

MULTI-VIEW COLLABORATIVE REPRESENTATION CLASSIFICATION

YINGSHAN TAO¹, HAOLIANG YUAN¹, CHUN SING LAI^{1,2}, LOI LEI LAI¹

¹Department of Electrical Engineering, School of Automation, Guangdong University of Technology, Guangzhou 510006, China

²School of Civil Engineering, Faculty of Engineering, University of Leeds, Leeds LS2 9JT, UK
E-MAIL: 2111704005@mail2.gdut.edu.cn, hunteryuan@126.com, c.s.lai@leeds.ac.uk, l.l.lai@gdut.edu.cn

Abstract:

With the increase popularity of multi-view data, multi-view learning has attracted vital attentions in pattern recognition as well as machine learning. Most of existing methods apply in traditional single view learning. However, these methods neglect the complementary information among the views. The aim of multi-view is to discover complementary information and enhance the single view learning result. Multi-view is capable of capture incomplete and different types of information from multiple sources. However, multi-views may contain redundant information. Many multi-view methods assume that multi-views are generated from various view-specific generation matrices. This paper proposes the multi-view collaborative representation classification (MVCRC) algorithm which contains the information of different views and the connection of view-to-view. Experimental results conducted on five practical databases are used to confirm the effectiveness of the proposed approach.

Keywords:

Multi-view learning; Supervised classification; Collaborative representation

1. Introduction

In pattern recognition applications, the same object can be represented by different features, such as image classification, image retrieval, and scene recognition. This kind of data is named as multi-view data, which feature representation corresponds to a view [1]. Since multi-view learning discovers complementary information with different views [2], and enhances the classification performance or clustering performance, it has become a significant research area.

At present, multi-view learning has developed to different learning areas, like transfer learning [3], metric learning [4] as well as subspace learning [5]. According to the availability of label information, multi-view learning can be divided into

multi-view unsupervised learning, multi-view semi-supervised learning, and multi-view supervised learning [6]. In this paper, we discuss multi-view classification, which is included in the supervised learning method.

Currently, multi-view classification algorithms can be divided into about three groups. The first one is the classical support vector machine (SVM) classifier [7]. For instance, multiple kernel learning (MKL) belongs to SVM, and for the combination of multiple kernels, it optimizes to avoid selecting the specific kernel for SVM. MKL is naturally used in multi-view classification because the inputs of kernels come from different representations [8]. The second one is subspace learning. Two essential works are multi-view Fisher discriminative analysis (MvFDA) [9] and the SVM-2K [10]. The third one is the least square regression (LSR) for multi-view classification. For instance, Zheng et al. [11] proposed a multi-view low rank regression (MLRR) model to dig out the low-rank structure of data as well as gave a closed-form solution to the problem.

In multi-view data, every view contains the identical part of information, at the same time each view also contains the distinct information [12]. Strictly speaking, since the characteristic of multi-view data, the existing algorithms mainly discover either the consensus principle, or the complementarity principle [13]. The consensus principle mainly aims to maximize the agreement of all views. The complementarity principle [14] means that the complementary information from different views in multi-view data can help to improve the classification performance [6].

Although the existing algorithms have a good performance on multi-view data, they are just only considered either the consensus principle or the complementarity principle [15]. Therefore, in order to achieve better performance, we propose the multi-view collaborative representation classification (MVCRC) method which not only can be used in multi-view data but also utilize the properties of cohesion and diversity on

these view-specific generation matrices. In addition, MVCRC theoretically can further guarantee the correct rate of multi-view classification.

The rest of the paper is organized as follows. In Section 2, we briefly introduce the related works about collaborative representation classification. Section 3 describes the formulation of our proposed method, and the details of the optimization of MVCRC are presented. Section 4 presents the experiments results. Finally, the conclusion of this paper is shown in Section 5.

2. Related works

In this section, we briefly describe the related works about Collaborative representation based classifier (CRC) [16]. CRC model is one of the common classification methods which has been widely applied in pattern recognition and machine learning [17]. The mechanism of CRC model mainly includes two steps: first, using the training samples across all classes to collaboratively represent the test sample; second, assigning the test sample to the class minimizing the residual [18]. In essence, the first step exploits the collaborative ability of all training sample, and the second step utilizes the competitive ability of the training samples in each class [19].

2.1. CRC

The training set of the k -th class is $\mathbf{X}_k = [\mathbf{x}_1^k, \dots, \mathbf{x}_{N_k}^k]$ and the number of the classes is C . Thus, all training samples are consisted of $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_C] = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$, where d denotes the number of the spectral bands and $N = \sum_{k=1}^C N_k$. For CRC, they all use the training samples across all classes to collaboratively reconstruct the test sample. Concretely, given a test sample $\mathbf{y} \in \mathbb{R}^{d \times 1}$, the general model of CRC can be formulated as:

$$\min_{\mathbf{a}} \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_2^2, \quad (1)$$

where λ is the regularization parameter and $\mathbf{a} = [a_1, \dots, a_N]^T \in \mathbb{R}^{N \times 1}$ is the representation of \mathbf{y} over \mathbf{X} . For CRC, it can yield a closed-form solution, which is formulated as:

$$\mathbf{a} = \mathbf{H}_1 \mathbf{y}, \quad (2)$$

where $\mathbf{H}_1 = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T$ and \mathbf{I} is the identity matrix. After obtaining the representation \mathbf{a} , the residual in k -th class

is computed as:

$$r_k = \|\mathbf{y} - \mathbf{X}_k \mathbf{a}_k\|_2, \quad (3)$$

where \mathbf{a}_k is the part of coefficients corresponding to the k -th class in \mathbf{a} . The class label of the test sample \mathbf{y} is decided as:

$$\text{label}(\mathbf{y}) = \min_k r_k. \quad (4)$$

3. Proposed method

In this section, we introduce the proposed method MVCRC. Although the collaborative representation exploits the collaborative ability of all training sample, and utilizes the competitive ability of the training samples in each class, it is just only applied for single view data. It may not have a very good effect on multi-view data. Therefore, we propose the multi-view collaborative representation classification (MVCRC) method which not only can be used in multi-view data but also utilize the properties of cohesion and diversity on these view-specific generation matrices.

3.1. Objective function

Suppose we are given data points from V views, $V = \sum_{v=1}^V v$. \mathbf{X} presents all training sample with all the views, \mathbf{X}^v presents the v -th view of \mathbf{X} . \mathbf{y} presents all test sample with all the views, \mathbf{y}^v presents the v -th view of \mathbf{y} . \mathbf{s} is the representation of \mathbf{y} over \mathbf{X} . \mathbf{s}^v presents the v -th view of \mathbf{s} . $\bar{\mathbf{s}}$ presents the average of \mathbf{s} .

The objective function is formulated as:

$$\begin{aligned} \min_{\mathbf{s}^v} \sum_v \|\mathbf{y}^v - \mathbf{X}^v \mathbf{s}^v\|_2^2 + \lambda_1 \|\mathbf{s}^v\|_2^2 + \lambda_2 \|\mathbf{s}^v - \bar{\mathbf{s}}\|_2^2 \\ \text{s.t. } \bar{\mathbf{s}} = \frac{1}{V} \sum_v \mathbf{s}^v \end{aligned} \quad (5)$$

According to Equation (5), the first term of the objective function shows the diversity of the data of each view which is used by minimizing the sum of residuals. In this part, i means the different view of data. The complementarity principle is used in this term. And the third term of the objective function reflects the cohesion when each representation tries to approach its center.

3.2. Computing \mathbf{s}

We first discuss how to compute \mathbf{s} while fixing \mathbf{X}, \mathbf{y} . To effectively solve \mathbf{s} , we need to define

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} \mathbf{y}^1 \\ \mathbf{y}^2 \\ \vdots \\ \mathbf{y}^V \end{bmatrix} \in \mathbb{R}^{D \times 1}, \\ \mathbf{s} &= \begin{bmatrix} \mathbf{s}^1 \\ \mathbf{s}^2 \\ \vdots \\ \mathbf{s}^V \end{bmatrix} \in \mathbb{R}^{NV \times 1}, \\ \mathbf{X} &= \begin{bmatrix} \mathbf{X}^1 & & \\ & \cdots & \\ & & \mathbf{X}^V \end{bmatrix} \in \mathbb{R}^{D \times NV} \end{aligned} \quad (6)$$

N means the number of all samples, $NV = N \times V$, D means the sum of dimension of all sample. Then, the loss function in (5) is equivalent to

$$\|\mathbf{y} - \mathbf{X}\mathbf{s}\|_2^2 \quad (7)$$

Then, we denote

$$\begin{aligned} \mathbf{s}^v &= \mathbf{A}^v \mathbf{s} \\ \mathbf{A}^v &= \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{I} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}^T \in \mathbb{R}^{N \times NV} \\ \mathbf{A} &= \begin{bmatrix} \mathbf{I}^1 \\ \mathbf{I}^2 \\ \vdots \\ \mathbf{I}^v \\ \vdots \\ \mathbf{I}^V \end{bmatrix}^T \in \mathbb{R}^{N \times NV} \end{aligned} \quad (8) \quad (9) \quad (10)$$

Among them, $\mathbf{I} \in \mathbb{R}^{N \times N}$. We also denote that

$$\bar{\mathbf{s}} = \frac{1}{V} \sum_v \mathbf{s}^v = \frac{1}{V} \sum_v \mathbf{A}^v \mathbf{s} = \frac{1}{V} \mathbf{A} \mathbf{s} \quad (11)$$

So

$$\begin{aligned} \sum_v \|\mathbf{s}^v - \bar{\mathbf{s}}\|_2^2 &= \sum_v \|\mathbf{A}^v \mathbf{s} - \frac{1}{V} \mathbf{A} \mathbf{s}\|_2^2 \\ &= \sum_v \|(\mathbf{A}^v - \frac{1}{V} \mathbf{A}) \mathbf{s}\|_2^2 \\ &= \sum_v \mathbf{s}^T (\mathbf{A}^v - \frac{1}{V} \mathbf{A})^T (\mathbf{A}^v - \frac{1}{V} \mathbf{A}) \mathbf{s} \\ &= \mathbf{s}^T \mathbf{T} \mathbf{s} \end{aligned} \quad (12)$$

Therefore (5) equals to

$$\|\mathbf{y} - \mathbf{X}\mathbf{s}\|_2^2 + \lambda_1 \|\mathbf{s}\|_2^2 + \lambda_2 \mathbf{s}^T \mathbf{T} \mathbf{s} \quad (13)$$

Taking the derivative with respect to \mathbf{s} to zero, we have

$$\begin{aligned} 2(-\mathbf{X}^T)(\mathbf{y} - \mathbf{X}\mathbf{s}) + 2\lambda_1 \mathbf{s} + 2\lambda_2 \mathbf{T} \mathbf{s} &= 0 \\ \Rightarrow -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \mathbf{s} + 2\lambda_1 \mathbf{s} + 2\lambda_2 \mathbf{T} \mathbf{s} &= 0 \\ \Rightarrow (\mathbf{X}^T \mathbf{X} + \lambda_1 \mathbf{A} + \lambda_2 \mathbf{T}) \mathbf{s} &= \mathbf{X}^T \mathbf{y} \\ \Rightarrow \mathbf{s} &= (\mathbf{X}^T \mathbf{X} + \lambda_1 \mathbf{A} + \lambda_2 \mathbf{T})^{-1} (\mathbf{X}^T \mathbf{y}) \end{aligned} \quad (14)$$

4. Experiment

In this section, we evaluate the performance of our proposed method and compare it with two other methods in terms of overall classification accuracy. In addition, the best classification results will be highlighted in bold. Then, experimental results about parameter sensitivity, convergence, and time complexity and advantages of the proposed method are presented.

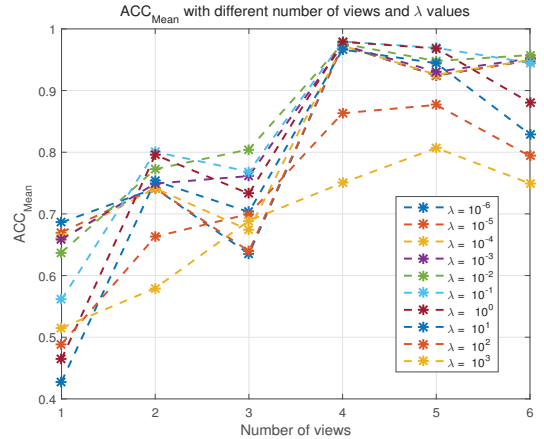


FIGURE 1. λ selecting from Caltech1 data set in CRC

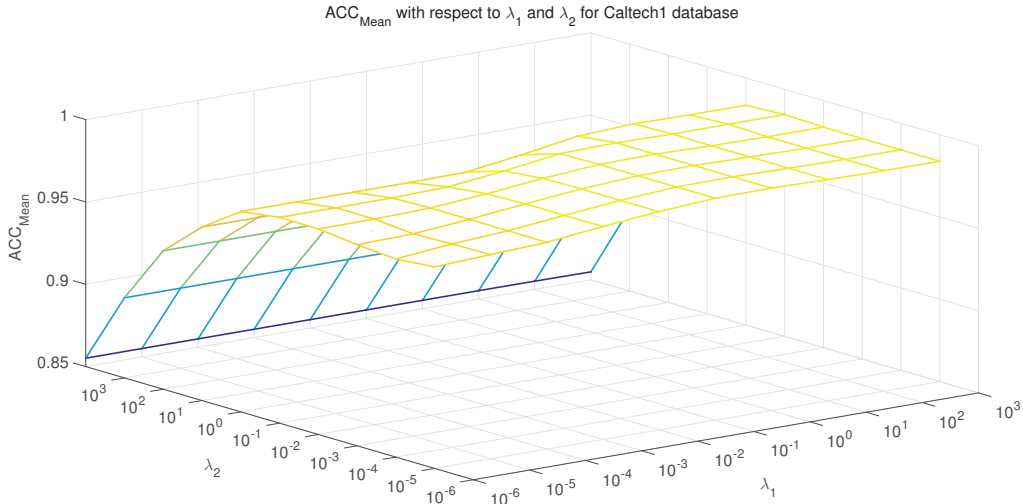


FIGURE 2. λ_1 and λ_2 selecting from Caltech1 data set in MVCRC

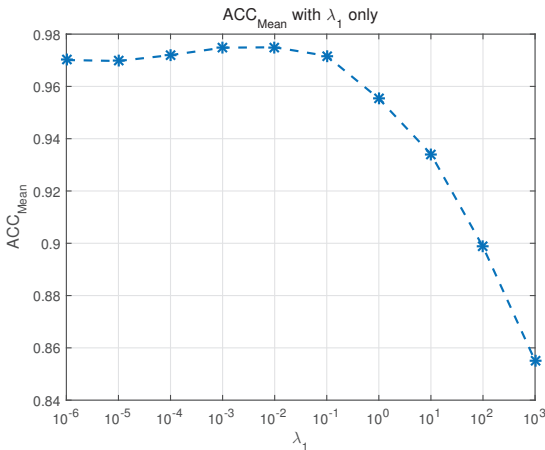


FIGURE 3. λ_1 selecting from Caltech1 data set in MVCRC with $\lambda_2=0$

There are five data sets used in our experiments. The statistics of five data sets are summarized in Table I.

- Caltech1¹ is an object recognition data set containing 8677 images, belonging to 101 categories. We chose the widely used 7 classes, i.e., Faces, Motorbikes, Dolla-Bill, Garfield, Snoopy, Stop-Sign and Windsor-Chair. We extract the same visual features: LBP, HOG, GIST, CMT, CENTRIST and SIFT.

¹<https://www.dropbox.com/s/ulvato08gepcdfk/Caltech101-7.mat>

- Handwritten numerals (HW)² data set [20] is comprised of 2,000 digital images, 200 images for each class from 0 to 9. There are six public features are available: 76 Fourier coefficients of the character shapes (FOU), 216 profile correlations (FAC), 64 Karhunen-love coefficients (KAR), 240 pixel averages in 2 3 windows (PIX), 47 Zernike moment (ZER) and 6 morphological (MOR) features
- MSRA³ is commonly used for object recognition. It contains 30 classes, and each class has 30 images. The class features are tree, building, airplane, cow, face, car, bicycle, etc.
- BBCSport⁴ consists of news article data. We selected 544 documents from the BBC Sport website corresponding to sports news articles in five topical areas from 2004-2005. It contains 5 class labels, such as athletics, cricket, football, rugby, and tennis.
- ORL⁵ face dataset consists of 40 distinct subjects with 10 different images for each. The images are discovered by changing lighting conditions, facial expressions, and facial details for some subjects. There are three types of features- LBP features, and Gabor features, are extracted and utilized for testing.

²<http://archive.ics.uci.edu/ml/datasets/Multiple+Features>

³<http://research.microsoft.com/en-us/projects/objectclassrecognition/>

⁴<http://mlg.ucd.ie/datasets>

⁵<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>

TABLE 1. STATISTICS OF FIVE MULTI-VIEW DATA SETS

Name Data sets	Information			
	Data points	Class	View	Dimensionality of each view
Caltech1	1474	7	6	435/798/52/34/35/64/56
HW	2000	10	6	200 each
MSRA	210	7	6	30 each
BBCSport	544	5	2	62/104/193/124/61
ORL	400	40	3	10 each

TABLE 2. THE CLASSIFICATION ACCURACY

Name Data sets	Methods		
	CRC	Proposed method	(Proposed method) $\lambda_2=0$
Caltech1	0.9797	0.9852	0.9750
HW	0.9192	0.9358	0.9352
MSRA	0.9114	0.9638	0.9581
BBCSport	0.8727	0.9237	0.9186
ORL	0.9554	0.9679	0.9507

4.1. Classification results

Figures 1 to 3 respectively represent different values of accuracy from Caltech1 data set as example when choose different values of lambda. As control complexity. Figure 1 shows λ selecting in CRC, Figure 2 shows λ_1 and λ_2 selecting from Caltech1 data set in the proposed method(MVCRC), and Figure 3 shows the MVCRC with λ_2 equalling to zero. The λ (MVCRC with λ_1 and λ_2) values are selected when the classification accuracy reaches the best, and the values of λ from figures 1 to 3 are $\lambda = 10^{-1}$, $\lambda_1 = 10^{-2}$ and $\lambda_2 = 10^0$, $\lambda = 10^{-2}$, respectively. It also shows from figures 1 to 3 that as the values of λ increase, the classification accuracy increases too. When the values of λ unceasingly increase, the accuracy rather decreases, and this indicates that trapped in under fitting problem. Figures 1 to 3 indicate that the best accuracy situate in non extreme lambda value.

Figure 4 and Table 2 show the accuracy with 3 different methods. Figure 4 presents the classification results for the five data sets and Table 2 highlights the best results in bold. The yellow columns represents the MVCRC, which are taller than the other blue and green columns. It indicates that the proposed method always has the best performance compared to the other 2 methods. The result proves that λ_2 has an important role in multi-view classification.

5. Conclusions

In this paper, we propose a multi-view collaborative representation classification (MVCRC) method which not only can

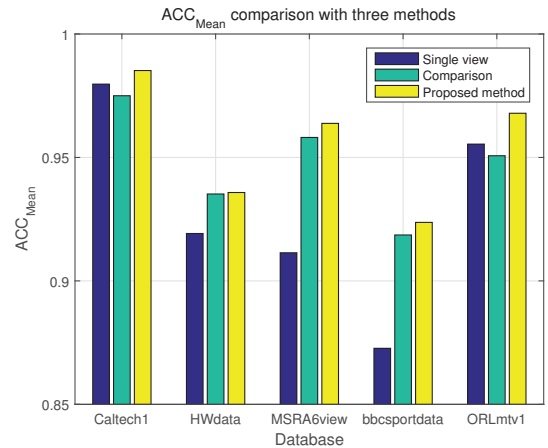


FIGURE 4. Accuracy comparison with 3 methods

be used in multi-view data but also utilize the properties of cohesion and diversity on these view-specific generation matrices. Our aim is to improve the accuracy of multi-view classification. Experiments on diverse real-world data sets verify the advantages of our proposed method when comparing with other two methods. For future research work, there will be additional experiments and comparison with state of art methods as well as more different data sets to further verify the validity and promotion of this algorithm.

Acknowledgements

This work is sponsored by the Department of Finance and Education of Guangdong Province 2016 [202]: Key Discipline Construction Program, China; and the Education Department of Guangdong Province: New and Integrated Energy System Theory and Technology Research Group [Project Number 2016KCXTD022].

References

- [1] C. Xu, D. Tao, C. Xu, A survey on multi-view learning, arXiv preprint arXiv:1304.5634.
- [2] X. Cao, C. Zhang, H. Fu, S. Liu, H. Zhang, Diversity-induced multi-view subspace clustering, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 586–594.
- [3] G. Cao, A. Iosifidis, K. Chen, M. Gabbouj, Generalized multi-view embedding for visual recognition and cross-modal retrieval, *IEEE Transactions on Cybernetics* 48 (9) (2018) 2542–2555. doi:10.1109/TCYB.2017.2742705.
- [4] C. Xu, D. Tao, C. Xu, Large-margin multi-view information bottleneck, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36 (8) (2014) 1559–1572. doi:10.1109/TPAMI.2013.2296528.
- [5] H. Tao, C. Hou, F. Nie, J. Zhu, D. Yi, Scalable multi-view semi-supervised classification via adaptive regression, *IEEE Transactions on Image Processing* 26 (9) (2017) 4283–4296. doi:10.1109/TIP.2017.2717191.
- [6] H. Tao, C. Hou, D. Yi, J. Zhu, Multiview classification with cohesion and diversity, *IEEE Transactions on Cybernetics* (2018) 1–14doi:10.1109/TCYB.2018.2881474.
- [7] M. Gönen, E. Alpaydın, Multiple kernel learning algorithms, *Journal of machine learning research* 12 (Jul) (2011) 2211–2268.
- [8] F. R. Bach, R. Thibaux, M. I. Jordan, Computing regularization paths for learning multiple kernels, in: *Advances in neural information processing systems*, 2005, pp. 73–80.
- [9] T. Diethe, D. R. Hardoon, J. Shawe-Taylor, Multiview fisher discriminant analysis, in: *NIPS workshop on learning from multiple sources*, 2008.
- [10] J. Farquhar, D. Hardoon, H. Meng, J. S. Shawe-taylor, S. Szedmak, Two view learning: Svm-2k, theory and practice, in: *Advances in neural information processing systems*, 2006, pp. 355–362.
- [11] S. Zheng, X. Cai, C. Ding, F. Nie, H. Huang, A closed form solution to multi-view low-rank regression, in: *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [12] J. Wang, F. Tian, H. Yu, C. H. Liu, K. Zhan, X. Wang, Diverse non-negative matrix factorization for multiview data representation, *IEEE transactions on cybernetics* 48 (9) (2017) 2620–2632.
- [13] C. Xu, D. Tao, C. Xu, A survey on multi-view learning, arXiv preprint arXiv:1304.5634.
- [14] W. Wang, Z.-H. Zhou, Theoretical foundation of co-training and disagreement-based algorithms, *CoRR abs/1708.04403*.
- [15] S. Sun, X. Xie, M. Yang, Multiview uncorrelated discriminant analysis, *IEEE transactions on cybernetics* 46 (12) (2015) 3272–3284.
- [16] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: Which helps face recognition?, in: *2011 International conference on computer vision*, IEEE, 2011, pp. 471–478.
- [17] L. Zhang, M. Yang, X. Feng, Y. Ma, D. Zhang, Collaborative representation based classification for face recognition, arXiv preprint arXiv:1204.2358.
- [18] S. Cai, L. Zhang, W. Zuo, X. Feng, A probabilistic collaborative representation based approach for pattern classification, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2950–2959.
- [19] H. Yuan, X. Li, F. Xu, Y. Wang, L. L. Lai, Y. Y. Tang, A collaborative-competitive representation based classifier model, *Neurocomputing* 275 (2018) 627 – 635, <http://www.sciencedirect.com/science/article/pii/S0925231217315278>. doi:<https://doi.org/10.1016/j.neucom.2017.09.022>.
- [20] D. Dua, C. Graff, UCI machine learning repository, <http://archive.ics.uci.edu/ml> (2017).