

# CLUSTERING CONSISTENCY IN NEUROIMAGING DATA ANALYSIS

Chao Liu<sup>1</sup>, Basel Abu-Jamous<sup>1</sup>, Elvira Brattico<sup>2,3</sup>, Asoke Nandi<sup>1,4,5\*</sup>

1. Department of Electronic and Computer Engineering, Brunel University London, UK
2. Cognitive Brain Research Unit, Institute of Behavioural Science, University of Helsinki, Finland
3. Helsinki Collegium of Advanced Studies, University of Helsinki, Finland
4. Department of Mathematical Information Technology, University of Jyväskylä, Finland
5. Tongji University, Shanghai, China

**Abstract**—Clustering techniques have been applied to neuroscience data analysis for decades. New algorithms keep being developed and applied to address different problems. However, when it comes to the applications of clustering, it is often hard to select the appropriate algorithm and evaluate the quality of clustering results due to the unknown ground truth. It is also the case that conclusions might be biased based on only one specific algorithm because each algorithm has its own assumption of the structure of the data, which might not be the same as the real data. In this paper, we explore the benefits of integrating the clustering results from multiple clustering algorithms by a tunable consensus clustering strategy and demonstrate the importance and necessity of consistency in neuroimaging data analysis.

**Keywords:** *neuroimaging, clustering, consensus, fMRI*

## I. INTRODUCTION

The advances in data collection and storage have generated huge amount of data in research fields such as neuroscience. Clustering has gained great popularity in exploring and identifying the natural distribution of these datasets [1]–[5]. Being different from supervised classification, clustering techniques are designed to separate the input dataset into several categories without explicitly training the classifier using labeled training data. This nature of clustering makes it a powerful tool when researchers explore a new dataset. However, despite a lot of clustering algorithms having been developed [6], clustering itself is an ill-posed problem whose solution violates at least one of the common assumption regarding cluster consistency, richness, and scale invariance [7]. Almost every clustering algorithm explicitly or implicitly assumes a certain data model based on which the data is clustered [8]. Without a priori knowledge, each clustering algorithm could produce plausible results based on its underlying assumption of the data structures that might not be satisfied by the input data. To match the assumed data model and the true input data structure, it is crucial to obtain the correct a priori information to obtain good clustering results; however, such a priori information is hard to obtain even from experts. A recent main criticism relies on the wide variety of analysis strategies, combined with small sample sizes, used to investigate regional brain activity measured with functional

magnetic resonance imaging (fMRI) and leading to inconsistent findings [9].

To address the aforementioned problems, clustering ensemble techniques [8], [10] were proposed. An ensemble clustering technique uses various strategies to combine multiple clustering results from different algorithms to improve the overall quality of results such as stability, robustness, and consistency. Besides, an ensemble clustering technique will also benefit from the fact that it could reuse the existing results from individual algorithm without rerunning the experiment from the beginning. Even a total experiment reset is needed, the jobs can be spread to different computing facilities with each one handling one or several clustering experiments, since different algorithms do not need to communicate with each other during the single clustering experiment period. This feature would help significantly to reduce the computation time and requirement of computation power when analysing the neuroimaging data with increasing size and volume these days [11].

To demonstrate the necessity of employing an ensemble clustering technique in neuroimaging data analysis, we used a recently published tunable consensus clustering paradigm called Binarization of Consensus Partition Matrices (Bi-CoPaM) [2] to combine the clustering results on a real functional magnetic resonance imaging data by k-means, hierarchical, and self organizing map (SOM). The results show that there are obvious differences among the results by different clustering algorithms combination. while Bi-CoPaM can not only find the objects that consistently belong to same cluster across many clustering experiment but also is able to extract the clusters according to their tightness (quality) through tuning the consensus level. These selected clusters have very strong inner similarity as well as relatively large amount of members, making it more reliable to draw conclusions from ensembled clustering results.

The rest of the paper is organized as follows. In section II, we describe the fMRI data and briefly introduce the published methods used in this paper. The setup of the clustering experiment is described in section III. In section IV, the results of comparisons between different method combination are shown. We then had a discussion about the method and its benefit and draw conclusions in section V.

\* A. K. Nandi would like to thank TEKES for their award of the Finland Distinguished Professorship. Professor Nandi is a Distinguished Visiting Professor of Tongji University, Shanghai, China.

## II. METHOD

### A. fMRI data

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 ( $TR = 2s$ ) 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. The dataset is a subset of a larger data collection, parts of which have been published in [12], [13].

### B. Bi-CoPaM

The binarisation of consensus partition matrices (Bi-CoPaM) method is a tunable consensus clustering method that utilises various single clustering algorithms while considering various datasets to identify the subset of objects that are consistently correlated across many clustering experiments. Figure 1 demonstrates the procedure of Bi-CoPaM. For  $L$  datasets and  $C$  clustering algorithms, the typical Bi-CoPaM pipelines includes the following four steps:

1) Partition generation. This step generates the clustering results by single clustering algorithm. Note that this step can be skipped if one decides to reuse the existing results as discussed in the introduction part.

2) Relabeling. This step aligns the numbering of the clusters in each partition. After this step, the cluster numbers correspond across different partitions so that the same number represents the same cluster.

3) Fuzzy consensus partition matrix (CoPaM) generation. This step averages the relabeled partition matrices to produce the fuzzy consensus matrix where each object has a fuzzy membership value ranging from 0 to 1 with 0 representing absolutely no consistency and 1 representing absolute consistency.

4) Binarisation. This step binarise the fuzzy membership value to 0 and 1 in CoPaM in step 3) using Different threshold binarisation (DTB) technique by the thresholds  $\delta$  with  $\delta = 0$  yielding the least tight clusters and  $\delta = 1$  yielding the tightest clusters.

One important feature of Bi-CoPaM is that the results are tunable and clusters with different tightness can be obtained by tuning the parameter  $\delta$ . Users can generate consensus clusters based on their needs for the tightness ranging from high to low in terms of cluster inner similarity.

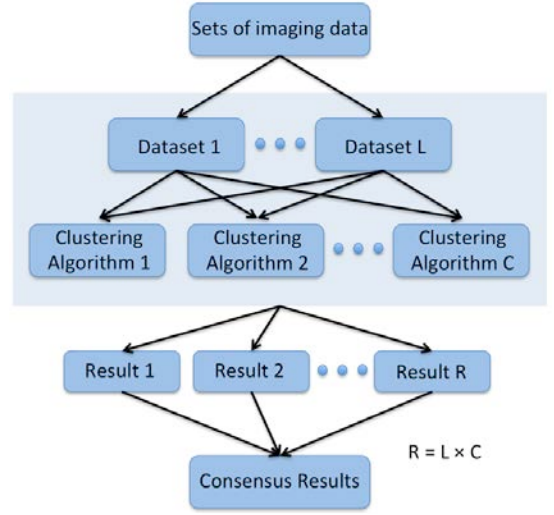


Figure 1. Flowchart of Bi-CoPaM

### C. Extracting Clusters with Good Quality

Bi-CoPaM combines clustering results from different clustering algorithms and different clustering parameters such as the number of cluster,  $k$ . The M-N scatter plot technique [14] is applied to optimize the final cluster results. Each time that Bi-CoPaM is applied with different  $k$  values, DTB binarisation is performed with a range of  $\delta$  values (e.g. from 0.0 to 1.0 with 0.1 steps). All of the individual clusters appearing in the results are plotted on a 2-D plot 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. Figure 2 is an example of the scatter plot.

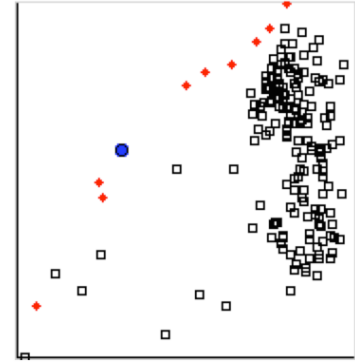


Figure 2. Example of M-N scatter plot

The cluster closest to the top left corner of the plot 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.

### III. CLUSTERING EXPERIMENT

The  $k$ -means, Hierarchical, and Self Organizing Map (SOM) were applied on each of 1856 excerpts data (normalized to 0 mean and unit variance) was clustered by with  $k$  equals to 10, 25, 50, and 100. These clustering results generated by different algorithms with four different cluster numbers were combined using the Bi-CoPaM paradigm for each possible combination of three methods and selected by the M-N scatter plot, yielding the consensus clustering results. In total, seven sets of final clustering results were obtained, i.e. clustering results from  $k$ -means, hierarchical, SOM,  $k$ -means and hierarchical,  $k$ -means and SOM, hierarchical and SOM, and all three methods. The brain regions within clusters were extracted out by using automated anatomical labeling (AAL) atlas [15] and these regions names were used to identify the neuroscience background of clusters.

Two types of filtering were used in order to make the results more reliable. The original clusters were first filtered by discarding those voxels with weak responses (voxels whose time series have a small variance), since the data used to be clustered were normalized and thus, to some extent, lost the signal magnitude information. In this analysis, the voxels whose variance corresponded to less than half of the mean of the variance for all the voxels from one subject were discarded. Note that the chosen threshold does not necessarily mean this is the optimal value. It only serves as removing the voxels with weak signal change. After repeating this process for all the subjects, we obtained 29 thresholded partitions. Then if more than seventy percent of the subjects showed a strong response at a certain voxel, this voxel was retained for the following analysis. After this step, the clusters only contained voxels having strong blood oxygen level dependent (BOLD) responses. Secondly, the resulting clusters from the previous step were filtered by using the hypergeometric distribution test which discarded isolated voxels, i.e., any voxel above the  $p$  value threshold of 0.001.

The M-N scatter plot was used to select the first 20 clusters for each CoPaM generated by corresponding combination of methods. For each combination of methods, we picked out the final clusters covering three important brain regions, namely visual area, reward system and auditory system. Then comparisons among different method combination were made regarding each of the three important clusters. Then in order to give a straightforward demonstration on the differences among different clustering combinations, we visualize some of the final cluster results in 3D space to see the brain areas that consistently have highly correlated BOLD activities.

### IV. RESULTS

Jaccard index was used to compare the similarity of two clusters. Table I shows the Jaccard index between different method combinations for cluster covering visual area. Table II shows the Jaccard index between different method combinations for cluster covering reward system and Table III shows the Jaccard index between different method combinations for cluster covering auditory system.

From the three Tables, it is easy to spot the differences between any combinations of two or three clustering algorithms on all three important clusters. Note that for those methods or method combinations having all zeros Jaccard index, it means this method or method combination was not able to detect the corresponding cluster. For example, in Table III, the method combination of  $k$ -means and hierarchical clustering (KM & HC) did not group the voxels within the auditory system into a cluster. So, all the members for clusters covering the visual areas end up producing a value of zero in the Jaccard index between KM & HC and any other method combinations.

TABLE I. JACCARD INDEX FOR CLUSTER COVERING VISUAL AREA.

	KM	HC	SOM	KM+HC	KM+SOM	HC+SOM	KM+HC+SOM
KM		0.48	0.70	0.70	0.78	0.73	0.40
HC	0.48		0.38	0.57	0.53	0.56	0.26
SOM	0.70	0.38		0.58	0.66	0.62	0.50
KM+HC	0.70	0.57	0.58		0.76	0.84	0.32
KM+SOM	0.78	0.53	0.66	0.76		0.82	0.37
HC+SOM	0.73	0.56	0.62	0.84	0.82		0.35
KM+HC+SOM	0.40	0.26	0.50	0.32	0.37	0.35	

TABLE II. JACCARD INDEX FOR CLUSTER COVERING REWARD SYSTEM.

	KM	HC	SOM	KM+HC	KM+SOM	HC+SOM	KM+HC+SOM
KM		0.00	0.56	0.41	0.53	0.52	0.53
HC	0.00		0.00	0.00	0.00	0.00	0.00
SOM	0.56	0.00		0.43	0.57	0.55	0.57
KM+HC	0.41	0.00	0.43		0.33	0.33	0.37
KM+SOM	0.53	0.00	0.57	0.33		0.67	0.61
HC+SOM	0.52	0.00	0.55	0.33	0.67		0.60
KM+HC+SOM	0.53	0.00	0.57	0.37	0.61	0.60	

TABLE III. JACCARD INDEX FOR CLUSTER COVERING AUDITORY SYSTEM.

	KM	HC	SOM	KM+HC	KM+SOM	HC+SOM	KM+HC+SOM
KM		0.49	0.00	0.00	0.41	0.71	0.62
HC	0.49		0.00	0.00	0.35	0.51	0.56
SOM	0.00	0.00			0.00	0.00	0.00
KM+HC	0.00	0.00			0.00	0.00	0.00
KM+SOM	0.41	0.35	0.00	0.00		0.48	0.40
HC+SOM	0.71	0.51	0.00	0.00	0.48		0.65
KM+HC+SOM	0.62	0.56	0.00	0.00	0.40	0.65	

There are in total 18 final clusters with 7 representing visual areas, 6 representing reward system, and 5 representing auditory system by different single methods and method combinations. To save the space and make a brief demonstration of how different these clustering results look when they are mapped on to the three dimensional human brain model, we illustrated only 6 different clustering results by six different method combinations as shown in Table II. In Figure 3, the red parts represent the cluster covering the reward system by six different method combinations.

### V. DISCUSSION AND CONCLUSION

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

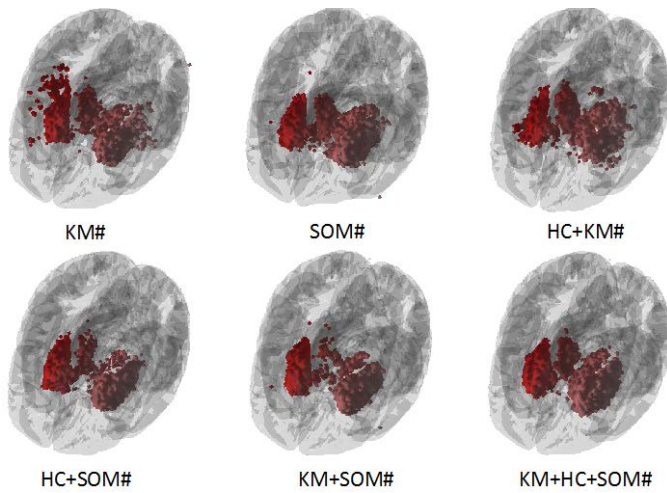


Figure 3. 3D illustrations of final clusters generated by six different method combinations.

solution violates at least one of the common assumption regarding cluster consistency, richness, and scale invariance. Thus it is often very hard to determine an optimal clustering algorithm for a specific problem. In this paper, we utilised a tunable consensus clustering paradigm Bi-CoPaM to combine the clustering results on a real functional magnetic resonance imaging data with multiple clustering algorithms to improve the consistency and stability performance of clustering in neuroimaging data analysis. The results show that the proposed method has great capability and potential to help address the consistency issue in neuroimaging data analysis.

From Table 1 to Table 3, which show the mutual differences between any two method combinations, and the three dimensional demonstration of one set of clusters covering the reward system in human brain, we can see the obvious differences among different clustering schemes. For the cluster covering the visual system, although all of the method combinations were able to detect this area, details of the regions such as the size and shape vary among each other. For the cluster covering the reward system and another cluster covering the auditory system, not only the results are different from each combination of clustering algorithms but also some single methods failed to detect the important brain structures based on their BOLD response time profiles. However, when it comes to the results obtained from any combined results from different single methods, these two important clusters could be identified even though one method failed to form the corresponding important cluster. This demonstrates the powerfulness of integrating the results from multiple clustering methods. Furthermore, the Bi-CoPaM paradigm distinguishes from existing clustering ensemble techniques. It features the capability of tuning the tightness of generated clusters and in conjunction with the M-N scatter plot technique, one can obtain the non-trivial clusters with high inner similarity and a large number of objects within the cluster. Such clusters often include very important information regarding the structure and underlying distribution of the data from which these clusters are formed.

In real clustering applications, due to the lack of sufficient a

priori information, it is hard for the researchers to determine which clustering result is the best. Although one can try to run the clustering experiment exhaustively to find the optimal solution, this approach becomes more and more infeasible when the size of the neuroimaging data keeps increasing [16]. In addition, in neuroimaging studies, the signal to noise ratio is often very low [17], making it hard to draw any conclusion based on a single or even several experiments. The Bi-CoPaM used in this paper manages to exploit the results from single clustering method and find the clusters having consistent patterns across many clustering algorithms and many datasets. By integrating the results from many independent clustering processes means that the results are reproducible. Indeed, if certain clusters would only appear in few clustering trials, then it is very likely that the appearance of these clusters is due to random error or other factors and would not be included in the final results. This ensures that the final clusters from Bi-CoPaM are more stable and robust than the results from single method, so that researchers can be more confident on the interpretations or conclusions made from the data.

## REFERENCES

- [1] R. Fa, D. J. Roberts, and A. K. Nandi, "SMART: unique splitting-while-merging framework for gene clustering," *PLoS One*, vol. 9, no. 4, p. e94141, Jan. 2014.
- [2] 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, Jan. 2013.
- [3] A. Venkataraman, K. R. A. Van Dijk, R. L. Buckner, and P. Golland, "Exploring functional connectivity in fMRI via clustering," in *ICASSP 2009*, 2009, pp. 441–444.
- [4] R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg, "A whole brain fMRI atlas generated via spatially constrained spectral clustering," *Hum. Brain Mapp.*, vol. 33, no. 8, pp. 1914–28, Aug. 2012.
- [5] 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, Mar. 1999.
- [6] V. Estivill-castro, "Why so many clustering algorithms — A Position Paper," *ACM SIGKDD Explor. Newsl.*, vol. 4, no. 1, pp. 65–75, 2002.
- [7] J. Kleinberg, "An impossibility theorem for clustering," *Adv. Neural Inf. Process. Syst.*, pp. 446–453, 2002.
- [8] A. Strehl and J. Ghosh, "Cluster Ensembles – A Knowledge Reuse Framework for Combining Multiple Partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, 2002.
- [9] K. S. Button, J. P. a Ioannidis, C. Mokrysz, B. a Nosek, J. Flint, E. S. J. Robinson, and M. R. Munafò, "Power failure: why small sample size undermines the reliability of neuroscience," *Nat. Rev. Neurosci.*, vol. 14, no. 5, pp. 365–76, May 2013.

- [10] A. Topchy, A. K. Jain, and W. Punch, "Clustering ensembles: models of consensus and weak partitions.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1866–81, Dec. 2005.
- [11] D. Boyd and K. Crawford, "Critical Questions for Big Data," *Information, Commun. Soc.*, vol. 15, no. 5, pp. 662–679, Jun. 2012.
- [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, p. 308, Jan. 2011.
- [13] S. Saarikallio, G. Luck, B. Burger, M. Thompson, and P. Toiviainen, "Dance moves reflect current affective state illustrative of approach–avoidance motivation.," *Psychol. Aesthetics, Creat. Arts*, vol. 7, no. 3, pp. 296–305, 2013.
- [14] B. Abu-Jamous, R. Fa, D. J. Roberts, and A. K. Nandi, "UNCLES: method for the identification of genes differentially consistently co-expressed in a specific subset of datasets," *BMC Bioinformatics*, vol. 16, no. 184, 2015.
- [15] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.," *Neuroimage*, vol. 15, no. 1, pp. 273–89, Jan. 2002.
- [16] C. Liu, R. Fa, B. Abu-Jamous, E. Brattico, and A. K. Nandi, "Scalable Clustering Based on Enhanced-Smart for Large-Scale fMRI Datasets," in *ICASSP*, 2015.
- [17] M. Welvaert and Y. Rosseel, "On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data.," *PLoS One*, vol. 8, no. 11, p. e77089, Jan. 2013.