

LARGE SCALE FUNCTIONAL CONNECTIVITY FOR BRAIN DECODING

Orhan Firat¹, Itir Onal¹, Emre Aksan¹, Burak Velioglu¹, Ilke Oztekin² and Fatos T. Yarman Vural¹

¹Department of Computer Engineering, Middle East Technical University, Turkey,

²Department of Psychology, Koc University, Turkey,

Emails: {orhan.firat, itir, emre.aksan, velioglu, vural}@ceng.metu.edu.tr, ioztekin@ku.edu.tr

ABSTRACT

Functional Magnetic Resonance Imaging (fMRI) data consists of time series for each voxel recorded during a cognitive task. In order to extract useful information from this noisy and redundant data, techniques are proposed to select the voxels that are relevant to the underlying cognitive task. We propose a simple and efficient algorithm for decoding the brain states by modelling the correlation patterns between the voxel time series. For each stimulus during the experiment, a separate functional connectivity matrix is computed in voxel level. The elements in connectivity matrices are then filtered out by making use of a minimum spanning tree formed using a global connectivity matrix for the entire experiment in order to reduce dimensionality. For a recognition memory experiment with nine subjects, functional connectivity matrices are computed for encoding and retrieval phases. The class labels of the retrieval samples are predicted within a k-nearest neighbour space constructed by the traversed entries in the functional connectivity matrices for encoding samples. The proposed method is also adapted to large scale functional connectivity tasks by making use of graphics boards. Classification performance in ten categories is comparable and even better compared to both classical and enhanced methods of multi-voxel pattern analysis techniques.

KEY WORDS

Data mining and Machine Learning, Magnetic Resonance Imaging, Medical Image Processing, Brain State Decoding, Functional Connectivity, MVPA.

1 Introduction

Understanding and modelling the human brain is one of the greatest challenges of our century across many fields in neuro-science, cognitive science and computer science. Neuro-imaging mediums, in particular, functional Magnetic Resonance Imaging (fMRI) is the premier choice for the analysis of the brain in order to reach this goal. However, the bulk amount of incomplete, noisy and yet redundant data requires intensive image processing and machine learning techniques to extract useful information about the nature of the brain. The techniques employed for analysing and learning the patterns of brain activity in fMRI data are called Multi-voxel pattern analysis (MVPA). Under the umbrella of MVPA techniques, many problems, such as

hypothesis validation [1], diagnosing disorders [2, 3] and recently brain state decoding, also known as mind reading [4, 5], are studied to be resolved and expected to have a huge impact on understanding the intrinsics of the brain.

Brain decoding tasks are challenging in nature, because of the scarcity of the labelled data compared to the high dimensional feature spaces. Therefore, researchers try to reduce the dimensionality of the data by selecting the *relevant* voxels in order to improve significance. An alternative to this approach is to incorporate various information sources and modalities to improve the accuracy and the generalization performances.

One way to extract useful information from the fMRI data is to employ all the temporal measurements and develop a model for the temporal structure between the voxel pairs. There exist various techniques to represent the relationship between the time series data. A popular model which is widely used in fMRI data is known as functional connectivity [6,7]. Functional connectivity is defined as the statistical dependency between the neural elements or regions across time, and widely used for decoding problems recently [8–10]. Extracting the functional connectivity patterns from the fMRI measurements is not a well defined task and requires many assumptions about the data. First of all, the number of time samples which fall into a cognitive task is very few. Therefore, it is not easy to estimate the correlations other than zero-lag in voxel level. Secondly, due to the many unexplained characteristics of the brain it is not clear which similarity metric suits to model the time series between the voxel pairs.

There are several approaches proposed in the literature to model the functional connectivity among the voxels for brain decoding. In [9, 10] region level functional connectivity matrices for two classes are used for classification by subtracting them to reveal discriminative relations. In [11] functional connectivity is used to select a neighbourhood for a mesh model (MAD) [12] around each voxel to extract features that compress neighbouring structures for further classification. Most of the above mentioned methods utilize functional connectivities among the brain regions, but not among the voxels, resulting a far fewer number of elements in connectivity matrix calculation steps. The major drawback of these inter regional connectivity analysis is the smoothing effects introduced by averaging all the voxel intensity values within a region.

Also, in most cases, the region priors from anatomical atlases are quite coarse or not detailed enough for the specific activation patterns caused by the cognitive process in voxel level. These downsides motivates our research to further explore voxel level connectivity patterns resulting large connectivity matrices that are difficult to compute and store. However, this holistic approach may provide a fine and informative representation for a cognitive process. An example of using whole brain connectivity can be found in [13] where whole brain functional connectivity matrices are used but along with MAD for functional neighbour selection in a two class classification task again for linear-spatial feature extraction.

Another soft spot on fMRI brain decoding tasks as mentioned above, is the scantiness of the labelled samples compared to the high dimensional feature spaces. Such regimes are known to be prone to severe deficiencies for classification problems such as over-fitting, insignificance and curse of dimensionality. In order to overcome these problems, dimensionality reduction and/or voxel selection methods are employed by researchers such as Principal Component Analysis [14–16], Independent Component Analysis [17, 18] and Low-Dimensional Embeddings [19]. The commonality between these methods is the transformation of original feature space into a low dimensional feature space. Experimental priors are hard to inject in these approaches and further voxel level interpretations transformed feature space is not trivial. A convenient and emerging tool to attack dimensionality reduction problem is network coding where the network structure regarding to cognitive states are compressed using graph theoretical approaches [20–23], as also employed in our proposed method.

In this study, we employ voxel-level functional connectivity matrices directly to classify cognitive processes. The difference of our approach with [11, 13] is that we discarded any complicated feature extraction steps after calculating the connectivity matrices and used them directly for matching the corresponding class label. Furthermore, we calculated the pair-wise correlation measures from signal windows adjusted according to the hemo-dynamic response function. We also employed graphics boards to further improve the capability of calculating whole brain functional connectivity matrices. The quadratic size of the connectivity matrices that form our feature space is reduced by taking the pairwise correlation measures that are spanned by a template Minimum Spanning Tree (MST). Template MST is formed using a separate functional connectivity matrix that is calculated by the pairwise correlations accounting all the time points, not within small signal windows, which further reduced dimensionality to the number of voxels in the experiment.

This paper is organised as follows, in the next section the fMRI experiment design and pre-processing steps are explained. Third section clarifies the details of the employed method. Experiments on fMRI data and classification results are analysed in fourth section followed by the

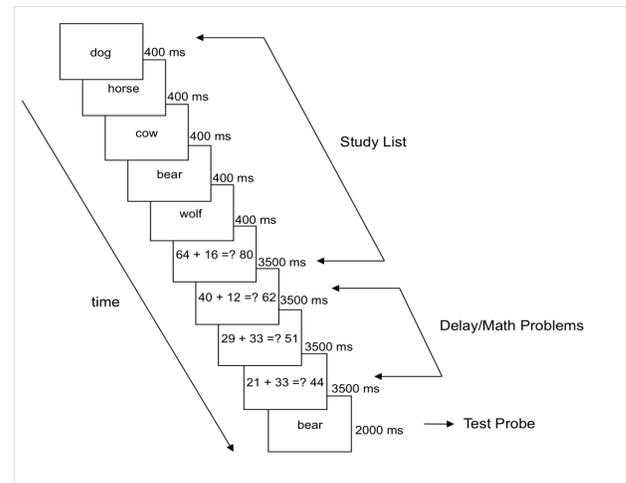


Figure 1. A sample trial for conducted fMRI recognition memory experiment. In each trial participant is exposed to a separate study list belonging to one of ten categories (animal category is shown). Each trial proceeded with the presentation of 5-word study list for 2 seconds in encoding phase, followed by a delay period of math problem solving. In the retrieval phase a test probe is presented and indicated whether the word was a member of the current study list.

discussion and conclusion in the fifth and sixth section.

2 Dataset and fMRI Experiment

2.1 fMRI Experiment

In the current study, fMRI recording was conducted during a recognition memory task for nine healthy subjects in eight runs per subject. Each participant is shown a list of words belonging to a specified category in the encoding phase for 2 seconds as a study list. In the study list, 5 words belonging to the same category is presented sequentially for 400 ms each. Following the study list, participants solved math problems 3500 ms each as a delay period. Employing a delay period for 14 seconds, allows independent assessment of encoding related (i.e. study list period) brain activation from retrieval related (i.e. during the test probe) activity patterns. Finally a test probe is presented as the last step, and the participant executes a yes/no response indicating whether the word belongs to the current study list for 2 seconds (e.g., see Figure 1). A total of ten semantic categories were used in the study, which are *animals*, *colors*, *furniture*, *body parts*, *fruits*, *herbs*, *clothes*, *chemical elements*, *vegetables* and *tools*. Recording was conducted using a 3T Siemens scanner with a 2 seconds TR, meaning that we obtained a separate brain volume each 2 seconds (see [1] for further experimental details). For the MVPA analyses we focused on the lateral temporal cortex region having 8142 voxels total. fMRI data consists of 2400 time points in eight runs with 240 class labels for the encoding phase, and 240 class labels for the retrieval phase (for each of the ten classes, 24 samples are obtained for both encoding and retrieval). The encoding samples are used for train-

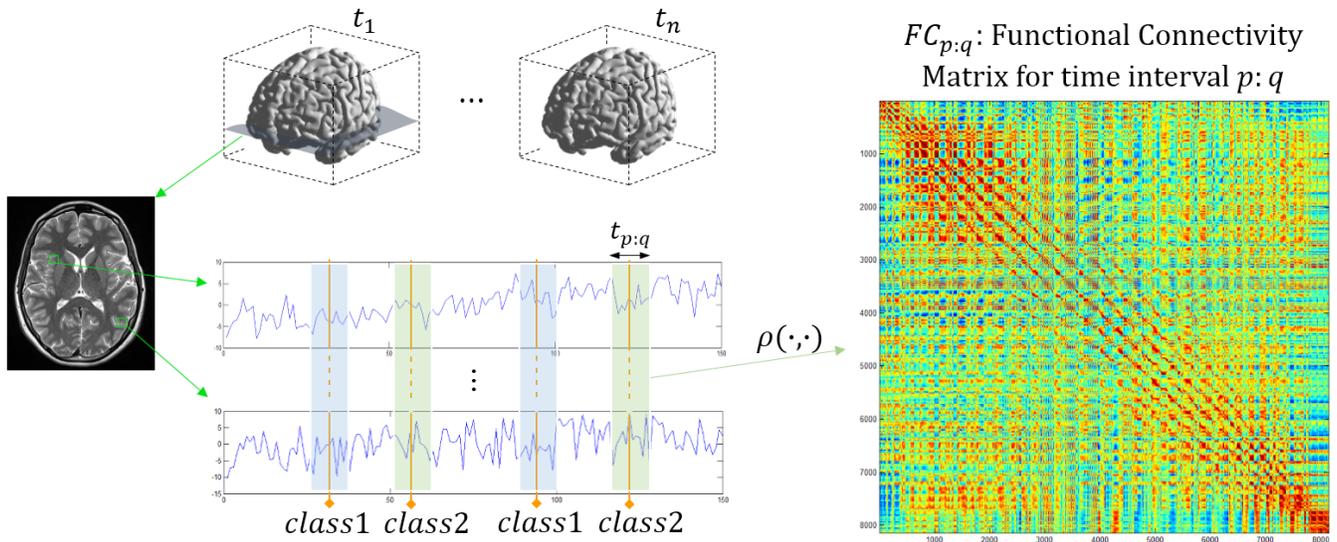


Figure 2. Example of a typical fMRI experiment for brain state decoding. 4D fMRI data consist of several volumes across time, some of which have assigned to a class labels(indicated by vertical orange lines in time axis). For each class either encoding or retrieval, a separate functional connectivity matrix is formed by considering a suitable time window that encapsulates indicated class labels (indicated by highlighted intervals with different colors for classes).

ing and decoding samples are used for testing a classifier further in our method.

2.2 fMRI Data Pre-Processing

Image processing and data analysis were performed using SPM5 (<http://www.fil.ion.ucl.ac.uk/spm/>). Following quality assurance procedures to assess outliers or artefacts in volume and slice-to-slice variance in the global signal, functional images were corrected for differences in slice acquisition timing by re-sampling all slices in time to match the first slice, followed by motion correction across all runs (using sinc interpolation). Functional data were then normalized based on MNI stereotaxic space using a 12-parameter affine transformation along with a nonlinear transformation using cosine basis functions. Images were resampled into 2-mm cubic voxels and then spatially smoothed with an 8-mm FWHM isotropic Gaussian kernel. Further detrending and normalization was avoided because we employed connectivity measures within small time windows.

3 Method

fMRI data consists of 3-dimensional brain volumes across time $\{t_i\}_{i=1}^n$, where each 3D volume is formed by stacking several 2D scans (slices). Each pixel in these 2D images represents the intensity of a small volume of brain tissue (voxel) at a time instant t_i . We represent the intensity of a voxel at location \bar{s}_j in a time instant t_i as $v(t_i, \bar{s}_j)$ and note that \bar{s}_j is a three dimensional vector indicating the position of a voxel in the volume. We indicate the time signal vector for a voxel at location \bar{s}_j as $\bar{v}(t_{p:q}, \bar{s}_j)$ where time points are delimited starting from p and ending at q . A typical fMRI experiment consists of several runs, where in

each run the subject is exposed to some task specific stimulus $\{c_j\}_{j=1}^S$ at the predefined time instants, where S is the total number of semantic classes in the experiment and n is the total length of the experiment across runs (see Figure 2). Within a large amount of samples across time, only few have assigned class labels (orange vertical lines in Figure 2). The rest of the samples that are not having a class label t_i/c_j are generally discarded in MVPA tasks or a pre-determined number of time points are averaged according to the prior knowledge of the peaks of hemo-dynamic response function (e.g. 2-3 time points after the stimulus).

3.1 Functional Connectivity for Feature Representation

In this study we utilize the entire temporal structure of voxels for each stimulus and none of the samples are discarded. For this purpose, first we determine a time window delimited by p and q for each separate stimulus. It is well known that, Hemo-dynamic Response Function (HRF) peaks around 4-6 seconds after a stimulus and returns to a baseline after 10-12 seconds [24]. In order to capture the temporal structure, we specified time windows that span 6 samples in the fMRI recordings (see Figure 2). Considering 2 seconds TR as the experimental protocol, 6 samples span a full hemo-dynamic response function with 12 seconds. After determining a window size in time, we extract a distinct functional connectivity matrix for each time window in voxel-level (blue and green highlighted bars in Figure 2). Note that starting from a time window adjusted for a single stimulus delimited by p and q , we obtain a single connectivity matrix for each stimulus. As we expand the delimiters to the experimental limits where $p = 1$ and $q = n$, we obtain a single connectivity matrix for the whole

experiment. The elements in the functional connectivity matrix are represented by symmetric dependence measures in the time domain. It has been suggested that correlation based measures are well suited for functional connectivity analysis [25]. Consequently, we use zero-order correlation (cross-correlation) to measure the functional similarity between time-series. The zero-order correlation coefficient ρ_{jk} , between two voxels at location \bar{s}_j and \bar{s}_k , within the time window started at time point p and ended at time point q is calculated by,

$$\rho_{jk} = \frac{cov_{jk}(\bar{v}(t_{p:q}, \bar{s}_j), \bar{v}(t_{p:q}, \bar{s}_k))}{\sqrt{var_j(\bar{v}(t_{p:q}, \bar{s}_j)) \cdot var_k(\bar{v}(t_{p:q}, \bar{s}_k))}}, \quad (1)$$

where cov_{jk} is the covariance of the signals measured at two voxels within the time window, and var_j is the variance of the signals measured at a voxel $\bar{v}(t_{p:q}, \bar{s}_j)$ and $\rho_{jk} \in [-1, 1]$. Note that, p is set to the beginning of a stimulus, q is the end point of the window as $p + 6$ and the time window is indicated by $p : q$. Here, the voxel intensities within a time window $p : q$ comprise a vector of intensity values \bar{v} and used to calculate ρ_{jk} . Each stimulus corresponds either to encoding (training) or retrieval (test) phase and accommodates a class label. By calculating connectivity matrices $FC_{p:q}$ for all stimulus, we obtain a training and test set for our further classification task. Considering the nature of zero-order correlation which is symmetric, resulting functional connectivity matrices are also symmetric. Therefore it is sufficient to take into account only the lower diagonals of each connectivity matrix. In this study we constructed our feature space for classification, by extracting lower diagonals of connectivity matrices and converting them into column vectors for further comparison. Each connectivity matrix is converted into a column vector and represented in the feature space by these vectors where we finally used a k-nearest neighbours method to determine the class labels for test samples.

3.2 Large Scale Connectivity and Parallelism

The functional connectivity matrices grow quadratically as we increase the number of voxels considered in the experiment. As an example, a single connectivity matrix for 8k voxels has 64M pairwise elements and is time and space consuming to calculate. For 80k voxels, 6.4B entries reserve around 50GB of memory which is just for one connectivity matrix. The practical solution for time and space complexity can be achieved by making use of graphics boards to speed-up computation time and storing resulting matrix into disk as chunks for the space complexity. We attacked this problem by first storing a portion of the time signals data on the texture memory of GPU then computing only the corresponding chunk of connectivity matrix and storing result to disk in binary format to further space gain and speed. By iterating over different time signal portions repeatedly, lower diagonal chunks of resulting functional connectivity matrix is calculated and stored to disk. For speeding up the computations, correlation measure is

expanded to exploit parallelism. Let us simplify the notation of correlation and denote x and y as our time window intensity vectors for two distinct voxels. The correlation measure in (1) can be re-written as follows with a slight abuse of notation,

$$\rho_{xy} = \frac{\sum x_i y_i - m \left(\frac{\sum x_i}{m} \right) \left(\frac{\sum y_i}{m} \right)}{\sqrt{\left(\sum x_i^2 - m \left(\frac{\sum x_i}{m} \right)^2 \right) \left(\sum y_i^2 - m \left(\frac{\sum y_i}{m} \right)^2 \right)}}, \quad (2)$$

where m is the temporal window size, which is 6 in our case.

The parallelism can be achieved easily by converting a complex problem into a matrix-vector multiplication problem which is very efficient to calculate on GPUs. Analysing the re-arranged formula (2), we observe that $\sum x_i$, $\sum y_i$, $\sum x_i^2$ and $\sum y_i^2$ can be calculated with a single matrix-vector multiplication. Only the summation $\sum x_i y_i$ needs to be calculated in a kernel as the last step. We can achieve further computation enhancements by storing other summation results in texture memory of GPU. The overall algorithm to compute a large scale functional connectivity matrix is given as follows:

1. Load data chunk to device memory (time series)
2. Calculate $\sum x_i$, $\sum y_i$ using CUBLAS, save resulting vectors on device memory
3. Square data matrix with a transform routine on Thrust and calculate $\sum x_i^2$, $\sum y_i^2$ using CUBLAS, save resulting vectors on device memory
4. Load source and destination time-series into texture memory (pairwise cliques)
5. Calculate pairwise correlation on GPU, save resulting matrix chunk on disk
6. Return to step 4

In our CUDA implementation for large scale functional connectivity calculation, we employed CUDA Basic Linear Algebra Subroutines (CUBLAS) and CUDA Thrust libraries for speed up and simplicity.

3.3 Dimensionality Reduction using Minimum Spanning Trees

The quadratic growth of the connectivity matrices drastically increases the dimension of feature space when we use all the correlation measures directly in classification. As an example 8k voxels results in a connectivity matrix that has 64M pairwise elements and a feature vector having 32M dimensions (considering only lower diagonals). Due to the scarcity of the labelled samples which is 240/240 for training/test, we need to employ some dimension reduction techniques. In such large scales, conventional dimensionality reduction approaches become obsolete because of their

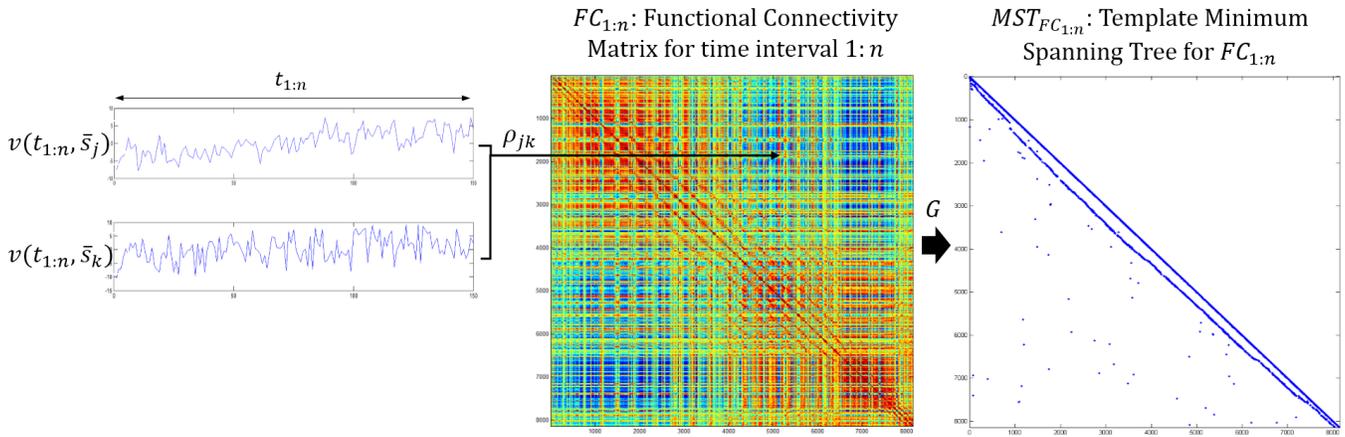


Figure 3. Template Minimum Spanning Tree calculation process. Each pairwise relations within the entire experiment is used to calculate a global connectivity matrix $FC_{1:n}$ with a time window equal to the length of experiment. Template MST is then formed by converting connectivity matrix into an intermediate distance matrix G .

time and space complexities (even inverting a big matrix or singular value decomposition is costly). Therefore one needs an efficient approach to resolve this problem. Graph theory is a well structured and developed field in computer science where it becomes handy in large scale network analysis problems. Now let us reconsider our dimensionality reduction problem with a graph theoretical perspective. A functional connectivity matrix is an affinity matrix which can be converted to a distance matrix easily. For each pairwise element jk of a connectivity matrix FC , we can convert a correlation measure ρ_{jk} into a distance measure d_{jk} by calculating,

$$d_{jk} = 1 - |\rho_{jk}|, \quad \text{for all pairs } jk, \quad (3)$$

where both positive and negative correlation are assumed to have equal information, and zero correlation indicating largest distance.

After defining a distance measure, we can transform our dimensionality reduction problem into a graph traversal problem using a Minimum Spanning Tree (MST). An MST spans all the nodes of the graph (voxels) with minimum total distance (regarding to maximum correlation after our transformation). The suggested approach enables us to compress the information encoded in a correlation matrix by taking into account only the most informative (highly positive or negative correlations) pair-wise relations, spanning entire region of interest in the brain for cognitive processes. Furthermore, we use the path traversed by the MST to filter out elements in any other correlation matrix by considering only the pairwise relations reside on the path. Note that, in a graph with V nodes, MST of the graph has $V - 1$ edges, meaning that we can transform a functional connectivity matrix as a distance matrix for 8k voxels having 64M elements then representing it with a spanning tree having only 8k-1 edges (elements of the distance matrix) out of 64M elements. The reduction of the dimensionality

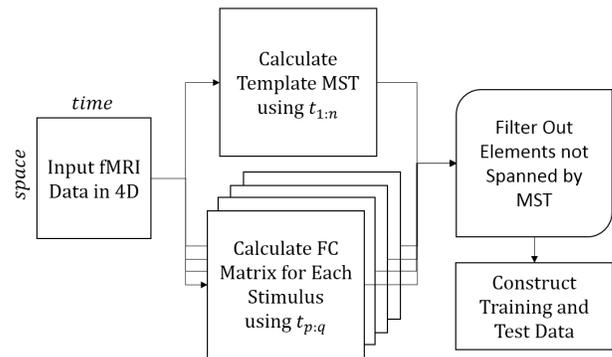


Figure 4. Algorithm flow of proposed method. Input fMRI data is used simultaneously to compute functional connectivity matrices for each stimulus within small time windows $t_{p:q}$ and calculating the template MST within entire time domain $t_{1:n}$. Resulting template MST is then used to filter out elements of each $FC_{p:q}$, constructing training and test data further fed to classifier.

is colossal and easy to compute with iterative MST routines [26].

Extraction of a unique MST from a set of functional connectivity matrices is not a trivial problem; because in our model, for each stimulus, a different connectivity matrix is calculated and correlations are extracted within the time interval of that stimulus. If we extract a different MST for each connectivity matrix, then each MST will span a different feature space. A practical solution to overcome this problem is to calculate a unique MST that is shared among all the connectivity matrices extracted from distinct stimulus windows. This global template MST should encapsulate all the shared structure among different stimulus and must be shared to extract comparable features that are localized with its unique path. As mentioned before, we

Method	Subject-1	Subject-2	Subject-3	Subject-4	Subject-5	Subject-6	Subject-7	Subject-8	Subject-9
MVPA	0.39	0.54	0.46	0.47	0.49	0.47	0.44	0.52	0.43
MAD [12]	0.51	0.65	0.60	0.64	0.67	0.58	0.56	0.63	0.55
FC-Euc	0.64	0.68	0.63	0.60	0.63	0.67	0.61	0.61	0.64
FC-Cos	0.63	0.68	0.62	0.60	0.62	0.64	0.61	0.60	0.64
FC-Cor	0.65	0.68	0.64	0.62	0.61	0.68	0.62	0.63	0.64
FC-Abs	0.64	0.68	0.63	0.60	0.62	0.66	0.61	0.61	0.64
FC-MST-Euc	0.54	0.61	0.58	0.55	0.56	0.61	0.57	0.56	0.62
FC-MST-Cos	0.50	0.59	0.54	0.54	0.54	0.60	0.55	0.54	0.59
FC-MST-Cor	0.66	0.67	0.64	0.62	0.61	0.65	0.63	0.64	0.64
FC-MST-Abs	0.57	0.63	0.59	0.58	0.57	0.63	0.57	0.58	0.62
Max - FC	0.65	0.68	0.64	0.62	0.63	0.68	0.62	0.63	0.64
Max - FC-MST	0.66	0.67	0.64	0.62	0.61	0.65	0.63	0.64	0.64
Impr.wrt MAD	0.15	0.03	0.03	-0.01	-0.04	0.10	0.07	0.01	0.09
Impr.wrt MVPA	0.26	0.14	0.18	0.15	0.14	0.21	0.19	0.12	0.21

Table 1. Classification performances of proposed method compared with classical methods for fMRI machine learning tasks (MVPA) and a local-linear feature extraction method (MAD). Proposed method is indicated by FC and followed by the distance metric used.

can extend the time window of a stimulus into the limits and obtain a single connectivity matrix for the whole experiment by setting delimiters $p = 1$ and $q = n$. This yields a single connectivity matrix which can also be considered as correlation pattern of the entire experiment. By calculating a unique connectivity matrix for the entire experiment, we can extract a template MST to be used as a reference path to filter out not-spanned edges in each of the connectivity matrices calculated for each stimulus.

Formally speaking, a minimum spanning tree is connected sub-graph of an undirected weighted graph, containing all the nodes in the original graph with no cycles. Such a sub-graph with no cycles is a spanning tree and with the minimum total weights it is called minimum spanning tree. We define a fully connected intermediate graph G as $G = (V, E)$ where the nodes of the graph $s_j \in V$, represent the voxel coordinates and edge weights are defined as d_{jk} calculated using (1) by setting $p = 1$ and $q = n$ then transforming into distances using (3). Note that in the time interval $t_{1:n}$, sample windows that belong to retrieval phases (test samples) are excluded at all steps for fairness. Computation of the template MST for whole time series $MST_{FC_{1:n}}$, is straightforward after nodes and arc weights are defined, using Kruskal’s method. Resulting $MST_{FC_{1:n}}$ has $V - 1$ edges that are further used to filter each connectivity matrix extracted using small windows of stimulus. This process is illustrated in Figure 3, where functional connectivity matrix is calculated using all voxel pairs including entire time points this time. Resulting template minimum spanning tree $MST_{FC_{1:n}}$ of connectivity matrix $FC_{1:n}$ is then computed with the help of intermediate graph G . In Figure 3, template MST is visualised only with the edges marked in lower diagonal since it is undirected. The marked elements in the template MST is further used to filter each connectivity matrices to reduce

the dimension of the feature vectors. The algorithm flow is illustrated in Figure 4. In the last step of the algorithm, the selected elements in the lower diagonals by template MST of each stimulus FC matrix, are used to form training and test data, depending on the records obtained at the encoding or retrieval phase of the experiment.

4 Experiments

In our experiments on fMRI brain decoding task, we employ a region of interest having 8142 voxels for 9 subject where each subject has 8 experimental runs. The classification performance is measured for each run separately and then averaged for each subject. For a k-nearest neighbour classification regime, we need a distance metric for our feature space. In this study, we employ four different distance metrics to measure the distance between training and test sample vectors. These metrics are standard Euclidean distance (FC/FCMST-Euc), cosine between two vectors (FC/FCMST-Cos), 1-absolute correlation (FC/FCMST-Cor) and sum of absolute differences (FC/FCMST-Abs). We compare our method, first with a standard MVPA approach where voxel intensity values are directly fed into the classifiers. As a second comparison we used Mesh Learning approach (MAD) to extract local-linear features from voxel intensity values, which exploit spatial structure among the voxels. The performance results are illustrated in Table 1. We observe a significant improvement up to 25% compared to classical MVPA method in all of the subjects. This improvement is expected as we incorporated temporal information into classification process. Calculating correlation and using coefficients as features smooths out noise in voxel intensities.

MAD approach exploits spatial structure of the voxel intensities for each time instant separately and enhances the classification performance compared to the classical MVPA approach. We observed a better performance up

to 14% compared to MAD approach. For two out of nine subjects, MAD approach give better results. For the other seven subjects proposed method substantially outperforms the MAD approach.

5 Discussion

The suggested approach is distinguished from the other comparable methods in the way of accounting the fMRI data modalities. MAD exploits spatial modality of fMRI data whereas our proposed method exploits temporal modality. Considering the noisy spatial acquisition process of fMRI data, MAD approach is more prone to noise with respect to proposed method which slightly smooths-out noise component by correlation analysis. Another advantage of the proposed method is its simplicity. The model order selection or neighbour selection problems of MAD approach are simply avoided by incorporating correlation structures. However, the problem of selecting the informative pairwise relations is emerged. This problem is resolved by calculating a template MST to span the most informative pair-wise relations and accounting only the ones that are spanned by template MST. Final step reduced the dimensionality drastically down to 8141 from 32M with a minor performance drawback in three subjects but equal or better performance in other subjects. The performance gain observed by reduced dimensionality is also expected. Avoiding the curse of dimensionality and overfitting by reduced dimensions results a higher generalization performance with a better significance. The dimensionality of the feature spaces for each method is also illustrated in Table 2. By analysing Table 2 we observe that improved performance for enhanced methods are encouraged by higher dimensions in the feature spaces. The baseline MVPA method employs around 8k feature dimensions. With such enhanced methods (MAD) in order to achieve better performance the trade off reveals it self as increased feature dimensions. The proposed method aids this problem by retracting the high number of dimensions with improved classification accuracy.

Another observation regarding to the proposed method is the structure of the calculated template MST. In Figure 3, illustrated MST has the majority of the connections near diagonals meaning that the most correlated and informative relations reside in a close vicinity of a voxel which is known from the capillary structure of the brain and point spread function of the fMRI medium. But the interesting connections also observed off the diagonals which are remote informative relations. Further studies should reveal these connectivities and examine the hub and clustering structures.

In our experiments we also employed a stress test for correlation analysis on GPU, by calculating a pairwise functional connectivity matrix for 80.000 and 220.000 voxels. We observed that the developed approach is capable of calculating and storing in such large-scales which is just not possible with a straight-forward approach or by using tools such as SPM, Functional Connectivity Toolbox [27]

Method	Dimension of Feature Space
MVPA	8.142
MAD	16.284-154.698
FC	33 M
FC-MST	8.141

Table 2. Dimensionality of the feature spaces for different method employed and proposed method.

or FSL [28]. The developed approach implemented with CUDA is available on <http://ceng.metu.edu.tr/~e1697481/SCT/>.

6 Conclusion

In this study we introduced a simple approach for brain decoding using temporal structures of voxel activations. In the proposed method, voxels are represented in the feature space by their pairwise correlation coefficients with all other voxels within a time window determined by hemodynamic response function. Proposed method is tested in a recognition memory experiment having 10 classes with 9 subjects and results are compared with standard methods and methods that exploit spatial structure. Experimental results suggest that the proposed method is more robust to noise by incorporating temporal structure and comparable with the current methods for MVPA.

Further studies should focus on selecting the informative relations encoded in the functional connectivity matrices in order to discriminate semantic classes for higher classification accuracy. For a better significance, dimensionality of the feature representations should be further reduced by either model order selection methods or by analysing between class connectivity matrices. As a future work we take into consideration to combine the proposed method with methods exploiting spatial structures.

Acknowledgement

This work is supported by TUBITAK under the grant number 112E315.

References

- [1] I. Oztekin and D. Badre, "Distributed Patterns of Brain Activity that Lead to Forgetting." *Frontiers in human neuroscience*, vol. 5, 2011.
- [2] T. Schmah, G. E. Hinton, R. S. Zemel, S. L. Small, and S. C. Strother, "Generative versus discriminative training of rbms for classification of fmri images." in *NIPS*, 2008.
- [3] M. N. Coutanche, S. L. Thompson-Schill, and R. T. Schultz, "Multi-voxel pattern analysis of fMRI data predicts clinical symptom severity." *NeuroImage*, vol. 57, 2011.

- [4] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fMRI data." *Trends in cognitive sciences*, vol. 10, 2006.
- [5] J.-D. Haynes and G. Rees, "Decoding mental states from brain activity in humans." *Nature reviews. Neuroscience*, vol. 7, 2006.
- [6] K. J. Friston, "Functional and effective connectivity: a review," *Brain connectivity*, vol. 1, no. 1, pp. 13–36, 2011.
- [7] O. Sporns, "Networks of the Brain," Nov. 2010.
- [8] B. Chai, D. B. Walther, D. M. Beck, and F.-F. L., "Exploring Functional Connectivity of the Human Brain using Multivariate Information Analysis," *Advances in neural information processing systems*, 2009.
- [9] J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. Van De Ville, "Decoding brain states from fMRI connectivity graphs." *NeuroImage*, vol. 56, no. 2, pp. 616–26, May 2011.
- [10] W. R. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and M. D. Greicius, "Decoding subject-driven cognitive states with whole-brain connectivity patterns." *Cerebral cortex (New York, N.Y. : 1991)*, vol. 22, no. 1, pp. 158–65, Jan. 2012.
- [11] O. Firat, M. Ozay, I. Onal, I. Öztekin, and F. T. Yarman-Vural, "Functional mesh learning for pattern analysis of cognitive processes," in *IEEE 12th International Conference on Cognitive Informatics and Cognitive Computing, ICCI*CC*, 2013, pp. 161–167.
- [12] M. Ozay, I. Öztekin, U. Öztekin, and F. T. Y. Vural, "Mesh Learning for Classifying Cognitive Processes," *Arxiv:1205.2382*, 2012.
- [13] O. Ekmekci, O. Firat, M. Ozay, I. Öztekin, F. T. Yarman-Vural, and U. Öztekin, "Mesh learning for object classification using fmri measurements," in *IEEE International Conference on Image Processing, ICIP*, 2013, pp. 2631–2634.
- [14] L. K. Hansen, J. Larsen, F. Å. Nielsen, S. C. Strother, E. Rostrup, R. Savoy, N. Lange, J. Sidtis, C. Svarer, and O. B. Paulson, "Generalizable patterns in neuroimaging: How many principal components?" *NeuroImage*, vol. 9, no. 5, pp. 534–544, 1999.
- [15] R. Viviani, G. Grön, and M. Spitzer, "Functional principal component analysis of fmri data," *Human brain mapping*, vol. 24, no. 2, pp. 109–129, 2005.
- [16] G. S. Sidhu, N. Asgarian, R. Greiner, and M. R. G. Brown, "Kernel principal component analysis for dimensionality reduction in fmri-based diagnosis of adhd," *Frontiers in Systems Neuroscience*, vol. 6, no. 74, 2012.
- [17] V. D. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar, "Ica of functional mri data: An overview," in *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, 2003, pp. 281–288.
- [18] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fmri: a tutorial overview," *Neuroimage*, vol. 45, no. 1, pp. S199–S209, 2009.
- [19] P. Mannfolk, R. Wirestam, M. Nilsson, F. Ståhlberg, and J. Olsrud, "Dimensionality reduction of fmri time series data using locally linear embedding," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 23, no. 5-6, pp. 327–338, 2010.
- [20] S. L. Bressler and V. Menon, "Large-scale brain networks in cognition: emerging methods and principles," *Trends in cognitive sciences*, vol. 14, no. 6, pp. 277–290, 2010.
- [21] A. F. Alexander-Bloch, N. Gogtay, D. Meunier, R. Birn, L. Clasen, F. Lalonde, R. Lenroot, J. Giedd, E. T. Bullmore, A. Alexander-Bloch *et al.*, "Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia," *Frontiers in systems neuroscience*, vol. 4, p. 147, 2010.
- [22] Y. He and A. Evans, "Graph theoretical modeling of brain connectivity," *Current opinion in neurology*, vol. 23, no. 4, pp. 341–350, 2010.
- [23] O. Firat, M. Ozay, I. Onal, I. Oztekin, and F. Yarman Vural, "Representation of cognitive processes using the minimum spanning tree of local meshes," in *Engineering in Medicine and Biology Society (EMBC), 35th Annual International Conference of the IEEE*, 2013, pp. 6780–6783.
- [24] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional magnetic resonance imaging*. Sinauer Associates Sunderland, MA, 2004, vol. 1.
- [25] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich, "Network modelling methods for FMRI." *NeuroImage*, vol. 54, no. 2, pp. 875–91, Jan. 2011.
- [26] J. B. Kruskal, "On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem," *Proceedings of the American Mathematical Society*, vol. 7, no. 1, pp. 48–50, Feb. 1956.
- [27] D. Zhou, W. K. Thompson, and G. Siegle, "MATLAB toolbox for functional connectivity." *NeuroImage*, vol. 47, no. 4, pp. 1590–607, Oct. 2009.
- [28] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.