Classification of Cognitive State Using Statistics of Split Time Series

J. Siva Ramakrishna
M.S.Ramaiah University of Applied Sciences
Bangalore, India.
Email: jsivaramakrishna003.et.et@msruas.ac.in

Hariharan Ramasangu
M.S.Ramaiah University of Applied Sciences
Bangalore, India.
Email: hariharan.ec.et@msruas.ac.in

***Abstract*—Functional MRI (fMRI) data comprises of a set of trials, each trial is described in terms of a group of 20 to 25 anatomical Region Of Interests (ROI). Each ROI consists of neuroimage sequence information in terms of a set of voxels. Extracting features from ROIs and classifying cognitive states is a challenging task. In this work, average of voxel time horizon for each ROI is considered as an input to the Support Vector Machine (SVM) classifier, which gives the classification performance of each ROI. The proposed method is discussed in the context of ROIs in Starplus fMRI data, and classifier performance of each ROI has been analysed. The ROI, CALC and the combination of ROIs CALC & LIPS provides a good classification accuracy compared to remaining ROIs in Starplus fMRI data. Features formed using second half of time horizon of voxel time series gives a good classification accuracy and it is consistent over six subjects in Starplus fMRI data.**

I. Introduction

Statistical measures, such as mean, variance, mode, and moments have been used to analyse data for decades. All these statistical measures may not be suitable for all kinds of data for analysis. It is depending on the nature of the data, these measures are used for the analysis. Among all these measures statistical mean provides a central tendency for data, in terms a single value. Classification of cognitive states in functional MRI (fMRI) data provides illation about brain data analysis. Brain imaging modalities such as MRI, PET, fMRI are used in determining brain diagnosis, and disorders. fMRI uses functional imaging modality, for diagnosing Alzheimer's disease, brain tumors, Parkinson's disease and epilepsy. fMRI uses $T2^*$ weighted image sequence used to diagnose the brain in a non-invasive fashion. The neural activity in the brain is represented as Blood Oxygen Level Dependent (BOLD) signal. BOLD signal in fMRI is captured at different time instants, while subject performing a specific task. The image sequence in fMRI comprises of BOLD signal intensity at specific time points.These image sequences are represented as voxel time series in the organization fMRI data. The total fMRI data for a particular subject comprises a set of anatomical Regions Of Interest (ROI).

The brain concerned cognition activities can be analysed using machine learning techniques, such as classification, and clustering. The important footfall in the classification of cognitive states in fMRI is creating a constant doorway, depicting the regions that are active for a specific task. A cognitive state analysis of fMRI data includes extraction of voxel time series signal from a specific ROIs. There are few significant grounds for ROI-based analysis, which includes different conjunctions, and presumptions. One is to analyse data thoroughly in complex designs. In complex designs recognizing the pattern of activity from the overall statistical map can be hard. ROI analysis can be useful to see the signal in the orbits of interest, drawn for each one condition. The second is to restrict the statistical run to elite ROIs. The third is to examine the regions the that are functionally well-defined, based on some stimuli. Alternative approach to ROI analysis is to research the fundamental signal hidden in the whole brain voxel-wise analysis. Though ROI analysis helps in an analysing pattern of activations, it can also be helpful to explore the cause for deficiency in activation. Statistical test performed on a specific region or ROI can bring down the hardness of rectification for various examinations, rather than correct a great number of voxels in the entire brain, it is to correct voxels in few ROIs. In this paper features are extracted based on statistical measure on voxels selected from a particular ROI, the combination of few ROIs before building a classifier framework, and the classification accuracy of features extracted from each ROI have been studied.

The rest of the paper consists of, related work in section II. The proposed method for ROI-based analysis for classification of cognitive states has been discussed in section III. The fundamental information on StarPlus fMRI data set has been explained in section IV. The ROI-based analysis of the proposed method on StarPlus fMRI and obtained results for each ROI are discussed in section V, and conclusions for the proposed method have been demonstrated in section VI.

II. Related Work

ROIs can be specified in terms of structural or functional features. Structural ROIs are mostly founded on macro anatomy. ROIs for each subject can be defined based on their own anatomy. Functional ROIs are mostly found on analysis of data from the same person. A separate localizer scan can be used to identify voxels in a specific region, for a specific response. Classification of cognitive states as the subject doing a particular job leaves a substantial amount of information for fMRI data analysis. Cognitive state classification using fMRI image sequence depends on the selection of voxels as features

978-1-5090-3646-2/16/$31.00 ©2016 IEEE

from a specific region. fMRI consists of a number of ROI's and each ROI constitute a great number of voxels. Feature vectors formed using these voxels contribute large dimension vector space. Thus several people explored and aimed for simplification of feature space [1] [2] [3]. Selection of voxels and formation of feature vectors using genetic algorithm have been examined [4]. The computational complexity of voxel selection is greatly reduced, while the size of attributed for classification is still large in size. Voxels selected from all ROIs and processing as features to Gaussian Naive Bayes (GNB) classifier or Support Vector Machine (SVM) classifier contributes over-fitting of the classifier. The problem of overfitting can be defeated by selecting voxels from specific active ROIs. Principle Component Analysis (PCA) can also be used as a solution for overfitting problem. In this case, PCA is applied prior to classifier [5]. Nevertheless, neither of the approache does not consider valid information encoded by the voxels. Hence both approache fall below the optimal level for classification. Space regularization methods have recently been applied on functional neuroimage data. It includes the application of zero weights to irrelevant voxels and considered only relevant voxels. Nevertheless, the approach fails to classify fMRI data, because of spatially spurious classification weights [6]. Spectral analysis of fMRI data produces useful phase information to form feature vectors. Application of Fourier transform or Hilbert transform along with random sieve function produce phase information in fMRI data. GNB classifier framework accuracy attains 93.7%, for the feature vectors formed by phase information in fMRI data, whereas SVM classifier accuracy attains 99% for the same feature vectors [7]. The high-dimensional and noisy fMRI data are considered as a classification problem. For this reason, maximum voxel distance between two cognitive is used for attribute selection. In this method, the voxels in the whole brain are considered for the analysis. The voxel distance in two cognitive states is calculated using statistical mean and standard deviation [8]. The average classification accuracy of 85.83% is achieved for the voxels selected from four noteworthy ROIs in Starplus fMRI data set. The individual ROI classification performance is not examined in the regularized estimation method [9].

From the point of view of high dimensional, noisy, and sparse fMRI data, feature selection and classification seem to have been looked into separately. In this work, it is proposed that statistical mean of voxel time series for an individual and different combination of ROIs have been calculated and a classifier is trained based on these features. The classification performance of each ROI in Starplus fMRI data has been examined.

III. PROPOSED METHOD

The functional block diagram of the proposed Region Of Interest (ROI) based analysis of cognitive state classification is shown in Figure 1. The proposed method is the classification of cognitive states for the statistical mean of voxel time series data as features, and analysing the classification accuracy for the voxels selected from different ROIs. The features are given as input to the SVM classifier. The proposed method provides analysis for the ROIs selected from Starplus fMRI data. Starplus data comprises of fMRI data for six subjects. Each subject consists of 25 ROIs, expressed in terms of a set of trials. Every trial involves a subject performing cognitive task reading a sentence, and viewing a picture. Seven earmarked ROIs as specified in the website [10] are used for the analysis of cognitive state classification.

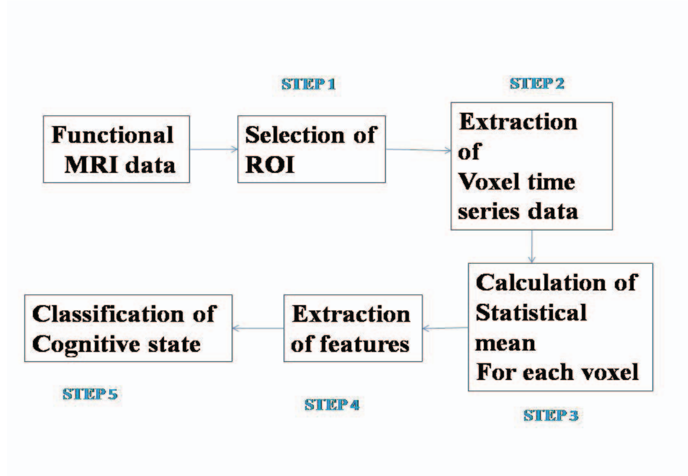

Fig. 1. Proposed ROI based analysis of cognitive state classification

Step 1: In this step, an ROI is selected from the group of ROI's for the classification of cognitive state.

Step 2: Each ROI consists of a set of trials performing a cognitive task, such as reading a sentence or viewing a picture. Each trial in ROI comprises of a set of voxels. Depending on the experimental design each voxel will have definite time points, expressed as time series data for each voxel. In this step, the time series data for the voxels selected from ROI are extracted

Step 3: In this step the statistical mean for the voxel time series data are calculated. In general the activation of voxel time series is delayed, because of the BOLD contrast, caused by stimulus response. Hence the voxel time series is split into two half. Let voxel V consists of L time points, $V = V(1), V(2), ..., V(L)$. The statistical mean for the voxel time series V calculated as described in equations 1 and 2, where m_1 is the statistical mean for the first half of the voxel V time series and m_2 is the statistical mean for the second half of the time series. The mean vector M of the voxel V is obtained by appending m_1 and m_2 i.e M=[m_1 m_2].

$$m_1 = (2/L)\Sigma_{i=1}^{L/2} V(i) \tag{1}$$

$$m_2 = (2/L)\Sigma_{i=(L/2)+1}^{L} V(i) \tag{2}$$

Step 4: In this step features vectors are formed from the mean vector M, obtained in step 3. These feature vectors act as input to the classifier.

Step 5: In this step, the feature vectors generated in step 4, supplied as input to train a classifier. SVM classifier is used for classification of cognitive states.

A. Dataset

StarPlus data was downloaded from publicly available website [10]. The data set describes experiments performed on human subjects in terms of number trials and observations. StarPlus data consists of cognitive fMRI data for six subjects. Subjects were shown a picture ($\frac{+}{*}$) or ($\frac{*}{+}$) and then sentence: The $Plus$ above the $Star$ or The $Star$ above the $Plus$ for the first half of the trials, and the order of the tasks was varied for rest of the trials. Each stimulus last for 8sec, 16 image sequences were captured within that time. The data set comprises of a movie of activation levels in the brain called voxels. Each three-dimensional image consists of approximately 5000 number of voxels. A two-dimensional image of the voxel activation is shown in Figures 3, 4, and 5. Figure 3 describes the snapshot of voxel activation for the 10^{th} time interval,of 5^{th} trial. Figure 4 describes the snapshot of voxel activation for the 20^{th} time interval,of 4^{th} trial. Figure 5 describes the movie of voxel activation for the 4^{th} trial. The StarPlus experiment conducted for 54 trials, during 40 number of trials the subject was performed a cognitive task, and for rest of the trials, the subject was rested or gazed at a fixation point on the screen. Therefore there are 40 trials each for the cognitive task of reading a picture and sentence respectively. In this work, a single trial data for each cognitive task is considered and formed two data sets $dataP$ and $dataS$, where $dataP$ consists of voxel time series for the trial reading picture and $dataS$ consists of voxel time series for the trial reading sentence. The downloaded StarPlus fMRI data was already preprocessed to remove artifacts due to head motion, signal drifts. The data set consists of 25 anatomical Regions of Interests(ROIs) for each subject. The voxels from seven ROIs - CALC, LIPL, LIPS, LOPER, LDLPFC, LT, LTRIA - are used for classification of cognitive states.

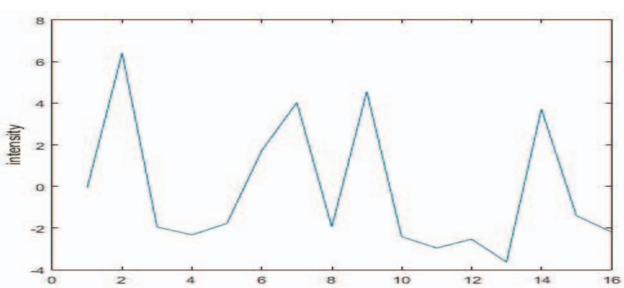

Fig. 2. Voxel time series for the picture task

IV. DISCUSSION

A measure of a statistical mean is a single value that attempts to describe a set of fMRI data time series by identifying the central position in the data set. Although statistical mean has been used to measure central tendency for a data set, their utility in a classification of cognitive fMRI, through the statistical mean of time series fMRI data as features to a classifier is a novel contribution. fMRI data has a definite time series, specified in terms of different time instants. Splitting

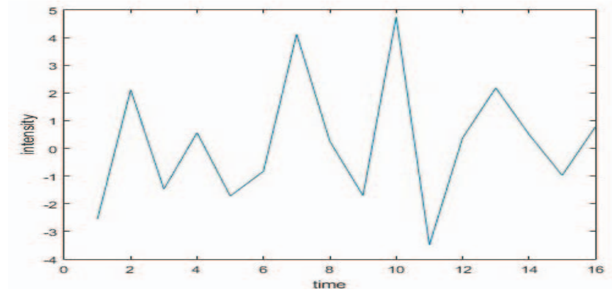

Fig. 3. Voxel time series for the sentence task

the time series into two halves and calculation of mean for the first half and second half time series will provide two features m_1 and m_2 for each voxel, where m_1 is the mean for the first half time series and m_2 is the mean for the second half time series. Statistical mean values of each voxel time series m_1 and m_2 applied as an input to SVM classifier. Starplus fMRI data comprises of 25 ROIs for each subject. Experts have specified seven ROIs: CALC, LIPL, LIPS, LT, LTRIA, LOPER, LDLPFC for classification of cognitive tasks for each subject [10]. In this work the statistical mean values m_1 and m_2 are calculated for the voxels selected from each of the seven ROI's and applied as features to SVM classifier. This method provide a great analysis for the cognitive state classification power of each ROI. SVM classifies test data in a binary linear classifier. SVM model represents test data as points in space, the examples of training data are divided by a clear gap. Test examples are then mapped to same space and predict the category. The main focus of the present work is to analyse the classification power of each ROI and choosing the best ROI combination for cognitive state classification.

Each subject in Starplus fMRI data comprises all the seven ROIs: CALC, LIPL, LIPS, LT, LTRIA, LOPER, LDLPFC in common as specified in the website [10]. The tabulations in Table I,II show the typical classification accuracy for each ROI across different subjects. These classification accuracies are obtained by drawing features from statistical mean values m_1 and m_2. It is observed that ROI- CALC, gives better classification accuracy compare to other six ROIs. It provides maximum accuracy 98.75% for the subjects 04847, 04799, 05680, and it is 97.5% for the subject 97.5. It is 92.5%, 85% for the subjects 05675,04820 respectively. The classification accuracy for SVM classifier for the features drawn from statistical mean m_2 alone is shown in Table III, IV. It is observed that the classification accuracy of the ROI-CALC is far better across all the subjects, compare to other ROIs. It is also observed that the classification accuracy for the features selected from statistical mean m_2 alone, for the voxels present in ROI-CALC is increasing 2 to 5 percent depending on the subjects. The tabulations in Table V shows the classification accuracy for the voxels selected from the ROIs CALC and LIPS. C1 represents accuracy stands for the features generated using statistical mean m_1 and m_2, and C2 represents the accuracy stands for the features generated using statistical mean m_2 alone. It is observed that the classification accuracy stands in C2 are consistent for the subjects 04847, 04799, and 05680 with 100% accuracy. Various combinations of ROIs used for classification of cognitive states, out of them the two ROI combination CALC and LIPS provides better classification accuracy. C3 and C4 show the accuracy stands for the voxels selected from ROI's CALC, LIPS, LIPL, and LOPER. C3 is the accuracy for features generated using mean m_1 and m_2, C4 is the accuracy for features generated using mean m_1. It is observed that accuracy stands in C4 are better than C3. C5 and C6 show the accuracy stands for the complete data set(includes voxels in all ROIs). C5 and C6 are the accuracy for features generated using mean m_1 and m_2, and mean m_2.

The tabulations in Table XI and XII show the accuracy stands for the voxels selected randomly, for various combination of ROI's. C7 and C8 are the accuracy stands for 25% of a total number of voxels selected randomly from the ROI combination CALC, LIPS, LIPL,LOPER. C7 and C8 are the accuracy for features generated using mean m_1 and m_2, mean m_2 respectively. Similarly, C9 and C10 are the accuracies for 50% of a total number of voxels selected randomly from the same ROI combination. C11 and C12 are the accuracies stands for 25% of voxels selected randomly from ROI combination CALC and LIPS. C11 is for the features generated using mean m_1 and m_2. C12 is for the features generated using mean m_2. Similarly, C13 and C14 are the accuracies stands for 50% of voxels selected randomly from the same combination of ROI. The tabulations in Table XIII show the average classifiers for various features, where R1 is the ROI combination {CALC,LIPS}, and R2 is the ROI combination {CALC,LIPS,LIPL,LOPER}.

The above discussion provides a detailed analysis for the discriminating ROI's used to classify cognitive tasks present in fMRI data.

Table I: Classifier accuracy for features generated using statistical mean m_1 and m_2 for the specified ROI.

SUBJECT	CALC	LIPL	LIPS	LT
04820	85	50	69.75	67.5
04847	98.75	63.75	86.25	73.75
04799	98.75	50	58.75	53.75
05675	92.5	55	70	70
05680	98.75	48.75	63.75	78.75
05710	97.5	63.75	56.25	77.5

Table II: Classifier accuracy for features generated using statistical mean m_1 and m_2 for the specified ROI.

SUBJECT	LTRIA	LOPER	LDLPFC
04820	57.5	41.25	57.5
04847	51.25	77.5	70
04799	65	52.5	47.5
05675	61.25	77.5	63.75
05680	57.5	725	60
05710	63.75	52.5	60

Table III: Classifier accuracy for features generated using statistical mean m_2 for the specified ROI.

SUBJECT	CALC	LIPL	LIPS	LT
04820	88.75	50	73.75	71.25
04847	100	62.5	65	67.5
04799	98.75	53.75	60	53.75
05675	97.5	53.75	83.75	68.75
05680	100	48.75	58.75	86.25
05710	98.75	52.5	65	81.25

Table IV: Classifier accuracy for features generated using statistical mean m_2 for the specified ROI.

SUBJECT	LTRIA	LOPER	LDLPFC
04820	52.5	43.75	58.75
04847	36.25	67.5	58.75
04799	68.75	40	50
05675	63.75	65	62.5
05680	60	65	56.25
05710	63.75	38.75	56.25

Table V: Description of average classifiers and corresponding feature formation.

Classifier	Formation of features
C1	m1&m2 from ROI combination R1
C2	m2 from ROI combination R1
C3	m1&m2 from ROI combination R2
C4	m2 from ROI combination R2
C5	m1&m2 from whole data set
C6	m2 from whole data set
C7	m1&m2 from 25% of voxels drawn from R2
C8	m2 from 25% of voxels drawn from R2
C9	m1&m2 from 50% of voxels drawn from R2
C10	m2 from 50% of voxels drawn from R2
C11	m1&m2 for 25% of voxels drawn from R1
C12	m2 for 25% of voxels drawn from R1
C13	m1&m2 for 50% of voxels drawn from R1
C14	m2 for 50% of voxels drawn from R1

Table VI: Classifier accuracy for features generated using statistical mean for the voxels selected from the ROI combination CALC and LIPS.

SUBJECT	C1	C2
04820	82.5	92.5
04847	98.75	100
04799	98.75	100
05675	97.5	97.5
05680	97.5	100
05710	96.25	97.5

Table VII: Classifier accuracy for features generated using statistical mean for the voxels selected from the ROI combination CALC, LIPS, LIPL, and LOPER.

SUBJECT	C3	C4
04820	82.5	88.75
04847	98.75	100
04799	98.75	100
05675	97.5	97.5
05680	96.25	97.5
05710	95	95

Table VIII: Classifier accuracy for features generated using statistical mean for the complete data set.

SUBJECT	C5	C6
04820	88.75	92.5
04847	97.5	100
04799	93.75	98.75
05675	93.75	97.5
05680	91.25	95
05710	93.75	96.25

Table IX: Classifier accuracy for features generated using statistical mean for the voxels selected randomly from the ROI combination CALC, LIPS, LIPL, and LOPER.

SUBJECT	C7	C8	C9	C10
04820	79.25	79.38	81.62	85.5
04847	98.87	98.63	98.75	99.62
04799	88.75	91.25	93.75	95
05675	94	95.75	95.88	96.62
05680	91.63	92.63	93.25	94.25
05710	93.25	93.25	94.75	94.75

Table X: Classifier accuracy for the features generated using statistical mean for the voxels selected randomly from the ROI combination CALC, and LIPS.

SUBJECT	C11	C12	C13	C14
04820	82	81.88	83.87	83.75
04847	98.75	98.63	98.88	99.5
04799	93	93.25	96	97.12
05675	93.5	93.88	95	96
05680	95	97	96.25	97.37
05710	94	94.87	95.88	96.5

Table XI: Classification accuracy as reported in [9]

Subject	% Accuracy
04799	80
04820	95
04847	85
05675	95
05680	70
05710	90

Comparing Table VI and XI. the following observations are made:

For the subjects 04847, 04799, 05675, 05680, 05710, the proposed method gives 100%, 100%, 97.5%, 100%, 97.5% classification accuracy respectively. These classification accuracy are much better than results presented in [9]. The classification accuracy for the subject 04820 is 92.5%, which is almost close to [9].

V. Conclusion

A novel feature selection technique of mixed mean method has been proposed. The voxel time series has been split into two half and calculated mean for both the time series. The mean values of time series have been used to train SVM classifier. The classification performance of seven noteworthy ROIs in Starplus fMRI data have been analysed. The results show the mean values of second half of the time series for the voxels selected from the ROI-CLAC and combination of ROI-CALC,LIPS achieve good classification accuracy.

References

[1] T. M. Mitchell, R. Hutchinson, M. A. Just, R. S. Niculescu, F. Pereira, and X. Wang, "Classifying instantaneous cognitive states from fmri data," in *American medical informatics association annual symposium*, 2003.

[2] M. Palatucci, "Temporal feature selection for fmri analysis," Working paper (unpublished), Tech. Rep., 2007.

[3] V. Michel, C. Damon, and B. Thirion, "Mutual information-based feature selection enhances fmri brain activity classification," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*. IEEE, 2008, pp. 592–595.

[4] S. Parida, S. Dehuri, and S.-B. Cho, "Application of genetic algorithms and gaussian naïve bayesian approach in pipeline for cognitive state classification," in *Advance Computing Conference (IACC), 2014 IEEE International*. IEEE, 2014, pp. 1237–1242.

[5] T. A. Carlson, P. Schrater, and S. He, "Patterns of activity in the categorical representations of objects," *Journal of Cognitive Neuroscience*, vol. 15, no. 5, pp. 704–717, 2003.

[6] B. Ng, A. Vahdat, G. Hamarneh, and R. Abugharbieh, "Generalized sparse classifiers for decoding cognitive states in fmri," in *Machine Learning in Medical Imaging*. Springer, 2010, pp. 108–115.

[7] H. Ramasangu and N. Sinha, "Cognitive state classification using transformed fmri data," in *Signal Processing and Communications (SPCOM), 2014 International Conference on*. IEEE, 2014, pp. 1–5.

[8] N. Alamdari, S. Akbari, and E. Fatemizadeh, "Unsupervised versus supervised methods for categorizing mental states from fmri data," *Journal of Medical Devices*, vol. 9, no. 2, p. 020949, 2015.

[9] X. Chen, M. Xu, and W. B. Wu, "Regularized estimation of linear functionals for high-dimensional time series," *arXiv preprint arXiv:1506.03832*, 2015.

[10] M. Just and T. Mitchell, "Starplus fmri data," *URL http://www. cs. cmu. edu/afs/cs. cmu. edu/project/theo-81/www*, 2001.