9	14	12	4	60.9
7	6	3	73.9	15	15	4	60.9
8	9	3	73.9	16	3	5	-

Table 3-8: Results generated from the third input data set D_3

Classifier	Boost trial	MCT	Accuracy	Classifier	Boost trial	MCT	Accuracy
1	3	2	78.3	9	12	3	69.6
2	6	2	78.3	10	15	3	69.6
3	9	2	73.9	11	3	4	69.6
4	12	2	73.9	12	6	4	73.9
5	15	2	73.9	13	9	4	69.6
6	3	3	73.9	14	12	4	73.9
7	6	3	73.9	15	15	4	69.6
8	9	3	69.6	16	3	5	-

Table 3-9: The results generated from the fourth input data set D_4

Classifier	Boost trial	MCT	Accuracy	Classifier	Boost trial	MCT	Accuracy
1	3	2	65.2	9	12	3	56.5
2	6	2	65.2	10	15	3	56.5
3	9	2	65.2	11	3	4	65.2
4	12	2	65.2	12	6	4	60.9
5	15	2	60.9	13	9	4	60.9
6	3	3	69.6	14	12	4	60.9
7	6	3	60.9	15	15	4	60.9
8	9	3	60.9	16	3	5	-

According to the prototype CPS, a range of five trial numbers (3, 6, 9, 12, and 15) for adaptive boosting have to be conducted with each MCT value, where the experimentation with the MCT value have to start from an MCT value of 2 (i.e. a full tree) until a single-node tree results. Following these instructions, 15 candidate classifiers are generated from each input data set. A total of 60 candidate classifiers are developed from the four input data sets.

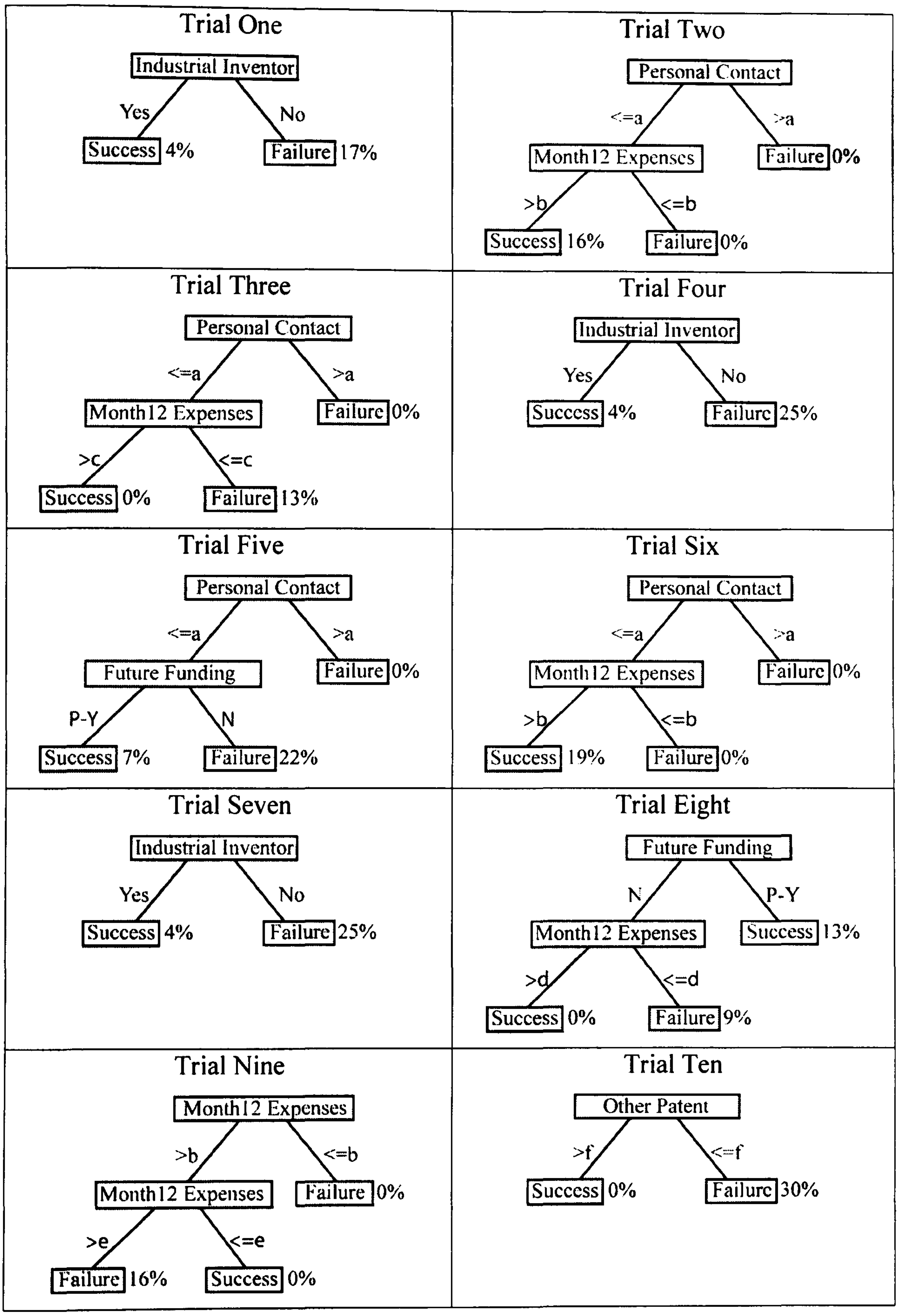
The fourth and eighth columns of each table showed the estimated accuracies for the candidate classifiers. It is found that the 16^{th} candidate classifier for every input data set is not valid, since the candidate classifier generated is a single-node tree.

The fifth step of the prototype CPS is to choose the best performing candidate classifier. Among the 60 candidate classifiers, 3 candidate classifiers have achieved the highest level of estimated accuracy of 78.3%. Among these 3 candidate classifiers, the one generated using the second input data set is the best performing candidate classifier, according to the definition described in the prototype CPS. In short, the best performing candidate classifier is therefore the 10^{th} candidate classifier generated using the second input data set, as shown in Table 3-7.

The final step of the prototype CPS is to generate the final classifier according to the definition of the best performing candidate classifier, using every case of the respective input data set as the training data. Consequently, the final classifier is generated using every case of the second input data set as training data, 15 trials for adaptive boosting, and an MCT value of 3.

Consequently, the final classifier generated consists of 13 trees (micro-classifiers), where this classifier determines the class labels of invention cases by counting the votes from these 13 micro-classifiers. These micro-classifiers are displayed in Figure 3-5. Though the number of trials is predefined to be 15, only 13 trials are conducted, as the 14^{th} trial is stopped due to an extreme inaccuracy of the 14^{th} tree.

Figure 3-5: The micro-classifiers of the resultant classifier



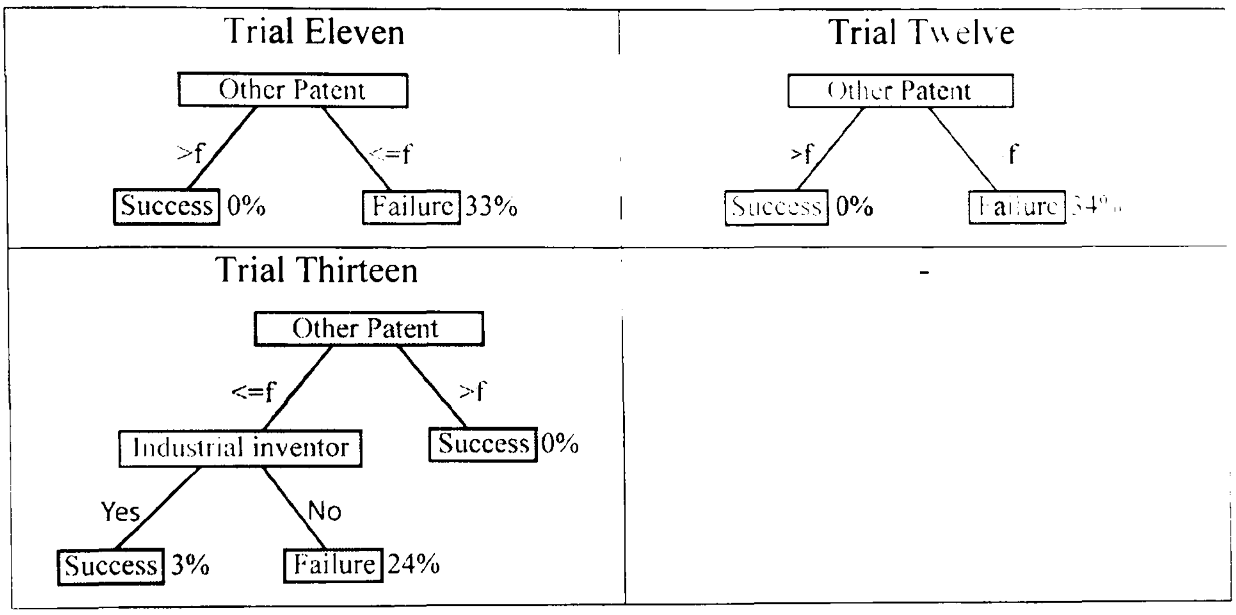


Figure 3-5 showed 13 micro-classifiers resulted from 13 trials of tree construction for adaptive boosting. Each micro-classifier is constructed using all 23 cases of the second input data set as training data set. Due to confidentiality issue, the actual values for node splitting have been replaced with labels 'a', 'b', 'c', 'd', 'e', and 'f'. The percentage values next to the leaf nodes represent the percentage of misclassified cases. The estimated classification accuracy of this final classifier is 78.3%.

Discussions

The first step of the prototype CPS is the generation of an input data set according to the specification that comprised of a list of literature suggested critical factors. During data collection, it was found that only a small subset of the data required by the list of critical factors is available from the source. The list of critical factors therefore represents an ideal rather than realistic list of invention attributes for data collection from technology transfer offices. However, extending the data source to other information bases may alleviate the problem of unavailable data.

Furthermore, due to the ambiguous nature of the suggested critical factors, further effort was required to develop coding systems for data collection. For instance, the coding system for the critical factor 'historical research leading to the invention was funded by the private sector' was comprised of the following four invention attributes, namely the first, second, third, and fourth 'source of funding leading to the current development of the invention'. The value for each of these four attributes for a particular case was coded as one of the followings: 'Firm', 'Government', 'Charity', 'Spinout', 'Other university', 'Department', 'Technology-transfer office', or 'Not applicable'.

In addition, a large amount of subjective judgement was required to determine the categorical value for a number of attributes because of the ambiguous nature of the critical factors. The data collected therefore may not be reproducible, since subjective judgements vary from person to person. This represents a drawback of using qualitative and ambiguous invention attributes.

In addition, because no electronic database was used at the data source, the data collection process was a time-consuming and labour intensive task. The manual codification of information from hard copies of documents into electronic data might also have given rise to an input data set more prone to errors. Moreover, the absence of electronic database also means the absence of predefined data fields. Thus, when certain data needed by the prototype CPS was missing from the hard copies, it was not clear whether the information was indeed missing or merely did not exist. This problem caused the discarding of a subset of potentially important invention attributes, and may also have had an impact on the classification accuracies of the resultant classifiers.

Due to the small size (with only 23 cases) of input data set derived from the data source, leave-one-out sampling method was used to compute a candidate classifier's classification accuracy. As a result, 23 trees were generated for each candidate classifier. In addition, the use of adaptive boosting further increases the number of trees generated in order to arrive at a candidate classifier. For instance, the use of adaptive boosting with 3 trials would result in the generation of 69 (3 x 23) trees in order to compute the accuracy of the candidate classifier.

Moreover, the experimentation with MCT values was time-consuming, and therefore would not be feasible for large input data sets consisting of many cases.

Consequently, 60 candidate classifiers were generated prior to the generation of the final classifier. Most of the candidate classifiers achieved satisfactory classification accuracy, and the average accuracy was 66%. However, the individual classification accuracies ranged from 48% to 78%. The low end of the range showed that the prediction accuracies of certain candidate classifiers were unacceptably low. Nonetheless, this supports the design of the prototype CPS, in which it generates more than one candidate classifier in order to determine the appropriate parameter values for the final classifier.

Among the 60 candidate classifiers, the best candidate classifier accurately predicted the class labels for 18 out of the 23 cases, which is equivalent to an accuracy rate of 78%. This classifier used only 5 attributes in 'Month 12' to compute the class labels. This resultant classifier is therefore suggested to be capable of giving reasonably accurate prediction based on only 5 attributes obtained in the first year. 4 of the 5 attributes were similar to the transferable critical factors identified in the literature review, which shows that the transferable critical factors are potentially important attributes for invention assessment.

However, the final classifier generated was comprised of 13 micro-classifiers, where each of those generally consisted of 2 invention attributes. Meaning that only two attributes were used at a time to arrive at investment decisions, which is unlikely to be correct.

In view of the achievements and limitations of the prototype CPS, it is concluded that although the prototype CPS seemed to be capable of providing reasonably accurate predictions, it requires significant improvements.

3.5 CONCLUSIONS

This chapter presented the prototype CPS and the results generated. The limited data collected from the technology transfer office implied that extending the data source to other information bases may be beneficial. Due to the ambiguous nature of literature-suggested critical factors, the development of coding systems and the use of subjective judgement were needed, which have prolonged the data collection process and given rise to non-reproducible data. Also, the absence of electronic database at the source has resulted in slow data collection, and potentially unreliable input data set. In addition, the use of leave-one-out sampling method, adaptive boosting, and the need to experiment with MCT values have resulted in the generation of a large number of trees in order to arrive at one candidate classifier. Such classifier generation mechanism is therefore not feasible for input data sets consisting of a large number of cases.

Nonetheless, the variable classification accuracies of the candidate classifiers supported the use of candidate classifiers to determine appropriate parameter values for the final classifier. The final classifier was estimated to be capable of providing reasonably accurate classification based on only five attributes obtainable in the first year. The similarity between these five attributes and the transferable critical factors also demonstrated that the transferable critical factors can be used for the final classifier. Lastly, the limited attribute nodes contained in the micro-classifiers suggests that the classification logic employed by the final classifier was not likely to be correct. It is concluded that although the prototype CPS seemed to be capable of providing reasonably accurate predictions, the prototype CPS requires significant improvements.

CHAPTER FOUR

BACKGROUND SURVEYS

4.1 INTRODUCTION

Findings from the literature review and prototype CPS showed the need to further investigate the invention assessment practices in order to improve the design for the final CPS. Three surveys with university technology transfer offices are therefore conducted in this chapter. To start with, an invention disclosure form study involving various technology transfer offices is conducted (Section 4.2). A questionnaire survey is then carried out to several other technology transfer offices that were not covered by the previous study (Section 4.3). Finally, interviews with technology transfer managers from different universities are carried out (Section 4.4). While the first two surveys concentrated on what invention attributes are required during assessment, the third survey investigated how these attributes are used during the assessment process. In Section 4.5, conclusions from these surveys are presented, in which relevant findings are incorporated into the final CPS.

4.2 INVENTION DISCLOSURE FORM STUDY

Invention disclosure form is a common tool employed by university technology transfer offices to collect necessary invention information, upon which intuitive invention assessments are conducted. During the several invention assessments conducted over the course of the technology transfer process for an invention, invention disclosure forms are often used repeatedly for the first as well as later rounds of assessment. Since the assessment decisions are based on the information recorded on these forms, the information requested by the forms therefore represents important invention attributes used for invention assessment considerations.

Despite the importance of invention disclosure forms, very little research has been done on this topic. In order to identify important invention attributes used for invention assessment, this study endeavoured to examine the invention disclosure forms used by different universities in the United Kingdom. The methodology adopted for this study, and the resultant findings are presented below.

Data Source

Initially, invention disclosure form requests were sent to several universities. Consequently, only one university provided a sample of their invention disclosure form (ID form). Other universities denied the requests, and replied that their ID forms are either available from the university websites, or that such information is confidential and can not be provided. An extensive search is therefore carried out on university websites in the United Kingdom. In this study, an ID form is defined as a form or a set of forms where a university directs their inventors to complete for invention disclosure purpose, before the application of any intellectual property rights protection. Eventually, 16 sets of forms are located and downloaded from various university websites.

Measurement

In order to investigate the choice of invention attributes by these universities in relation to their technology transfer performance, two measures are adopted. They are the latest Intellectual Property Scoreboard (IP-Scoreboard), and the 2005 Times Good University Guide (TGU). IP-Scoreboard is a league table of organisations ranked by their number of intellectual property registration within the UK jurisdiction. Universities listed on the IP-Scoreboard are thus considered to have higher commercialisation performance. TGU is a well-known league table ranking 100 universities in terms of a series of factors, such as research assessment and graduate prospects. Although these are indirect measures for the technology transfer performances of universities, they are adopted because the direct technology transfer outcome data are confidential and unavailable.

Among the 16 universities, 5 universities are listed on both the IP-Scoreboard and the TGU; 10 are listed on the TGU only; and 2 are listed on neither the IP-Scoreboard nor the TGU. The 5 universities listed on the IP-Scoreboard are within the top 10, and are also ranked in the top 30 of the TGU. Among the 10 universities listed only on the TGU, 5 are ranked within the 11th to the 30th; 3 are ranked within the 31st to the 60th; and 3 are ranked within the 61st to the 100th. Based on the rankings on the two measures, the 16 universities are divided evenly into three classes representing varying levels of technology transfer capability, as shown in Table 4-1.

Table 4-1: Classification of sample universities

Sample University	1	2	3	4	5	6	7	8
IP-Scoreboard	5	10	10	-	-	10	5	-
TGU	5	5	10	20	20	25	30	30
Not listed	-	-	-	-	-	-	-	-
Class	U	U	U	M	M	U	U	M

Continuation of Table 4-1:

Sample University	9	10	11	12	13	14	15	16
IP-Scoreboard	-	-	-	-	-	-	-	-
TGU	35	35	55	55	75	80	-	-
Not listed	-		-	-	-	-	X	X
Class	M	M	L	L	L	L	N	N

Where 'U' denotes universities with upper-class technology transfer performance, 'M' denotes the middle-class, 'L' denotes the lower-class, and 'N' denotes the class of unclassified.

As shown in Table 4-1, the sample universities are sorted by the TGU rankings. In particular, the first row shows the reference number for each sample university. The second and the third rows show their rankings on the IP-Scoreboard and TGU respectively. Due to the confidentiality issue, only the upper limit of the range of every five rankings is shown, as opposed to the actual ranking. For instance, '5' denotes a ranking position located between the 1st and the 5th, and '10' denotes a ranking between the 6th and 10th. Moreover, the two sample universities not listed on neither IP-Scoreboard nor TGU are marked with 'X' in the forth row. The class label for each sample university is showed on the fifth row. This sample therefore consists of 5 upper-class, 5 middle-class, 4 lower-class, and 2 unclassified universities.

Potential attributes for the invention assessment

The lengths of the 16 ID forms collected ranged from 1 to 6 pages, and each form covers a varying number of invention attributes. After removing repeating attributes, 127 invention attributes are identified from the 16 ID forms. A list containing the 127 attributes is provided in Appendix A4.1.

This list is then shortened by the removal of redundant and irrelevant attributes. Examples of irrelevant attributes include 'the ethical consents of using biological material from humans', and 'previous invention disclosure record'. These attributes refers to the basic legal requirements for patent application. They are irrelevant to this study, since this thesis concerns the assessment of patentable inventions. Consequently, the shortened list contains 29 attributes, as presented in Table 4-2.

The first column of Table 4-2 shows the 29 attributes derived from the 16 sample ID forms. The second column shows the number of sample ID forms that have included the attribute. The third to sixth columns show the inclusion of the attribute by different classes of university. The last column shows the similar attributes adopted by the prototype CPS (denoted as Prototype). In addition, these 29 attributes are sorted in descending order by the second column. For instance, the first attribute is included in all the 16 sample ID forms, which consist of ID forms from four 'upper class' universities, 6 'middle class' universities, four 'lower class' universities, and two 'unclassified' universities. Moreover, the last column shows that no attributes used for the prototype CPS resemble the first attribute in Table 4-2.

Table 4-2: Attributes derived from the sample Invention Disclosure forms

Invention attributes derived from the 16 sample ID forms	**ID forms**	**U**	**M**	**L**	**N**	**Proto-type**
1. Contributors/Inventors – the number and employment status of contributors and inventors	16	4	6	4	2	-
2. Technical description	15	4	6	3	2	-
3. Advantages - the advantages of the invention over the existing solution	14	3	6	3	2	-
4. Support source – source of funding and in-kind support lead to the invention	13	3	5	3	2	1st - 4th
5. Company list – a list of potentially interested firms (contacted).	12(5)	2	5	3	2	9th

6. Uses – commercial uses and application of the invention.	11	2	5	3	1	-
7. Drawing attachment – drawings and sketches relating to the explanation of the invention are to be attached	10	1	4	4	1	-
8. Contractual agreements – the number of contractual agreements linked with the invention, including MTA[7]	10	3	3	2	2	25th
9. Competitor list – the list of potentially competing existing products.	8	2	1	3	2	18th
10. Existing solution - publicly acknowledged existing solution	8	1	3	3	1	-
11. Records – record details of the prior arts such as patent number and verbal records	8	3	3	1	1	-
12. Novel feature – novelty of the invention	8	2	3	1	2	-
13. Developmental stage – whether the invention is an early stage invention, a proven concept, or a prototype	6	2	1	2	1	12th
14. Disadvantages – disadvantages of the existing solution	6	1	3	1	1	-
15. Background intellectual properties (IP) – the list of pre-existing IP used as a basis for the invention	5	1	2	2	0	-
16. Search – whether a prior art search has been conducted (the date of such a search)	4	2	1	1	0	7th
17. Current activity – activities regarding invention exploitation, such as current commercial interest and marketing activity.	4	1	1	2	0	19th

[7] MTA (Material Transfer Agreements) are the basis for university-industry collaborations involving transactions of 1) proprietary material and or information; 2) information that cannot be obtained without a secrecy agreement; 3) a substance that embodies a trade secret; 4) no true collaboration anticipated; data only may be provided. http://www.fda.gov/oc/ofacs/partnership/techtran/criteria.htm 17/11/05

18. Self evaluation – the invention's commercial value rated by the inventors themselves	4	1	0	1	2	-
19. Invention testing – whether the invention has been tested in any ways and the results of the testing	4	1	1	1	1	-
20. Other application – whether the invention can be applied in other areas outside the inventor's special interests.	3	0	1	1	1	-
21. Market size – target market size.	2	1	0	0	1	18th
22. Type of invention – whether it is a device, a process, a drug or software, and so on.	2	0	0	1	1	5th
23. Further development – whether further development is needed and the funding requirement for that, and whether such funding has been secured.	2	1	0	1	0	8th, 17th
24. Research period – period of research relevant to the creation of invention	2	1	1	0	0	-
25. Field – the field or area in which the invention has application.	3	0	2	0	1	
26. Country list – the list of countries where the invention is likely to be used	1	0	1	0	0	18th
27. Compatibility – the invention's compatibility with existing products.	1	0	1	0	0	-
28. Non-confidential summary	1	0	0	1	0	-
29. Replace possibility – whether the invention will replace certain existing applications and whether existing applications can be improved to compete with the invention.	1	0	1	0	0	-

With a total of 16 sample ID forms, attributes included by 8 or more sample ID forms are thus considered as frequently used attributes. As indicated by column two of Table 4-2, the first 12 attributes are frequently used attributes. These attributes represent popular attributes commonly adopted by universities of different classes. These 12 attributes therefore are potential attributes for invention assessment, and they are added to a list so-called 'potential attributes for invention assessment'.

Among the 16 universities, 4 universities were identified as 'upper class', based on the technology transfer performance. Attributes included by 2 or more ID forms from these 'upper class' universities are therefore attributes frequently used by the 'upper class' universities. By sorting the 29 attributes in Table 4-2 in descending order by the third column, it is found that 12 of the 29 attributes are attributes frequently used by the 'upper class' universities. These 12 attributes are reproduced from the first and third columns of Table 4-2, which is shown in Table 4-3. Among the 12 attributes in Table 4-3, only the last 2 attributes are not covered by the frequently used attributes (the first 12 attributes in Table 4-2). These 2 attributes are therefore added to the list of 'potential attributes for invention assessment'. As a result, the list of 'potential attributes for invention assessment' currently contains 14 attributes, where these potential attributes consists of attributes frequently used by all sample universities, as well as attributes frequently used by the 'upper class' universities.

Lastly, out of the 29 attributes, 11 of them are similar to the attributes used by the prototype CPS, as indicated by the last column in Table 4-2. These eleven attributes are therefore similar to the critical factors suggested by the University Technology Transfer literature. The 11 attributes resembling attributes used by the prototype CPS are reproduced from the first and last columns in Table 4-2, as shown in Table 4-4. Among these 11 attributes, the last 5 attributes are covered by neither the frequently used attributes (the first 12 attributes in Table 4-2) nor the attributes frequently used by the 'upper class' universities (the attributes shown in Table 4-3). These five attributes are therefore added to the list of 'potential attributes for invention assessment'. As a result, this list of potential attributes now contains 19 attributes.

Table 4-3: Attributes frequently used by 'upper class' universities

Invention attributes derived from the 16 sample ID forms	U
1. Contributors/Inventors – the number and employment status of contributors and inventors	4
2. Technical description	4
3. Advantages - the advantages of the invention over the existing solution	3
4. Support source – source of funding and in-kind support lead to the invention	3
5. Company list – a list of potentially interested firms (contacted).	2
6. Uses – commercial uses and application of the invention.	2
8. Contractual agreements – the number of contractual agreements linked with the invention, including MTA	3
9. Competitor list – the list of potentially competing existing products.	2
11. Records – record details of the prior arts such as patent number and verbal records	3
12. Novel feature – novelty of the invention	2
13. Developmental stage – whether the invention is an early stage invention, a proven concept, or a prototype	2
16. Search – whether a prior art search has been conducted (the date of such a search)	2

Table 4-4: Attributes resemble attributes used for prototype CPS

Invention attributes derived from the 16 sample ID forms	Prototype
4. Support source – source of funding and in-kind support lead to the invention	1st - 4th
5. Company list – a list of potentially interested firms (contacted).	9th
8. Contractual agreements – the number of contractual agreements linked with the invention, including MTA	25th
9. Competitor list – the list of potentially competing existing products.	18th
13. Developmental stage – whether the invention is an early stage invention, a proven concept, or a prototype	12th
16. Search – whether a prior art search has been conducted (the date of such a search)	7th
17. Current activity – activities regarding invention exploitation, such as current commercial interest and marketing activity.	19th
21. Market size – target market size.	18th
22. Type of invention – whether it is a device, a process, a drug or software, and so on.	5th
23. Further development – whether further development is needed and the funding requirement for that, and whether such funding has been secured.	8th, 17th
26. Country list – the list of countries where the invention is likely to be used	18th

4.3 QUESTIONNAIRE SURVEY

As encountered in the invention disclosure form study, the requests of invention disclosure form (ID form) are often denied by universities due to the issue of confidentiality. In order to further investigate invention attributes used for invention assessment, a questionnaire survey with technology transfer offices has been conducted. In addition to the investigation of invention attribute, the questionnaire also investigates the technology transfer offices' usage of any invention assessment tools other than the ID form, and their assessment context. The invention attributes queried in the questionnaire consist of both the attributes from the list of 'potential attributes for invention assessment' (identified earlier by the invention disclosure form study), and other attributes revealed in the invention disclosure form study. A copy of the questionnaire is shown in Appendix A4.2. To make sure that the most critical questions are covered while maintaining the questionnaire to be reasonably short, all questions queried in the questionnaire have been consulted with an experienced university technology transfer office director. In order to target universities that operate technology transfer activities, questionnaires have been sent to the member universities of AURIL (The Association for University Research and Industry Links) – a central association for the University Technology Transfer industry in the United Kingdom. Consequently, 15 questionnaires were completed and returned.

The questions of the questionnaires can be divided into two categories. The first category concerns the invention attributes included by the technology transfer office's ID form. The second category is about the office's use of invention assessment tools and the assessment context.

Measurement

Similar to the invention disclosure form study, the responded universities are divided into classes of technology transfer performance using the two indirect measures for technology transfer performance: Intellectual Property Scoreboards (IP-Scoreboard) and the Times Good University Guide (TGU). Upon which the frequently-used attributes by universities of different classes are then identified.

Among the 15 responded universities, 3 are listed on both the IP-Scoreboard and the TGU; 9 are listed on the TGU only; and 3 are listed on neither the IP-Scoreboard nor the TGU. The 3 universities ranked on the IP-Scoreboard are within the top 15, whereas those ranked on the TGU are quite evenly distributed over different range of ranks.

The class labels of the 15 responded universities are shown in the fourth row in Table 4-5, where 'U', 'M', 'L' and 'N' denote upper-class, middle-class, lower-class, and unclassified level of technology transfer performances respectively. The first three rows show the reference number, and ranking positions for the two measures respectively. Again, only the upper limit of the range of every five positions of ranking is shown, due to the confidentiality issue. Lastly, the universities in this table are sorted by the third row.

Table 4-5: Class labels for the 15 responded universities of the questionnaire survey

University	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
IP-Scoreboard	-	-	15	15	-	10	-	-	-	-	-	-	-	-	-
TGU	5	10	10	10	15	25	40	40	45	50	55	60	-	-	-
Class	U	M	U	U	M	U	M	M	L	L	L	L	N	N	N

The 15 sample universities therefore consist of 4 upper-class, 4 middle-class, 4 lower-class, and 3 unclassified universities.

Potential attributes for the invention assessment

Table 4-6: Attributes included in the questionnaire

Invention attributes included in the questionnaire	Frequency	U	M	L	N	Potential Attribute
1. Invention description in terms of functions	13	3	4	4	2	X
2. Research funding source leading to the invention	11	3	4	2	2	X
3. Invention description in terms of the developmental stage	10	2	2	4	2	X
4. Interested third parties who have contacted the inventors	10	2	2	4	2	X
5. End product envisaged	9	2	3	3	1	X
6. Potentially competing products	9	2	3	3	1	X
7. Involvement of other parties during invention development	9	3	2	3	1	-
8. Plan for future development of the invention	8	1	2	3	2	X
9. Contacts for potential third parties provided by inventors	8	2	2	3	1	-
10. Patent search	7	1	1	3	2	X
11. Non-confidential summary	7	1	2	2	2	-
12. Market research conducted	3	1	1	1	0	X
13. Inventor's agreed share of income	3	2	0	0	1	-
14. Inventor's publication record	3	0	2	1	0	-
15. Whether future funding is secured	2	1	0	1	0	X
16. Plan for Licensing	2	1	0	1	0	-
17. Plan for setting up a spinout company	2	1	0	1	0	-
18. Future development timetable	1	0	0	0	1	-
19. Development funding source	1	0	0	1	0	-
20. The estimated future cost	1	1	0	0	0	X

A number of invention attributes are listed in the questionnaire, and the respondents are asked whether those attributes are included in their ID forms. These attributes are shown in the first column in Table 4-6, and the table is sorted using the second column.

The second column of Table 4-6 shows the number of respondent technology transfer offices which have included the attribute in their ID forms. Column three to six show the numbers of such inclusion by universities of different classes. In the seventh column, the attribute is marked with an 'X' if it is one of the attributes from the list of 'potential attributes for invention assessment' identified earlier in the invention disclosure form study. For instance, the first attribute is used by 13 universities, and is one of the attributes from the list of 'potential attributes for invention assessment'.

Since there are 15 respondents, attributes used by 8 or more respondents are identified as the frequently used attributes, based on these 15 sample universities. Column two shows that the first 9 attributes in Table 4-6 are the frequently used attributes. As indicated by column seven, the 7th and the 9th attributes are not included in the list of 'potential attributes for invention assessment'. These two attributes are therefore added to the list, so that the list currently contains 21 attributes.

Similarly, 4 of the sample universities are classified as having 'upper class' technology transfer performance. Attributes included by 2 or more of these 'upper class' universities are therefore attributes frequently used by the 'upper class' universities. By sorting the 20 attributes in Table 4-6 in descending order by the third column, it is found that 9 of the 20 attributes are attributes frequently used by the 'upper class' universities. These 9 attributes are reproduced from the first and third columns of Table 4-6, which is shown in Table 4-7.

The last attribute shown in Table 4-7 is not covered by the frequently used attributes (the first 9 attributes in Table 4-6). This attribute is therefore added to the list of 'potential attributes for invention assessment'. In effect, this list is extended to contain 22 attributes, as showed in Table 4-8, where column one displays the potential attributes and column two shows the origins of the attribute. For instance, 'ID-1' denotes the first attribute derived from the invention disclosure form study shown in Table 4-2, and 'Q-7' represents the seventh attribute derived from the questionnaire survey shown in Table 4-6.

Table 4-7: Attributes frequently used by the upper class universities

Invention attributes included in the questionnaire	U
1. Invention description in terms of functions	3
2. Research funding source leading to the invention	3
3. Invention description in terms of the developmental stage	2
4. Interested third parties who have contacted the inventors	2
5. End product envisaged	2
6. Potentially competing products	2
7. Involvement of other parties during invention development	3
9. Contacts for potential third parties provided by inventors	2
13. Inventor's agreed share of income	2

Table 4-8: Potential attributes for invention assessment

Potential attributes for invention assessment	Origins
1. Contributors/Inventors – the number and employment status of contributors and inventors	ID-1
2. Technical description	ID-2
3. Advantages - the advantages of the invention over the existing solution	ID-3
4. -Support source – source of funding and in-kind support lead to the invention	ID-4
5. Company list – a list of potentially interested firms (contacted).	ID-5
6. Uses – commercial uses and application of the invention.	ID-6

7. Drawing attachment – drawings and sketches relating to the explanation of the invention are to be attached	ID-7
8. Contractual agreements – the number of contractual agreements linked with the invention, including MTA	ID-8
9. Competitor list – the list of potentially competing existing products.	ID-9
10. Existing solution - publicly acknowledged existing solution	ID-10
11. Records – record details of the prior arts such as patent number and verbal records	ID-11
12. Novel feature – novelty of the invention	ID-12
13. Developmental stage – whether the invention is an early stage invention, a proven concept, or a prototype	ID-13
14. Search – whether a prior art search has been conducted (the date of such a search)	ID-16
15. Current activity – activities regarding invention exploitation, such as current commercial interest and marketing activity.	ID-17
16. Market size – target market size.	ID-21
17. Type of invention – whether it is a device, a process, a drug or software, and so on.	ID-22
18. Further development – whether further development is needed and the funding requirement for that, and whether such funding has been secured.	ID-23
19. Country list – the list of countries where the invention is likely to be used	ID-26
20. Involvement of other parties during invention development	Q-7
21. Contacts for potential third parties provided by inventors	Q-9
22. Inventor's agreed share of income	Q-13

Invention assessment tools and assessment context

A subset of the questionnaire queries concern the use of invention assessment methods other than invention disclosure forms, and some assessment contextual aspects, their findings are presented below.

The queries in this subset are shown in the first column of Table 4-9. The second column shows the total frequency of all sample universities, where the third to sixth columns show the frequencies with respect to universities of different classes.

Table 4-9: Other questions included in the questionnaire

Questions	Frequency	U	M	L	N
1. Whether Invention Disclosure forms (ID forms) or similar forms are used	13	3	4	4	2
2. Invention assessment method: Face-to-face meetings	15	4	4	4	3
3. Invention assessment method: Presentations	7	0	2	3	2
4. ID forms must be completed before patent applications	6	2	2	1	1
5. Invention database created using Microsoft Access	2	0	1	0	1
6. Invention database created using Microsoft Excel	2	0	0	1	1
7. Invention database created using other software	7	1	3	2	1
8. Filing the hard copies of invention information	10	3	3	2	2
9. The use of analytical software for invention assessment	0	0	0	0	0

The findings presented in Table 4-9 shows that the use of ID form is very common but is not generally a strict requirement for patent applications. Face-to-face assessment is more important than ID form as an invention assessment method. Lastly, it is found that all sample offices store the invention information either in electronic databases or in the form of hard copies, and no offices employ analytical software for invention assessment. These findings together showed that the invention assessment practices are highly intuitive.

Furthermore, the questionnaire also asked the year when the use of ID form was started. Among the 13 offices that use ID form, 12 provided the year information, which is displayed in Table 4-10.

Table 4-10: The year when the offices started using ID form

Year	1996	1998	2001	2002	2003	2004	2005
Class	M	U	M	U, M, N	U, M	N	L, L, L

As shown, there is a mixture of offices with class labels 'U', 'M', and 'N' started using ID forms in different years, while the latest year only consists of lower-class offices.

4.4 INTERVIEW SURVEY

The literature review in Chapter Two revealed a rather strange side of the University Technology Transfer sector. While both practitioners and researchers upheld the importance of the invention assessment, the details of invention assessment practices in real life is typically avoided and rarely researched. Despite the great number of critical-factor studies, hardly ever do these studies discuss the finding's application, and the conclusions often stopped abruptly at the correlation between critical factors. Literature interviewing practitioners often found that the practitioners usually depend on the business partners to make the invention selection decisions. It is therefore suspected that the invention assessment process does not exist in some or even most university technology transfer offices, where the role of these offices is to assist rather than to judge/assess. Besides, according to the limited literature provided information about the invention assessment process, the process is mainly described as intuitive and the general practice is to subjectively give scores (such as on a scale of 1 to 10) to the technical and market aspects of an invention. In order to investigate the intuitive assessment process as well as the context for invention assessment, for the purpose of improving the design of the CPS, an interview survey is performed with technology transfer managers.

Sample

The previous two studies revealed the frequently used invention attributes, and several aspects of the assessment contexts in practice. These findings established a good foundation to conduct structured interviews with technology transfer managers. Interview requests were sent to the 15 universities responded to the questionnaire survey, 7 of these universities agreed to be interviewed. The class labels of these universities (based on the definitions adopted by the questionnaire survey) are shown in Table 4-11.

Table 4-11: Class label for offices of the interview survey

University	1	2	3	4	5	6	7
Class	U	U	M	L	L	N	N

Queries asked during interviews

When considering the invention assessment process, it is necessary to understand the background setup for the process as well as the aftermath actions carried out caused by the assessment outcome. The queries asked during the interview survey are hence based on these three aspects: 1) the background setup, 2) the invention assessment process, and 3) actions. Under the first and the third aspects, questions regarding several dimensions were asked. The dimensions are listed in Table 4-12.

Table 4-12: Dimensions of two of the three aspects queried during the interviews

Background Setup	Actions
• Source of invention projects • Funding • Assessment circumstances • Division of assessment staff • Perceptions	• Actions regarding chosen and not-chosen invention projects • Information storage and maintenance

Findings

The questions asked during each interview therefore include the query of the invention assessment process adopted as well as the queries presented in Table 4-12. The findings generated show that the queries can be divided into two types, namely discriminative queries and non-discriminative queries. Findings from the discriminative queries showed the division of the seven offices into two groups, namely Group I and Group II. Offices within the same group are very similar in terms of issues covered by the discriminative queries. The first row of Table 4-13 displays the group labels for each office. The second and the third rows respectively present the offices' background information including their class labels and the years when the offices first started to use invention disclosure form (such information is derived from the questionnaire study). Group I consists of four offices, and Group II comprises of the other three offices. As shown in the third row of Table 4-13, Group I offices generally started using invention disclosure forms earlier than Group II offices.

Table 4-13: Group labels for offices of the interview survey

Group	1	1	1	1	2	2	2
Class	U	U	N	N	M	L	L
ID year	2002	2003	2002	2004	2005	2005	2005

However, findings generated by the non-discriminative queries do not assist the differentiation of offices into groups. The seven offices either generally agree upon the issues or each office adopts different approach for the issues.

The distribution of the discriminative (marked 'X' under 'D') and non-discriminative (marked 'X' under 'ND') queries is displayed in Table 4-14.

Table 4-14: Distribution of discriminative and non-discriminative queries

Aspects	**Queries**	**D**	**ND**
Background Setup	Source of invention projects	X	-
	Funding	X	-
	Assessment circumstances	-	X
	Division of assessment staff	-	X
	Perception	-	X
Invention Assessment Process		X	-
Actions	Actions regarding chosen and not-chosen invention projects	X	-
	Information storage and maintenance	-	X

The findings regarding discriminative and non-discriminative queries are presented sequentially next.

Findings from discriminative queries

Source of Invention Projects

The invention source refers to the mechanisms transferring inventions from academic departments to the technology transfer office. For Group I offices, a reactive approach is adopted, where invention projects are generally initiated by inventors contacting the offices. Whilst for Group II offices, a proactive approach is adopted, where events are conducted to increase the awareness among the academics regarding various technology transfer issues, and technology transfer managers often take the initiative to visit almost every academic staff to identify potential invention projects. For one of the Group II offices, representative personnel are available in every academic department for academic staff to report their inventions.

Funding

Regarding funding issue, only Group II offices have been recently granted various Government technology transfer funding for office expansion. One Group I offices, however, are often self-funded (i.e. based on the proceeds generated from technology-transfer), or internally funded (i.e. funded by the university).

Invention Assessment Process

The invention assessment processes adopted by the seven offices can be summarised as the execution of a three-step loop, which is executed under certain background setups and produces certain actions, as illustrated in Figure 4-1. The loop starts with the distribution and completion of Invention Disclosure forms through a channel. followed by holding meetings between the inventors and assessment panel, and ends with information evaluation and assessment decisions. While Group I offices commonly execute the three-step loop once. Group II offices generally execute the three-step loop two to three times.

For Group II offices, each of the three-step loop acts like a gate which filter out a subset of inventions considered to be of relatively low potential. The number of remaining inventions decreases as the process is moving towards the final gate. inventions reaching the final gate are considered to have the highest potential. The assessment panel involved in the meeting at the final gate normally consists of a variety of personnel, including technology transfer office staff, technology transfer experts, business advisors, and external technology transfer agents. At every gate, invention evaluations are mainly based on information provided by the inventors. For two of the three Group II offices, inventions at every gate are generally graded using scoring systems developed by professional technology-assessment related agencies.

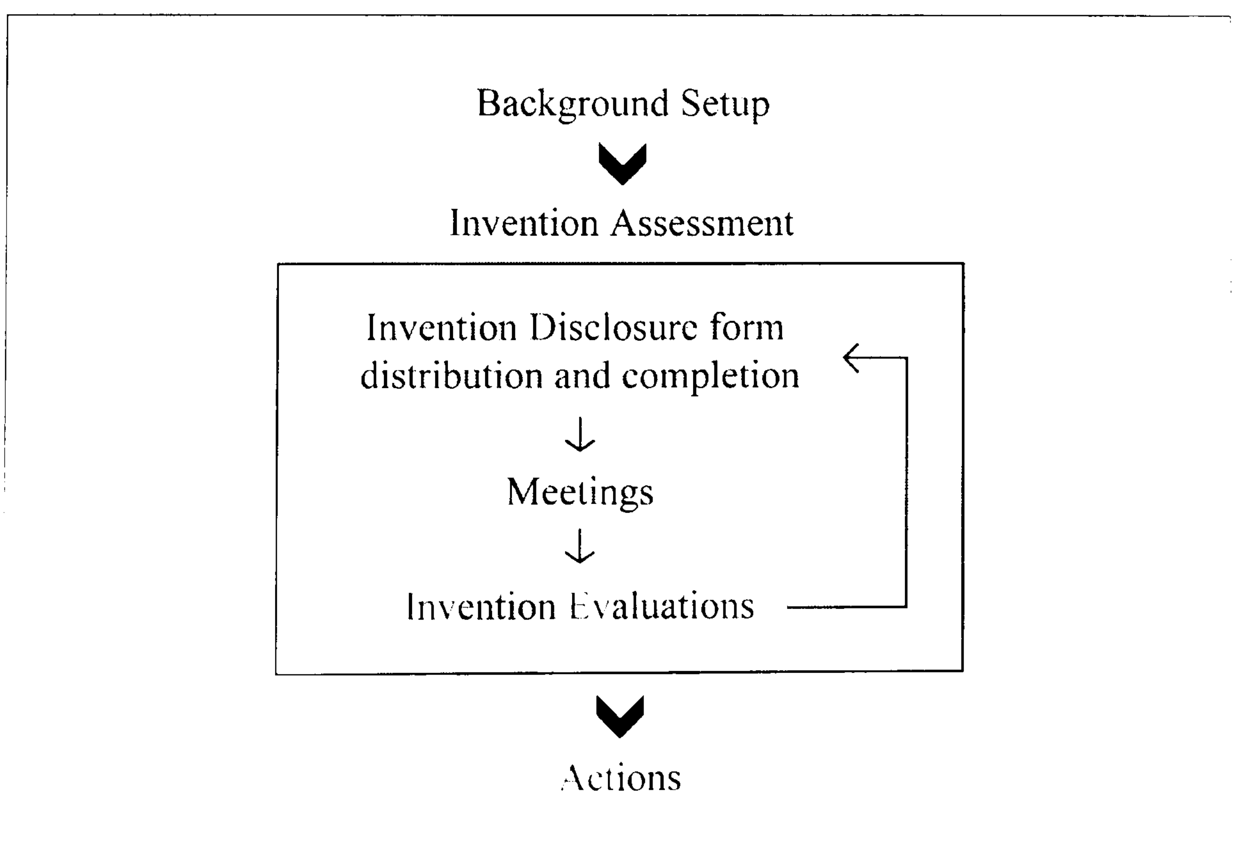


Figure 4-1: The three-step loop of invention assessment

These two Group II offices also supplied their scoring systems to assist the current interview survey. Due to confidential issue, certain details such as individual invention attributes covered by the scoring systems will not be discussed here. By studying the scoring systems, it is found that both scoring systems are a list of qualitative invention attributes organised into categories, where a qualitative score (such as 'good' or 'bad') is allocated to each attribute by the assessor. After giving scores to all attributes of a particular category, the user then allocates a score (such as on a scale of 1 to 10) to the category. Assessors determine the category score based on a subjective summation of all the attribute scores. All category scores are then summed up to form the final score. This final score denotes the potential of the invention project.

For Group I offices, though the three-step loop is only executed once, the requirements are often more rigorous. During the meeting (the second step of the loop), inventors are usually required to demonstrate the inventions or display the invention prototypes. Instead of having a variety of personnel as the assessment panel, the panel only consists of about 1 or 2 technology transfer staff. In addition to the invention information provided by the inventors, technology transfer personnel generally conduct prior art search using public intellectual property search engines as well as general internet search engines. Finally, Group I offices evaluate inventions using general methods such as the SWOT analysis, or based on personal judgement of the assessors.

Actions regarding Chosen and Not-chosen inventions

For Group I offices, the actions taken after invention assessment include filing patent applications, conducting market research, and funding applications. Three of the four Group I offices emphasised that while actions will be taken for inventions considered to be of high potentials, lower potential inventions are usually revisited after a period of time such as 6 to 12 months. In addition, Group I offices often seek funding from general sources such as Research Councils and Charities to accomplish the actions. They emphasised that it is important to minimise the resources spent on individual invention projects. Most actions, such as filing patent applications and conducting market research, are thus done by internal technology transfer office staff in order to avoid expenses.

For Group II offices, actions taken include filing patent applications and conducting market research. These actions are generally taken merely for inventions considered to be of high potentials. Moreover, internal funding is normally provided to accomplish the actions. Lastly, for two of the three Group II offices, the actions are generally done by employing external professional agents such as patent attorneys and marketing consultants.

Findings from non-discriminative queries

Assessment circumstances

Assessment circumstances refer to the conditions when invention assessments are performed. Five of the offices expressed that not every invention project are assessed. Most managers surveyed admitted that the perceived status and/or the perceived scientific capability of the inventor affect the rigor of the invention assessment process. Also, inventions are not necessarily assessed before patent application, although this is normally the case. The five offices stated that informal patent applications or formal patent applications prepared internally are often adopted without assessment in times of emergency, such as when the inventors are due to disclose the invention in public or due to meet potential licensees. Invention assessments are often performed for occasions when inventors require resources from the technology transfer office. The occasions are structured around the patent expenses schedule, including times prior to critical points in time like the first patent filing date ('Month 0'), 12 months away from 'Month 0', and 30 months from 'Month 0', and so on. These occasions are when inventors are looking for internal funding from the universities or seeking help from the technology transfer office to locate external funding.

Division of assessment staff

Division of assessment staff refers to whether technology transfer office staff responsible for invention assessment is divided into teams of scientific disciplines, such as dividing into the bioscience and engineering teams. The findings from the interviews show that except one Group I office that only target bioscience inventions, other 6 offices deal with inventions from different disciplines. The numbers of divisions for these offices and their group labels are shown in Table 4-15. It indicates that half of these offices do not adopt division of assessment staff.

Table 4-15: Division of assessment-staff

Group	1	1	1	2	2	2
Number of divisions	8	3	1	3	1	1

Perceptions

Several perception-related questions are asked during the interviews. The questions are regarding the manager's perception towards the followings:

- The managers' own expertises and responsibilities during invention assessment
- The inventors' expertise
- The role of invention disclosure form

The managers are often the sole assessor or one of the assessment panel members, and invention assessment is largely based on the assessors' subjective judgement. The aim of these perception-related queries is to find out how they perceive various issues related to invention assessment. The findings are presented below.

The seven managers generally perceive themselves as experts in judging an invention's potential in terms of two dimensions, namely the invention's technical novelty, and marketability.

They expressed that they rely on the inventors to provide the relevant technical and market information for the inventions. During invention assessment, they perceive their responsibilities to be collecting the maximum relevant information from the inventors, and conducting comparisons among invention projects to choose the higher potential inventions. They expressed that they are not necessarily familiar with the invention's field, the comparisons among inventions are therefore often based on non-technical aspects (aspects that do not involve specialised scientific knowledge), such as the *number* of existing prior arts, rather than the differences in *technical specifications* between the prior arts and the invention. A Group I office manager commented that it was impossible to choose a good invention as every invention often appeared good in the first place. He suggested that the strategy is to pick the relatively weak inventions, such as those with small market sizes.

The managers perceive the inventors as experts in the invention fields, but not necessarily familiar with intellectual property issues. Generally, the managers perceive themselves to be more knowledgeable than inventors in terms of intellectual property issues, yet they perceive inventors to be more knowledgeable in terms of the invention-specific technical and market issues.

The managers generally perceive invention disclosure form as a standard tool to collect from the inventors, technical invention information and signed documents for legal purposes. The managers expressed that the invention attributes contained in the invention disclosure form cover almost all the information they use during invention assessment. However, the interviewees often do not request inventors to spend time to provide many details on the invention disclosure forms. Some managers consider the form to be a tool just to better prepare the manager to extract information during later face-to-face meeting with the inventors.

Information storage and maintenance

The seven offices commonly store all the invention attribute information for all the invention projects. Two of them (one Group I and one Group II offices) use specialised database software for University Technology Transfer, another Group I office use general-purpose database software, the remaining offices do not use any software and only store the hard copies of the information.

Discussion

Discriminative Queries

The findings generated from the discriminative queries are summarised in Table 4-16, which shows that Group I offices are possibly more experienced and more confident with invention assessment. This is reflected in their class labels under Table 4-13 as well as their usage of simple assessment process based on personal judgement and the utilisation of internal staff. They also seem more self-contained and the inventions they dealt with are possibly of higher quality. These can be traced from the facts that the offices are able to self-fund, and the inventions often come with future funding secured and are subjected to more rigorous requirement.

Table 4-16 also shows that Group II offices are possibly more inexperienced and cautious with invention assessment. This is revealed by their class labels in Table 4-13, and the fact that they often draw on external resources. They are also less self-contained, since the offices generally rely on the government's third stream funding. The inventions they dealt with may be of lower quality, and their inventions rarely have secured external funding and often depend on internal funding.

Table 4-16: Summary of the findings from discriminative queries

Findings	Group One	Group Two
Table 4-13	Classes of Technology transfer performance: upper-class and unclassified	Classes of Technology transfer performance: middle-class and lower-class
	More experience regarding the use of Invention Disclosure form	New to the use of Invention Disclosure form
Background	Less initiative	More initiative
	Self-funding through the proceeds earned from technology transfer activities	University-funded and secured government funding for office expansion funded
Invention Assessment	Simpler process with more rigorous requirement	More complicated process without specific requirement
	1 or 2 assessors who are all office staff	Larger number of assessors consists of a variety of background

	Invention information derived from inventors and public search engines	Invention information solely derived from inventors
	Assessment using SWOT analysis or personal judgement	Assessment using scoring systems developed by external agents
Actions	Seek external funding	Use internal funding
	Utilise internal staff to minimise expenses	Employ external agents

Nonetheless, given the unavailability of performance data, the findings generated from Table 4-16 do not qualify as definitive evidences to prove that Group I offices are more experienced nor the inventions they dealt with are of higher quality. The most important conclusion derived from these findings, however, is the typology of technology transfer offices (Group I and Group II), which showed that members of different groups behave very differently. This difference further supports the assumption from the literature review that certain critical factors are non-transferable among universities, which therefore can only be learned from an office's own experience.

Invention Assessment Process

The two scoring systems employed by the two Group II offices can be viewed as a list of attributes organised into categories, where each category score is a summation of its attribute scores and the final score is a summation of all category scores. This is a subjective and abstract scoring process, where the accuracy and consistency both depend on the user.

Besides, Group I offices use general analysis methods such as SWOT analysis (Strength-Weakness-Opportunity-Threat analysis). Such analysis methods are indeed analogous to the scoring systems used by Group II offices, where the final decision is derived from a number of categories/groups of categories. Furthermore, the number of attributes for each category and the scores given are defined by the analyst. In other words, both the scoring systems used by Group II offices and the SWOT analysis used by Group I offices are based on a summation of attribute scores. They differ in terms of the attributes included, the organisation of attributes into categories, and the scores allocated to each attributes. Thus, the analysis method used by both groups of offices can be described as a formula of the final score, *final_score*, as shown in Equation 4-1.

Equation 4-1: Formula of the final score

$$final\_score = \sum_{1}^{n} s_1, \ldots s_i \ldots, s_n$$

Where *final_score* denotes the value of the final score, s_i represents the value of the score for the i-th category, and n represents the number of categories.

Since each category score is a summation of its attribute scores, s_i is therefore:

Equation 4-2: Category score

$$s_i = \sum_{1}^{k} x_1, \ldots, x_j, \ldots, x_k$$

Where x_j represents the value of the score for the j-th attribute, and k represents the number of attributes. The formula for the final score is therefore a function of x.

Furthermore, each attribute score is proportional to the value of the attribute. For instance, let one of the attribute be '*the number of publication by the inventor*' and that it is a positive attribute. If the values of this attribute for invention A and invention B are '5' and '10' respectively. This attribute score for invention A is therefore lower than that for invention B, such as having attribute score for invention A being 'medium' and that for invention B being 'good'. Thus, attribute score can be view as a factor of the attribute value.

Moreover, since there are more than one attribute, it is therefore not likely that the user considers every attribute to be equally important. More emphasis is thus given to the attribute scores of important attributes. This concept of favouring over certain attributes is reflected in the final score computation of the scoring system used by the two Group II offices, where the weights attached to certain category scores are higher than others.

Hence, the attribute score can be defined as:

Equation 4-3: Attribute score

$$x_j = w_j p_j$$

Where x_j, w_j and p_j denote the score value, weight value and attribute value for the j-th attribute respectively. The formula for the final score is therefore a function of w and p.

While the analysis methods of both Group I and Group II offices can be described using Equation 4-3, the difference between their methods lies in the two ingredients of Equation 4-3: the weight values attached to the attributes, and the list of attributes used.

The determination of the weight values is highly likely to be related to the learning process through trial and error, as explained below. Group I offices use methods like SWOT analysis, which is based on personal judgements rather than explicit weight values. The personal judgements are highly likely to be derived from the experience of the assessor, which is developed through learning by trial and error. Furthermore, Table 4-16 showed that Group II offices may be less experienced than Group I offices, which could be why Group II offices need to depend on scoring systems developed by external professional agents, who also probably developed the methods based on their knowledge gained from experience through years of trial and error.

Non-discriminative Queries

Based on the query 'Assessment Circumstances', it is found that the rigor of invention assessment is not evenly applicable to every invention project, and is dependent on the perceived status and/or the perceived scientific capability of the inventors. Therefore, evaluative bibliometric indicators reflecting researchers' scientific capability (such as publication counts and patent citations) may serve as good candidates for invention attributes used for invention assessment. Moreover, the interviewees explained that rounds of assessment are often structured around the patent expenses schedule. Those critical points in time for patent expenses, such as 'Month 0' and 'Month 12', therefore serve as appropriate data collection points for the input data for the CPS.

According to the findings from the query 'Division of assessment staff', half of the offices have only one assessor (manager), and most offices with divisions of staff have 3 divisions (previously shown in Table 4-15). Each manager therefore generally carries out invention assessment for all or one-third of the numerous disciplines within a university. However, it is highly unlikely that the manager is familiar with the knowledge involved such a large number of disciplines.

Furthermore, according to the findings from the query 'Perceptions', managers expressed that they rely on inventors to provide technical and market information of the inventions and they perceive the inventors to be more knowledgeable than them in terms of the invention-specific technical and market issues. They also described that the comparisons of inventions are based on non-technical aspects. such as the market size and the number of prior arts.

Based on the findings from the queries 'Division of assessment staff' and 'Perceptions', it is suggested that the invention attributes used for assessment purposes are often quantitative or categorically measurable. and non-technical (non-technical attributes refers to attributes that do not involve specialised scientific knowledge.).

According to the finding from 'Perceptions', the managers agreed that the information required by the invention disclosure form (ID form) covers almost all the information they use during invention assessment, though they often do not request inventors to spend time to provide much details on the ID form. This further supports the effort of the Invention Disclosure form survey and the questionnaire survey in identifying important invention attributes used by the ID form.

Lastly, a finding shared by the questionnaire survey and the interview survey showed that although less than half of the surveyed technology transfer offices use database software currently. the usage has been increasing with time. This trend increases the applicability of the CPS to the sector. since the data contained in the database can efficiently be used as the input data for the CPS.

4.5 CONCLUSIONS

The first two surveys identified the important invention attributes requested in the invention disclosure forms (ID forms) of about 30 universities. which are then combined to form a list of 'potential attributes for invention assessment'. Findings from the interview survey also confirmed that ID forms cover most of the information used for invention assessment purposes. Additionally. findings from the interview survey also suggest that evaluative bibliometric indicators reflecting inventors' scientific capability may serve as important invention attributes for invention assessment. Interview findings also suggest that the patent expenses schedule should be used for data collection for the final CPS.

Moreover, the interview findings showed that the assessment methods adopted by the interviewees to compute an invention's potential could be understood as formulas of weighted attributes. It is also showed that non-technical attributes that are quantitatively or categorically measurable could be used for the final CPS. Lastly, it is suggested that technology transfer offices are increasingly using database software to store invention information, which enhances the applicability of the CPS to the sector.

CHAPTER FIVE

COMMERCIAL OUTCOME PREDICTION SYSTEM

5.1 INTRODUCTION

This chapter presents the Commercial Outcome Prediction System (CPS), which is based on the findings derived from a number of background research, including literature review, prototype CPS, and three surveys with various technology transfer offices. The CPS is based on three well known classification related methods, namely decision tree induction, principal component analysis, and neural network analysis. This chapter starts with a description of the last two methods in Section 5.2 (the decision tree method has already been described in Chapter Three and therefore is not covered here). The chapter then presents the steps involved in the CPS in Section 5.3, and ends with a summary of the CPS in Section 5.4

5.2 PCA, AND NEURAL NETWORKS

5.2.1 Principal Component Analysis

Principal Component Analysis (PCA), also called the Karhunen-Loeve method, is among the oldest and the most widely used multivariate analysis techniques. It was originally introduced by Pearson (1901) and separately by Hotelling (1933). The main idea of PCA is to represent the variation of a data matrix with n original-attributes in terms of a set of n uncorrelated alternative-attributes, the so-called *principal components*. Each of the principal components is a particular linear combination of a proportion of each of the original-attributes. These principal components are derived in descending order of 'significance' – their account for the proportion of variation in the original data. For example, the first principal component (PC1) accounts for most of the variation in the original data, and the second principal component (PC2) accounts for more variation in the original data than every other principal component except PC1. This method is illustrated below.

By applying the PCA algorithm to a data matrix (**X**), three outputs are generated, as illustrated in Equation 5-1:

1. A matrix containing the principal components of **X**, denoted as $\mathbf{X}_{pca}$
2. A matrix containing the original-attribute coefficients for $\mathbf{X}_{pca}$, denoted as **E**.
3. A vector containing the variance of **X** represented by each principle components, denoted as **v**.

Equation 5-1: Definitions of matrices X and $\mathbf{X}_{pca}$

$$\mathbf{X}=\begin{pmatrix} x_{11} & x_{1j} & \cdots & x_{1n} \\ x_{i1} & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{kj} & \cdots & x_{kn} \end{pmatrix} \qquad \mathbf{X}_{pca}=\begin{pmatrix} x_{pca_{11}} & x_{pca_{1j}} & \cdots & x_{pca_{1n}} \\ x_{pca_{i1}} & x_{pca_{ij}} & \cdots & x_{pca_{in}} \\ \vdots & \vdots & & \vdots \\ x_{pca_{k1}} & x_{pca_{kj}} & \cdots & x_{pca_{kn}} \end{pmatrix}$$

$$\mathbf{E}=\begin{pmatrix} e_{11} & e_{1i} & \cdots & e_{1n} \\ e_{j1} & e_{ji} & \cdots & e_{jn} \\ \vdots & \vdots & & \vdots \\ e_{n1} & e_{ni} & \cdots & e_{nn} \end{pmatrix} \qquad \mathbf{v}=\begin{bmatrix} v_1 \ldots v_j \ldots v_n \end{bmatrix}$$

Where both **X** and $\mathbf{X}_{pca}$ are of dimension k by n, x_{ij} is the value of the j-th original-attribute for the i-th case, $x_{pca_{ij}}$ is the value of the j-th principal component for the i-th case. Furthermore, **E** are of dimension n by n, e_{ji} is the value of the i-th coefficient for the j-th principal component, v_j denotes the portion of variance of **X** represented by the j-th principal component.

For example, if n = 5, and $\mathbf{v}=\begin{bmatrix}0.6 & 0.3 & 0.05 & 0.03 & 0.02\end{bmatrix}$, it means the first and second principal components have covered 60% and 30% of the original variation respectively. In other words, only two principal components already represent most (90%) of the variation accounted by **X**.

The computation of $\mathbf{X}_{pca}$, which requires **E**, is explained as follows:

Equation 5-2: Definition of $x_{pca_{ij}}$

$$x_{pca_{ij}} = x_{i1}e_{j1} + x_{i2}e_{j2} + \cdots x_{ij}e_{ji} + \cdots + x_{in}e_{jn}$$

Where e_{j1} to e_{jn} denote the coefficients of the n original-attributes needed for the computation of $x_{pca_{ij}}$.

When $\mathbf{e}_j = \left[e_{j1} \ldots e_{ji} \ldots e_{jn} \right]$ and $\mathbf{x}_{pca_j} = \left[x_{pca_{1j}} \ldots x_{pca_{ij}} \ldots x_{pca_{kj}} \right]$, then:

Equation 5-3: Definition of $\mathbf{x}_{pca_j}$

$$\mathbf{x}^{\mathrm{T}}_{pca_j} = \mathbf{X}\mathbf{e}^{\mathrm{T}}_j$$

Where $\mathbf{x}_{pca_j}$ is a vector denoting the values of the j-th principal component.

Therefore, the values of each principal component can be computed if the respective $\mathbf{e}_j$ is known.

Since the variance of $\mathbf{x}_{pca_j}$ could be increased without limit simply by increasing the elements of $\mathbf{e}_j$, a restriction is placed on $\mathbf{e}_j$ that its sum-of-square ($\mathbf{e}^{\mathrm{T}}_j \mathbf{e}_j$) has to be set at a value of unity. Also, in order to create uncorrelated principal components, the sum-of-square of $\mathbf{e}_j$ and $\mathbf{e}_{j-1}$ must be zero.

As a result, the first principal component, $\mathbf{x}_{pca_1}$ is:

Equation 5-4: Definition of the first principal component

$$\mathbf{x}^{\mathrm{T}}_{pca_1} = \mathbf{X}\mathbf{e}^{\mathrm{T}}_1$$

Subject to the constraint: $\mathbf{e}_1\mathbf{e}^{\mathrm{T}}_1 = 1$

The j-th principal component, $\mathbf{x}_{pca_j}$ is defined as:

Equation 5-5: Definition of the j-th principal component

$$\mathbf{x}^{T}_{pca_j} = \mathbf{X}\mathbf{e}^{T}_{j}$$

Subject to the constraints: $\mathbf{e}_j\mathbf{e}^{T}_{j} = 1$, and $\mathbf{e}_{j-1}\mathbf{e}^{T}_{j} = 0$

By using the standard procedure for maximising a function of several variables subject to one or more constraints – Lagrange Multiplier, the variance of $\mathbf{x}_{pca_j}$ can be maximised, subject to the constraints(s) stated in the above definitions. This gives the result that $\mathbf{c}_j$ is the eigenvector of $\mathbf{CC}$ corresponding to the j-th largest eigenvalue[8], where $\mathbf{CC}$ is the correlation coefficient matrix of $\mathbf{X}$. The definition of the correlation coefficient matrix is explained below.

The correlation coefficient matrix consists of the correlation coefficients of individual pairs of attributes. Where the mean of one attribute (X) is denoted as μ_{X_i}, the covariance, cov_{ij}, of a pair of attributes (X_i and X_j) is: $\mu_{X_iX_j} - \mu_{X_i}\mu_{X_j}$. To avoid the problem of X_i and X_j being measured in different units, cov_{ij} is then standardised to become the correlation coefficient. The correlation coefficient, denoted as cc_{ij}, of this pair of attributes is:

Equation 5-6: Definition of correlation coefficient

$$cc_{ij} = \frac{cov_{ij}}{\sqrt{cov_{ii}cov_{jj}}}$$

[8] More details can be referred to Chatfield, C., and Collins, A.J., (1980).

5.2.2 Neural Networks

Similar to decision tree induction, neural networks are a classification method based on supervised learning. It is based on the assumption that there is an underlying unknown function governing the relationship between the attribute values and the class labels. A classifier is an approximation of such a function, with which the class labels of future inventions can then be predicted. By supplying an attribute matrix (which contains a number of attribute values for a number of cases) and a class vector (which contain the actual class label of a number of cases) to a neural network training algorithm, the relationship between the two can be learned during training, and a neural network classifier is produced. Given sufficient training, the resultant classifier will be able to predict the class labels of new cases given the availability of the respective attribute matrix.

In order to generate a classifier, a neural network has to be trained. During training, attributes of cases are supplied to the neural network and it predicts the class labels of these cases based on an initial classifier. The difference between the predicted and actual class labels suggests adjustment to the initial classifier. Iterative adjustment goes on until the initial classifier has been properly transformed. This training process is based on iterative operations of feed-forwarding and weight update to a network of neurons (processing units), which is presented below.

Network architecture

Network architecture is defined by the number of *hidden layers*, and the number of *neurons* in each layer. By having different network architectures, classifiers of varying performances (as measured in terms of classification accuracy and speed) can be generated. A typical two-layer network is shown in Figure 5-1:

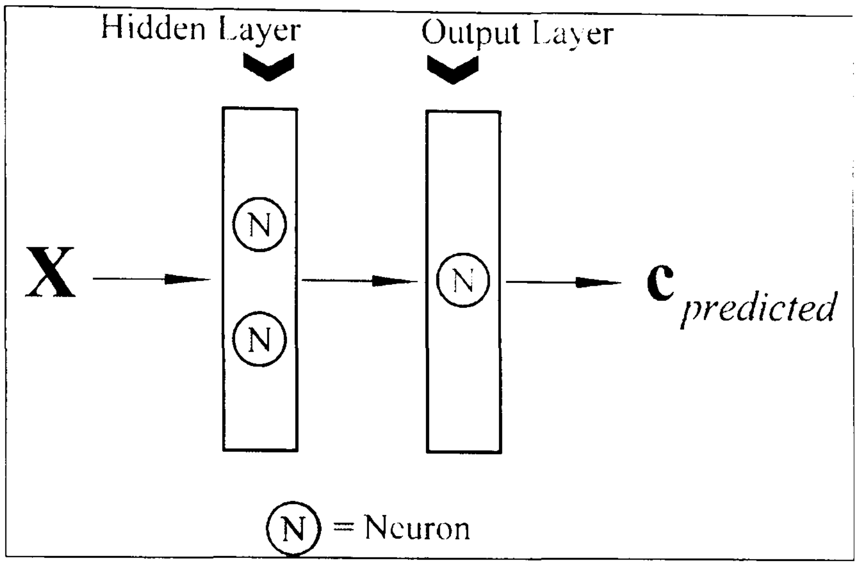


Figure 5-1: A neural network of 2-1 network architecture

As shown, this is a network connecting an attribute matrix (**X**), a hidden layer, and an output layer which computes the predicted class vectors ($\mathbf{c}_{predicted}$). While network definition varies, some may address this network as a three-layer network consisting of an input layer, a hidden layer, and an output layer. This research adopted the definition that a layer is defined by the existence of neurons. Without any neurons, the input (**X**) is therefore not regarded as a layer, and the network displayed is hence a two-layer network. This architecture of this network is denoted as 2-1, as in two neurons in the first layer and one neuron in the second layer.

Hidden layers are defined as the layers locating between the input and the output layer. While there is only one hidden layer in this example, there is no upper limit on the number of hidden layers. However, a network must contain at least one hidden layer. Lastly, there is always only one output layer with at least one output neuron. Again, there is no upper limit on the number of output neurons.

Figure 5-2 shows a standard neuron connected with an input and output. A neuron is defined by the following three elements: 1) weight vector, 2) bias value, 3) and transfer function.

Neuron Computation

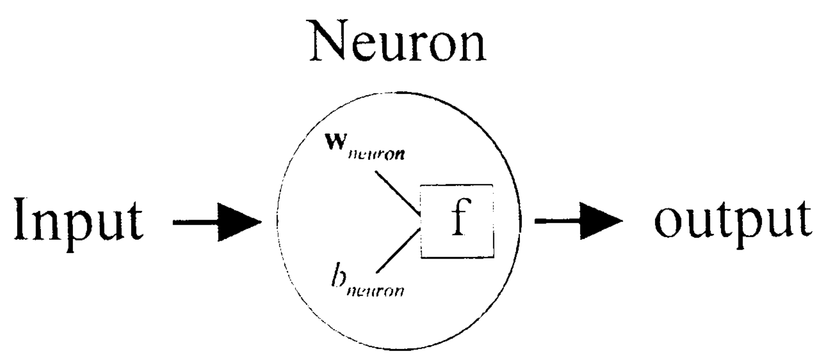


Figure 5-2: Neuron definition

Where $\mathbf{w}_{neuron}$ denotes the weight vector containing a weight for each attribute – $\mathbf{w}_{neuron} = \left[w_{neuron_1} \ldots w_{neuron_j} \ldots w_{neuron_n} \right]$, b_{neuron} represents the bias value, and f denotes the transfer function employed by the neuron.

These three elements jointly compute the neuron's output for a given input, as shown below:

Equation 5-7: Definitions of $\mathbf{X}_{neuron}$ and $\mathbf{y}_{neuron}$

$$\mathbf{X}_{neuron} = \begin{pmatrix} x_{neuron_{11}} & x_{neuron_{1j}} & \cdots & x_{neuron_{1n}} \\ x_{neuron_{i1}} & x_{neuron_{ij}} & \cdots & x_{neuron_{in}} \\ \vdots & \vdots & & \vdots \\ x_{neuron_{k1}} & x_{neuron_{kj}} & \cdots & x_{neuron_{kn}} \end{pmatrix}$$

$$\mathbf{y}_{neuron} = \left[y_{neuron_1} \ldots y_{neuron_i} \ldots y_{neuron_k} \right]$$

Equation 5-8: Formula to compute the output of a neuron

$$\mathbf{y}'_{neuron} = \mathrm{f}\left(\mathbf{X}_{neuron} \mathbf{w}^{\mathrm{T}}_{neuron} + b_{neuron} \right)$$

Here matrix $\mathbf{X}_{neuron}$ denotes the input matrix to the neuron. $x_{neuron_{ij}}$ represents the input value of the j-th attribute for the i-th case. vector $\mathbf{y}_{neuron}$ is the output computed by the neuron, and y_{neuron_i} is the output value for the i-th case. As shown in Equation 5-8, the neuron output ($\mathbf{y}_{neuron}$) can be computed by passing the summation of a linear combination of weighted inputs ($\mathbf{X}_{neuron}\mathbf{w}^{\mathrm{T}}_{neuron}$) and the bias ($b_{neuron}$) through a transfer function $\mathrm{f}(\cdot)$.

Network Computation – Feed-forward

While a neuron's output can be computed by Equation 5-8, a network's output is the joint effort of layers of neurons. For multiple-layer networks, the input to the current layer is the neuron output from the previous layer. Each neuron of the current layer then computes an output using Equation 5-8, which the neuron outputs in the form of an input for the next layer. This process goes on until the output layer computes the final output vector ($\mathbf{c}_{predicted}$). This network computation is presented below with respect to a 2-1 network.

Equation 5-9: Definitions of $\mathbf{X}_O$, $\mathbf{w}_O$, and $\mathbf{y}_O$

$$\mathbf{X}_O = \begin{pmatrix} x_{H1_1} & x_{H2_1} \\ x_{H1_i} & x_{H2_i} \\ \vdots & \vdots \\ x_{H1_k} & x_{H2_k} \end{pmatrix}$$
$$\mathbf{w}_O = \begin{bmatrix} w_{O_1} & w_{O_2} \end{bmatrix}$$
$$\mathbf{y}_O = \begin{bmatrix} y_{O_1} \cdots y_{O_i} \cdots y_{O_k} \end{bmatrix}$$

Equation 5-10: Formula to compute the output of an output-neuron

$$\mathbf{y}^{\mathrm{T}}_O = \mathrm{f}_O\left(\mathbf{X}_O\mathbf{w}^{\mathrm{T}}_O + b_O\right)$$

Where x_{H1_i} and x_{H2_i} are the values of the first and second hidden neuron output for the i-th case respectively. They then form matrix $\mathbf{X}_O$ – the input for the output layer, where there is only one output neuron in the case of the CPS. Vector $\mathbf{w}_O$, b_O, and f_O represent the weight vector, bias value, and transfer function of the output neuron, $\mathbf{y}_O$ is the output vector computed by the 2-1 network using Equation 5-10, which is equivalent to $\mathbf{c}_{predicted}$, given the network input matrix $\mathbf{X}$.

In the above illustration, the network input ($\mathbf{X}$) is fed through the hidden layer and the output layer, where the network output ($\mathbf{c}_{predicted}$) is computed. Such an output computation is called feed-forwarding, given a 1-1 network architecture, $\mathbf{c}_{predicted}$ is computed through Equation 5-11:

Equation 5-11: Formula to compute $\mathbf{c}_{predicted}$

$$\mathbf{c}_{predicted} = \mathrm{f}_O\left(\mathrm{f}_H\left(\mathbf{X}\mathbf{w}^{\mathrm{T}}{}_H + b_H\right)\mathbf{w}^{\mathrm{T}}{}_O + b_O\right)$$

Where f_H and f_O denote the transfer functions for the hidden layer and the output layer respectively, vectors $\mathbf{w}_H$ and $\mathbf{w}_O$ denote the weights for the hidden layer and the output layer respectively, b_H and b_O represent the bias values for the hidden layer and output layer respectively.

In order to simplify the notation for each neuron, the bias value has been inserted into the weight vector as the last element, and a unity vector has been inserted into the input matrix as the last column. These new definitions for weight vector ($\mathbf{w}_{Neuron}$) and neuron input ($\mathbf{X}_{Neuron}$) are represented below.

Equation 5-12: Definitions of $\mathbf{w}_{Neuron}$ and $\mathbf{X}_{Neuron}$

$$\mathbf{w}_{Neuron} = \begin{bmatrix} w_{neuron_1} \cdots w_{neuron_j} \cdots w_{neuron_n} & b_{neuron} \end{bmatrix}$$

$$\mathbf{X}_{neuron} = \begin{pmatrix} x_{neuron_{11}} & x_{neuron_{1j}} & \cdots & x_{neuron_{1n}} & 1 \\ x_{neuron_{i1}} & x_{neuron_{ij}} & \cdots & x_{neuron_{in}} & 1 \\ \vdots & \vdots & & \vdots & 1 \\ x_{neuron_{k1}} & x_{neuron_{kj}} & \cdots & x_{neuron_{kn}} & 1 \end{pmatrix}$$

When the definitions of $\mathbf{w}_{Neuron}$ and $\mathbf{X}_{Neuron}$ applies to neurons in both the hidden layer and output layer, Equation 5-11 hence becomes:

Equation 5-13: The shortened formula to compute $\mathbf{c}_{predicted}$

$$\mathbf{c}_{predicted} = \mathrm{f}_O\left(\mathrm{f}_H\left(\mathbf{X}\mathbf{w}^{\mathrm{T}}{}_H\right)\mathbf{w}^{\mathrm{T}}{}_O\right)$$

Training

After one operation of feed-forwarding, if the predicted class labels ($\mathbf{c}_{predicted}$) are very similar to the actual class labels ($\mathbf{c}_{actual}$), it means the network has approximated the underlying function governing $\mathbf{X}_{predicted}$ and $\mathbf{c}_{actual}$ successfully. The particular specification of network architecture, transfer functions and weight vectors are therefore correct. This, however, is often impossible, as the weight vectors are merely random values. Training, using a process of iterative weight update, is therefore needed.

Weight Update by Gradient Descent

During training, $\mathbf{c}_{predicted}$ is first computed by feed-forwarding. $\mathbf{c}_{predicted}$ is then compared with $\mathbf{c}_{actual}$, and any difference between them represents a classification error. In order to concentrate on the magnitude of the error instead of the sign of the error, the error function widely adopted is the average of the error squared.

In order to minimise the error, the network adjusts each weight vector in the direction that the error function is decreasing most rapidly, which is where the gradient of the error function is negative. This weight update process is therefore called gradient descent.

At the next iteration (also called epoch), feed-forwarding is performed and $\mathbf{c}_{predicted}$ is computed with the adjusted weight, the error is then computed and the weight update is performed again.

Convergence and Learning Rate

While weight update is accomplished by adjusting the weight vector based on the gradient descent, the amount of adjustment is the learning rate. This reflects the magnitude of each weight update. With the learning rate, l, the weight change is therefore defined as:

Equation 5-14: Formula to compute new weights

$$\mathbf{w}_{e+1} = \mathbf{w}_e - l \cdot g_e$$

Where $\mathbf{w}_e$ and $\mathbf{w}_{e+1}$ denote the weight vectors for the e-th epoch and the $e+1$ epoch respectively, l denotes the learning rate, and g_e represents the gradient of the error function at the e-th epoch.

With a proper learning rate, the error function will eventually converge to a minimum (reaching the error goal), given a sufficient number of epochs.

Weight update algorithms (also called training algorithms) are therefore adopted to speed up the convergence of basic backpropagation. The basic backpropagation algorithm is a fundamental training algorithm from other training algorithms were developed upon.

Backpropagation rule

The weight update for the output layer involves comparing $\mathbf{c}_{predicted}$ (the layer output) with $\mathbf{c}_{actual}$ (i.e. the proper output). In other words, by knowing the proper output of the layer, the layer weights can be updated accordingly.

The same principle applies to the hidden layers. Hidden layer weights can be updated if the proper layer output of the hidden layer is known. By means of the backpropagation algorithm, the proper output of the hidden layer can be calculated. The term 'backpropagation' is based on the process that during training, an error must be propagated from the output layer back to the hidden layer in order to perform update of the hidden layer weights.

Backpropagation is a natural extension of the Least Mean Squared (LMS) algorithm (also called the Widrow Hoff learning rule). The LMS rule is used for the single-layer networks to evaluate the error for each output neuron, where the error is proportional to the square of the difference between the predicted output and the actual output. Backpropagation works by generalising the LMS rule to multiple-layer networks. It is based on the application of the calculus chain rule for continuous functions, which allows the computation of derivatives of the error function with respect to all neuron weights. The first part of the backpropagation rule (Equation 5-15) is the computation of the output layer weight update, and the second part (Equation 5-16) is the hidden layer weight update. The thorough derivation of the rules can be found in Simon (1999), the final algorithms are given below.

Equation 5-15: The first part of backpropagation rule

$$\mathbf{w}_{O_{e+1}} = \mathbf{w}_{O_e} + l\left(\mathbf{c}_{actual} - \mathbf{c}_{predicted_e}\right)\mathbf{y}_{H_e}\, \mathbf{f}'_O\left(\mathbf{y}_{H_e}\mathbf{w}^{\mathbf{T}}{}_{O_e}\right)$$

Equation 5-16: The second part of backpropagation rule

$$\mathbf{w}_{H_{e+1}} = \mathbf{w}_{H_e} + l\left(\mathbf{w}_{O_e}\left(-\frac{\partial R_e}{\partial\left(\mathbf{y}_{H_e}\mathbf{w}^{T}{}_{O_e}\right)}\right)\right)\mathbf{f}'_H\left(\mathbf{X}\mathbf{w}^{\mathbf{T}}{}_{H_e}\right)\mathbf{X}$$

Where $\mathbf{w}_{H_e}$ and $\mathbf{w}_{H_{e+1}}$, $\mathbf{w}_{O_e}$ and $\mathbf{w}_{O_{e+1}}$ denote the hidden layer weight vectors for the current epoch and the next epoch, and the output layer weight vector for the current epoch and the next epoch respectively. l denotes the learning rate, $\mathbf{y}_{H_e}$ denotes the neuron outputs from the hidden layer. $\mathbf{f}_O$ and $\mathbf{f}_H$ denote the transfer functions for the output neurons and the hidden neurons respectively. R denotes the squared error of the current epoch.

Gradient Descent implementation

Training Mode

Gradient descent can be implemented in two different ways: incremental mode and batch mode. The incremental mode updates weights after the feeding of each case of the training data, while the batch mode performs weight updates only after all the training data cases are fed to the network in the training data. In the case of batch mode, the weight update is based on the summation of the gradients calculated at every training case.

Momentum

Momentum is a technique used to speed up batch training convergence. The concept is loosely based on the notion in physics that objects move slower when acted upon by outside forces. It is analogous to a low-pass filter in digital signal processing, in a way that momentum smoothens the error function by causing the networks to respond to the local gradient while ignoring small features in the error surface.

The formalisation of momentum is to include a momentum fraction, *mf*, to the weight update function, where *mf* represents a fraction of the last weight change. The weight change with momentum is therefore:

Equation 5-17: Weight change defined by momentum

$$\Delta \mathbf{w}_{e+1} = mf \Delta \mathbf{w}_{e-1} + \Delta g_e$$

Subject to the constraint of 1>= *mf*>=0 .

Where $\Delta \mathbf{w}_{e+1}$, $\Delta \mathbf{w}_{e-1}$, and Δg_e denote the weight change for the next epoch, the weight change for the last epoch, and the gradient change suggested by the backpropagation rule respectively.

The impact of the last weight change is hence mediated by *mf*. In extreme cases, the new weight change completely depends on the last weight change if *mf* = 1; and solely based on the gradient computed by backpropagation if *mf* = 0.

Advanced Training Algorithms

The following four training algorithms are proven to be ten to a hundred times faster than the basic backpropagation with momentum. They are:

1. Adaptive learning rate backpropagation
2. Resilient backpropagation
3. Conjugate gradient
4. Levenberg-Marquardt

Among these, the first two algorithms are heuristic techniques that still depend on the gradient of the error function to determine weight update, whereas the third and fourth algorithms are mainly based on standard optimisation techniques. Each of them is presented briefly below.

Adaptive learning rate backpropagation

The basic gradient descent function works with a constant learning rate. However, as discussed earlier, too high a learning rate may give rise to an unstable performance, whereas too low a learning rate can lead to slow convergence. Yet, keeping the learning optimal but constant is in practice infeasible because the learning rate changes during training.

Adaptive learning rate backpropagation overcomes this problem by producing a learning rate that reacts to the local error surface. Rather than using a constant learning rate, it updates the learning rate at each epoch. The learning rate is increased whenever the error of the current epoch is less than the error of the last epoch, and the learning rate is decreased when the current error exceeds the previous error. The learning rate is therefore constantly getting larger under stable performance, but is decreased whenever the error increases.

Resilient backpropagation

As shown in Figure 5-3, the gradient of a sigmoid transfer function approaches zero when the input approaches extreme values.

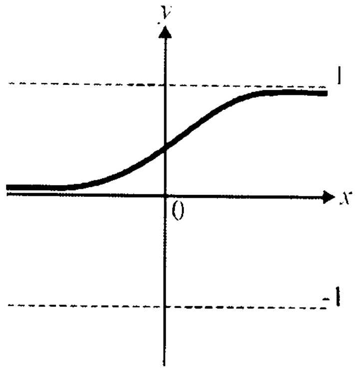


Figure 5-3: The graph of a sigmoid transfer function

Where $y = f(x)$, x denotes the input to the $f(\cdot)$ sigmoid transfer function, and y denotes the output of the transfer function.

After scaling using the sigmoid function, the differences among extreme-valued inputs become very small, the gradient descent would in turn produce weight updates of very small magnitude while the weight vectors can be far from optimal. The convergence is thus slowed down.

Resilient backpropagation fixes this problem by ignoring the gradient magnitude of the error function completely, with the weight update value only depending upon the sign of the error function's derivative. The weight update value is increased every time the sign of the error function remains the same for two epochs, and is decreased if the sign of the error function is different from that of the previous epoch. Finally the weight update is discarded if the error gradient is zero. The weight update hence only gets larger during stable performance.

Instead of repeatedly relying on the error function's gradient to compute weight update, the following two algorithms, conjugate gradient and Levenberg-

Marquardt algorithms, employ numerical optimisation techniques to speed up convergence.

Conjugate gradient

Conjugate gradient training methods start the first weight update by following the negative gradient of the error function. subsequent weight updates are then determined by searching along the conjugate gradient directions by means of optimisation methods. The first weight change is hence:

Equation 5-18: The first weight change for conjugate gradient training

$$\Delta \mathbf{w}_1 = -g_e$$

The weight update of each subsequent iteration is computed by adding a value that is determined by a line search (α_e) along the current direction (γ_e):

Equation 5-19: The formula of subsequent weight change

$$\Delta \mathbf{w}_{e+1} = \alpha_e \gamma_e$$

Where γ is computed by the summation of the new steepest descent direction (-g) and the previous direction (γ_{e-1}):

Equation 5-20: The definition of γ

$$\gamma_e = -g + \beta \gamma_{e-1}$$

Where β is a constant that can be determined through various algorithms, two of which are adopted in the CPS and are shown below:

1. Fletcher-Reeves update algorithm (Hagan et al. 1996) – β is the ratio of the norm squared of the current gradient, to the norm squared of the previous gradient.
2. Polak-Ribi閕e update algorithm (Hagan et al. 1996) – β is the inner product of the previous change in the gradient, with the current gradient divided by the norm squared of the previous gradient.

Levenberg-Marquardt Algorithm

The Levenberg-Marquardt algorithm (Hagan et al. 1996) shares the characteristic of high speed convergence with Newton's method, without the expensive computation of the Hessian matrix (the second-order derivatives matrix summation of the squared error function). Newton's method provides fast convergence by means of the Hessian matrix. The method is, however, computationally very complex and expensive for backpropagation neural networks as it requires computing, storing and inverting the n by n Hessian matrix (given a network with n weights).

The Levenberg-Marquardt training algorithm overcomes the drawback of Newton's method by approximating the Hessian matrix using the Jacobian matrix. The approximated Hessian matrix (**H**) is simply:

Equation 5-21: Definition of the approximated Hessian matrix

$$\mathbf{H} = \mathbf{J}^{\mathrm{T}}\mathbf{J}$$

Where the gradient g is:

Equation 5-22: Definition of the error-gradient in Levenberg-Marquardt algorithm

$$g = \mathbf{J}^{\mathrm{T}}\mathbf{r}$$

The $\mathbf{J}$ is the Jacobian matrix that contains first derivatives of the classification errors, and $\mathbf{r}$ is a vector of network errors.

According to the Levenberg-Marquardt algorithm, the new weight is defined as:

Equation 5-23: Formula to compute new weight using Levenberg-Marquardt algorithm

$$\mathbf{w}_{e+1} = \mathbf{w}_e - \left(\mathbf{J}^{\mathrm{T}}\mathbf{J} + \varepsilon\mathbf{I}\right)\mathbf{J}^{\mathrm{T}}\mathbf{r}$$

Where $\mathbf{I}$ is an identity matrix.

If the approximate Hessian matrix is replaced with the actual Hessian matrix and that scalar ε is zero, this definition is actually equivalent to the Newton's method. When ε is large, the definition becomes gradient descent with a small size of weight change. In order to achieve fast convergence, ε is decreased whenever the error function decreases and is increased if the error function increases. Hence, the error function will always be reduced at each iteration of the algorithm.

5.3 COMMERCIAL OUTCOME PREDICTION SYSTEM

This section presents the Commercial Outcome Prediction System (CPS). The CPS is designed to predict the likely future monetary return of an invention. It is also designed to be used by university technology transfer offices for invention assessment purposes.

Briefly speaking, the CPS works by identifying and learning discriminative invention attributes from historical invention cases, from which it produces a classifier which is capable of commercial outcome prediction. The CPS consists of three steps, namely input data set specification, data reduction, and classifier generation. These steps are explained sequentially in the following sections.

5.3.1 Input Data Set Specification

The first step of the CPS is to prepare the input data set, which is comprised of an attribute matrix ($\mathbf{X}_2$) and a class vector ($\mathbf{c}_2$), as illustrated in Equation 5-24. The attribute matrix contains the values of a number of (n) invention attributes for a number of (k) historical invention cases, in which the matrix consists of rows of cases and columns of attributes. The class vector contains the class labels of these invention cases.

Equation 5-24: Definitions of the input data set for the CPS

$$\mathbf{X}_2 = \begin{pmatrix} x_{11} & x_{1j} & \cdots & x_{1n} \\ x_{i1} & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{kj} & \cdots & x_{kn} \end{pmatrix} \qquad \mathbf{c}_2 = \left[c_1 \ldots c_i \ldots c_k\right]$$

Where x_{ij} represents the value of the j-th attribute for the i-th case, n is the number of attributes, and k is the number of cases.

The attributes to be recorded in the attribute matrix will from now on be called the CPS attributes. The CPS attributes are based on the data available, which can be sought from sources such as public databases, and the database of the user technology transfer office (TTO database). Since the data stored or retained in the TTO database varies from office to office, the CPS attributes also vary from office to office.

A number of invention attributes potentially important for invention assessment have been identified in Chapter Two, Three and Four, which covers attributes widely used by various technology transfer offices, and attributes available from public databases. These attributes (will from now on be called 'potentially predictive attributes') represent important attributes for invention assessment purposes, and will be presented below with Box 5-1 and Table 5-1. The list of potentially predictive attributes, however, does not represent an exhaustive list of

CPS attributes, but represents a number of attributes which are proposed to be potentially critical for invention assessment based on the findings generated in Chapter Two, Three and Four. While the data for some of the potentially predictive attributes are available from public databases on the Internet, some are based on confidential invention information which would only be available from a university's database. Additionally, the TTO database may contain useful invention attributes which are not covered in Table 5-1. In order to obtain maximum number of attributes for data mining purpose, the CPS attributes therefore include as many invention attributes as possible, based on the potentially predictive attributes listed in Table 5-1 that are available, and any other additional invention attributes available from the TTO database as well as attributes considered as appropriate by the user technology transfer office.

The potentially predictive attributes are based on the findings derived from the surveys / studies detailed in Chapter Two, Three, and Four. These findings are briefly restated below in Box 5-1.

Box 5-1: Findings leading to the potentially predictive attributes

The literature survey in Chapter Two identified five transferable critical factors as important basis to develop invention attributes for invention prediction. These five transferable critical factors are:

1. Significant inventor's scientific capacity and reputation
2. Invention discipline
3. The inventor's / university's personal relationship with firms or investors
4. Developmental stage of the invention
5. Secured external funding

The prototype CPS (as presented in Chapter Three) identified five invention attributes to perform classification, and four out of which are attributes based on the five transferable critical factors. These five attributes are:

1. The level of expense consumed by the invention
2. Whether there are industrial inventors
3. The number of personal contacts for potential source of investment
4. Whether funding for the next stage has been secured
5. The number of other patent applications owned by the inventors.

Furthermore, the first two background surveys presented in Chapter Four identified the potential attributes for invention assessment, and they were listed in Table 4-8. In addition, the third background survey in Chapter Four found that only quantitative / categorically measurable, and non-technical variables were used by technology transfer managers for invention assessment. Thus, only quantifiable and non-technical variables from Table 4-8 are adopted here. These variables are reproduced from Table 4-8 and are listed below:

1. The number and employment status of contributors and inventors
2. Support source – source of funding and in-kind support lead to the invention
3. Company list – a list of potentially interested firms (contacted)
4. Uses – commercial uses and application of the inventions
5. Drawing attachment – drawings and sketches relating to the explanation of the invention are to be attached
6. Contractual agreements – the number of contractual agreements linked with the invention, such as MTA
7. Competitor list – the list of potentially competing existing products
8. Existing solution – publicly acknowledged existing solution
9. Records – record details of the prior arts such as patent number and verbal records
10. Developmental stage – whether the invention is an early-stage invention, a proven concept, or a prototype
11. Search – whether a prior art search has been conducted
12. Current activity – activities regarding invention exploitation, such as current commercial interest and marketing activity.
13. Market size – target market size

14. Further development – whether further development is needed and the funding requirement for that, and whether such funding has been secured
15. Country list – the list of countries where the invention is likely to be used
16. Involvement of other parties during invention development
17. Contacts for potential third parties provided by inventors

The findings from the third survey in Chapter Four also showed that bibliometric indicators reflecting an inventor's status and scientific capability are potentially useful for invention assessment.

Moreover, the data collection experience of the prototype CPS showed that the use of ambiguous critical factors had resulted in the needs to develop coding system and to use subjective judgement, giving rise to non-reproducible data. Thus, it is important to specify explicit coding definition for each attribute. By developing explicit coding definitions for the findings in Box 5-1, a list of potentially predictive attributes is developed and showed in Table 5-1, and they are organised into seven categories.

Table 5-1: Seven categories of potentially predictive attributes

Category I: Inventor (Bibliometric indicators estimating the inventor's capability)
The number of publications (journal/conference) authored by inventor A[9]
The number of journal papers published by inventor A
The number of journals where inventor A had published

[9] The number of inventors involved in each invention is different. In order to collect data for all/some of the inventors, each attribute denoted with 'inventor A' will be collected *n* times, resulting in *n* attributes, the value of *n* is dependent on the user preference. For instance, if the invention, among other inventions in the university's portfolio, having the maximum number of inventors has 7 inventors, then the value of *n* can be any integer in the range of 1 to 7. The user may collect the attribute 7 times for the 7 inventors for each invention in the portfolio. This however, generates 6 empty attributes for each invention that has only 1 inventor. Alternatively, the user can collect the data once for the chief inventor.

The number of countries where inventor A had published
The total number of co-authors associated with inventor A's publications
The total number of class/subject codes covered by inventor A's publications
The total number of control vocabularies used by inventor A's publications
The year when inventor A first published
The year of the latest publication authored by inventor A
The sum of the normalised citation values[10] received by inventor A's publications
The sum of the normalised impact factors[11] of the journals where inventor A had published
The total amount of research grants (from private / public sources) awarded to inventor A
The number of research grants awarded to the inventor per year
The number of research grants awarded to inventor A
The number of year when inventor A's research was funded by research grants
Inventor A's rank position, among other academics of the university, in terms of the total amount of research grants awarded
The number of patent applications (invented) by inventor A
The number of US patent applications (invented) by inventor A
The number of granted patent (invented) by inventor A
The number of granted US patent (invented) by inventor A
The total number of normalised patent citations[12] received by inventor A's patents
The total number of international patent classes (based on the first 3 digits) where inventor A's patents belong
The total number of international patent classes (based on the first 5 digits) where inventor A's patents belong
The total number of patents existed in the international patent classes where

[10] Normalised citation value refers to the citation received by a publication divided by the average citation received by other publications of the same journal.
[11] Normalised impact factor refers to the impact factor achieved by a journal divided by the average impact factor achieved by other journals of the same subject/sub-field.
[12] Normalised patent citation refers to the citation received by a granted patent divided by the average citation received by other patents of the same international patent class.

inventor A's patents belong
The total number of external collaborating organisations of inventor A's patented/patent-filed inventions
The status of inventor A (student, internal academic staff, external academic staff, industrial personnel[13])
The annual income of inventor A
The age of inventor A at 'Month 0'
The number of years that inventor A had worked in the private sector at 'Month 0'
The number of industrial collaborative projects that inventor A had been involved in previously at 'Month 0'
Whether inventor A had previous experiences of setting up own business at 'Month 0'
Category II: Basic Information (Information relating to the invention)
The year when the invention project ended
The year of 'Month 0'
Life of the invention project (in years), which is the difference between the previous two attributes
Whether the invention project is active or not
Whether the first / primary source of funding leading to the current development of the invention is from the private sector
Whether the second / secondary source of funding leading to the current development of the invention is from the private sector
The first (the major) academic department where the invention belongs[14]
The number of academic departments which had collaborated to generate the invention

[13] A number of dummy variables are generated for this attribute. For instance if the status value for all the invention cases is either 'internal academic staff' or 'industrial personnel', then 2 dummy variables will be generated. The value for the dummy variables is either '1' or '0'.
[14] A number of dummy variables can be created for this attribute. This is analogous to the arrangement detailed in footnote 11.

The number of external organisations which had collaborated to generate the invention
The number of personal contacts provided by the inventor as potential licensees or investors
Whether the developmental stage of the invention is proven design made using manufacturing tooling
Whether the developmental stage of the invention is proven final design
Whether the developmental stage of the invention is optimisation by prototype variation
Whether the developmental stage of the invention is prototype
Whether the developmental stage of the invention is proof of concept
Whether the developmental stage of the invention is early stage
Whether an invention disclosure form has been completed
The number of pages of the completed invention disclosure form
Whether a market research has been conducted
The number of pages of the market research report
Whether the market research was produced by external organisation
Whether the market research was produced by internal staff
Whether a prior art search has been conducted prior to patent application
Whether the prior art search was produced by external organisation
Whether the prior art search was produced by internal staff
Whether an estimation for future costing has been done
The amount of funding required by the next point in time for data collection
The number of pages of the future costing estimation report
The number of external inventors
The number of inventors who are staff of external companies
The number of firms that have contacted the technology transfer office/inventor and have shown a positive interest on the invention
The number of licensing deals in the process of negotiation
The number of spin-out companies planned

Category III: Communication
The number of pages of communication (such as email, letter, phone message, and so on) from the inventors to the technology transfer office (TTO)
The number of pages of communication from the TTO to the inventors
The number of pages of communication to the TTO/inventors from external organisations which are potential customers or investors
The number of external organisations involved in the communication to the TTO/inventors
Category IV: Expenses
Patent/patent related service expenses consumed by the invention project
Total expenses consumed by the invention project
Category V: Agreements
The number of internal agreements (draft)
The number of internal agreements (signed)
The number of pages of internal agreements (draft)
The number of pages of internal agreements (signed)
The number of external agreements (draft)
The number of external agreements (signed)
The number of pages of external agreements (draft)
The number of pages of external agreements (signed)
The number of external collaborating organisations involved in the signed agreements
The number of external investor organisations involved in the signed agreements

Category VI: Technical documents
The number of internal technical documents (such as 'technical summary of the invention')
The number of authors involved in the internal technical documents
The total number of pages of the internal technical documents
The total number of figures of the internal technical documents
The total number of references of the internal technical documents
The number of patent specification drafts
The total number of pages of the patent specification drafts
The total number of figures of the patent specification drafts
The total number of references of the patent specification drafts
The total number of invention-relevant[15] external publications authored by the inventor
The total number of authors involved in such external publications
The total number of pages of such external publications
The total number of figures of such external publications
The total number of references of such external publications
The number of invention-relevant granted-patents (invented) by the inventors before 'Month 0'
Category VII: Patent documents (patent documents for the invention)
The number of patent applications filed
The number of ex-PCT patent applications filed
Whether an ex-PCT patent application is filed in country A[16]

[15] Since the relevancy of a publication to the invention requires expert knowledge of the invention field, a publication is only considered relevant when it is provided / supported by the inventor.

[16] (This is analogous to the arrangement for 'inventor A' which was explained earlier in footnote 7.) The number of ex-PCT patent applications of each invention is different. In order to collect data for all/some of the ex-PCT patent applications, each attribute denoted with 'country A' will be collected *n* times, resulting in *n* attributes, the value of *n* is dependent on the user preference.

The number of international patent classes (based on the first 3 digits) where the patent application filed in country A belongs
The number of international patent classes (based on the first 5 digits) where the patent application filed in country A belongs
The number of international patent classes (based on all digits) where the patent application filed in country A belongs
The total number of patents existed in the international patent classes (based on all digits) where the patent application filed in country A belongs
The number of times the patent application filed in country A is cited
The number of references cited by the patent application filed in country A
The average distance (in year) between the references cited by the patent application filed in country A
The number of foreign references cited by the patent application filed in country A
The number of patents granted
The year when the priority patent application was granted
The number of claims contained in the priority patent application
The average number of claims contained in other patent applications (patent applications other than the priority patent application)
The number of figures contained in the patent applications
The number of prior arts stated in patent office's search reports of the patent applications

For the convenience of the acquisition of the potentially predictive attributes, an 'Innovation Evaluation Form' has been formulated. Such form, as displayed in Appendix A5.1, could be used as an efficient information gathering tool by the CPS user for the purpose of building an input data set for the CPS. Since the current practices of invention assessment are largely intuitive, the Innovation Evaluation Form also acts as a bridge linking the qualitative norms of technology transfer offices and the quantitative requirements of the CPS. Effectively, the

For instance, if the invention having the maximum number of ex-PCT applications has 10 ex-PCT patent applications, then the value of n can be any integer in the range of 1 to 10.

content requested by the form should contain both qualitative and quantitative information, where the content should be comprised of a sufficient level of details where the data for the potentially predictive attributes can be sought. However, it should be noted that this form is not equivalent to a conventional invention disclosure form where issues regarding due diligence and patentability may be included. Designed within the scope of the CPS, the Innovation Evaluation Form is therefore only applicable to patentable inventions developed by inventors from the user university.

Data Collection Points in time

According to the findings from the third survey in Chapter Four, invention assessments are generally carried out in accordance with the patent expense schedule. The value of each of the CPS attributes is therefore to be recorded at a number of points in time, namely 'Month 0', 'Month 12', 'Month 30', 'Month 36', 'Month 48', and so on until the end of an invention case's life. 'Month 0' is defined as the first patent filing date of an invention case. Other points in times are titled according to the number of months away from 'Month 0'. For instance, 'Month 12' refers to 12 months from 'Month 0'. After 'Month 30', attribute data is to be recorded at the end of each year. These points in time are structured according to the patent expense schedule. Briefly speaking, 'Month 0' requires patent filing related expense in the United Kingdom, 'Month 12' requires PCT (patent cooperation treaty) related expense, 'Month 30' requires patent filing related expense in other countries, 'Month 36' onward require other expenses including examination fees and renewal fees. The recording of attribute values at these points in time is based on the rationale that invention assessment is often performed when expenses are required.

The attribute value recorded at 'Month 0' is the start value of an attribute. Attribute values recorded at the remaining points in time are the marginal values accounting from the previous points to the current points. For example, the value of attribute 'Expense consumed by the invention project' recorded at 'Month 30' is the value of the increase in expense from between 'Month 12' and 'Month 30'.

The class label refers to the final technology transfer outcome of an invention, which is represented as either 'success' or 'failure'. The outcome is measured in terms of the monetary return generated by the invention. The range of monetary return represented by each class label is determined by the user of the CPS.

5.3.2 Data Reduction

This step generates a reduced attribute matrix by applying data reduction operations to the original attribute matrix ($\mathbf{X}_2$). The goal is to reduce the volume of attribute data to be processed in the last step of the CPS (classifier generation), in a way that the resultant distribution of the reduced matrix remains almost unchanged, or very close to the original matrix. Instead of developing a reduced matrix using alternative attributes, the methods used in this step generate a reduced matrix using the original attributes. The benefits of using a reduced attribute matrix for classifier generation are:

- Faster algorithm calculations.
- An improvement in the classification accuracy by the removal of irrelevant attributes (noise).
- The production of a more understandable/presentable result, since there are fewer attributes involved.
- Fewer future attributes would need to be collected for classification.

Nonetheless, the necessity of data reduction depends on the size of the attribute matrix. Data reduction is more important for a larger attribute matrix, yet may do more harm than good for a small attribute matrix. As the number of class increases, more attributes would be needed to discriminate the finer differences.

Thus data reduction is more applicable to input data set consisting of large attribute matrix and small number of class. As a rule of thumb, the data reduction step of the CPS is performed when the attribute-class ratio is larger than 10. Since 2 classes are adopted here, the data reduction step is performed when the attribute matrix contains more than 20 (i.e. 20/2 = 10) attributes.

Two methods are employed to generate a reduced attribute matrix, namely decision tree induction, and principal component analysis. Subsets of original attributes are identified by each method, which are then combined to produce a reduced attribute matrix. The decision tree method is chosen due to its non-metric nature, as well as its capability of identifying critical attributes. Principal component analysis is employed because of its widely proven capability of projecting high-dimensionality data onto a much lower dimensional space. These methods are explained below.

Decision Tree

A decision tree is generated from the input data set using the impurity measure Gini Index, and the following parameters: monothetic decision tree, a branch factor of two, and a Minimum Case Threshold of two. The attributes appear on the nodes of the resultant tree then form an attribute list, denoted as L_{tree}.

Information regarding the details of the decision tree classification method can be referred to Chapter Three. Since the impurity measure Gini Index was not explained in Chapter Three, it is presented below.

Gini Index is one of the most popular impurity measure adopted for classification, it is therefore used to generate a decision tree from the input data set. Let $i(N)$ denote the impurity of node N. $i(N)$ equals 0 if all the cases that reach the node bear the same class label, and $i(N)$ will be maximum if the both classes are equally represented. The Gini Index is defined in Equation 5-25:

Equation 5-25: Gini Index

$$i(N) = -\sum_{i \neq j} \hat{P}(c_i)\hat{P}(c_j) = 1 - \sum_j \hat{P}^2(c_j)$$

Where $\hat{P}(c_i)$ and $\hat{P}(c_j)$ are the fractions of cases at node N that belong to class c_i and c_j respectively. Since only two classes are used in the CPS, the Gini Index used would be in its simplest polynomial form, as shown in Equation 5-26:

Equation 5-26: Gini Index in its simplest polynomial form

$$i(N) = \hat{P}(c_1)\hat{P}(c_2) \quad .$$

Where $\hat{P}(c_1)$ and $\hat{P}(c_2)$ are the fractions of cases at node N that belong to class c_1 and c_2 respectively.

Principal Component Analysis

Data Normalisation

Due to the use of attributes measured in different units, the first step is to standardise each column of the attribute matrix so that each have unit variance. This is done by computing the standardised Z scores for each column of the attribute matrix ($\mathbf{X}_2$). This process is illustrated below.

Equation 5-27: Definition of the standardised attribute matrix

$$\mathbf{X}_s = \left(\frac{\mathbf{X}_2 - \mu_X}{\sigma_X} \right)$$

Where $\mathbf{X}_s$ is the standardised version of matrix $\mathbf{X}_2$. μ_X is the arithmetic mean of $\mathbf{X}_2$, and σ_X is the standard deviation of $\mathbf{X}_2$.

Computation of Principal Components

Next, the principal component analysis algorithms (presented in Section 5.2) are applied to the standardised matrix ($\mathbf{X}_s$), which produces three outputs as listed below:

1. A matrix containing the principal components of $\mathbf{X}_s$, denoted as $\mathbf{X}_{s\_pca}$
2. A matrix containing the original-attribute coefficients for $\mathbf{X}_{s\_pca}$, denoted as $\mathbf{E}$
3. A vector containing the variance of $\mathbf{X}_s$ represented by each principle components, denoted as $\mathbf{v}$.

PCA-Based Data Reduction Operations

Principal component analysis is generally used to reduce data size by first computing the principal components, and then using a subset of which to represent the original data. However, since each principal component is a linear combination of different proportion of each of the original-attributes, the reduced data thus comprises of attributes that are not immediately recognisable. The interpretation of the effect of individual invention attribute is therefore less straight forward. Hence, two methods have been developed to reduce data size by choosing a subset of original attributes, rather than generating alternative attributes. Namely, Surrogate PCA Limit, and Single PCA Attributes.

Surrogate PCA Limit

Since each principal component is a linear combination of different proportion of each of the original-attributes, higher proportions are needed from certain attributes than others. For instance, as shown on Equation 5-28, let the number of original attributes be 5, and vector $\mathbf{e}_1$ be the original-attribute coefficient vector needed to compute the first principal components. Vector $\mathbf{e}_1$ shows that the first principal components require the highest proportion from the fifth original attribute, as the fifth coefficient is the highest. In other words, the fifth original attribute is the best representation (surrogate attribute) of the first principal components, as compared to all other original attributes.

Equation 5-28: Definitions of $\mathbf{e}_1$ and $x_{s\_pca_{11}}$

$$\mathbf{e}_1 = \begin{bmatrix} 0.4 & 0.5 & 0.3 & 0.2 & 0.6 \end{bmatrix}$$

$$x_{s\_pca_{11}} = 0.4x_{s11} + 0.5x_{s12} + 0.3x_{s13} + 0.2x_{s14} + 0.6x_{s15}$$

Where $\mathbf{e}_1$ is the original-attribute coefficient vector for the first principal components, and $x_{s\_pca_{11}}$ is the first principal component for the first case.

Furthermore, since the same attribute can be the surrogate attribute for more than one principal components, only an attribute that has *not* been chosen before would be selected for subsequent components. Also, because coefficient values can be either positive or negative, the sign of the coefficient value thus is not taken into account. In this way, only one attribute will be associated with each principal component. In other words, a surrogate attribute is the attribute that has not already been chosen and has the highest coefficient in absolute value of a particular principal component, as illustrated using Table 5-2 as an example.

Table 5-2: Coefficients for the first three principal components

	PC1	PC2	PC3
A1	0.4	0.2	0.15
A2	0.5	0.4	0.23
A3	0.3	-0.1	-0.2
A4	0.2	0.3	-0.4
A5	0.1	0.09	0.1

Where A1 to A5 refers to the five original attributes, PC1 to PC3 denotes the first three principal components.

Table 5-2 shows the coefficients defining PC1 to PC3 based on the linear combinations of the original-variables (A1 to A5). As shown, A2 is the most representative original attribute for both PC1 and PC2. Nonetheless, according to the definition of surrogate attribute stated above, the surrogate attribute for PC2 should be A4, as it is the attribute with highest coefficient and has not already been chosen as surrogate attribute before.

By using the surrogate attributes of *m* principal components, a subset of *m* attributes is then selected. The number of surrogate attributes resulted depends on the number of principal components considered, which in turn depends on the portions of variance of the original attribute matrix represented by the principal components.

Since the principal components are sorted in decreasing order of 'significance', the number of principal components used to represent a certain variance of the attribute matrix can be controlled by eliminating principal components representing less than a certain variance portion.

For instance, by selecting only principal components individually representing more than 1% of variance of the attribute matrix, the selected principal components together will represent about 99% of the attribute matrix. However, if every principal component represents a variance portion exceeding 1%, then every attribute would eventually be selected as surrogate attributes. and no data reduction is resulted.

In order to discard at least half of the attributes, the following measure is adopted. If first half principal components all individually represents a variance portion of more than 1%, then the surrogate attributes of them would be selected. Otherwise, only the surrogate attributes of principal components individually representing a variance portion of more than 1% would be selected. The surrogate attributes selected are therefore based on less than or half of the top principal components. These attributes then form a list of discriminative attributes, denoted as L_{surro}.

Single PCA attributes

In extreme, only one attribute would be selected by using the Surrogate PCA Limit method. This method therefore chooses a number of surrogate attributes on a ratio basis.

Since each principal component is a particular combination of every original attribute, any one of the principal components is therefore a certain representation of the original data. Typically, when principal component analysis is used to choose only one alternative-attribute. the first principal component would naturally be chosen as it offers the maximum discrimination between the cases. However, the first principal component may not be the most appropriate index for every discrimination task. For instance. the first principal component may be a good indicator for the expenses incurred in an invention case when the expense attribute involve the highest level of variation. This. however, does not necessarily mean the first principal component/the expense attribute is also the most predictive for an invention's class label.

Since there is no guarantee that the first principal component will be the best index, this method considers the following measures. Only the top 1% (round up to the nearest integer) principal components are considered, and the 10 most representative attributes (the 10 attributes with the highest coefficient in absolute value) for each of these principal components will be selected. Similar to the concept of surrogate attribute, any of the 10 attributes has to be an attribute that has not already been chosen. The attributes selected using this method then form an attribute list, denoted as L_{single}.

Reduced Attribute Matrix

Three lists of attributes (L_{tree}, L_{surro}, and L_{single}) would be generated using the decision tree method and the two PCA-based methods. By combining the attributes from these lists and removing the redundant attributes, a final list of attributes is generated. By extracting these attributes of every case from the attribute matrix, a reduced attribute matrix is then generated (denoted as $\mathbf{X}_R$), which would be used in the next step to generate a classifier.

5.3.3 Classifier Generation

This part of the CPS generates a classifier based on the input data set, which consists of the reduced attribute matrix ($\mathbf{X}_R$) and the class vector ($\mathbf{c}_2$). Multi-layer backpropagation neural network is chosen as the method for classifier generation on the basis of the following reasons.

Based on the discussion regarding the findings on invention assessment methods in Chapter Four, it was concluded that the scoring mechanism (i.e. the method used to compute a score, the perceived potential to succeed, for each invention) employed by the interviewed technology transfer offices can be expressed in terms of weights and attributes, where the weights and attributes are derived based on learning through trial and error from experience with historical invention projects, as explained using equations 4-1, 4-2, 4-3. Effectively, the score given to each invention by the managers equates to a function of weighted attributes (Equation 4-3). This scoring mechanism is very similar to the classification / prediction mechanism of neural networks, which is also a function of weighted attributes through learning. Neural networks compute the final class label of a case based on weight updates (learning through trial and error based on historical cases) to generate the optimal weights vector for the attributes (a function of weighted attributes). In addition, it has been widely proven (Cybenko 1989; Hornik et al. 1989; Jones 1990; Kurkov? 1991, 1992) that a multi-layer backpropagation neural network can approximate any continuous function from input to output, given sufficient number of hidden neurons, proper non-linear transfer functions, and sufficient network training. Consequently, the multi-layer backpropagation neural networks method is chosen to generate the classifier, mainly due to its universal function approximation ability as well as its similarity with the scoring systems employed by technology transfer offices.

Moreover, backpropagation is one of the simplest and most general training methods for neural networks. Backpropagation is also the most popular network training method, and many other training methods are essentially modifications of it. Backpropagation neural network has also been proven by numerous empirical studies that it out-performs other classification methods in many real-world problems. Last but not least, most classification methods require a large quantity of training data in order to achieve accurate classification. Backpropagation neural network, however, is found frequently producing high performances even when the amount of training data is limited.

Neural network is a classification method based on supervised learning. The basic approach is to generate a classifier (the trained network) based on iterative training using a set of training data. Instead of merely producing the final classifier (the CPS classifier) using the input data set ($\mathbf{X}_R$ and $\mathbf{c}_2$) as training data, this part of the CPS adopts a number of steps to generate the classifier. Briefly speaking, the input data set is divided into training, validation, and testing data, upon which a number of candidate classifiers will be generated and evaluated. The training algorithm producing the best performing candidate classifiers will then be selected, and applied to the original input data set to produce the CPS classifier. The adoption of these steps is based on a number of considerations, which will be explained with the presentation of each step.

The steps involved in this part are listed below:

1. Data normalisation and randomisation
2. Data division
3. Generation of initial weights
4. Candidate classifier generation
5. Network training to generate the CPS classifier

The resultant classifier generated from a given training data varies depending on the choice of network architecture (such as the number of hidden layer), and a number of training specification (such as the use of transfer function). Therefore, prior to the detail explanations for the five steps listed above, the network architecture and training specifications adopted are presented first. Due to the close relation between transfer function and the computation of network output, the method for class computation is also presented in the section explaining transfer function.

Network Architecture

The number of output neurons is governed by the output representation. Since the classification task is to predict whether an invention will be a 'success' or a 'failure', these class labels can be represented as '1' and '0'. Thus, one variable is sufficient for such representation. As a result, one output neuron is adopted.

The number of hidden neurons, however, is not governed by such obvious rules. The output of each hidden neuron is governed by its respective transfer function (i.e. the respective decision boundary), which are then combined to form the final network output (i.e. the final decision boundary of the network). The number of hidden neurons therefore determines the complexity of the network's decision boundary. In other words, more hidden neurons are needed for more complicated underlying-functions (i.e. the underlying function relating the attribute matrix and class vector that the network is approximating), and the 'function approximating power' of a network is sensitive to the number of hidden neurons. However, since the complexity of the underlying function is likely to vary among university technology transfer offices, a fixed number of neurons may be sufficient for one technology transfer office but not another. Too few neurons can lead to under-fitting (i.e. insufficient learning, hence high classification error), whereas too many neurons can contribute to over-fitting.

Although there is no universal method to determine the complexity of the underlying function for a particular technology transfer office, over-fitting is very unlikely, if the number of data points (i.e. the number of attributes n multiplying the number of cases k) is a lot larger than the number of hidden neurons. Therefore, a ratio measure (based on the number of attributes and cases) is adopted to determine the number of neurons to be used. Given an input data set with n attributes and k cases, $(n+k)/10$ neurons will be used.

Finally, only one hidden layer will be adopted in a network. This is because it has been widely accepted that backpropagation consisting of one hidden layer with non-linear transfer functions should be sufficient to approximate any function (given enough hidden neurons). Empirical studies also found that networks with multiple hidden layers are more prone to getting caught in undesirable local minima. The network architecture adopted therefore comprised of one hidden layer.

In summary, the network architecture adopted here is comprised of one hidden layer with $(n+k)/10$ neurons, and one output layer with one neuron.

Training Specifications

General specifications

- The error function adopted here is the average of the error squared, which is also the most widely used error function for neural networks, due to the function's concentration on the magnitude of the error instead of the sign of the error.

- To avoid infinitive training, network training will be stopped when any of the following criteria are met: 1) training time exceeds 60 seconds, 2) the number of epoch exceeds 100, and 3) the error falls below the error goal of 0.01.

- Between the two training modes, namely incremental mode and batch mode, batch mode is adopted here. This is related to the adoption of the five training algorithms for candidate classifier generation (Adaptive Learning Rate backpropagation, Resilient backpropagation, Conjugate Gradient (Fletcher-Reeves), Conjugate Gradient (Polak-Ribière), and the Levenberg-Marquardt algorithm). These algorithms have been proven to be capable of significantly improving classification accuracy and speeding up network training, and they cannot be easily incorporated in incremental training.

Transfer Function

Since a neuron computes its output by passing the input through a transfer function, whether the input is being mapped linearly or nonlinearly therefore depends on the linearity of the transfer function. This is illustrated in Figure 5-4.

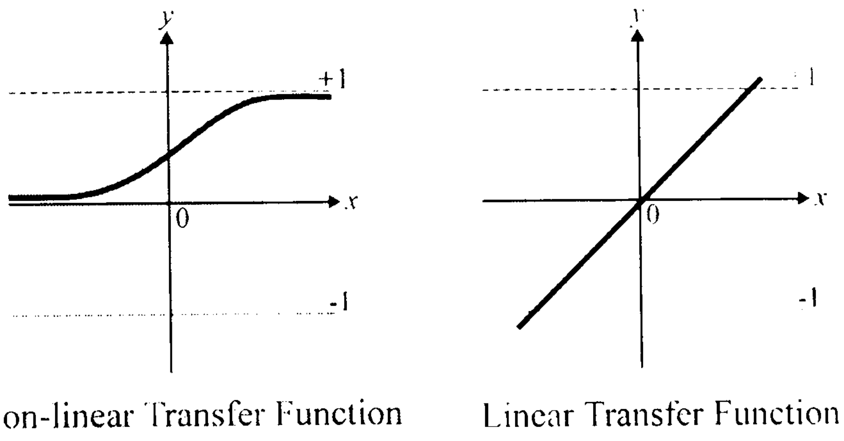


Figure 5-4: Non-linear and linear transfer functions

In other words, the transfer function at a neuron defines the decision surface that separates inputs into different regions of output. While the decision surface of a linear transfer function is a hyperplane (as shown in Figure 5-5), the decision surface for a non-linear transfer function will be a curved surface (as shown in Figure 5-6).

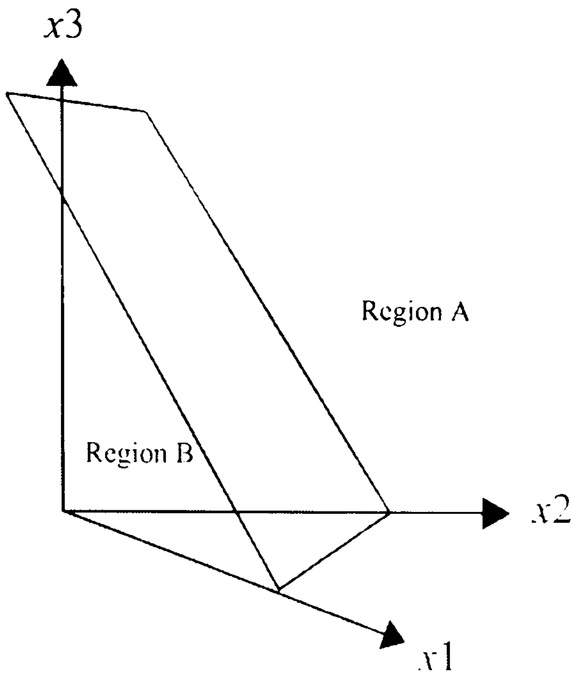


Figure 5-5: The hyperplane decision surface caused by a linear transfer function

Where $x1$, $x2$, and $x3$ denote three attributes of the input, Region A and Region B denote regions of the output separated by the transfer function.

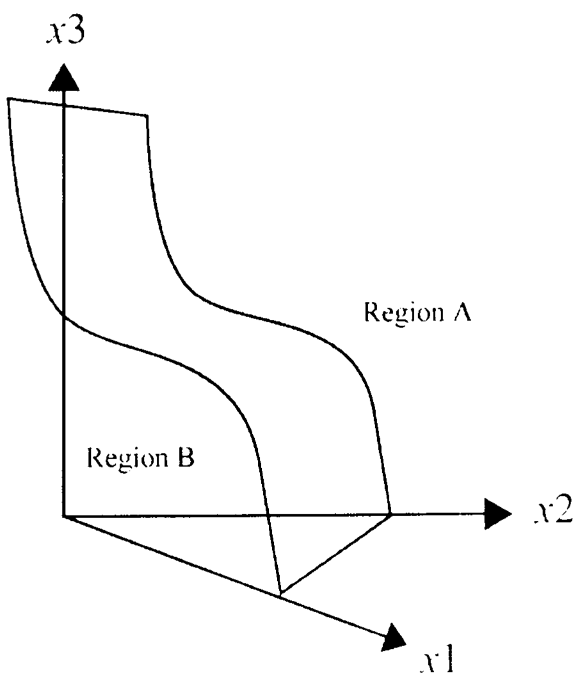


Figure 5-6: The curved plane decision surface caused by a non-linear transfer function

Since the outputs from hidden neurons are inputs for the output neurons, the decision surface of an output neuron is therefore an arbitrary surface, as illustrated in Figure 5-7.

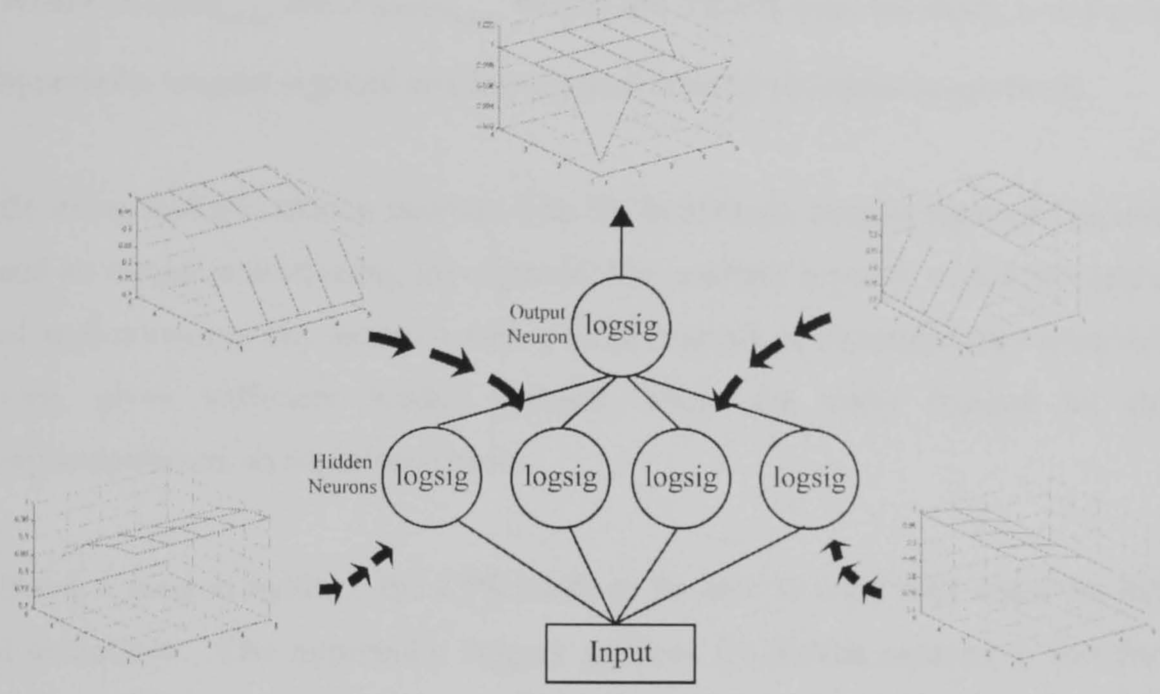


Figure 5-7: The decision surfaces caused by various non-linear transfer functions

One important point is that the transfer function must be differentiable, as the derivation of the backpropagation algorithm requires the derivative of any transfer function used. Moreover, since neurons may use any differentiable transfer function, and the final network output is dependent on the hidden neurons' output, the choice of transfer function for hidden neurons becomes crucial.

In the CPS, the transfer function adopted for hidden neurons is the hyperbolic tangent sigmoid (Equation 5-29), and the log-sigmoid (Equation 5-30) transfer function is employed for the output neuron.

Equation 5-29: Definition of the hyperbolic tangent sigmoid transfer function

$$y_{neuron_{\tan sig}} = 2/\left(1+e^{-2x}\right)-1$$

Equation 5-30: Definition of the log-sigmoid transfer function

$$y_{neuron_{\log sig}} = 1/\left(1+e^{-x}\right)$$

Where $y_{neuron_{\tan sig}}$ and $y_{neuron_{\log sig}}$ denote the outputs from the input, x, using the hyperbolic tangent sigmoid and log-sigmoid transfer functions respectively.

By using multiple hidden neurons with the hyperbolic tangent sigmoid function, and an output neuron using log-sigmoid, the resultant network would be capable of approximating any function with a finite number of discontinuities arbitrarily well, given sufficient hidden neurons. There are many reasons for this implementation, as explained below.

Being a generic method, the CPS needs to be able to cope with unknown data distributions. The hyperbolic tangent sigmoid for hidden neurons is therefore adopted since it possesses a number of desirable properties (differentiable, saturating, monotonic, centred at zero and antisymmetric), which makes the resultant classifier a good general approximator. These properties are described below.

First of all, it is a continuous function that is differentiable. Second, it saturates, in the sense that it restricts the output's maximum and minimum. According to the adopted definition (Equation 5-29), the output must range between -1 and 1. By keeping the neuron outputs bounded, the network training time can be kept limited. Third, it is monotonic, in the sense that the derivative does not change sign throughout the function. This property helps avoiding the introduction of additional local extrema in the error surface. Lastly, it is centred on zero, and is antisymmetric. With this property accompanying a normalised input data set, faster learning can therefore be achieved.

Log-sigmoid has been adopted for the output neuron, due to the fact that it shares most (except that it is not antisymmetric) of the desirable properties of the hyperbolic tangent, and that it is more suited for binary classification task since it saturates at the output values of 0 and 1.

Computation of the class label

Since the transfer function adopted for the output neuron is a continuous function, the output generated would also be a continuous value, within the range from 0 to 1. While the class label adopted by the CPS is either 'success' or 'failure', it can also be represented as a categorical value of either '1' or '0'. In order to convert the continuous values into the two categorical values, the following rules are adopted:

If $0 \leq y < 0.5$ then the class label is set to 0.
Else if $0.5 \leq y \leq 1$ then the class label is set to 1.

Where y represents the output value generated by the output neuron.

Validation (Measure to avoid Over-fitting)

Over-fitting is one of the major problems that hinder a network's generalisation. It happens when the network has over-learned so much during the training phase, that the network has only 'memorised' the training data, failing to learn the underlying function that governs the relationship between the attribute and the class data. An over-fitted classifier is analogous to a look-up table of the training data. Thus an over-fitted network has a limited ability to classify unseen cases, unless they are almost identical to the training cases. In other words, an over-fitted network has low generalisation ability.

Over-fitting is evident when there is a decreasing and very small training error, combined with an increasing and large test error. In order to improve network generalisation, and reduce the possibility of over-fitting, validation is carried out to detect the optimal time to stop the network training before over-fitting occurs. The use of validation is explained below, and is illustrated in Figure 5-8.

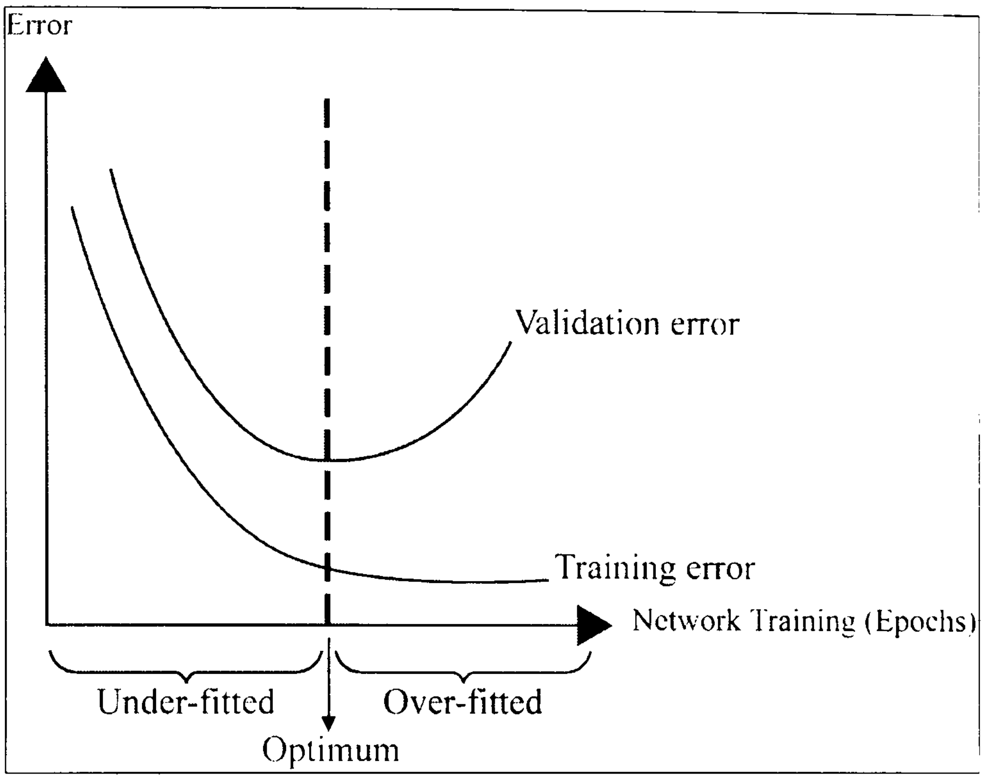


Figure 5-8: The use of validation data to detect over-fitting

During training, instead of merely using training data, both training data and validation data are used. Training data is used for the gradient computation of the training error function. Validation data is used as 'pretend testing data' in order to monitor the validation error (the pretend test error). Initially, the validation error typically falls together with the training error. When over-fitting occurs, the validation error begins to rise. Training is therefore stopped at the point where the validation error is at minimum (the optimum point). The network derived at the optimum point is used to classify data unseen during training (testing data).

Data Normalisation and Randomisation

The first step is to normalise the attribute matrix since each attribute is measured in different units. The convergence of the error will be much quicker with standardised data. To perform normalisation, the attribute matrix ($\mathbf{X}_R$) is first transposed. Each row of the matrix ($\mathbf{X}^{\mathrm{T}}{}_R$) is then normalised so that the minimum value is -1 and the maximum 1, using the algorithm shown in Equation 5-31. The normalised matrix is denoted as ($\mathbf{X}^{\mathrm{T}}{}_{RN}$).

Equation 5-31: Normalisation algorithm for classifier generation

$$\mathbf{X}^{\mathrm{T}}{}_{RN} = 2\left(\mathbf{X}^{\mathrm{T}}{}_R - \mathbf{X}_{min}\right) / \left(\mathbf{X}_{max} - \mathbf{X}_{min}\right) - 1$$

Where $\mathbf{X}_{min}$ is a matrix consists of identical columns, in which each column contains the minimum values of each row of $\mathbf{X}^{\mathrm{T}}{}_R$. Similarly, $\mathbf{X}_{max}$ is a matrix consists of identical columns, in which each column contains the maximum values of each row of $\mathbf{X}^{\mathrm{T}}{}_R$.

Next, randomisation of cases is carried out for both the normalised matrix ($\mathbf{X}^{\mathrm{T}}{}_R$) and the class vector ($\mathbf{c}_2$). This is due to the fact that cases in the input data set may have been organised in a certain unknown sequence during data collection, such as the order of time. While the cases will be divided into subsets later, it is important that all the subsets, be it the training data, or the testing data, are representative of the unknown distribution of the input data. All cases therefore have to be randomised, in which the cases will be sorted according to a computer generated random order. However, it is possible that the random order produces an uneven distribution which hinders classification, such as a distribution with a certain class dominating the training data.

To lower the possibility of such uneven distribution, 10 randomised version of the input data set are generated. Given an input data set containing k cases, 10 vectors of random permutation ($\mathbf{r\_order}_j$) of the integers from 1 to k are generated. The input data set ($\mathbf{X}^{\mathbf{T}}{}_R$ and $\mathbf{c}_2$) is then randomised as shown in Equation 5-32.

Equation 5-32: The vector of random permutation

$$\mathbf{r\_order}_j = \left[r_{j1} \ldots r_{ji} \ldots r_{jk} \right]$$

$$\mathbf{P}_j = \begin{pmatrix} p_{1r_{j1}} & \cdots & p_{1r_{j1}} \\ p_{ir_{ji}} & \cdots & p_{ir_{ji}} \\ \vdots & & \vdots \\ p_{nr_{jk}} & \cdots & p_{nr_{jk}} \end{pmatrix} \qquad \mathbf{c}_{random\_j} = \left[c_{r_{j1}} \ldots c_{r_{ji}} \ldots c_{r_{jk}} \right]$$

Where $\mathbf{r\_order}_j$ is the j-th vector of random permutation, $\mathbf{P}_j$ is the randomised version of $\mathbf{X}^{\mathbf{T}}{}_R$ using $\mathbf{r\_order}_j$, and $\mathbf{c}_{random\_j}$ is the randomised version of $\mathbf{c}_2$ using $\mathbf{r\_order}_j$.

For instance, let the number of cases of the input data set be 5, and $\mathbf{r\_order}_1$ be $[2 \quad 1 \quad 4 \quad 5 \quad 3]$. The first case in matrix $\mathbf{P}_1$ is therefore equivalent to the second case in $\mathbf{X}^{\mathbf{T}}{}_R$, and the first class label in $\mathbf{c}_{random\_1}$ is equivalent to the second class label in $\mathbf{c}_2$.

Data Set Partition

As previously stated, the technique of validation will be used to improve network generalisation by avoiding the problem of over-fitting. Also, the CPS classifier will be generated based on the evaluation of a number of candidate classifiers, which is based on their classification results on unseen data. The input data set is therefore partitioned into training data, validation data, and testing data. The training data will be used to generate candidate classifiers using several network training algorithms. The validation data will be used to improve the network generalisation of these candidate classifiers. The testing data will be used to evaluate the classification accuracy of the candidate classifiers. This partition of input data set into is explained below.

After case randomisation, 10 randomised versions of the input data set are generated. Each randomised data set is here partitioned using the ratio of 8:1:1 into the training data, validation data, and testing data. In other words, the first 80% of the cases in each randomised data set will be used as training data, the next 10% will be used as validation data, and the remaining 10% will be used as testing data. The training data set, validation data set, and testing data set derived from the j-th randomised input data set are denoted as $\mathbf{P}_{j\_train}$ and $\mathbf{c}_{random\_j\_train}$, $\mathbf{P}_{j\_val}$ and $\mathbf{c}_{random\_j\_val}$, and $\mathbf{P}_{j\_test}$ and $\mathbf{c}_{random\_j\_test}$ respectively. For instance, the training data set, validation data set, and testing data set derived from the first randomised input data set ($\mathbf{P}_1$ and $\mathbf{c}_{random\_1}$) are denoted as $\mathbf{P}_{1\_train}$ and $\mathbf{c}_{random\_1\_train}$, $\mathbf{P}_{1\_val}$ and $\mathbf{c}_{random\_1\_val}$, and $\mathbf{P}_{1\_test}$ and $\mathbf{c}_{random\_1\_test}$ respectively. Given n attributes and k cases in a randomised input data set, the dimensions of these matrices and vectors are shown in Table 5-3. These dimensions apply to every randomised input data set.

Table 5-3: The dimensions of the training data, validation data, and testing data

	Attribute matrix	Dimension	Class Vector	Dimension
Training Data Set	$\mathbf{P}_{1\_train}$	n by $0.8k$	$\mathbf{c}_{random\_1\_train}$	1 by $0.8k$
Validation Data Set	$\mathbf{P}_{1\_val}$	n by $0.1k$	$\mathbf{c}_{random\_1\_val}$	1 by $0.1k$
Testing Data Set	$\mathbf{P}_{1\_test}$	n by $0.1k$	$\mathbf{c}_{random\_1\_test}$	1 by $0.1k$

Generation of Initial Weights

In the same way as other gradient-descent algorithms, the behaviour of backpropagation training depends on the starting point – the initial values of weight vectors. If they are set to zero at the start, the outputs of the layer (which is the weighted sum) will be zero, the backpropagated error will also be zero, and the input-to-hidden weights will therefore never change. To avoid this, random values are used for the initial weights.

However, because of the shape of the hyperbolic tangent sigmoid, the output of neurons will saturate if the random weights are very large. Yet if the random weights are very small, the neurons' output will be close to zero since the antisymmetric transfer functions pass through the origin. In order to avoid random weights of extreme values, the Nguyen-Widrow initialisation algorithm (Nguyen and Widrow 1990) is adopted. It has been demonstrated that initial weights generated using this method result in an initial network leading to better function approximation result.

However, the problem with random weight generation is that it is different every time. Initialising with random weights each time means that even the same algorithm will perform differently each time. Since comparisons of candidate classifiers based on different training algorithms are to be performed in the CPS, each training algorithm generating the candidate classifiers will be initiated with the same set of random weights.

Nevertheless, comparing models based on one set of initial random weights may produce prejudiced results, as this particular set of random weights may be by chance favouring certain training algorithms. The error surface of a non-linear network is very complex, as it contains many local minima introduced by non-linear transfer functions. Depending on the weight initialisation, training algorithms may become trapped in a local minimum. This can be harmful if the local minimum is far away from the global minimum. Therefore, 100 sets of different initial random weights are generated, so that each training algorithm will run 100 times using these initialisations, in order to achieve robust and averaged network results for each model.

Each set of initial random weights consists of one weight matrix each for the hidden layer and output layer respectively. The dimensions of the two weight matrices depend on the number of neurons employed for the network, and the number of attributes contained by the attribute matrix of the training data. Given the use of q neurons for the hidden layer, and a training data consisting of a attributes, the dimensions of the weight matrices for the hidden layer and output layer are therefore q by a, and 1 by q respectively.

Candidate Classifier Generation

As demonstrated by the findings from the prototype CPS, the use of candidate classifiers has helped identifying the desirable parameters for classifier generation. Candidate classifiers will therefore be generated by applying different training algorithms to the same training data. The candidate classifiers will then be tested using unseen testing data. Based on the classification accuracies achieved by these candidate classifiers on the testing data, the desirable training algorithm can then be identified for the generation of the CPS classifier.

The choice of training algorithms is related to the learning rate used for network training. As illustrated in Figure 5-9, too small a learning rate takes the algorithm a long time to converge, whereas too large a learning rate may cause the algorithm to oscillate, giving an unstable error function.

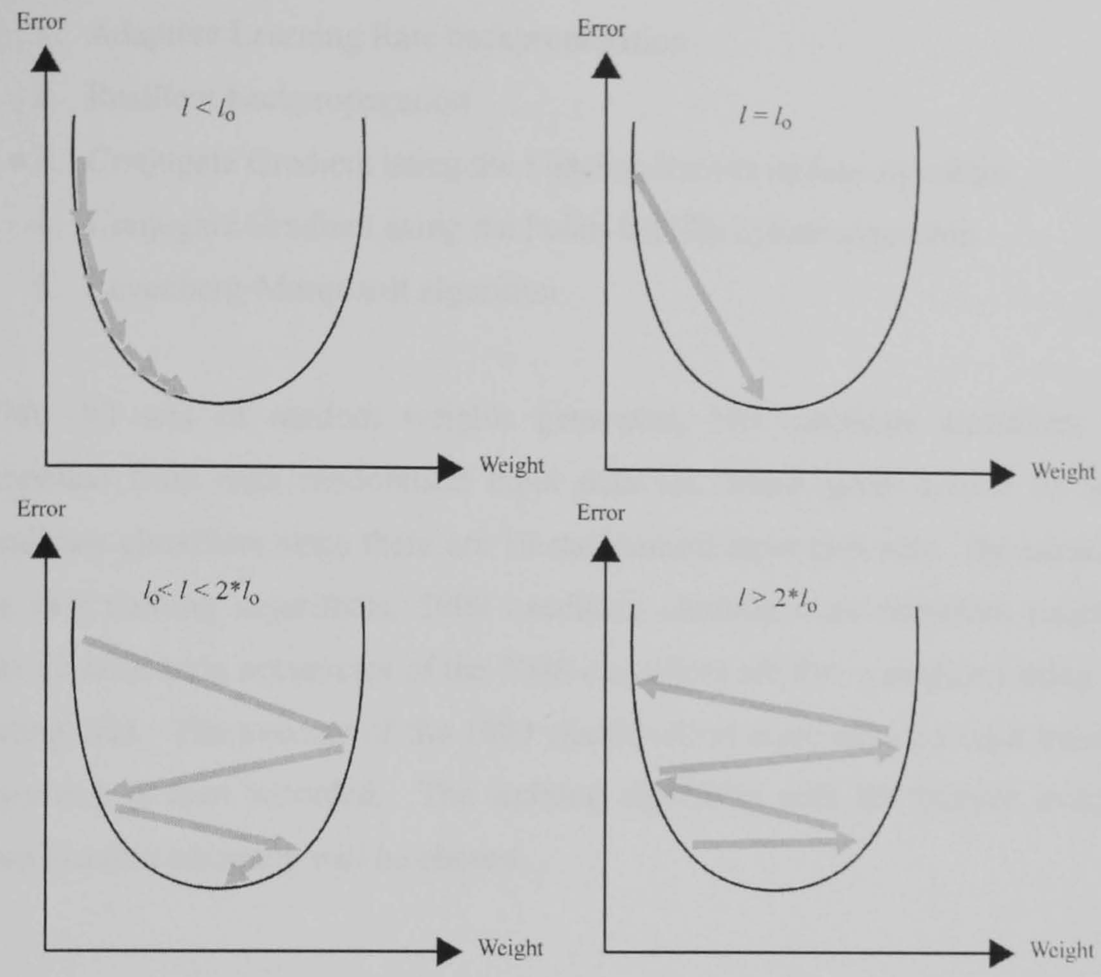


Figure 5-9: Four different learning rates

Where l denotes a learning rate, and l_o denotes the optimal learning rate.

In order to speed up network convergence, different training algorithms based on various implementations of learning rate have been proposed. However, it is very difficult to know before hand which training algorithm will be the most accurate and the fastest for a given data set. The choice of training algorithm generally depends on many factors, including the complexity of the underlying function, the size of the attribute matrix, and the network architecture adopted. Each training algorithms suits certain situations better than others. In order to identify the best training algorithms for the CPS classifier, five training algorithms are used to generate candidate classifiers using the same training data. The algorithms adopted are listed as follows:

1. Adaptive Learning Rate backpropagation
2. Resilient backpropagation
3. Conjugate Gradient using the Fletcher-Reeves update algorithm
4. Conjugate Gradient using the Polak-Ribière update algorithm
5. Levenberg-Marquardt algorithm.

With 100 sets of random weights generated, 100 candidate classifiers are generated from each randomised input data set, which gives a total of 1000 candidate classifiers since there are 10 randomised input data sets. By means of the five training algorithms, 5000 candidate classifiers are therefore resulted. The classification accuracies of the 5000 classifiers are then computed using the testing data. The average of the 1000 classification accuracies for each training algorithm is then recorded. The training algorithm with the highest average classification accuracy will be chosen.

<u>**Network Training to Generate the CPS Classifier**</u>

Finally, the chosen training algorithm is applied to the original input data (normalised) of the CPS ($\mathbf{X}^{T}{}_{RN}$ and $\mathbf{c}_2$), in which the whole input data set is used as the training data in order to generate the CPS classifier. The average classification accuracy (computed based on its 1000 candidate classifiers) of the chosen training algorithm therefore represents the estimated classification accuracy of the CPS classifier.

5.4 Summary

This section presents a summary of the CPS. The CPS stands for the Commercial Outcome Prediction System. This system predicts the likely future monetary return of an invention. The CPS consists of three steps, namely input data set specification, data reduction, and classifier generation. The first step is to generate an input data set according to the specification. The input data set would consist of an attribute matrix and a class vector. The second step is to perform data reduction, for the purpose of generating a reduced attribute matrix for the third step. The decision tree method and two methods based on principal component analysis are employed to arrive at the reduced attribute matrix. The last step is to generate a classifier using an input data set comprised of the reduced attribute matrix and the class vector. This input data set is first partitioned into training and testing data. Five neural network training algorithms are then employed to generate a large number of candidate classifiers using the training data, and their classification accuracies are computed using the testing data. The training algorithm generated the best performing candidate classifiers would then be chosen to generate the final classifier using every case of the input data set for network training.

CHAPTER SIX

RESULTS

6.1 INTRODUCTION

This chapter presents the results generated from using the CPS, and evaluates the classification capability of the CPS. First of all, the data collection results, and the data set partition arrangement are presented in Section 6.2. The results generated by the data reduction step of the CPS are presented in Section 6.3, followed by the results of the classifier generation step in Section 6.4. The classification capability of the CPS is then evaluated in Section 6.5.

6.2 DATA COLLECTION RESULTS

Data for the 7 categories of potentially predictive attributes were collected from the same source as the input data set for the prototype CPS. Consequently, the data for 55 invention cases was collected, and the total number of invention attributes resulted is 294. While the data of certain attributes at certain points in time were not available, the information of which attribute data was unavailable is not disclosed here, due to the issue of confidentiality.

The distributions of these 294 attributes among the 7 categories of potentially predictive attributes are shown in Table 6-1. As shown, most attributes belong to the *Inventor* category, as it covered 110 of the 294 attributes. The category covering the least attributes is the *Basic* category.

Table 6-1: Distribution of the 294 attributes among the 7 categories of potentially predictive attributes

Attribute Category	Number of Attributes Collected
Inventor	110
Basic	5
Communication	25
Expense	20
Agreement	16
Technical Document	57
Patent	35

The 294 invention attributes for the 55 invention cases is stored in matrix $\mathbf{X}_2$, and the outcome class data is stored in vector $\mathbf{c}_2$. The definitions for matrix $\mathbf{X}_2$ and vector $\mathbf{c}_2$ are illustrated in Equation 6-1.

Equation 6-1: Definitions of matrix $\mathbf{X}_2$ and vector $\mathbf{c}_2$

$$\mathbf{X}_2 = \begin{pmatrix} x_{11} & x_{1j} & \cdots & x_{1n} \\ x_{i1} & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{kj} & \cdots & x_{kn} \end{pmatrix} \qquad \mathbf{c}_2 = [c_1 \ldots c_i \ldots c_k]$$

Where x_{ij} represents the value of the j-th attribute for the i-th case, n is the number of attributes, and k is the number of cases. For this data set, n is 294 and k is 55. The value of c_i is either 'success' or 'failure', which refers to the class label for the i-th case.

During data collection, the cases of this data set ($\mathbf{X}_2$ and $\mathbf{c}_2$) are sorted in descending order of their age. In order to evaluate the classification capability of the CPS later in Section 6.5 using younger cases, the last 8 cases of this data set are reserved as the evaluation data set. The attribute data and class data of these 8 cases are stored in matrix $\mathbf{X}_{2eval}$ and vector $\mathbf{c}_{2eval}$ respectively. Removing the last 8 cases, the remaining 47 cases are used as the input data set for the CPS. The attribute data and class data of these 47 cases are stored in matrix $\mathbf{X}_{2IPS}$ and vector $\mathbf{c}_{2IPS}$ respectively.

The dimensional information for the three data sets is presented in Table 6-2.

Table 6-2: Dimensional information for the three data sets

	Attribute matrix	Dimension	Class Vector	Dimension
Original Data Set	$\mathbf{X}_2$	55 by 294	$\mathbf{c}_2$	1 by 55
CPS Input Data Set	$\mathbf{X}_{2IPS}$	47 by 294	$\mathbf{c}_{2IPS}$	1 by 47
CPS Evaluation Data Set	$\mathbf{X}_{2eval}$	8 by 294	$\mathbf{c}_{2eval}$	1 by 8

6.3 DATA REDUCTION

With 294 attributes and 2 classes, the attribute-class ratio is larger than 10, the data reduction step is therefore performed. The results generated from the data reduction step using the CPS input data set are presented in this section.

Two data reduction methods, decision tree induction and principal component analysis, are applied to the CPS input data set in order to generate a reduced list of invention attribute. The attribute list resulted from each method are reported in Section 6.3.1 and Section 6.3.2 respectively. The two attribute lists are then combined to form the reduced attribute list, as presented in Section 6.3.3.

6.3.1 Decision Tree Induction

The first step of data reduction by decision tree induction is the normalisation of each column of the attribute matrix ($\mathbf{X}_{2IPS}$), so that the minimum value is -1 and the maximum is 1.

Next, a decision tree is generated using the adopted impurity measure, Gini Index, with the parameter values namely monothetic decision trees, a branch factor of two, and a Minimum Case Threshold of two. Base on these specifications, a decision tree is generated, as shown in Figure 6-1.

The tree is comprised of 4 nodes, where the description next to each node denotes its respective attribute test. For instance, the attribute test at the root node is whether the 4th attribute is smaller than the value of -0.997395. This value is within the boundary of 1 and -1 because the attribute matrix ($\mathbf{X}_{2IPS}$) was normalised. However, the values at the attributes tests are not important, since the data reduction step only concerns the selection of attributes from the 294 attributes.

The attributes selected using this tree are therefore the 4^{th}, 20^{th}, 260^{th}, and the 31^{st} attributes. These four attributes are stored in a list denoted as L_{tree}.

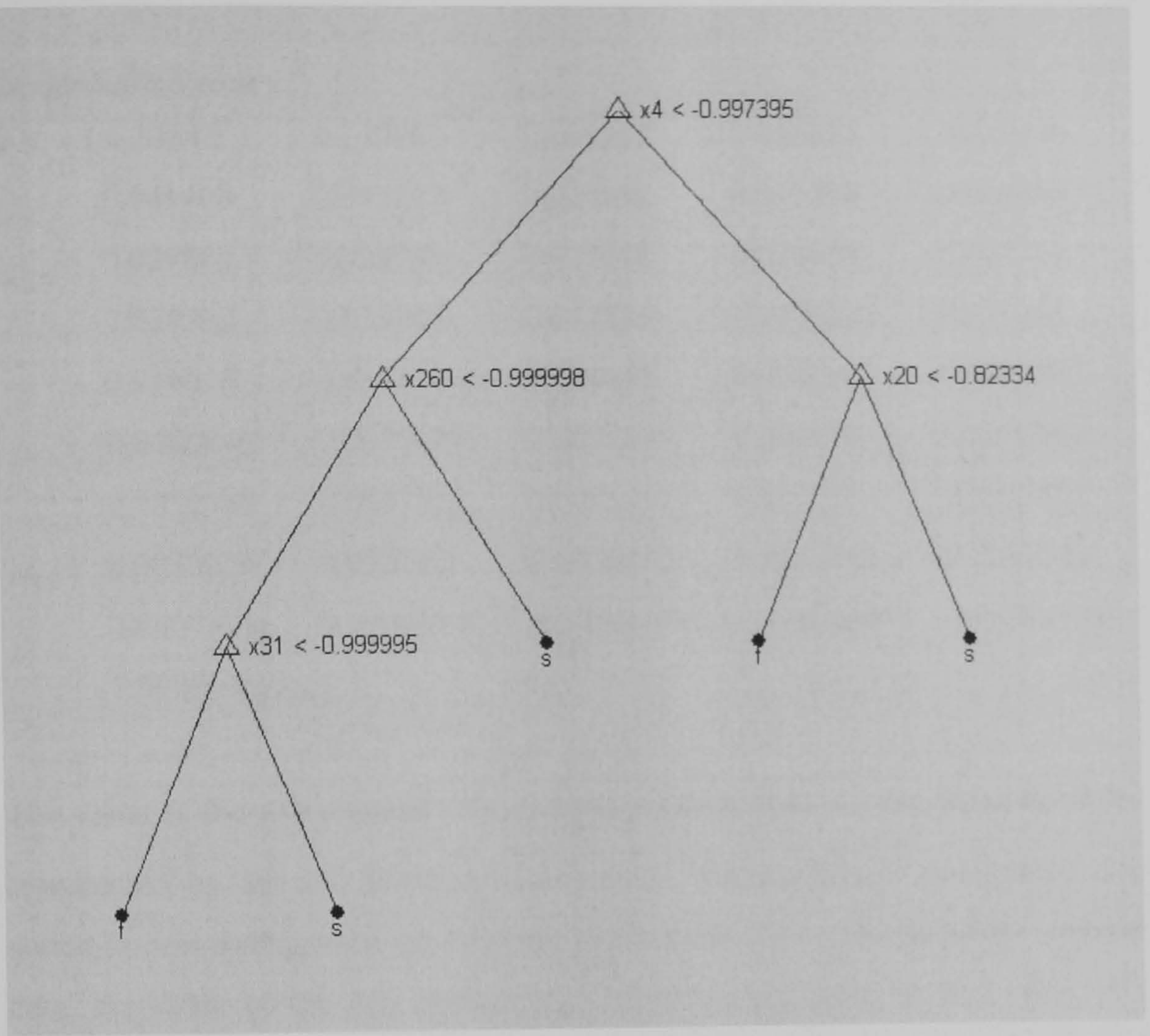


Figure 6-1: A full tree generated using Gini Index

6.3.2 Principal Component Analysis

The first step of data reduction by principal component analysis is to standardise the attribute matrix ($\mathbf{X}_{2IPS}$) by computing the standardised Z scores for each row. The standardised version of the attribute matrix is denoted as the $\mathbf{X}_s$ matrix. Equation 5-4 and 5-5 are then applied to the $\mathbf{X}_s$ matrix to produce the matrix of principal components (denoted as $\mathbf{X}_{s\_pca}$). In addition, a vector containing the portions of variance of $\mathbf{X}_s$ matrix represented by each of the 294 principal components in $\mathbf{X}_{s\_pca}$ matrix is also computed. This vector is denoted as $\mathbf{v}$ and the values of its elements are shown in Equation 6-2.

Equation 6-2: Vector v

v =	(0.14543	0.13078	0.064812	0.058825	0.053986
	0.049408	0.045393	0.038692	0.033298	0.030661
	0.029869	0.028994	0.028018	0.025804	0.023716
	0.019543	0.018009	0.017505	0.015624	0.014231
	0.014018	0.013532	0.011097	0.010063	0.0090873
	0.0086846	0.0078159	0.0072899	0.006654	0.0064984
	0.0050388	0.0047012	0.0041461	0.003739	0.0033852
	0.0033054	0.002295	0.0020102	0.0013965	0.0013523
	0.00074316	0.00045142	8.5704e-005	1.3187e-005	3.6121e-007
	1.3905e-032)				

The value of the *i*-th element of **v** represents the portion of the variance of **X** represented by the *i*-th principal component. Since principal components are sorted in descending order of their representations of the variance of the original data, the value of the *i*-th element of **v** therefore decreases as the value of *i* increases. While there are 294 principal components, **v** should contain 294 elements. Nevertheless, only the first 41 elements in **v** are displayed in Equation 6-2 since only these elements are non-zero. Other elements are zero because **v** is calculated to within a machine precision of 40 decimal places.

The attribute list resulted from the two principal component analysis based techniques, namely Surrogate PCA Limit and Single PCA Attributes, are presented below.

Surrogate PCA Limit

Based on the Surrogate PCA Limit method, only the first 24 principal components are selected. This is due to the fact that only the first 24 principal components individually represent more than the portion of 0.01 of the variance of the original data ($\mathbf{X}_s$), as shown in **v**.

To find out the 24 surrogate attributes, the original-attribute coefficients of the first 24 principal components are needed. The first 24 columns of the original-attribute coefficient matrix shows the original-attribute coefficients for the first 24 principal components respectively. Based on the definition of surrogate attributes defined in the CPS, the surrogate attributes for the first 24 principal components are respectively the 215th, 22nd, 254th, 195th, 114th, 279th, 164th, 39th, 286th, 281st, 289th, 37th, 47th, 61st, 284th, 189th, 23rd, 90th, 258th, 27th, 288th, 206th, 31st, 32nd attributes. The attribute list generated using Surrogate PCA Limit is therefore comprised of these 24 attributes. This list is denoted as L_{surro}.

Single PCA Attributes

Since there are 294 attributes, 30 (10% of 294) attributes will therefore be selected based on the Single PCA Attributes method. These 30 attributes consist of the 10 most representative attributes for each of the first three principal components. With reference to the original-attribute coefficients of the first three principal components, the attributes selected are displayed in Table 6-3. The attribute list produced by Single PCA Attributes is thus comprised of these 30 attributes, and the list is denoted as L_{single}.

Table 6-3: The attributes selected using Single PCA Attributes

The first principal component	The 99th, 207th, 211th 212th, 213th, 214th, 215th, 216th, 217th, 218th attributes
The second principal component	The 13th, 14th, 22nd, 51st, 53rd, 71st, 72nd, 73rd, 74th, 75th attributes
The third principal component	The 2nd, 35th, 55th, 87th, 146th, 156th, 160th, 161st, 162nd, 163rd attributes

6.3.3 Reduced Attribute matrix

Three lists of attributes are generated using the two data reduction method, namely L_{tree}, L_{surro}, and L_{single}. By combining the attributes from the three lists and removing the redundant attributes, the final list consists of 55 attributes. They are the 2nd, 4th, 13th, 14th, 20th, 22nd, 23rd, 27th, 31st, 32nd, 35th, 37th, 39th, 47th, 51st, 53rd, 55th, 61st, 71st, 72nd, 73rd, 74th, 75th, 87th, 90th, 99th, 114th, 146th, 156th, 160th, 161st, 162nd, 163rd, 164th, 189th, 195th, 206th, 207th, 211th, 212th, 213th, 214th, 215th, 216th, 217th, 218th, 254th, 258th, 260th, 279th, 281st, 284th, 286th, 288th, 289th attributes. By extracting these 55 attributes of every case from the attribute matrix of the CPS input data set ($\mathbf{X}_{2IPS}$), the reduced attribute matrix is resulted. Its dimension is 47 by 55, and this matrix is denoted as $\mathbf{X}_{2IPS_R}$.

The distribution of these 55 attributes among the 7 categories of invention attributes is shown in Table 6-4. The original number of attributes in each category is also displayed in column three.

Table 6-4: Distribution of attributes of the reduced set among the 7 categories

Attribute Category	Attributes from the reduced matrix	Attributes from the original matrix
Inventor	19	110
Basic	2	5
Communication	5	25
Expense	5	20
Agreement	16	42
Technical Document	4	57
Patent	4	35

6.4 CLASSIFIER GENERATION

This section presents the results generated from the third part of the CPS, where the CPS classifier is generated using neural networks. The input data set used in this part of the CPS is comprised of the reduced attribute matrix ($\mathbf{X}_{2IPS_R}$), and the class vector ($\mathbf{c}_{2IPS}$).

The steps involved in this part of the CPS are listed below:

1. Data normalisation and randomisation
2. Data set partition
3. Generation of initial weights
4. Network training and classification simulation

The results generated from each step are presented below.

6.4.1 Data Normalisation and Randomisation

First of all, the attribute matrix is transposed. Each row of the attribute matrix is then normalised, so that the minimum value is -1 and the maximum is 1. The normalised attribute matrix is denoted as $\mathbf{X}^{\mathrm{T}}{}_{2IPS_{RN}}$.

Next, 10 vectors of random permutation of the integers from 1 to 47 are generated. Each of these 10 vectors is denoted as $\mathbf{r\_order}_i$, where i represents the index of the vector. The first vector ($\mathbf{r\_order}_1$) is shown in Equation 6-3 as an example.

Equation 6-3: The first vector of random permutation

$$\mathbf{r\_order}_1 = (24 \quad 31 \quad 8 \quad 42 \quad 21 \quad 25 \quad 15 \quad 27 \quad 30 \quad 39 \quad 26 \quad 2 \quad 29 \quad 22 \quad 44 \quad 16 \quad 19 \quad 36 \quad 10 \quad 33 \quad 7 \quad 35 \quad 4 \quad 46 \quad 38 \quad 28 \quad 3 \quad 11 \quad 40 \quad 43 \quad 47 \quad 14 \quad 32 \quad 6 \quad 12 \quad 23 \quad 9 \quad 45 \quad 41 \quad 37 \quad 5 \quad 20 \quad 18 \quad 13 \quad 34 \quad 17 \quad 1)$$

The attribute matrix ($\mathbf{X}^{\mathrm{T}}{}_{2IPS_{RX}}$) and class vector ($\mathbf{c}_{2IPS}$) are then randomised using the 10 vectors of $\mathbf{r\_order}_i$. Each randomised input data set is denoted as $\mathbf{P}_i$ and $\mathbf{c}_{random\_i}$. As an example, the randomised input data set generated using $\mathbf{r-order}_1$ is presented as follows. $\mathbf{P}_1$ and $\mathbf{c}_{random\_1}$ denote the randomised version of $\mathbf{X}^{\mathrm{T}}{}_{2IPS_{RX}}$ and $\mathbf{c}_{2IPS}$ generated using $\mathbf{r-order}_1$, where $\mathbf{P}_1$ and $\mathbf{c}_{random\_1}$ store the cases of $\mathbf{X}^{\mathrm{T}}{}_{2IPS_{RX}}$ and $\mathbf{c}_{2IPS}$ in the order of $\mathbf{r-order}_1$. In $\mathbf{P}_1$, the first and second columns of the matrix are equivalent to the 24th and 31st columns of the $\mathbf{X}^{\mathrm{T}}{}_{2IPS_{RX}}$ matrix respectively. Similarly, the first element of the $\mathbf{c}_{random\_1}$ vector is the same as the 24th element of the $\mathbf{c}_{2IPS}$ vector. By means of 10 vectors of **r-order**, 10 randomised input data sets are resulted.

6.4.2 Data Set Partition

Each of the 10 randomised input data sets generated in the previous section is here partitioned using the ratio of 8:1:1 into the training data set, validation data set, and testing data set. The training data set, validation data set, and testing data set derived from the *i*-th randomised input data set are denoted as $\mathbf{P}_{i\_train}$ and $\mathbf{c}_{random\_i\_train}$, $\mathbf{P}_{i\_val}$ and $\mathbf{c}_{random\_i\_val}$, and $\mathbf{P}_{i\_test}$ and $\mathbf{c}_{random\_i\_test}$ respectively. For instance, the training data set, validation data set, and testing data set derived from the first randomised input data set ($\mathbf{P}_1$ and $\mathbf{c}_{random\_1}$) are denoted as $\mathbf{P}_{1\_train}$ and $\mathbf{c}_{random\_1\_train}$, $\mathbf{P}_{1\_val}$ and $\mathbf{c}_{random\_1\_val}$, and

$\mathbf{P}_{1\_test}$ and $\mathbf{c}_{random\_1\_test}$ respectively. The resultant dimensions of these matrices and vectors are shown in Table 6-5. These dimensions apply to every randomised input data set.

Table 6-5: The dimensions of the training data, validation data, and testing data

	Attribute matrix	Dimension	Class Vector	Dimension
Training Data Set	$\mathbf{P}_{1\_train}$	55 by 37	$\mathbf{c}_{random\_1\_train}$	1 by 37
Validation Data Set	$\mathbf{P}_{1\_val}$	55 by 5	$\mathbf{c}_{random\_1\_val}$	1 by 5
Testing Data Set	$\mathbf{P}_{1\_test}$	55 by 5	$\mathbf{c}_{random\_1\_test}$	1 by 5

6.4.3 Generation of Initial Weights

100 sets of random weights are then generated using the Nguyen-Widrow method for initialisation purposes. Based on the specification in CPS, 9 neurons are used for this input data set since (37+55)/10 equals to 9.

Since each neural network is based on 9 neurons, and the attribute matrix of the training data set contains 55 attributes, the dimensions of the resultant weight matrices for the hidden layer and output layer are therefore '9 by 55' and '1 by 9' respectively. These dimensions of weight matrices apply to all the 100 sets of random weights.

6.4.4 Network Training and Classification Simulation

Table 6-6 and Figure 6-2 display the distribution of the 5000 accuracies achieved by the candidate classifiers generated using the 5 training algorithms. The second column of Table 6-6 displays the average of the 1000 accuracies for each algorithm. The third column of Table 6-6 displays the distribution of the accuracies, where '40-50 (1)', for instance, is a notation showing that there is 1 classifier whose classification accuracy is larger than or equal to 40% and smaller than 50%.

Table 6-6: Accuracy scores of the five training algorithms

Training algorithm	Average classification accuracy (%)	Distribution of the 1000 classification accuracies
Adaptive Learning Rate backpropagation	71.54	40-50 (1); 50-60 (1); 60-70 (22); 70-80 (976)
Resilient backpropagation	76.04	70-80 (968); 80-90 (32)
Conjugate Gradient based on the Fletcher-Reeves update algorithm	73.46	60-70 (1); 70-80 (997); 80-90 (2)
Conjugate Gradient based on the Polak-Ribi閛e update algorithm	72.26	60-70 (1); 70-80 (979); 80-90 (20)
Levenberg-Marquardt algorithm	67.16	60-70 (683); 70-80 (315); 80-90 (2)

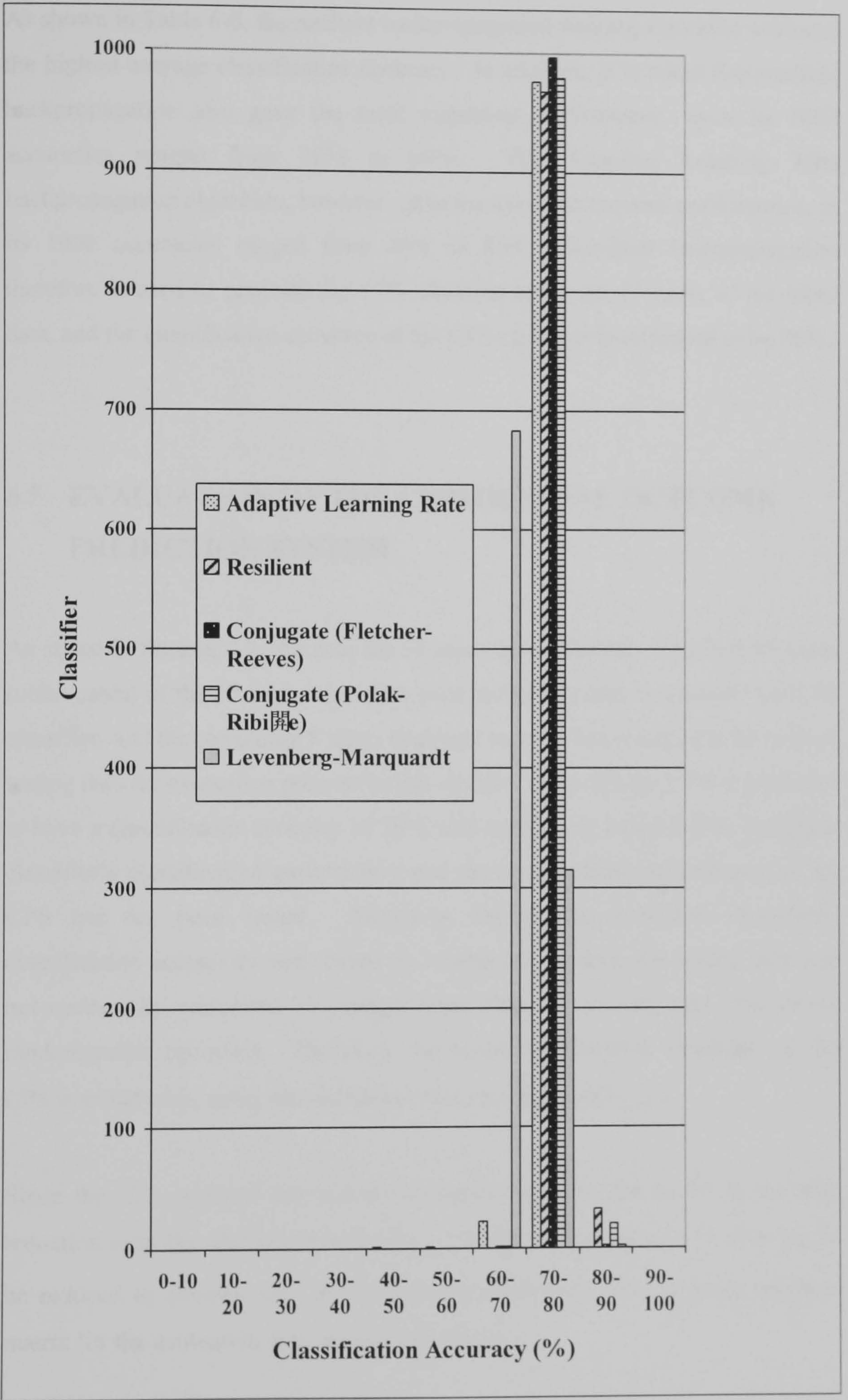


Figure 6-2: Distribution of the accuracies of the 5000 candidate classifiers

As shown in Table 6-6, the resilient backpropagation training algorithm achieved the highest average classification accuracy. In addition, it is noted that resilient backpropagation also gave the most consistent performance, since its 1000 accuracies ranged from 70% to 90%. The Adaptive Learning Rate backpropagation algorithm, however, gave the most inconsistent performance, as its 1000 accuracies ranged from 40% to 80%. Resilient backpropagation therefore is used to generate the CPS classifier using all 47 cases of the input data, and the classification accuracy of the CPS classifier is estimated to be 76%.

6.5 EVALUATION OF THE COMMERCIAL OUTCOME PREDICTION SYSTEM

As stated in Section 6.2, the data for 55 cases was collected. The first 47 cases (older cases) of the 55 cases were then used as training data to generate the CPS classifier, and the remaining 8 cases (younger cases) were reserved to be used as testing data for evaluation purpose in this section. Although the CPS is estimated to have a classification accuracy of 76%, this accuracy is based on the candidate classifier's classification performance and the classification performance of the CPS has not been tested. Moreover, though the candidate classifiers' classification accuracies were based on unseen testing data, the testing data was not necessarily comprised of younger cases than the training data, due to the randomisation operation. Therefore, the actual classification capability of the CPS is tested here, using the evaluation data set ($\mathbf{X}_{2eval}$ and $\mathbf{c}_{2eval}$).

Since the CPS reduced the number of attributes from 294 to 55 in the data reduction step, the attribute matrix ($\mathbf{X}_{2eval}$) of the evaluation data set also has to be reduced to contain only the 55 selected attributes. This reduced attribute matrix for the evaluation data set is denoted as $\mathbf{X}_{2eval_R}$.

Since the initial weights generated for network training are different each time, the classification accuracy resulted will also be different each time. Thus, 100 CPS classifiers are generated using the input data set ($\mathbf{X}_{2IPS_R}$ and $\mathbf{c}_{2IPS}$) as training data. Each of these 100 CPS classifiers is then tested using the evaluation data set ($\mathbf{X}_{2eval_R}$ and $\mathbf{c}_{2eval}$) as testing data. Consequently, the actual classification accuracies of the 100 CPS classifiers are: 39 classifiers achieved 75% accuracy (2 misclassifications), 52 classifiers achieved 87.5% accuracy (1 misclassification), and 9 classifiers achieved 100% accuracy (no misclassification). The average of these 1000 accuracies is 83.75%. The distribution of these accuracies is illustrated in Figure 6-3.

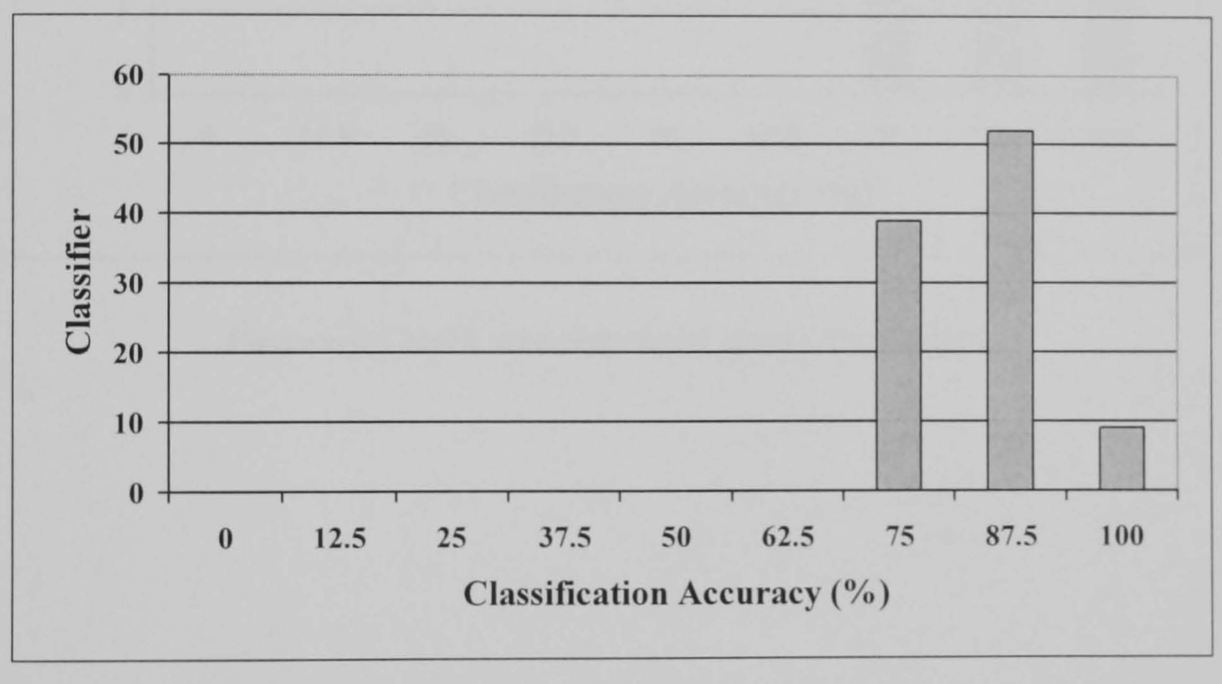


Figure 6-3: Classification accuracies of the 100 CPS classifiers

In order to evaluate the impact of the data reduction step on classification performance, CPS classifiers are generated using the original attribute matrix ($\mathbf{X}_{2IPS}$) that contains 294 instead of 55 attributes. Due to the random initial weight effect, 10 CPS classifiers are generated, and each tested using the evaluation data set ($\mathbf{X}_{2eval}$ and $\mathbf{c}_{2eval}$) as testing data. The 10 classification accuracies resulted are: 4 classifiers achieved 75% accuracy (2 misclassifications), 4 classifiers achieved 87.5% accuracy (1 misclassification), and 2 classifiers achieved 100% accuracy (no misclassification). The average of these 10 accuracies is therefore 85%. The distribution of these 10 accuracies is illustrated in Figure 6-4. These 10 classification accuracies are similar to those resulted using the reduced attribute matrix ($\mathbf{X}_{2IPS_R}$), in which the accuracies ranged from 75% to 100%, with a majority achieving 75% or 87.5%.

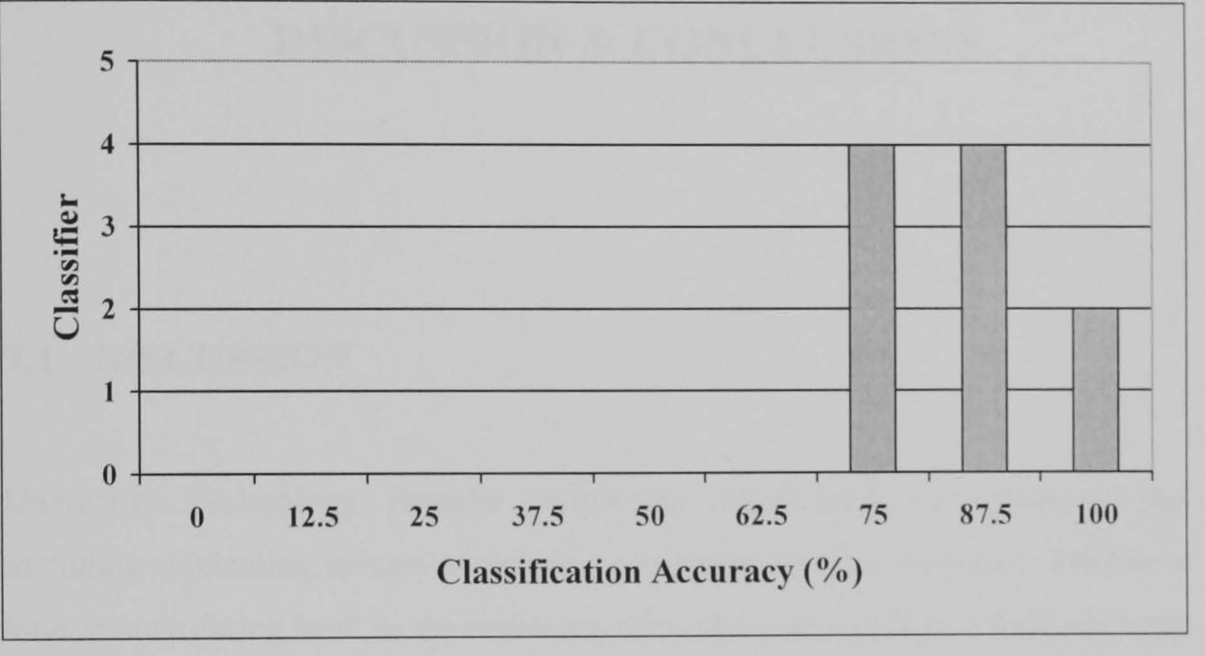


Figure 6-4: Classification accuracy of the 10 CPS classifiers

CHAPTER SEVEN

DISCUSSION & CONCLUSIONS

7.1 DISCUSSION

University Technology Transfer contributes significantly to various parties including academics, private companies, consumers and the economy. Despite a long history dating back to the beginning of modern universities, it has only been heavily promoted by Governments since the 1980s. Led by the United States, universities all over the world are increasingly establishing university technology transfer offices to maximise the return generated from University Technology Transfer activities. Simultaneously, research literature dedicated to improving the understanding and productivity of university technology transfer offices has burgeoned. Nonetheless, while University Technology Transfer has proven to be a source of lucrative income for a minority of universities, most of their counterparts are still struggling to break even as they fail to locate winning inventions from their portfolios. Despite the widely acknowledged importance of invention assessment for project selection, very little research has focused on its current practices or formal invention assessment methods.

The literature review performed here found that little has been done regarding formal invention assessment methods. The usual approach was to identify the critical factors affecting the final technology transfer outcome through qualitative methods, such as interviews, and/or apply latent variable models to analyse the correlation and the potential predictive power of these critical factors. Due to the lack of non-intuitive invention assessment methods in the University Technology Transfer domain, studies from two other domains have been reviewed. Namely, studies regarding the use of evaluative bibliometric indicators for the research proposal selection process from the Science and Research Assessment domain, and quantitative financial classification applications for investment option selection or outcome prediction purposes. Analogous to the invention selection task in University Technology Transfer, research proposal selection and investment option selection/outcome prediction also involve a high degree of uncertainty. While intuitive assessment remains as the chief method, quantitative methods like evaluative bibliometric indicators and financial classification applications have been used to achieve better performance and more informed assessments in the Science and Research Assessment and financial sectors. In particular, it was found that the financial sector went through a stage which was similar to the current situation in University Technology Transfer, where risky investment decisions were based solely on intuitive assessment using critical factors. Nowadays, classification applications making use of quantitative financial indicators, and the knowledge from qualitative critical factors have been widely adopted in the financial sector, and are producing promising results. This research therefore assumes that predictive classification applications, making use of evaluative bibliometric indicators and critical factors, can lead to fruitful results when used for invention assessment.

During the review of critical-factor studies for invention assessment, inconsistent findings regarding a subset of critical factors were frequently located. The inconsistency was caused by various studies using data provided by different technology transfer offices. This research therefore assumes that a subset of critical factors are non-transferable among universities, and can only be identified based on a university's own invention data. The notion of non-transferable critical factors is also compatible with the general understanding of invention assessment as a process which can only be learned through accumulative experience, and can not be taught. This compatibility further supports the assumption that classification applications may be fruitful for invention assessment, as supervised classification algorithms, such as decision tree induction or neural networks, also share this self-learning proposition, in that these algorithms develop classification rules through learning from past cases.

Simultaneously, a subset of critical factors was consistently supported by studies using different office data. This implied that in addition to the subset of non-transferable critical factors, there exists another subset of critical factors that is transferable among universities despite the use of different data. These transferable critical factors therefore represent potentially important predictor variables that can be used commonly by different universities for invention outcome prediction purposes. This research therefore assumes that the predictor variables used for invention outcome prediction can be derived from both transferable and non-transferable critical factors. Particularly, transferable critical factors can be learned from others, such as through research literature or experts, non-transferable critical factors can be learned from an office's own historical data.

Based on the above assumption, a prototype CPS based on decision tree induction was developed in Chapter Three. The predictor variables specified by the prototype CPS consisted of: 1) invention attributes based on critical factors suggested in University Technology Transfer literature, and 2) existing invention attributes available from an office's database. The dependent variable was the final outcome class of an invention. The prototype CPS comprised of six steps. They start from input data set specification, candidate classifier generation, to the generation of the prototype CPS classifier.

Based on a data source, an input data set containing 23 invention cases is obtained. Following the steps of the prototype CPS, 60 candidate classifiers were generated. Most of the candidate classifiers achieved satisfactory classification accuracy, at an average of 66%. However, the individual classification accuracies of the candidate classifiers ranged from 48% to 78%. The low end of the range showed that the prediction accuracies of certain classifiers were unacceptably low. Nonetheless, this supports the design of the prototype CPS, in which it generated a number of candidate classifiers in order to arrive at the optimum parameter values to be used for the generation of the prototype CPS classifier. Among the 60 candidate classifiers, the best one accurately predicted the class labels in 18 out of 23 cases, which is equivalent to an accuracy rate of 78%. This candidate classifier used only 5 predictor variables to compute the class labels. These 5 variables all belonged to 'Month 12' and 4 of them were similar to the transferable critical factors identified in the literature review. The prototype CPS classifier was therefore generated based on the definitions of the best candidate classifier, and is estimated to have an accuracy of 78%.

However, due to the use of the adaptive boosting technique, each tree classifier resulted would be comprised of several micro-classifiers. Consequently, the prototype CPS classifier comprised of 13 micro-classifiers, with each generally consisting of merely two nodes. Meaning that only two attributes were used at a time to arrive at decisions, such classification logic is unlikely to be correct. In view of the achievements and limitations of the prototype CPS, it is believed that although the prototype CPS classifier seemed to be capable of providing reasonably accurate predictions, it requires significant improvements.

In order to improve the design of the final CPS, three surveys were carried out in Chapter Four. Namely, the invention disclosure form survey, the questionnaire survey, and the interview survey. The first two surveys identified a list of potential attributes for invention assessment purposes, based on the important invention attributes requested in the invention disclosure forms of about 30 universities. These attributes were then added to the list of potentially predictive attributes proposed by the final CPS as potential predictor variables. The interview survey investigated the intuitive invention assessment process, in doing so it captured several insightful findings, which were then incorporated into the design of the final CPS.

Based on the findings derived from the interview survey, it was found that the process of invention assessment was not evenly applicable to different invention projects, and was largely dependent on the stage of commercialisation development. Most practitioners surveyed also admitted that the perceived status and/or the perceived scientific capability of the inventor, affected the rigor of the invention assessment process. This research therefore proposed that evaluative bibliometric indicators reflecting a researchers' scientific capability (such as publication counts and patent citations) may serve as good candidates of predictor variables for invention assessment purposes. Moreover, the interviewees explained that multiple rounds of assessment were often conducted at critical points in time structured around the patent expenses schedule, including points in time like 'Month 0', 'Month 12', 'Month 30'. In light of the reliance on perceived status/scientific capability of inventors, and the use of the patent expenses schedule to structure assessment rounds, relevant bibliometric indicators were added to the list of potentially predictive attributes proposed by the CPS as potential predictive variables, and those critical points in time were set to be the data collection points for the CPS.

Furthermore, it is found that the invention attributes used by the interviewees for invention assessment were often quantitatively or categorically measurable and non-technical. Examples of these attributes are 'number of years of research experience', 'the number of patents owned by the inventor', and 'the number of prior arts identified'. The interviewees also expressed that they often do not/are not able to judge qualitative and technical attributes such as 'what the invention is' and 'the advantage of the invention over the current method', as they do not possess more expertise than the inventors regarding the invention field. Because of this, only quantifiable and non-technical invention attributes were retained from the list of attributes proposed by the CPS. Moreover, based on the interview survey, it is found that the scoring system used by the managers for invention assessment can be expressed as a function of weighted attributes. Consequently, the multi-layer backpropagation neural networks method was chosen for the CPS, due to its classification mechanism of weighted attributes, its universal function approximation ability, and its proven classification performance against other methods in many real-world problems.

Based on a data source, 294 invention attributes for 55 cases were collected. From which 47 out of the 55 cases were used as the input data set for the CPS, and the remaining 8 cases were reserved as evaluation data set. Following other steps of the CPS, 5000 candidate classifiers were generated using 5 training algorithms. These candidate classifiers were generated and tested using the input data set for the CPS. The average classification accuracy (which is an average of the classification accuracies of 1000 candidate classifiers, each based on unseen cases) achieved by each of the 5 training algorithms was satisfactory, ranging from 67% to 76%. Removing the Levenberg-Marquardt algorithm (whose average classification accuracy was 67%), the remaining 4 training algorithms achieved similar average classification accuracies ranging from 72% to 76%. This demonstrated that the CPS is capable of producing candidate classifiers that give stable and reasonably accurate classification performances.

Among the remaining 4 training algorithms, resilient backpropagation achieved the highest average classification accuracy, at 76%. Adaptive learning rate backpropagation, conjugate gradient (Fletcher-Reeves), and conjugate gradient (Polak-Ribi閕e) produced average classification accuracies of 72%, 73% and 72% respectively. Additionally, resilient backpropagation also produced the most consistent performance. Among the 1000 candidate classifiers generated using resilient backpropagation, 968 of them achieved classification accuracies within the range of 70% to 80%, and the accuracies of the remaining 32 classifiers ranged from 80% to 90%. The 4000 (4 x 1000) candidate classifiers produced using the 4 other training algorithms achieved classification accuracies falling within the range of 40% to 90%. This showed that the candidate classifier generated using the resilient backpropagation training algorithm achieves the maximum prediction accuracy. The CPS classifier was therefore generated using resilient backpropagation, utilising the whole input data set (47 cases) as training data, its classification accuracy was estimated as 76%.

Although the CPS is estimated to have a classification accuracy of 76%, this accuracy is based on the candidate classifier's classification performance. The actual classification performance of the CPS had not yet been tested. Moreover, though the candidate classifiers' classification accuracies were based on unseen testing data, the testing data was not necessarily comprised of younger cases than the training data, due to the randomisation operation of the CPS. In order to test the actual classification accuracy of the CPS using new invention cases, the evaluation data set (the 8 of the 55 cases) would now be used as the testing data, in order to derive the actual classification accuracy of the CPS classifier.

Due to the use of random initial weights by neural networks, 100 CPS classifiers were generated using the input data set (47 cases) as training data. Each of these 100 CPS classifiers is tested using the evaluation data set (8 cases) as the testing data. Consequently, the actual classification accuracy achieved by each of these 100 classifiers ranged from 75% to 100%, with an average of 83.75%. Based on the actual classification accuracy, and the fact that the class data is measured by the monetary return generated, it was demonstrated that the CPS is capable of predicting an invention's likely future monetary return.

In addition, the number of predictor variables was reduced from 294 to 55 by the data reduction step of the CPS. In order to evaluate the effect of the data reduction step on the CPS's classification accuracy, 10 CPS classifiers were generated using the same settings, except that they used the original input data (with 294 variables), rather than the reduced input data set (with 55 variables) as training data. The actual classification accuracies of these 10 classifiers ranged from 75% to 100%, with an average of 85%. As shown, the accuracies achieved by the classifiers based on the original and reduced input data set are very similar. Therefore it can be said that the data reduction step has successfully captured the most discriminative predictor variables, without sacrificing the prediction accuracy of the resultant classifier.

The input data set collected consisted of 294 invention attributes stemming from the 7 categories. Although the data reduction step of the CPS reduced the full set (294 attributes) to a reduced set of 55 attributes, the relative importance of each category is not clear. For example, although it appears that the number of *Inventor* attributes (19 out of 55) is larger than that of *Expense* attributes (5 out of 55) in the reduced set, this does not necessarily mean that *Inventor* attributes are more important than the *Expense* attributes. There were many more *Inventor* attributes to choose from in the original input data set (110 *Inventor* attributes and 20 *Expense* attributes), so it was likely that the neural network would find a greater number of important *Inventor* attributes. In fact, as a percentage of the original number of attributes in a category, the network only found 17 % of *Inventor* attributes (19 out of 110) to be of importance, whilst it considered 25 % of the *Expense* attributes (5 out of 20) to be important. Therefore the proportion of attributes from each category within the reduced set should not be considered to represent the importance of the category. It should also be noted that the weight values assigned by the neural network will bias the importance of each attribute. Therefore, the relative importance of attributes or attribute categories has not been resolved due to the black box mechanism of neural networks.

The research design adopted in this thesis is iterative with respect to achieving the research aim. It is based on a cyclic process of prototyping, testing, analysing, and refining a system in progress. Through this iterative approach, surveys with practising technology transfer offices were conducted for the purpose of informing and improving the design of the CPS, as iterations of system design were implemented. By means of the iterative design process, new and unexpected findings emerged directly from both system implementation and evaluation. As the design of the CPS evolved, it defined and redefined the assumptions about the invention assessment process and the underlying distribution model where the system is based upon. This iterative research approach is radically different from the hypothesis testing/developing approach adopted by the majority of the domain literature. While the former generates and capitalises on intermediate findings, the latter stops at the generation of immediate findings. Due to the high level of uncertainty and unknown involved in the domain of university technology transfer, and the fact that actual system performance cannot be completely predicted in advance, iterative research process is therefore imperative for guiding modelling decisions and system optimisation. Consequently, the CPS was developed through an ongoing dialogue between the researcher, the problem domain, and the invention cases.

Nevertheless, no observational study regarding human-machine interaction was conducted in the current research, due to the limited availability of invention assessment practitioners. Such study could be useful as it might reveal important conflict issues which were not possible to be addressed, given the constraints faced by this research.

Finally, several limitations have been found with this research. The interview survey in Chapter Four generated valuable findings, but also highlighted some of the limitations inherent in qualitative methods, including interviewer bias, and small sample size (only 7 technology transfer offices agreed to be interviewed).

Furthermore, the task given to the classifiers in this research is a binary classification task, in which classifier predicts invention cases to be either a 'success' or 'failure'. While the classifiers achieved high levels of accuracy for the binary classification task, it was not clear if the accuracy level could be maintained when more than two classes were used. Nonetheless, binary classification is one of the most important classification arrangements, as it is commonly adopted in many imperative investment systems in the financial sector. Typical applications include corporate bankruptcy prediction and credit risk evaluation.

In addition, the input data sets for both the prototype CPS and the final CPS were largely generated by manual codification of hard copies of documents into numeric data, due to the lack of electronic data from the data source. When certain data needed by the CPS was missing from the hard copies, it was not clear whether the information was really missing or merely did not exist. This problem caused the discarding of a subset of potentially important invention attributes, and may also have had an impact on the classification accuracies of the resultant classifiers.

Finally, a major barrier to this research has been the shortage of data sources. When locating sources for data collection, it was found that universities were extremely reluctant in sharing their information due to the issue of confidentiality. Consequently, only one university agreed to provide invention information. This constituted a major limitation to the research, as all the input data sets were based on the information from one university. Nonetheless, while multiple data sources would be ideal, it is good that this research obtained genuine invention data rather than using computer generated data like much other research.

This research has contributed to the existing knowledge in the University Technology Transfer domain, filling some of the gaps. First of all, this research investigated the invention assessment process, which was one of the least researched areas in the University Technology Transfer domain. In addition, while literature investigating critical factors or indicators for University Technology Transfer purposes abound, very little has been done to test their predictive power or develop formal systems to capitalise on the findings. This research empirically tested the proposed factors, and established the CPS to make use of these factors. These contributions stem from the inter-disciplinary nature of the research. The research demonstrated that standard practices within the quantitative classification discipline, can be joined together, and be fruitfully applied to the University Technology Transfer domain. As such, the often intuitively handled invention assessment task is tackled quantitatively by means of the CPS.

7.2 CONCLUSIONS

The aim of this research was to develop a Commercial Outcome Prediction System (CPS) that can be used by individual technology-transfer offices. The method was based on the utilisation of data from past invention cases. The conclusions of the research can be summarised as follows:

- The CPS is a system which predicts the likely monetary return (success or failure) that would be generated by an invention.

- A list of potentially predictive attributes has been developed based on the background surveys conducted in this research. This list, although not exhaustive, is suggested to have covered important invention attributes for invention assessment purposes.

- Attributes to be recorded in the attribute matrix should include as much case data as possible, based on the list of potentially predictive attributes, and any other data available. An excess of data is not a problem, as the data reduction step of the CPS reduces the number of attributes by identifying the discriminative attributes.

- The classification accuracies remained almost unchanged after the data reduction step of the CPS, showing that the data reduction step successfully captured the most discriminative predictor variables, without sacrificing the prediction accuracy of the resultant classifier.

- Based on the available data, the 5 training algorithms employed by the CPS achieved stable and accurate classification performances, with prediction accuracies ranging from 67% to 76%.

- Among the 5 training algorithms, candidate classifiers based on the resilient backpropagation method, gave the highest classification accuracies and the most consistent performance.

- The actual classification accuracy of the CPS ranged from 75% to 100%, with an average of 83.75%. This demonstrated the potential of the CPS to predict an invention's likely future monetary return.

- Due to the black box mechanism of neural networks, the relative importance of various attributes has not been fully resolved.

- A limitation of this research was the small size of the input data set. However, this problem would be alleviated as the user technology transfer office adds more data over time. A larger input data set is also likely to result in a more accurate prediction system.

7.3 FURTHER WORK

The primary aim of developing the Commercial Outcome Prediction System (the CPS) has been realised. However, there are further refinements that could be achieved in various areas, which are presented below.

First of all, the classification results achieved by the CPS are based on one data source. Applying the CPS to other data sources would demonstrate any limitations that have not been found, due to this limited data source.

Although various surveys were performed to extract the knowledge of the technology-transfer experts, applying similar studies to a wider sample of universities may generate new insights. In addition, particular attention should be paid to investigating how invention attributes adopted by the experts are measured during the intuitive invention assessment process.

While this research has covered an enormous number of invention attributes, there are two more sources of attributes that might be worth examining. The first one is the UKSIC system (UK Standard Industrial Classification of Economic Activities), which could be studied with regard to the patent classification systems used by patent offices. The second source could be the generation of invention attributes by applying text mining technologies to patent documents and published papers. This may capture some of critical qualitative factors for invention assessment.

Data collection in the current research was done manually. Further work could be done on developing efficient data acquisition tools to generate attribute matrices from technology transfer offices' existing databases. An automatic case update function that regularly updates the existing attribute matrices with new invention cases would be a useful feature in such software.

Despite the good prediction performance of neural networks, the fundamental 'black box' problem still applies. This hinders the use/development of systems involving neural networks. Possible areas to further the research in this area include: the evaluation of derivatives between the input and output; the development of simplified network to ease interpretation; and the development of rule discovery during each learning cycle.

Finally, post-assessment functions would be useful to enhance the understanding of the causes of the invention outcomes. For instance, attribute-category insertion or extraction functions could be developed, to estimate the impact of attribute-categories on the predicted result. Or a particular category of attributes could be isolated in order to analyse its effect. The addition of knowledge-management functions would also be desirable. Possible knowledge derived from invention prediction systems includes the change of critical-invention-attributes/prediction rules, invention case properties, and prediction performance over time. Knowledge-management functions could include the capture, recording, analysis, and visualisation of this knowledge over time.

REFERENCE

Aernoudt, R. (1999). "Business angels, should they fly on their own wings?" Venture Capital – An International Journal of Entrepreneurial Finance 1(2): 187-195.

Aernoudt, R. (2005). "Business angels: the smartest money for starters? Plea for a renewed policy focus on business angels." International Journal of Business 10(3): 271-284.

Agrawal, A. and R. Henderson (2002). "Putting patents in context: Exporing knowledge transfer from MIT." Management Science 48(1): 44-60.

Alici, Y. (1995). "Neural networks in corporate failure prediction: the UK experience." Proceedings of the III International Conference on Neural Networks in the Capital Markets: 393-406.

Altman, E. I. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." Journal of Finance 23(4): 589-609.

Altman, E. I. and R. Haldeman (1995). "Corporate Credit-Scoring Models: Approaches and Tests for Successful Implementation.." The Journal of Commercial Lending 77: 10-22.

Altman, E. I., G. Marco, et al. (1994). "Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience)." Journal of Banking and Finance 18(3): 505-529.

Arora, A. (1996). "Contracting for tacit knowledge: the provision of technical services in technology licensing contracts." Journal of Development Economics 50: 233-256.

Atiya, A. F. (2001). "Bankruptcy prediction for credit risk using neural networks: A survey and new results." IEEE Transactions on Neural Networks 12(4): 929-935.

AURIL (2003a). AURIL Handbook of Intellectual Property Management. United Kingdom, AURIL.

AURIL (2003b). Continuing Professional Development Framework for Knowledge Transfer Practitioners. United Kingdom, AURIL.

AUTM (1997). The AUTM Licensing Survey, Fiscal Year 1996. Norwalk, CT, USA.

AUTM (2000). The AUTM Licensing Survey, Fiscal Year 1999. Norwalk, CT, USA.

Baesens, B., R. Setiono, et al. (2003). "Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation." Management Science 49(3): 312-329.

Beer, D. (2006). "In practice: business angels: angel eyes." Accountancy 137(1351): 72-75.

Bond, E. U. and M. B. Houston (2003). "Barriers to Matching New Technologies and Market Opportunities in Established Firms." Journal of Product Innovation Management 20: 120-135.

Boritz, J. and D. Kennedy (1995). "Effectiveness of Neural Network Types for Prediction of Business Failure." Expert Systems with Applications 9: 504-512.

Bozeman, B. (2000). "Technology transfer and public policy: a review of research and theory." Research Policy 29: 627-655.

Butler, L. (2004). ARC-Supported Research: The Impact of Journal Publication Output 1996-2000. A. R. Council. Canberra.

Campanario, J. M. (1995) Rejecting Nobel class articles and resisting Nobel class discoveries. (available on http://www2.uah.es/jmc/nobel.html) Volume, DOI:

Carlsson, B. and A. C. Fridh (2002). "Technology Transfer in United States Universities: A Survey and Statistical Analysis." Journal of Evolutionary Economics 12: 199-232.

Chan, G. K. L. (1976). "The foreign language barrier in science and technology." International Library Review 8: 317-325.

Chatfield, C. and A. J. Collins (1980). Introduction to Multivariate Analysis. London, Chapman and Hall.

Chen, A., M. T. Leung, et al. (2003). "Application of neural networks to an emerging market: forecasting and trading the Taiwan Stock Index." Computers & Operation Research 30: 901-923.

Colyvas, J., M. Crow, et al. (2002). "How do university inventions get into practice?" Management Science 48(1): 61-72.

Cordes, J. J., H. R. Hertzfeld, et al. (1999). A survey of high technology firms. Office of Chief Counsel for Advocacy, United States Small Business Administration: Contract No. SBA-8141-OA94.

Crown (2003). Lambert review of business-university collaboration – final report. London, Her Majesty's Stationery Office.

Cybenko, G. (1989). "Approximation by superpositions of a sigmoidal function." Mathematical Control Signals Systems 2: 303 – 314.

Davis, R. H., D. B. Edelman, et al. (1992). "Machine learning algorithms for credit-card applications." Journal of Mathematics applied in Business and Industry 4: 43-51.

Desai, V. S., D. G. Conway, et al. (1997). "Credit-scoring models in the credit-union environment using neural networks andgenetic algorithms." Journal of Mathematics Appliedin Business & Industry 8: 323-346.

Desai, V. S., J. N. Crook, et al. (1996). "A Comparison of Neural Networks and Linear Scoring Models in the Credit Union Environment." European Journal of Operational Research 95: 24-37.

DTI, H. Treasury, et al. (2002). Investing in Innovation: a strategy for science, engineering and technology.

Egghe, L., R. Rousseau, et al. (1999). "The "own-language preference": Measures of relative language self-citation." Scientometrics 45(2): 217-232.

Ehrlich, S., A. DeNoble, et al. (1999). "After the cash arrives: a comparative study of venture capital and private investor involvement in entrepreneurial firms" Journal of Business Venturing 9: 67-82.

Enke, D. and S. Thawornwong (2005). "The use of data mining and neural networks for forecasting stock market returns." Expert Systems with Applications 29: 927-940.

Fiet, J.O. (1995). "Risk avoidance strategies in venture capital markets" Journal of Management Studies 32: 551-574.

Foltz, J., B. Barham, et al. (2000). "Universities and Agricultural Biotechnology Patent Production." Agribusiness 16(1): 82-95.

Foster, D. P. and R. A. Stine (2004). "Variable Selection in Data Mining: Building a Predictive Model for Bankruptcy." Journal of the American Statistical Association 99: 303-313.

Freear, J., J.E. Sohl, et al. (2002). "Angels on angels: financing technology-based ventures – a historical perspective." Venture Capital 4(4): 275-287.

Friedman, J. and J. Silberman (2003). "University Technology Transfer: Do Incentives, Management, and Location Matter?" Journal of Technology Transfer 28(1): 17-30.

Gilbert, R. and C. Shapiro (1990). "Optimal patent length and breadth." RAND Journal of Economics 21(1): 106-112.

Gopalakrishnan, S. and M. D. Santoro (2004). "Distinguishing between knowledge transfer and technology transfer activities: the role of key organisational factors." IEEE Transactions on Engineering Management 51(1): 57-69.

Gracie, S. (1999). "Pennies from heaven." Director 52(7): 40-43.

Greenwald, A. G. and E. S. Schuh (1994). "An ethnic bias in scientific citations." European Journal of Social Psychology 24: 623-639.

Gregorio, D. D. and S. Shane (2003). "Why do some universities generate more start-ups than others?" Research Policy 32: 209-227.

Gustin, B. (1975). The Emergence of the German Chemical Profession, 1790 – 1867, University of Chicago. PhD.

Hagan, M. T., H. B. Demuth, et al. (1996). Neural Network Design. Boston, MA, PSW Publishing.

Hall, B. and R. Ham Ziedonis (2001). "The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995 " RAND Journal of Economics 32(1): 101-128.

Harhoff, D., F. Scherer, et al. (2002). "Citations, Family Size, Opposition and the Value of Patent Rights." Research Policy 1596: 1-21.

Harrison, R. and C. Mason (1996). “Developments in the promotion of informal venture capital in the UK.” International Journal of Entrepreneurial Behaviour & Research 2(2): 6-33.

Haykin, S. (1999). Neural Networks - A comprehensive foundation. New Jersey, Prentice-Hall.

Henley, W. E. and D. J. Hand (1997). "Construction of a k-nearest neighbour credit-scoring system." Journal of Mathematics applied in Business and Industry 8: 305-321.

Hornik, K., M. Stinchcombe, et al. (1989). "Multilayer feedforward networks are universal approximators." Neural Networks 2(5): 359-366.

Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components." Journal of Educational Psychology 24: 417-441.

Hsu, D. H. and T. Bernstein (1997). "Managing the university technology licensing process: findings from case studies." Journal of the Association of University Technology Managers 9: 1-33.

Hutchinson, M., A. W. Lo, et al. (1994). "A nonparametric approach to pricing and hedging derivative securities via learning networks." The Joumnal of Finance
XLIX: 851-889.

Ince, H. and T. B. Trafalis (2004). "Kernel Principal Component Analysis and Support Vector Machines for Stock Price Prediction." IEEE International Joint Conference on Neural Networks: 2053-2058.

Jaffe, A. B., R. Haldeman, et al. (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations." Quarterly Journal of Economics 108: 577-598.

Jaffe, A. B. and M. Trajtenberg (1996). Flows of knowledge from universities and federal labs: modelling the flow of patent citation over time and across institutional and geographic boundaries. Cambridge, MA, National bureau of Economic Research.

Jensen, H. L. (1992). "Using neural networks for credit scoring." Managerial Finance 18: 15-26.

Jensen, R. and M. Thursby (2001). "Proofs and prototypes for sale: the licensing of university inventions." American Economic Review 91(1): 240-259.

John, G. H., P. Miller, et al. (1996). "Stock selection using rule induction." IEEE Expert 11: 52-58.

Jones, L. K. (1990). "Constructive approximations for neural networks by sigmoidal functions." Proceedings of the IEEE, 78(10): 1586-1589.

Jungman, H., R. Harrison, et al. (2005). "The rise of knowledge investor: their investment process and complementarities with capital investors." National Technology Agency of Finland and The Finnish Ministry of Trade and Industry, ProACT.

Klemperer, P. (1990). "How broad should the scope of patent protection be?" RAND Journal of Economics 21(1): 113-130.

Korevaar, J. C. (1996). "Validation of bibliometric indicators in the field of mathematics." Scientometrics 37(1): 117-130.

Kostoff, R. N. (1997). "Citation Analysis Cross-Field Normalization: A New Paradigm." Scientometrics 39(3): 225-230.

Kurková, V. (1991). "Kolmogorov's theorem is relevant." Neural Computation, 3(4):617-622.

Kurková, V. (1992). "Kolmogorov's theorem and multilayer neural networks." Neural Computation, 5(3):501-506.

Lacher, R. C., P. K. Coats, et al. (1995). "A neural network for classifying the financial health of a firm." European Journal of Operational Research 85: 53-65.

Lanjouw, J. O. (1998). "Patent protection in the shadow of infringement: simulation estimations of patent value." Review of Economic Studies 65: 671-710.

Lazo, J. G., M. Maria, et al. (2000). A hybrid genetic-neural system for portfolio selection and management. Proceedings of the 7th International Conference on Engineering Applications of Neural Networks, Kingston Upon Thames, U.K.

Leigh, W., M. Paz, et al. (2002). "An analysis of a hybrid neural network and pattern recognition technique for predicting short-term increases in the NYSE." The International Journal of Management Science 30: 69-76.

Lerner, J. (1994). "The importance of patent scope: An empirical analysis." RAND Journal of Economics 25(2): 319-333.

Lindel鯽, P. and H. L鯽sten (2004). "Proximity as a Resource Base for Competitive Advantage: University-Industry Links for Technology Transfer." Journal of Technology Transfer, 29: 311-326.

Liu, M. X. (1993). "Progress In Documentation - The Complexities of Citation Practice – A Review of Citation Studies." Journal of Documentation 49(4): 370-408.

MacRoberts, M. and B. MacRoberts (1996). "Problems of Citation Analysis." Scientometrics 36(3).

Markman, G. D., P. H. Phan, et al. (2005). "Entrepreneurship and university-based technology transfer." Journal of Business Venturing 20: 241-263.

Mason, C. and R. Harrison (1995). "Closing the regional equity capital gap: the role of informal venture capital." Small Business Economics 7: 153-172.

Mason, C. and R. Harrison (1996). "Why 'business angels' say no: a case study of opportunities rejected by an informal investor syndicate" International Small Business Journal 14(2): 34-51.

Mason, C. and R. Harrison (1999). "An Overview of Informal Venture Capital Research" Venture Capital – An International Journal of Entrepreneurial Finance 1(2): 95-101.

Mason, C. and M. Stark (2002). "What do investors look for in a business plan? A comparison of bankers, venture capitalists and business angels." The Proceedings of the 25th ISBA National Small Firms Conference: Competing Perspectives of Small Business and Entrepreneurship.

Matutes, C., P. Regibeau, et al. (1996). "Optimal patent design and the diffusion of innovations." RAND Journal of Economics 27(1): 60-83.

McAdam, R., W. Keogh, et al. (2005). "Defining and improving technology transfer business and management processes in university innovation centres." Technovation 25: 1418-1429.

Merges, R. P. and R. R. Nelson (1990). "On the Complex Economics of Patent Scope." Columbia Law Review 90: 839-916.

Meseri, O. and S. Maital (2001). "A Survey Analysis of University-Technology Transfer in Israel: Evaluation of Projects and Determinants of Success." Journal of Technology Transfer 26(1-2): 115-126.

Meyer, M. (1999). "Does science push technology? Patents citing scientific literature." Research Policy 29(3): 409-434.

Mike, W., S. Birley, et al. (2004). "Entrepreneurship and University Technology Transfer." Journal of Technology Transfer 29: 235-246.

Mowery, D., B. Sampat, et al. (2002). "Learning to patent: institutional experience, learning, and the characteristics of US university patents after the Bayh-Dole Act, 1981-1992." Management Science 48(1): 73-89.

Narin, F. (1976). Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity, NSF C-637, National Science Foundation

Narin. F., M. P. Carpenter. et al. (1984). "Technological Performance Assessments Based On Patents And Patent Citations." IEEE Transactions On Engineering Management 31(4): 172-183.

Narin, F., K. S. Hamilton, et al. (1997). "The increasing linkage between US technology and public science." Research Policy 26: 317-330.

Narin, F., D. Olivastro, et al. (1994). "Bibliometrics/Theory, Practice, and Problems." Evaluation Review 18(1): 65-76.

Neely, C., P. Weller, et al. (1997). "Is technical analysis in the foreign exchange market profitable? A genetic programming approach." The Joumnal of Financial and Quantitative Analysis 32: 405-426.

Nelson, R. R. (2001). "Observations on the post Bayh-Dole rise of patenting at American universities." Journal of Technology Transfer 26: 13-19.

Nguyen, D. and B. Widrow (1990). "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights." Proceedings of the International Joint Conference on Neural Networks 3: 21-26.

Nordhaus, W. D. (1967). The Optimal Life of a Patent. New Haven, Yale University.

Odom, M. D. and R. Sharda (1990). "A neural network model for bankruptcy prediction." Proceedings of the International Joint Conference on Neural Networks 2: 163-168.

Owen-Smith, J. and W. W. Powell (2003). "The expanding role of university patenting in the life sciences: assessing the importance of experience and connectivity." Research Policy 32: 1695-1711.

Pakes, A. (1986). "Patents as options: some estimates of the value of holding European patent stocks." Econometrica 54(4): 755-784.

Pearson, K. (1901). "On lines and planes of closest fit to systems of points in space." Philosophical Magazine 2: 559-72.

Podolny, J. M. and T. E. Stuart (1995). "A Role-Based Ecology of Technological Change." American Journal of Sociology 100(5): 1224-1260.

Podolny, J. M. and T. E. Stuart (1996). "Networks, Knowledge, and Niches: Competition in the Worldwide Semiconductor Industry, 1984 - 1991." American Journal of Sociology 102(3): 659-689.

Refenes, A. N., A. Zapranis, et al. (1994). "Stock performance modeling using neural networks: a comparative study with regression models." Journal of Neural Networks 7: 375-388.

Rekha Pai, G. A. and R. Annapoorani (2002). Prediction of Industrial Sickness using Principal Component Analysis Augmented with Multiple Discriminant Analysis. Proceedings of International Conference On Operations Research for Development. Chennai.

Rekha Pai, G. A., R. Annapoorani, et al. (2004). "Performance analysis of a Statistical and an Evolutionary Neural Network based classifier for the prediction of industrial bankruptcy." Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems: 1033-1038.

Roberts, E. (1988). "Managing invention and innovation: what we've learned." Research Technology Management 31: 11-29.

Rogers, E. M., Y. Yin, et al. (2000). "Assessing the Effectiveness of Technology Transfer Offices at U.S. Research Universities." The Journal of the Association of University Technology Managers 12: 47-80.

Roman, J. and A. Jameel (1996). "Backpropagation and recurrent neural networks in financial analysis of multiple stock market returns." Proceedings of the Twenty-Ninth Hawaii International Conference on System Sciences 2: 454-460.

Ross, S. (1976). "The arbitrage theory of capital asset pricing." Journal of Economic Theory 13: 341-360.

Saad, E. W., D. V. Prokhorov, et al. (1998). "Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks." IEEE Transactions on Neural Networks 9: 1456-1470.

Salchenberger, L. M., E. M. Cianr, et al. (1992). "Neural networks: a new tool for predicting thrift failures." Decision Sciences 23: 899-916.

Scherer, F. M. and D. Harhoff (2000). "Policy implications for a world with skew-distributed returns to innvation." Research Policy 29: 559-566.

Schmiemann, M. and J. N. Durvy (2003). "New approaches to technology transfer from Publicly Funded Research." Journal of Technology Transfer 28(1): 9-15.

Schmoch, U. (1993). "Tracing the knowledge transfer from science to technology as reflected in patent indicators." Scientometrics 26(1): 193-211.

Schubert, A. and T. Braun (1996). "Cross-Field Normalization of Scientometric Indicators." Scientometrics 36(3): 311-324.

Schwartz, S. and J. Lopez Hellin (1996). "Measuring the impact of scientific publication. The case of the biomedical sciences." Scientometrics 35(1): 119-132.

Shane, S. (2001). "Technology opportunities and new firm creation." Management Science 47(2): 205-220.

Shane, S. (2002). "Selling university technology: Pattern from MIT." Management Science 48(1): 122-137.

Shane, S. and R. Khurana (2003). "Bringing individuals back in: The effects of career experience on new firm founding." Industrial and Corporate Change 12(3): 519-543.

Shane, S. and T. Stuart (2002). "Organisational Endowments and the Performance of University Start-ups." Management Science 48(1): 154-170.

Siegel, D. S., D. Waldman, et al. (2003a). "Assessing the impact of organizational practices on the relative productivity of university technology transfer offices: an exploratory study." Research Policy 32: 27-48.

Siegel, D. S., D. A. Waldman, et al. (2003b). "Commercial knowledge transfers from universities to firms: improving the effectiveness of university-industry collaboration." Journal of High Technology Management Research 14: 111-133.

Siegel, D. S., D. A. Waldman, et al. (2004). "Toward a model of the effective transfer of scientific knowledge from academicians to practitioners: qualitative evidence from the commercialisation of university technologies." Journal of Engineering Technology Management 21: 115-142.

Sine, W. D., S. Shane, et al. (2003). "The halo effect and technology licensing: the influence of institutional prestige on the licensing of university inventions." Management Science 49(4): 478-496.

Sorensen, E. H., K. L. Miller, et al. (2000). "The Decision Tree Approach to Stock Selection - An evolving tree model performs the best." Journal of Portfolio Management 27(1): 42-52.

Southon, M. and C. West (2006). "The magic numbers." Director 59(10): 18.

Sørheim, R. (2000). "Business angels as facilitators for further finance: an exploratory study." Journal of Small Business and Enterprise Development 12(2): 178-191.

Srivastava, R. P. (1992). Automating judgmental decisions using neural networks: a model for processing business loan applications,". Proceedings of the Annual Conference on Communications, . Kansas City, MO.

Sung, H. M., J. Lee, et al. (2005). "Hybrid genetic algorithms and support vector machines for bankruptcy prediction." Expert Systems with Applications 31(3): 652-660.

Sung, T. K., N. Chang, et al. (1999). "Dynamics of modeling in data mining: interpretive approach to bankruptcy prediction." Journal of Management Information System 16: 63-85.

Tam, K. Y. and M. Y. Kiang (1992). "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions." Management Science 38: 926-947.

Tang, T. C. and L. C. Chi (2005). "Neural networks analysis in business failure prediction of Chinese importers: A between-countries approach." Expert Systems with Applications 29: 244-255.

Tay, F. E. H. and L. J. Cao (2002). "Modified support vector machines in financial time series forecasting." Neurocomputing 48: 847-861.

Taylor, S. (1986). Modeling Financial Time Series. New York, Wiley.

Thursby, J. and S. Kemp (2002). "Growth and Productive Efficiency of University Intellectual Property Licensing." Research Policy 31: 109-124.

Thursby, J. G. and M. C. Thursby (2002). "Who is selling the ivory tower? Sources of growth in university licensing." Management Science 48(1): 90-104.

Thursby, J. G. and M. C. Thursby (2003). "Industry/University Licensing: Characteristics, Concerns and Issues from the Perspective of the Buyer." Journal of Technology Transfer 28: 207-213.

Tong, X. and J. D. Frame (1992). "Measuring national technological performance with patent claims data." Research Policy 23: 133-141.

Trafalis, T. B. and H. Ince (2002). "Benders decomposition technique for support vector regression." Proceedings of the 2002 International Joint Conference on Neural Networks 3: 2767-2772.

Trajtenberg, M. (1990). "A penny for your quotes: patent citations and the value of innovations." RAND Journal of Economics 21(1): 172-187.

Trajtenberg, M., R. Henderson, et al. (1997). "University versus corporate patents: a window on the basicness of invention." Economics of Innovation and New Technology 5(1): 19-50.

TSBS (2001). "Investment readiness – bidding guidance for demonstration projects." The Small Business Service, URN 01/1482.

UNICO, AURIL, et al. (2003). UK University Commercialisation Survey Financial Year 2002.

Van Osnabrugge, M. (1999). A comparison of business angel and venture capitalist investment procedures: an agency theory-based analysis. Frontiers of Entrepreneurial Research. Wellesley, MA: Babson College.

Van Osnabrugge, M. and R.J. Robinson (2000). Angel investing: matching start-up funds with start-up companies. In Mason, C. and M. Stark (2002). "What do investors look for in a business plan? A comparison of bankers, venture capitalists and business angels." 25th ISBA National Small Firms Conference: Competing Perspectives of Small Business and Entrepreneurship.

Walczak, S. (1999). "Gaining competitive advantage for trading in emerging capital markets with neural networks." Journal of Management Information System 16: 177-192.

Walczak, S. (2001). "An empirical analysis of data requirements for financial forecasting with neural networks." Journal of Management Information System 17: 203-222.

Wang, J. H. and J. Y. Leu (1996). "Stock market trend prediction using ARIMA-based neural networks." IEEE International Conference on Neural Networks 4: 2160-2165.

Weingart, P. (2005). "Impact of bibliometrics upon the science system: Inadvertent consequences?" Scientometrics 62(1): 117-131.

West, D. (2000). "Neural network credit scoring models." Computers & Operations Research 27: 1131-1152.

West, D., S. Dellana, et al. (2005). "Neural network ensemble strategies for financial decision applications." Computers & Operations Research 32: 2543-2559.

Wilson, R. L. and R. Sharda (1994). "Bankruptcy Prediction using Neural Networks." Decision Support Systems: 545-557.

Xu, L. and Y. M. Cheung (1997). "Adaptive supervised learning decision networks for traders and portfolios." Proceedings of the Computational Intelligence for Financial Engineering: 206-212.

Zacharakis, A. L. and G. D. Meyer (2000). "The potential of actuarial decision models: can they improve the venture capital investment decision?" Journal of Business Venturing 15: 323-346.

Zacharakis, A. L. and D. A. Shepherd (2005). "A non-additive decision-aid for venture capitalists' investment decisions." European Journal of Operational Research 162: 673-689.

Zacks, R. (2000). "The TR university research scorecard." Technology Review 103(4): 88-90.

Zhang, D. and L. Zhou (2004). "Discovering Golden Nuggets: Data Mining in Financial Application." IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews 34(4): 513-522.

Zucker, L. G., M. R. Darby, et al. (1997). "Intellectual human capital and the birth of US biotechnology enterprises." American Economic Review 88: 290-306.

APPENDIX

A1.1: A Sample Invention Evaluation Form Developed by AURIL

Note: This form is extracted from AURIL (2003a).

DISCLOSURE EVALUATION FORM

Project Name ....

	SCORE	WEIGHTING	TOTAL
IP PROTECTION IPR, Novelty, Inventive Step, Claims	....	5	....
MARKET SIZE Customers with money	....	5	....
INVENTOR SUPPORT Experience, contribution, Enthusiasm	....	4	....
MARKET IMPACT	....	3	....
URGENCY Imminent disclosure (yes = 1; no = 10)	....	1	....
TIMESCALE	....	1	....

TLO EXPERIENCE IN FIELD Agency support, deputise inventor, precedents	....	1	....
TOTAL	(70)		(200) High = Good

Additional Information;

....
....

Patent Search;

....
....

Market Research;

....
....

A4.1: The 127 Attributes Located from 16 Invention Disclosure Forms

Due diligence

1. Working title of the invention
2. Inventor's name
3. Whether the invention is related to the university's work
4. The period of research relevant to the creation of the invention
5. During the period, whether the inventor was a salaried employee of the university (if so, which department), was employed by others (if so, what is the employer's name and contact), or was a student (if so, which university and what is the supervisor's name)
6. Source of funding lead to the invention: source, period, reference number (such as grant number), name of the principal investigator
7. If there was no contract or grant, was there a significant use of university funds or facilities
8. Whether any background (pre-existing) information was used as a basis for the invention, if yes, who owns and who funded the research led to the background information
9. Relevant documents regarding the background information
10. Whether the details of such research had been publicised, in oral or written forms (if so, what are the details)
11. Source of in-kind support led to the invention during the period
12. Whether the invention was linked with any contracts or agreement not cited above, including past, present or agreements undergoing negotiation
13. Whether there any material used is subject to MTA (material transfer agreements), if so, attach the MTA
14. Whether any novel materials with other organisations have been shared
15. Whether any new material has been made
16. Whether the Cre-lox technology was used
17. Whether any bio material from human has been used (if so whether

ethical consents have been obtained, and must provide document proof)

18. Individuals who made identifiable active contribution: name and contact
19. Signature and date

Income distribution

20. Patent application number, patent title (if known)
21. Each contributor's name, share percentage, as an inventor or contributor (contributors refer to those who are not, under the patent law, formally recognised as inventors but had significant and identifiable contribution), signature, date
22. Details of each contributors: name, title, national insurance number, email, address, nationality, whether the person is a university employee, date of birth, signature, date

Invention record

23. Any relevant documents for the invention
24. Contact person
25. Research and development staff
26. Date / date received by the technology transfer office
27. Descriptive title
28. Forms of intellectual property (such as patents, trade marks, and so on)
29. Name, address/email of contributors
30. Name, address/email of inventors
31. Name of non-university staff inventors
32. The first recorded date of invention
33. Date of the employment commencement with the university
34. Types of invention: a process, a composite of matter, a device
35. Whether the invention is a new use or an improvement to an existing product/process
36. Type of research: basic, applied, product development
37. Stage of research: idea, proof of concept, prototype

38. Whether there is a prototype, or whether the invention is available for demonstration
39. Whether further development is needed for the invention (if so, what is the funding requirement and what is the funding secured so far)
40. Field of invention (i.e. area/field in which the invention has application)
41. Reason and objective for undertaking the research
42. Invention description
43. Non-confidential version of invention description
44. Relevant documents for detailed invention description (such as drawing, papers)
45. Whether the inventor/contact person had been in touch with a member of the technology transfer office (if so, who and when)
46. Description of the disclosure information
47. Who, and how many people will disclose the information
48. How and why the invention works
49. What is new about it
50. What is the publicly known current solution to the problem
51. Disadvantage of current solution
52. How does it improve the present situation/ advantage of the invention over the existing solution to the problem
53. What is the added value of the invention
54. What is the disadvantage of the invention, and can they be and how to overcome them
55. Whether the invention is compatible with existing products (such as Window), if yes state how and which existing product
56. Whether there exists any standardisation for this type of product, if yes how this invention meets the standards
57. Uses/application of the invention
58. Other applications outside the inventor/contact person's own special interests
59. Whether the inventor/contact person had read the instruction about

prior art

60. Whether a prior art search had been done
61. When was the prior art search done
62. Details of the prior arts
63. Whether the invention has been tested in laboratory or has been used (if so, give details)
64. When and where the invention was first conceived
65. Sites where research was conducted led to the invention
66. Any signed and dated laboratory records about the date and origin of the invention
67. Who the inventor/contact person first told about the invention
68. Details of different types of disclosure (such as journal, thesis, internet): date, planned date, any written disclosure
69. Planned date of future disclosure
70. The first disclosure date
71. When did you first describe the invention in writing
72. Has anything relevant to the invention been published, verbally or in writing (if so when and what)
73. Has anything been disclosed, fully/partially, verbally/in writing/by demo, to any other parties
74. The employer of each inventor
75. Was any part of the invention made under a contract, or under support by third parties - if yes, attach a copy of the agreement
76. Head of department comments – signed and dated
77. Form is signed and dated by every inventor

Spinout

78. Company / project name
79. Names of the researchers and department involved
80. Document date
81. Target investment date
82. Product or service type
83. Product / service: short, medium or long term

84. Location of product development / manufacture
85. Market
86. Any current commercial interest?
87. List of potentially interested firms: company name. contact. address and telephone number
88. Any potential commercial opportunities
89. List the potentially interested industries
90. If have already contacted them, provide contact information
91. Target market size
92. Is the market growing, static or declining?
93. How commercially valuable do you rate your own invention
94. Existing commercial applications (i.e. uses/application, manufacturers, license agreements)
95. Will the invention replace current marketed product or be entirely new
96. Could existing technology be improved to compete with the invention
97. How does the market cope without your invention - what technology is used instead?
98. List the potential competitors' products or inventions
99. Is statutory approval required before product sale?
100. Route to market (i.e. how will it be sold)
101. Which countries is the product likely to be used in?
102. Any current activity regarding this invention? (R&D activity / exploitation activity)
103. Is the invention relevant to the market place
104. Business projection
105. Expected sales, margin, overheads, cash requirement for first few years
106. Employees required? (give details)
107. Name of expected employees. full time or part time
108. Who is performing the R & D
109. Sales and marketing plans

110. Finance plan
111. Management plan
112. Names of the directors and company secretary
113. Names of lawyers, accountants and bankers

Intellectual Property

114. What existing university intellectual property will be needed
115. What is the cash evaluation of the intellectual property
116. Total cash investment by third parties
117. What equity is offered to the university and the investors

Risk

118. What are the major sources of risk and uncertainty? (e.g. technical risk, academic competition, commercial competition, management risk)

Expertise

119. Describe the basis of the science
120. Describe the experience of the scientists involved
121. Describe the likely developments in the field
122. Describe the intellectual property and how it will be used by the spinout company

Contract with University

123. Will the spinout replace the contract with the university, if yes give details
124. Is this agreed
125. Will royalty payments be involved
126. Name, date of the author
127. Any other information that is relevant

A4.2: A Copy of the Questionnaire Used in the Questionnaire Survey

Survey questionnaire regarding the Invention Disclosure forms

1. Does your organisation have an Invention Disclosure form or a similar form for inventors to disclose the details of their inventions?

   ☐ Yes ☐ No

2. In which year was the Invention Disclosure form introduced in your organisation? _____

3. Apart from the use of the Invention Disclosure form, are any of the following used as inputs for initial invention assessment?

   ☐ Face-to-face meetings ☐ Presentations ☐ Other, please specify: _____

4. Is it a strict requirement to have the Invention Disclosure form completed before a patent application can be filed by an employee of your organisation?

   ☐ Yes ☐ No

5. Is it a strict requirement to have the ownership of the invention assigned to the university before a patent application can be filed?

   ☐ Yes ☐ No

6. Is the Invention Disclosure form available for download from the university website?

   ☐ Yes ☐ No

7. Does your organisation have an Online Invention Disclosure system that allows inventors to submit the Invention Disclosure form online?
☐ Yes ☐ No

8. Is the following information required in the Invention Disclosure form?
☐ Inventor's publication record ☐ Inventor's agreed share of income

9. Please tick if the following issues are addressed in the Invention Disclosure form.
☐ The sources of funding/in-kind support for the research that led to this invention
☐ Whether a patent search has been conducted
☐ Whether future funding is secured for the forthcoming patent expenses
☐ Invention description (Please tick if this includes ☐ function of the invention, ☐ stage of invention, ☐ involvement of any other parties during the invention development process)
☐ The list of any interested third parties that have contacted the inventors
☐ The list of the potentially interested third parties that the inventors intend to contact
☐ Plan for Licensing
☐ Plan for setting up a spinout company
☐ The estimated future cost (Please tick if this includes ☐ patent expenses, ☐ prototype development cost)
☐ Whether market research has been conducted
☐ The envisaged end products
☐ The list of potentially competing products currently in the market
☐ Non-confidential summary of the invention

10. Please tick if any of the following instructional documents are available to assist filling out the Invention Disclosure form.

☐ Overview of the University's Patent Protection and Licensing Processes

☐ Guidelines for information to be included in different sections of the Invention Disclosure form

☐ Determining inventorship

11. Please tick if any of the following is/are used to archive the Invention Disclosure forms.

☐ Database created using Microsoft Access

☐ Database created using Microsoft Excel

☐ Database created by other software, please name: _____

☐ Other information system, please name: _____

☐ Filing the hard copies of Invention Disclosure forms

12. How many employees are there in your technology transfer office? _____

13. Please tick if any of the following is/are applicable to your organisation.

☐ The university *always* pays for informal British patent applications from a central patent fund*.

☐ The university *always* pays for formal British patent applications from a central patent fund.

☐ The university assesses on a case-by-case basis to consider paying for any patent applications from a central patent fund

☐ The university does not pay for any patent applications from a central patent fund.

☐ The university pays for patent applications through channels other than a central patent fund, please specify ___________

☐ The university shares the patent costs with academic schools/departments on a ____/____ (e.g. 50/50) basis.

* The central patent fund refers to the budget allocated to patent expenses and is managed by a central body such as the technology transfer office.

A5.1: Innovation Evaluation Form

Disclaimer: It should be noted that this form does not represent an exhaustive list of information dimensions, but a suggested list based of the findings of the current research.

Innovation Evaluation Form

Today's date: ....

Short title of the invention: ....

Major area of discipline: ....

Name and contact details of the lead inventor: ....

Non-confidential summary of the invention: ....

Inventor information

Details of each inventor, in terms of the following dimensions:

- Name and contact details
- The level of contribution towards the invention
- Historical employment details, including period and employer
- Historical technical experience, including funding source, outcome, period, and reference
- Historical business experience, including funding source, outcome, period, and reference
- Historical achievements, including publications, awards, and grants

Details of each contributor in terms of the first two dimensions listed above.

Market information

- Contact details of potentially interested firms
- Details of current commercial activities, such as licence negotiation
- Target market size

- Whether the market is growing, static, or declining
- Details of any market research conducted
- The commercial value of the invention rated by external bodies
- The commercial value of the invention rated by each inventor
- Details of potential competitor's, including the products and their market shares
- The countries where the invention is likely to be used in
- Plans regarding the route to market
- Names, contacts, and employment status of any marketing personnel

Technical information

- Types of invention, such as: a process, a software, a device, etc
- Developmental stage: proven design made using manufacturing tooling, proven final design, optimisation by prototype variation, prototype, proof of concept, early stage
- The publicly known current solution to the problem
- Whether the invention is a new use or an improvement to an existing product/process
- Type of research leading to the invention: basic, applied, product development
- Details of current developmental activities
- Whether there exists any standardisation for this type of product, if yes how this invention meets the standards
- References of relevant technical publications
- Requirements of further development
- Details of prior art
- Details of prior art searches, including the dates and the search personnel
- Names of the academic departments involved
- Names, contacts, and employment status of any research personnel

Financial information

- Details of future funding plan

- Funding secured
- The level of expenses incurred so far
- Contact details of potential investors
- Names, contacts, and employment status of any financial personnel

Support information

- Source of funding leading to the invention: source, period, reference
- Names, contacts, and employment status of any management personnel
- Source of risk regarding the change of any personnel relevant to the invention