

# EXPLORATION OF DISTANCE METRICS IN CONSENSUS CLUSTERING ANALYSIS OF FMRI DATA

Chao Liu<sup>1</sup>, Elvira Brattico<sup>2,3</sup>, Basel Abu-jamous<sup>1</sup>, Asoke Nandi<sup>1,4,\*</sup>

1. Department of Electronic and Computer Engineering, Brunel University London, UK

2. Center for Music in the Brain (MIB), Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

3. The Royal Academy of Music Aarhus/Aalborg, Aarhus, Denmark

4. The Key Laboratory of Embedded Systems and Service Computing, College of Electronic and Information Engineering Tongji University, Shanghai, P. R. China

## ABSTRACT

Clustering techniques have gained great popularity in neuroscience data analysis especially in analysing data from complex experiment paradigm where it is hard to apply traditional model-based method. However, when employing clustering analysis, many clustering algorithms are available nowadays and even with an individual clustering algorithm, choices like parameter settings and distance metrics are very likely to have impacts on the final clustering results. In our previous work, we have demonstrated the benefits of integrating clustering results from multiple clustering algorithms, which provides more stable, reproducible, and complete clustering solutions. In this paper, we aim to further inspect the possible influences from the choices of distance metrics in clustering analysis.

*Index Terms* — fMRI data, consensus clustering, Bi-CoPaM, distance metric

## 1. INTRODUCTION

Clustering has gained great popularity in exploring and identifying the natural distribution of huge amount of data in various research fields such as neuroscience [1–3], bioinformatics [4–7] and image processing [8, 9]. Clustering techniques are designed to partition the input dataset into several categories without explicitly training the classifier using labeled training data. This nature of clustering technique makes it a powerful tool when a new dataset without many ground truths is to be analysed, e.g. in biological or neuroimaging data analysis. However, clustering technique is described as an ill-posed problem whose solution violates at least one of the common assumption regarding cluster consistency, richness, and scale invariance [10]. Due to the fact that no a priori knowledge is used, the clustering algorithms could yield results based on its underlying assumption of the data structures that might be different from the structure of the input data. Thus, it is important to recognise that clustering results should be interpreted with great caution.

To address this, we have demonstrated the benefits of using consensus strategy in clustering analysis of fMRI dataset in our previous work [11]. We found individual clustering algorithm is exposed to the risk of failing to detect one or more clusters covering important brain regions while the scenario that clustering results from multiple algorithms are combined in a tunable manner is able to compensate the bias and incompleteness of individual clustering result, yielding more stable, reproducible, and complete consensus clustering solutions. Three clusters covering visual cortex, reward system, and auditory area were found to consistently show synchronised fMRI BOLD response patterns during affective processing of emotions conveyed in music listening [10, 11]. It is one of the few successful studies that have obtained such findings in the context of cognitive neuroscience of music.

In addition to the choice of various clustering algorithms, another factor that is likely to influence the clustering result is the distance metric or similarity metric, as it defines the space where objects are compared and assigned to corresponding clusters based on the chosen algorithms. In our previous work, we used Euclidean distance by default for all the clustering algorithm, i.e., K-means, hierarchical clustering and self organising map (SOM), aiming to mimic the real situation where many clustering algorithms have been used as well as the commonly used Euclidean distance metric.

In this paper, we set the scope to explore further the possible effect of the choice of distance/similarity metrics in consensus clustering analysis of fMRI data, thus enhancing the understanding of the performance of consensus clustering. We adopted the reverse engineering idea where we quantitatively compared the four distance/similarity measure within two clusters (Visual and Reward) reported in [11]. The distance/similarity measures we chose are L1 norm, L2 norm, Pearson correlation, and cross-correlation, all of which have been used in various neuroimaging studies [10–16]. We compared the distributions of these four distance/similarity values between objects in each cluster and the corresponding cluster centre. Results show that the clusters have very good quality and the intra cluster similarity conforms to the cluster selection criterion.

\* Professor Nandi is a Distinguished Visiting Professor at Tongji University, Shanghai, China. This work was partly supported by the National Science Foundation of China grant number 61520106006.

## 2. METHODS

### 2.1. fMRI experiment

The fMRI data comes from an experiment studying human brain responses to music with different preference and emotions, i.e. liked happy category (LH) [this represents the case of when a participant considers the music is “happy” and “likes” it], liked sad category (LS) [this represents the case of when a participant considers the music is “sad” and “likes” it], disliked happy category (DH) [this represents the case of when a participant considers the music is “happy” and “dislikes” it], and disliked sad category (DS) [this represents the case of when a participant considers the music is “sad” and “dislikes” it]. A total of 13 musicians and 16 non-musicians participated in the fMRI scanning sessions. The whole fMRI experiment for one participant has 450 scans ( $TR = 3s$ ) including 32 music categories with each one repeated twice and each scan contains 228,453 voxels after preprocessing. In total, there are 1856 fMRI data segments corresponding to 1856 ( $= 29 \times 64$ ) listening trials. The study was approved by the ethical committee of the Helsinki University Central Hospital and complied with the Helsinki Declaration.

### 2.2. Consensus partition matrix generation

There are in total 1856 data excerpts with each one representing an experimental condition. Each excerpts data (normalized to zero mean and unit variance) was clustered based on Euclidean distance ( $L_2$  norm) by K-means, Hierarchical and SOM with K equals to 10, 25, 50, and 100 respectively. The individual clustering generated 22272 partition matrices that will be fed into a consensus clustering paradigm called binarisation of consensus partition matrices (Bi-CoPaM) [18]. The following procedure was applied on these individual partition matrices to produce the fuzzy consensus partition matrix with different tightness (intra cluster similarity) level.

(i) Relabeling: because clustering is unsupervised, there are no labels for the clusters in the different partitions, i.e. the  $i$ -th cluster in one partition is not guaranteed to match the  $i$ -th cluster in another partition. Relabeling reorders the clusters in the partitions so that they become aligned. Min–min approach was used to perform relabeling. Note that this relabeling method works on the clustering results with the same number of clusters  $K$ . The relabeling does not yield a complete correspondence between clustering solutions, however, the following steps will attenuate the inconsistency and keep the most consistent cluster members across multiple clustering solutions.

(ii) Fuzzy consensus partition matrix (CoPaM) generation: the relabeled partitions are averaged to produce a fuzzy CoPaM in which each voxel has a fuzzy membership value in each of the clusters based on the number of individual partitions that assigned it to it.

(iii) Quenching/Binarization: the fuzzy CoPaM is binarised to produce the final binary partition. Different threshold binarization (DTB) technique assigns a voxel to a cluster if and only if its fuzzy membership value in that cluster is higher than its closest competing cluster fuzzy membership value by the value of the parameter  $\delta$ . The parameter  $\delta \in [0, 1]$  controls the tightness of the cluster where  $\delta = 0.0$  is the least tight (most sparse) and  $\delta = 1.0$  is the tightest, indicating the best cluster quality.

### 2.3. Extracting clusters with good quality

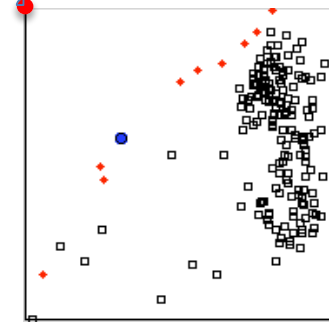


Figure 1. Example of M-N Scatter plot.

The above procedure yielded consensus clustering solutions. The M-N scatter plot technique [18, 19] is applied to optimize the final cluster results. All of the individual clusters appearing in the results are plotted on a 2-D plot (Figure 1) where the vertical axis (N) represents the logarithm of the number of voxels in the cluster and the horizontal axis (M) represents the average mean square error (MSE) values of the cluster over all of the dataset. The cluster closest to the top left corner of the plot (red dot) is selected as the best cluster (blue dot). This cluster consists of a large number of voxels (high vertical axis value), yet tight with high correlation (low horizontal axis value). Then the selected cluster and all of the other clusters that have overlaps with it, are removed from the plot. For the remaining clusters (dots), the closest remaining cluster to the top left corner of the plot is selected as the second best distinct cluster. The steps of selecting clusters and removing those with overlaps with the selected ones are repeated iteratively up to a preset maximum number of clusters or earlier when the scatter plots are empty.

### 2.4. Cluster topological filtering

The following filtering procedure was applied on the clusters generated by Bi-CoPaM and M-N scatter plot technique. Firstly, if a voxel’s variance is greater than half of the mean of the variance for all the voxels in a particular cluster, and if more than seventy percent of the subjects showed a strong response at this particular voxel, the voxel was labeled as having “strong response”. Secondly, the resulting clusters from the previous step were filtered by hypergeometric distribution test to exclude minor structures within each cluster. Voxels within a certain brain structure

covering large connected brain area would feature a very small p value (normally below 0.001 level) while those covering tiny isolated brain structures would result in a relatively high p value (normally above 0.1 level). We chose p less than 0.001 to distinguish the major brain structures from the minor ones within each cluster. Finally, we used the fMRIToolbox, developed at University of Jyväskylä [20] to remove the scattered small clusters.

## 2.5. Distance and similarity metric

Four distance and similarity metrics were used to inspect the cluster quality. Assume two BOLD time series vector  $\mathbf{u}$  and  $\mathbf{v}$ . The four metrics are defined as follows:

a) L1 norm

$$\|\mathbf{u} - \mathbf{v}\|^1 = \sum_{i=1}^N |\mathbf{u}_i - \mathbf{v}_i| \quad (1)$$

b) L2 norm

$$\|\mathbf{u} - \mathbf{v}\|^2 = \sqrt{\sum_{i=1}^N |\mathbf{u}_i - \mathbf{v}_i|^2} \quad (2)$$

c) Pearson correlation

$$\text{corr}(\mathbf{u}, \mathbf{v}) = \frac{\sum_{i=1}^N (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^N (u_i - \bar{u})^2} \sqrt{\sum_{i=1}^N (v_i - \bar{v})^2}} \quad (3)$$

$$\bar{u} = \frac{1}{N} \sum_{i=1}^N p_i, \bar{v} = \frac{1}{N} \sum_{i=1}^N q_i$$

d) Cross correlation (at delay  $d$ )

$$\text{xcorr}(\mathbf{u}, \mathbf{v})_d = \frac{\sum_{i=1}^N [(u_i - \bar{u})(v_{i-d} - \bar{v})]}{\sqrt{\sum_{i=1}^N (u_i - \bar{u})^2} \sqrt{\sum_{i=1}^N (v_{i-d} - \bar{v})^2}} \quad (4)$$

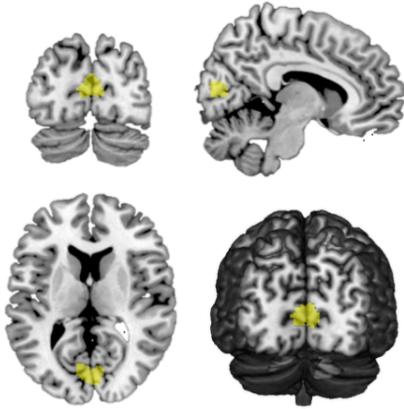


Figure 2. Topology of cluster Visual.

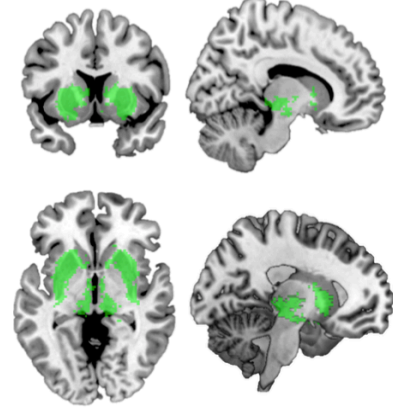


Figure 3. Topology of cluster

For each cluster (Visual and Reward) and each BOLD data excerpt, the above four metrics between every object in the cluster and the corresponding cluster centre (mean of all the objects) are computed. We then plotted the histogram of all the metric values with respect to each cluster and each metric to inspect the distribution of these distance and similarity values. Note that for the normalised cross correlation, the value is taken from the maximum value among all the delays, indicating the maximum correlation between two BOLD time series. The cross correlation metric is thus able to take the delay between two signals into account and find the best match at a certain delay.

## 3. RESULTS

### 3.1. Topology of clusters

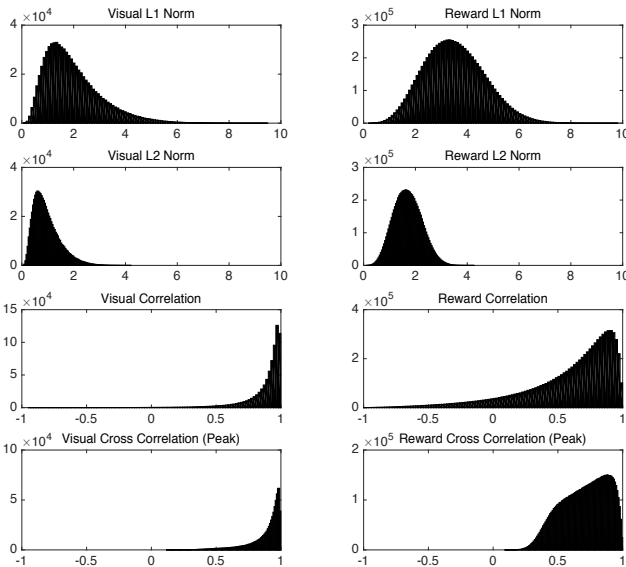
Figure 2. and Figure 3 are the topology of the two clusters we chosen to inspect. Cluster Visual is the second cluster selected by M-N scatter plot and cluster Reward is the ninth. These two clusters consistently show synchronised BOLD response patterns during affective processing. Visual area is believed to be responsible for imagination and reward system is capable of processing various emotions and preference people perceived from the music listening.

### 3.2. Distribution of distance/similarity within cluster

The histogram is plotted for each cluster and each distance/similarity measure as shown in Figure 4. The left column is the distribution of metric values in cluster Visual and the right column is the distribution of metric values in cluster Reward. From the second row, we can see the two clusters generated by Euclidean distance (L2 norm) has a very high tightness, where a majority of objects has a short distance to the cluster centre. Comparing cluster Visual with cluster Reward, Visual has an even tighter feature than Reward. This is a confirmation of the logic used in M-N scatter plot technique, which is to extract the clusters with a better quality (tight) before the rest of the clusters (less

tight). Thus this is a good validation of the obtained clusters with dense objects around the cluster centres.

L1 norm, compared to L2 norm, still reflect the general tightness of two clusters Visual and Reward. Visual is more tight in the sense that more objects are closer to the centre. However, in terms of the norm values along the horizontal range, the L1 norm gives a larger mean distance from objects to the centre. This may not influence the clustering results when L1 norm is used, as the distance between objects in different clusters are also larger accordingly.



**Figure 4. Histogram of distance and similarity values for cluster Visual and Reward.**

Similarly, the correlation and cross correlation metric in each cluster reflect the strong intra cluster similarity of the objects. Again, by comparing the two plots at the third row and fourth row respectively, the cluster Visual shows a tighter feature than cluster Reward, which conforms to the M-N plot strategy discussed before. An interesting point regarding the correlation and cross correlation metric (the last two plots on the second column) is there are some negative correlation values but no negative peak cross correlation values were obtained. This indicates a possible scenario that there is some phase lag between certain voxels' BOLD response patterns of a stimuli. Because the Pearson correlation measures the signal similarity at zero lag while cross correlation takes the signal lag into account and thus able to capture the most similar temporal features of two BOLD response patterns.

#### 4. DISCUSSIONS AND CONCLUSIONS

Clustering has gained great popularity in exploring and identifying the natural distribution in neuroimaging data. However, clustering itself is an ill-posed problem whose

solution violates at least one of the common assumption regarding cluster consistency, richness, and scale invariance. We have demonstrated the benefits of integrating clustering results from multiple algorithms instead of using only one clustering method. In this paper, we further explored another factor, i.e., similarity/distance metric, which is likely to introduce the differences between clustering results. Towards this aim, we adopted the reverse engineering idea where we quantitatively compared the four distance/similarity measure within two clusters (Visual and Reward) in our previous work [11]. That is to say, we did not carry out individual clustering experiments with respect to the similarity/distance metrics, rather we computed these four metric values for all the objects within each cluster already obtained by Euclidean distance (L2 norm). Here four similarity/distance metrics (L1 norm, L2 norm, Pearson correlation, and cross-correlation) were used to evaluate the intra cluster similarity of these consensus clustering results.

We found that these two clusters inspected (Visual and Reward) have good cluster quality in terms of the dense distribution of objects around the cluster centre. In addition, we further demonstrated the efficacy of cluster selection process by M-N scatter plot. The cluster Visual has a tighter cluster feature (higher intra cluster similarity) than the cluster Reward, which was selected after Visual, in all the four similarity/distance metrics. We spotted an interesting difference between correlation and cross-correlation metric in cluster Reward, where cross correlation can take the BOLD response pattern lag into account and being able to capture the most similar temporal features of two signals. This scenario is very possible as the coupling of brain regions are not always simultaneous.

Based on the results of the inspection, we can further confirm the quality of clusters obtained by consensus clustering paradigm Bi-CoPaM. By comparing the distributions of four metric values in each of the two clusters, we reckon that the similarity/distance metric selection will not significantly influence the consensus clustering results, despite the inevitable discordance due to the random behaviour of certain clustering algorithms, as they all reflect a similar cluster tightness feature. Hence the findings in this study enhance the understanding of consensus clustering analysis of fMRI data in terms of similarity/distance metric used in clustering.

#### 5. REFERENCE

- [1] C. Goutte, P. Toft, E. Rostrup, F. Nielsen, and L. K. Hansen, 'On clustering fMRI time series.', *Neuroimage*, vol. 9, no. 3, pp. 298–310, 1999.
- [2] M. van den Heuvel, R. Mandl, and H. Hulshoff Pol, 'Normalized cut group clustering of resting-state FMRI data.', *PLoS One*, vol. 3, no. 4, p. e2001, 2008.
- [3] Y. Golland, P. Golland, S. Bentin, and R. Malach, 'Data-driven clustering reveals a fundamental

- subdivision of the human cortex into two global systems’, *Neuropsychologia*, vol. 46, no. 2, pp. 540–553, 2008.
- [4] D. J. Higham, G. Kalna, and M. Kibble, ‘Spectral clustering and its use in bioinformatics’, *J. Comput. Appl. Math.*, vol. 204, no. 1, pp. 25–37, 2007.
- [5] D. M. Witten, ‘Classification and clustering of sequencing data using a poisson model’, *Ann. Appl. Stat.*, vol. 5, no. 4, pp. 2493–2518, 2011.
- [6] B. Abu-Jamous, C. Liu, D. J. Roberts, E. Brattico, and A. K. Nandi, ‘Data-driven analysis of collections of big datasets by the Bi-CoPaM method yields field-specific novel insights’, in *Frontiers in Electronic Technologies*, Springer Berlin Heidelberg, 2017.
- [7] B. Abu-Jamous, R. Fa, and A. K. Nandi, *Integrative cluster analysis in bioinformatics*. John Wiley & Sons, 2015.
- [8] J. Shi and J. Malik, ‘Normalized Cuts and Image Segmentation’, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.
- [9] N. Dhanachandra, K. Manglem, and Y. J. Chanu, ‘Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm’, *Procedia Comput. Sci.*, vol. 54, pp. 764–771, 2015.
- [10] J. Kleinberg, ‘An impossibility theorem for clustering’, *Adv. Neural Inf. Process. Syst.*, pp. 446–453, 2002.
- [11] C. Liu, B. Abu-Jamous, E. Brattico, and A. K. Nandi, ‘Towards Tunable Consensus Clustering for Studying Functional Brain Connectivity During Affective Processing’, *Int. J. Neural Syst.*, vol. 27, no. 2, p. 1650042, 2017.
- [12] E. Brattico, V. Alluri, B. Bogert, T. Jacobsen, N. Vartiainen, S. Nieminen, and M. Tervaniemi, ‘A functional MRI study of happy and sad emotions in music with and without lyrics’, *Front. Psychol.*, vol. 2, no. December, p. 308, 2011.
- [13] E. Brattico, B. Bogert, V. Alluri, M. Tervaniemi, and T. Eerola, ‘It ’ s Sad but I Like It: The Neural Dissociation Between Musical Emotions and Liking in Experts and Laypersons’, *Front. Hum. Neurosci.*, vol. 9, no. 676, 2016.
- [14] B. Yeo, F. Krienen, J. Sepulcre, M. Sabuncu, D. Lashkar, M. Hollinshead, J. Roffman, J. Smoller, L. Zollei, J. Polimeni, B. Fischl, H. Liu, and R. Buckner, ‘The organization of the human cerebral cortex estimated by intrinsic functional connectivity’, *J. Neurophysiol.*, vol. 106, no. 3, pp. 1125–1165, 2011.
- [15] H. Chen, H. Yuan, D. Yao, L. Chen, and W. Chen, ‘An integrated neighborhood correlation and hierarchical clustering approach of functional MRI’, *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 452–458, 2006.
- [16] X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, and P. Boesiger, ‘A new correlation-based fuzzy logic clustering algorithm for fMRI’, *Magn. Reson. Med.*, vol. 40, no. 2, pp. 249–260, 1998.
- [17] C. E. Davey, D. B. Grayden, G. F. Egan, and L. a. Johnston, ‘Filtering induces correlation in fMRI resting state data’, *Neuroimage*, vol. 64, no. 1, pp. 728–740, 2013.
- [18] B. Abu-Jamous, R. Fa, D. J. Roberts, and A. K. Nandi, ‘Paradigm of tunable clustering using Binarization of Consensus Partition Matrices (Bi-CoPaM) for gene discovery.’, *PLoS One*, vol. 8, no. 2, p. e56432, 2013.
- [19] B. Abu-Jamous, R. Fa, D. J. Roberts, and A. K. Nandi, ‘Yeast gene CMR1/YDL156W is consistently co-expressed with genes participating in DNA-metabolic processes in a variety of stringent clustering experiments.’, *J. R. Soc. Interface*, vol. 10, no. 81, p. 20120990, 2013.
- [20] I. Burunat, P. Toivainen, V. Alluri, B. Bogert, T. Ristaniemi, M. Sams, and E. Brattico, ‘The reliability of continuous brain responses during naturalistic listening to music’, *Neuroimage*, vol. 124, pp. 224–231, 2016.