Classifying Instantaneous Cognitive States from fMRI using Discriminant based Feature Selection and Adaboost

Tien Duong Vu*, Hyung-Jeong Yang**, Luu Ngoc Do*, Thao Nguyen Thieu*

Abstract

In recent decades, the study of human brain function has dramatically increased thanks to the advent of Functional Magnetic Resonance Imaging. This is a powerful tool which provides a deep view of the activities of the brain. From fMRI data, the neuroscientists analyze which parts of the brain have responsibility for a particular action and finding the common pattern representing each state involved in these tasks. This is one of the most challenges in neuroscience area because of noisy, sparsity of data as well as the differences of anatomical brain structure of each person. In this paper, we propose the use of appropriate discriminant methods, such as Fisher Discriminant Ratio and hypothesis testing, together with strong boosting ability of Adaboost classifier. We prove that discriminant methods are effective in classifying cognitive states. The experiment results show significant better accuracy than previous works. We also show that it is possible to train a successful classifier without prior anatomical knowledge and use only a small number of features.

Keywords : cognitive states|Adaboost|Fisher Discriminant Ratio|hypothesis testing|classification.

I. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) is a brain imaging technique which uses magnetic resonance imaging to measure the changes in local oxygenation blood related to the brain activity. Blood oxygen level dependent (BOLD) is used in fMRI as an indicator of neural activity changing in voxel intensity values. Because of the higher spatial resolution than any other earlier techniques, fMRI is the best technique for observing human brain activity that is available currently.

Advantages of fMRI recently open the dramatic development of human brain analysis in the neuroscience area. The final goal of its applications is not only for treating psychological diseases but also for simulating the logistic structure of human brain for machine leaning. Many tools and frameworks have developed for supporting the neuron signals analysis and related issues.

Traditional, fMRI technology has been used to detect what areas in the brain raise neural activity responses when a subject does specified actions, called localizing.

Another fMRI data analysis direction is classification, whose main goal is detecting the patterns of neural activity and determine the way of mapping them onto cognitive states. This is the most challenge problem in neuroscience, because of variety about anatomical structure among different people and the high dimensional feature with thousands of voxels. These critical problems cause difficulties for mining neural data unless there are appropriate preprocessing and reduced dimensional methods.

Therefore, the feature selection becomes the most important key for the performance of classifier. A good selection measurement helps to choose the most informative features and remove irrelevant features, so that the accuracy is increased and the processing time decreases significantly.

In this paper, we propose Fisher Discriminant Ratio and hypothesis testing as effective feature selection methods. Adaboost is used as classifier aiming to boost classifying accuracy as well as optimizing processing time. Different to Michell et al [1], which rounds up activity-based feature selection methods, our work takes advantage of naturally discriminant characteristics of the

* This study was financially supported by Chonnam National University, 2014.

*Student Member, **Member: Dept. of Electronics and Computer Engineering, Chonnam National University

Manuscript : 2015.1125
Revised : 2016.03.22
Confirmation of Publication : 2016.03.23

Corresponding Author : Hyung-Jeong Yang,
e-mail : hjyang@jnu.ac.kr

cognitive state classes. The experiment results show high accuracy within short processing time and require a relatively small number of features.

The remained parts of this paper are organized as follows: section II summarizes outstanding related works, section III describes the proposed method. In section IV, we present details of experiments and show the results. Finally, we conclude the result and future work in section V.

II. RELATED WORK

In the past, a variety of approaches is conducted to analyzing fMRI data. Bly et al. [2] used Generalized Linear Models (GLM) to predict voxel activities given the stimulus. Hidden Markov Model (HMM) was used by Hojen-Sorensen et al. [3] to learn a model of activity resulting from stimulus of flashing light. Cox and Savoy [4] applied Linear Discriminant Analysis and Support Vector Machine to classify successfully patterns of fMRI activation when subjects saw photographs. Wagner et al. [5] reported that predictions have been made better than random if a visually presented word will be remembered later.

Mitchell et al. [6] performed experiments to understand human cognitive states. Mitchell and his colleagues used feature selection approaches based on voxel activities and applied machine learning classifiers to analyze received fMRI data. Also on Mitchell's dataset, Hoang et al. [7] applied incremental principal component analysis (iPCA) as an efficient feature extraction method. The advantage of this method was not requiring domain experts to select Regions of Interest (ROIs). Do et al. [8] used Fisher discriminant ratio (FDR) to select most active voxels. This method showed high accuracy but required a relatively long processing time.

Adaboost is a meta-algorithm. It can be used in conjunction with many other types of learning algorithms to improve their performance. This flexible integration makes widespread application of Adaboost in many domains [9]. In this paper we propose using Adaboost classifier on the most discriminant voxels, obtained by choosing voxels through FDR or hypothesis testing. We also select the best ROIs by considering its effects on accuracy and processing time. Because the main goal of this paper is to demonstrate effectiveness of the approach without domain knowledge, ROIs selection is considered as optional step.

III. PROPOSED METHOD

In this section we describe the proposed method for classifying cognitive states. We introduce two discriminants based on feature selection methods: Fisher Discriminant Ratio and hypothesis testing. Then we present Adaboost algorithm and Regions of interest.

A. Fisher Discriminant Ratio

Fisher discriminant ratio (FDR) is used to measure the discriminatory power of individual features between two classes. FDR value is defined as:

$$FDR = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2} \quad (1)$$

where m represents mean, s^2 represents a variance, and the subscripts denote two classes. FDR is one of the best feature selection methods because it considers both mean and variance – two most important characteristics of a distribution of samples. The high FDR value maximizes the distance between the means of the two classes while minimizing the variance within each class. Hence, if a feature has high FDR value, its values are absolutely different in each class. Therefore, this feature has a discrimination power to classify classes. The higher FDR also means the more powerful feature. In our study, we choose n features having the highest FDR values from whole voxels on each fMRI image.

B. Hypothesis testing

The first step in feature selection is to look at each feature individually and check whether or not it is an informative one. If not, the feature is discarded. The idea is to test whether two mean values that a feature has in two classes differ significantly. Assuming that the data in the classes are normally distributed, the so-called t-test is a popular choice. [13]

The goal of the statistical t-test is to determine which of the following two hypotheses is true:

- H1: The mean values of the feature in the two classes are different.
- H0: The mean values of the feature in the two classes are equal.

The first is known as the alternative hypothesis

(the values in the two classes differ significantly); the second, as the null hypothesis (the values do not differ significantly). If the null hypothesis holds true, the feature is discarded. If the alternative hypothesis holds true, the feature is selected. The hypothesis test is carried out against the so-called significance level, ρ, which corresponds to the probability of committing an error in the decision. Typical values used in practice are ρ = 0.05 and ρ = 0.01. Depend on the how small of ρ, the number of features satisfied the alternative hypothesis also is different, i.e. smaller ρ, less features chosen. Therefore, the processing time also decreases corresponding.

C. Adaboost

Adaboost was discovered by Yoav Freund and Robert Schapire in 1995 [10]. The theory of Adaboost is: the output of the other learning algorithms is combined into a weighted sum that represents the final output of the boosted classifier. While every learning algorithm will tend to suit some problem types better than others, and will typically have many different parameters and configurations to be adjusted before achieving optimal performance on a dataset, AdaBoost (with decision trees as the weak learners) is often referred to as the best out-of-the-box classifier. [9]

We use original AdaBoost for binary classification, because of the simplicity, effective processing time and high classifying accuracy. Adaboost pseudocode is given below.

Adaboost algorithm

Given: $(x_1, y_1),...,(x_m, y_m)$

where $x_i \in X, y_i \in Y = \{-1,+1\}$

Initialize $D_1(i) = 1/m$

(2)

For $t=1,...,T$

- Train weak learner using distribution D_t.
- Get weak hypothesis $h_t : X \to \{-1,+1\}$ with error $\varepsilon_t = \Pr_{i-D_t}[h_t(x_i) \neq y_i]$

(3)

- Choose $\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$

(4)

- Update

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{if } h_t(x_i) \neq y_i \end{cases}$$

$$= \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

(5)

where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

Output the final hypothesis:

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \quad (6)$$

D. Regions of interest

Most neuroscientists suspect that information in the brain is stored in the patterns of activity across groups of neurons. Therefore, the whole brain is anatomically divided into many locations called ROIs for accessing mental tasks easily. We followed Mitchell et al. [1] to mark up the brain with 25 anatomical ROIs.

In order to create these ROIs, Mitchell et al. used structural images that capture the static physical brain structure at high resolution. For each subject, this structural image was used to identify the anatomical ROI, using the parcellation scheme of Rademacher et al. [11]. Then, the mean of fMRI images was co-registered to the structural image. Hence, individual voxels in fMRI images could be associated with the ROIs identified in the structural image.

It is generally preferable to perform classification using all voxels in each ROI, as this does not restrict the classifiers to specific spatial patterns. However, including all voxels may make analysis difficult since anatomical structures vary in size. ROIs size is critical for a classifier's performance since it is not good if there are many more dimensions than examples [12]. In this paper, we consider both cases when we include all voxels of ROIs for classifier and just select several ROIs which have the best performance. The main goal of using ROIs in our study is for processing time reduction.

IV. EXPERIMENT

A. Data description

We used StarPlus dataset collected by Michell et al.[1] for validation. Starplus included fMRI data collected many times for each human subject performing a set of trials. During each trial they were shown in sequence a sentence and a simple picture, then answered whether the sentence correctly described the picture. In half of trials, the sentence is presented first, following by a picture. In the remaining trials, the order of presenting stimulus is inversed. This experiment is also called 'sentence versus picture' experiment. In either case: a subject sees the first stimuli (sentence or picture) for 4 seconds, followed by a blank screen for 4 seconds. the second stimuli (picture or sentence) is presented for next 4 seconds, during which the subject must press a button for "yes" or "no", depending on whether the sentence correctly describes the picture seen or not. Finally, a rest period of 15 seconds is inserted before next trial begins. Therefore, each trial lasts approximately 27 seconds. Snapshots are made every 1/2 seconds. Thus, each trial involves 54−55 snapshots. Pictures are geometrically arrangement of simple symbols +,* and $, such as

$$\frac{+}{*}$$

Sentences are descriptions such as "It is true that the plus is below the dollar." Half of the sentences are negated (e.g., "It is not true that the star is above the plus.") and the other half involves affirmative sentences.

The learning task is to train a classifier to determine, given a particular 8−second interval of fMRI data, with probability, whether a subject is looking at a picture or a sentence. In other word, the expected individual classifier for each subject is as the following form:

f: fMRI − sequence(t, t+8) −> {Picture, Sentence} (7)

where t is the starting time of stimulus. Although the maximum duration of each stimuli presented is 4 seconds, but it's necessary to choose 8−second interval in order to capture the full fMRI activity associated with the stimulus. Because fMRI BOLD signal often extends for 9–12 seconds beyond the neural activity of interest. [1]

There are a total of 80 examples from each subject (40 examples per class). The average number of voxels is approximately 10,000 per subject, eight second interval contains 16 images, thus, the total features of an example can reach to 160,000![1] This is a huge number, so if there isn't an efficient feature selection method, the computation time is so large, especially when we use sophisticated algorithms.

B. Evaluation

1. Evaluating sentence/picture classification performance on each single subject using FDR and t−test

For evaluating the classification, we use k−fold cross validation with k=10. The average accuracy was computed and compared to other methods such as iPCA [7], ROIs [8], ActiveFDR [8] and the methods in Mitchell et al. [1] including Discrim, Active, roiActive and roiActiveAvg. Our proposed feature selection methods were described as follow :

- FDR: using FDR to choose n highest FDR features from all features of that example, where n is 30, 50 or 100.
- ROIs+FDR: first we choose voxels from seven best ROIs as in [1], then use FDR to choose n highest FDR features from the features generated by voxels chosen in previous step.
- ttest: using t−test calculated as in section III with $\rho = 0.01$
- ROIs+ttest: similar to ROIs+FDR, first we choose ROIs, then using t−test to decide whether a feature is chosen or rejected for features generated from previous chosen voxels, with $\rho = 0.01$

In figure 1, we show the detailed results of using 100 highest FDR features. It is the best selection (with best accuracy) among 30, 50 and 100. We see that using "All features" shows the better accuracy than random threshold (50%), but it's still the worst case, because no feature selection method is applied, so many irrelevant features used in training.

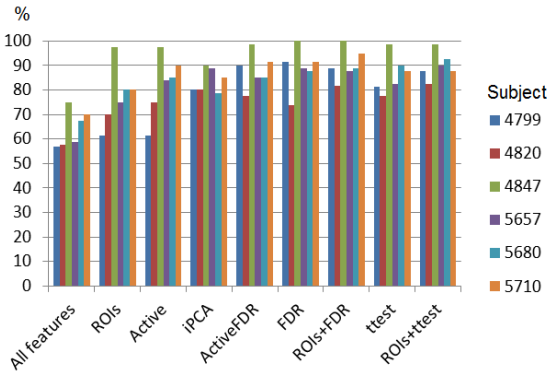

Fig. 1. Classification results for each subject

In general, we easily realize that the result of subject 4820 is always the lowest, even less much than the average, and subject 4847 is the highest – near absolutely 100%. This result is consistent with the fact that data of subject 4820 still involve high noisy rate and missing values after preprocessing, while 4847 data is the best preprocessed data among six subjects. Figure 2 shows the average accuracy over six subjects of proposed method and compared methods.

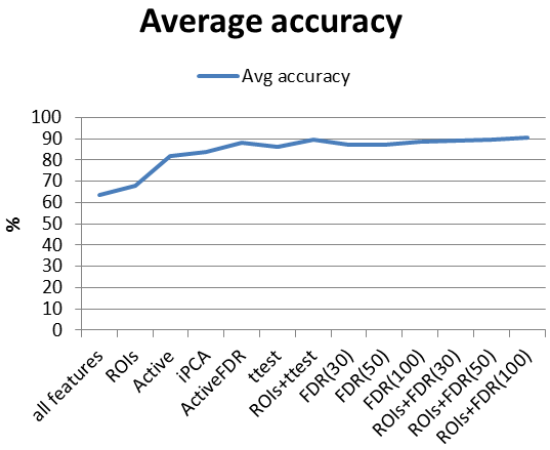

Fig. 2. Comparison of average accuracy cross all subjects of methods

ROIs and Active are two methods using anatomical knowledge (to determine what voxels belong to a particular ROI), so not only the number of required features decreases many times, but also the accuracy is improved significantly. In ROIs method, the features are generated by using the best ROIs in which still include irrelevant features, so the accuracy is lower than 70%. In Active method, the t-test between trials with stimuli and fixation trials helps selecting the most "outstanding" features, which are interpreted in neuroscience as voxels. They have clearly changing of BOLD values among three states: reading a sentence, watching a picture and completely relaxing. Therefore, the reduction of almost irrelevant features improves accuracy significantly.

iPCA has advantages of no domain knowledge required, not as two above methods, and provides relatively good results. The characteristic of a neural activity dataset can be the main reason makes iPCA not optimize when used in here.

Compared to ActiveFDR method – the best method in methods mentioned, our proposed strategies without ROIs selection provide accuracy approximately equal. When using additional ROIs selection, the accuracy is improved to 90.42%. These results show better effective classifying of Adaboost's methods than methods using Naïve Bayesian as classifier.

Figure 3 shows the average processing time of ActiveFDR and our proposed methods. Gathering statistics from figure 2 and 3, we see the correlation between the changes in accuracy and processing time thanks to select ROIs.

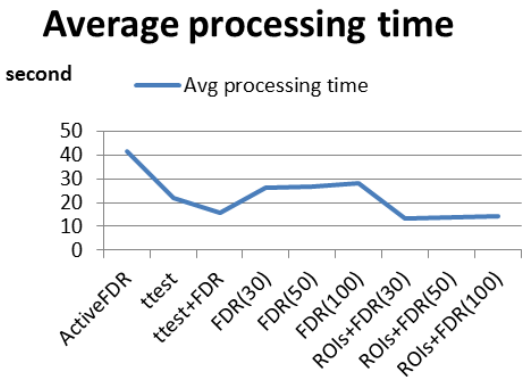

Fig. 3. Comparison of average processing time over all subjects of methods

In our proposed methods, using ROIs always makes the improvement on both accuracy and time. Especially in processing time aspect, our methods can help decreases time reach to 4 times compared to ActiveFDR, in both t-test and FDR cases. Choosing 30, 50 or 100 highest FDR features doesn't affect much processing time, but has a bit better in accuracy. We use only 30 features for classifying to get the accuracy 89%, this is meaningful, while we have to use 100, 240 and 250 for earlier methods. In addition, the best result without ROIs is 88.75% prove that discriminant

based on feature selection method is really effective without requiring domain knowledge. This is opposed to Michell's comment [1], in which he said that the discriminant methods are ineffective.

2. Evaluating individual PS dataset and SP dataset and new feature selection strategy.

We knew that there are two types of experiments distinguished by the order of stimulus: Sentence vs. Picture (SP) trials and Picture vs. Sentence (PS) trials. fMRI snapshots belong to these trials are organized to SP dataset and PS dataset respectively. We have conventions that fMRI snapshots taken when a subject is looking a picture in PS dataset called P2, and for sentence, it called S2. By a similar way, on SP dataset, we have S3 and P3 parts.

In this section, we consider sentence/picture classification of each pairs P2−S2, S3−P3 (pairs belong to same trial), P2−S3 and S2−P3 (pairs which have same time window in each trial). We compared the results of following methods: Active, ActiveFDR, ROIs+FDR and ROIs+ttest. They are showed in figure 4 as below.

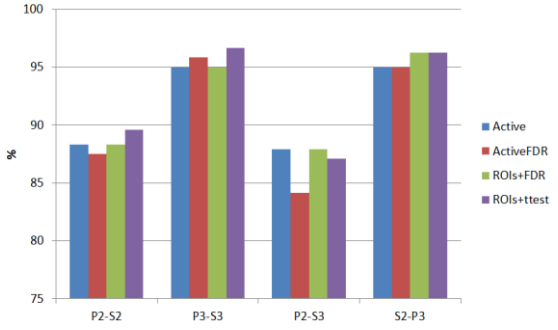

Fig. 4. Results of Picture vs. Sentence and Sentence vs. Picture studies

As shown in figure 4, the accuracies among considered methods are not different much, just 1−2%. ROIs+ttest shows the best performance, once again proves the effectiveness of discriminant feature selection strategies. This result absolutely differs to results in figure 1, for single subject, when the gap between Active and ROIs+FDR reaches to approximately 8%. We stated that the characteristic of each subset expressed in its study makes differences. For example, the S2 subset is somehow different to the S3 subset, because S2 taken given picture stimuli before, in same "picture vs. sentence" trial, but no prior stimuli in the case of S3. Thus, the discriminant of features as well as the active level of voxels is exhibited better in a particular study (i.e S3−P3 or S2−P3) for classifying than the mixed study. (Noted that the classification task in previous section B.1 is essentially performed on sentence set, involved S2, S3 and picture set, involved P2, P3. Therefore, It is the reason why it's considered as a mixed set.)

An another interesting point is the gaps between picture vs. sentence study P2−S2 and sentence vs. picture study S3−P3, about 7−8%, as well as between the first stimuli comparison P2−S3 and the second stimuli comparison S2−P3, also approximately 7−8%. Inspired by this important point, we proposed a new feature selection strategy based on FDR value so that the good discriminations of subsets, i.e. P2−S2 and S3−P3, help improving the discrimination of sentence/picture patterns of a single subject, i.e. S2&S3 − P2&P3. We assumed the weight of each subset is not equal, so the portion selected features followed each subset can be different.

For example, to select 7 highest discriminant features for classifying sentence (S2&S3) and picture (P2&P3) we choose the index of the highest FDR features {1, 2, 4, 5, 6} of pair P2−S2 and {6,9} of pair S3−P3. Then we intersect these 2 sets. The final feature index set chosen is {1, 2, 4, 5, 6, 9} with the contributed rate from two pairs is 5/2. Noted that we accepted overlap features, so the total number of features can be less than 7.

We performed on many rates of portion, from 0/10, 1/9, 2/8... to 9/1, 10/0 of each pair, in total 100 feature indexes from two pairs. Tables 1, and 2 show the performance of this selected feature strategies.

Surprisingly, this feature selection method is really effective, the accuracy is approximately 95%. This very high results confirm that the discrimination of each two sub sample set can contribute to the final discrimination of two full sample sets, boosting significantly the accuracy from 90.42% (ROIs+FDR) to 95.21%. The ideal rate is 3/7 or 4/6 when the majority part is from the pair which has better individual classification accuracy (i.e. S2−P3, S3−P3). This result affirmed again that the powerful ability of discriminant strategies on classification task, while activity based feature

selection methods can't over its highest threshold about 87%.

Together with the improvement of time processing, Adaboost also expressed as a good choice for classifier in this classification task.

Table 1. Classification results corresponding to different ratings of P2−S2 and S3−P3

Number of features from		Classification accuracy
P2-S2	S3-P3	
100	0	86.25%
90	10	87.08%
80	20	88.75%
70	30	90.42%
60	40	92.50%
50	50	92.50%
40	60	**95.83%**
30	70	92.71%
20	80	93.54%
10	90	90.83%
0	100	82.29%

Table 2. Classification results corresponding to different ratings of P2−S3 and S2−P3

Number of features from		Classification accuracy
P2-S3	S2-P3	
100	0	86.25%
90	10	89.38%
80	20	91.04%
70	30	92.92%
60	40	93.33%
50	50	95.00%
40	60	94.58%
30	70	**95.21%**
20	80	93.33%
10	90	91.88%
0	100	86.88%

V. CONCLUSION

In this paper, an effective method for classifying cognitive states of single subjects from fMRI is presented. By selecting the most discriminant features, together with strong boosting from Adaboost, the results showed a good performance with very high accuracy, shorter processing time, and a smaller number of selected features compared to other methods. In addition, a new feature selection strategy with contribution of the discrimination of sub sample sets improves dramatically the total accuracy of classifying sentence/picture on each subject. Our work also asserts the efficiency of discriminant based feature selection methods such as Fisher Discriminant Ratio and hypothesis testing. In addition, we don't need to choose Regions of Interest, which is related to anatomical neuroscience knowledge.

In future, we will research deeper about the other variants of Adaboost in human cognitive state researches. Morever, tensor decomposition in 3D fMRI data is a potential trend which can apply efficiently to feature selection problem. We expect that combination of tensor operators and Adaboost might improve performance of data analyzing on multi−dimension fMRI data.

References

[1] T. Mitchell, R. Hutchinson, M. Just, R.S. Niculescu, F. Pereira, X. Wang. "Classifying Instantaneous Cognitive States from fMRI Data", American Medical Informatics Association Symposium, October 2003.

[2] B.M. Bly, "When you have a General Linear Hammer, every fMRI time−series looks like independent identically distributed nails", Concepts and Methods in NeuroImaging Workshop, 2001

[3] Hojen−Sorensen, L.K. Hansen and C.E. Rasmussen, "Bayesian modeling of fMRI time series", Proc. Conf. Advances in Neural Information Processing Systems, NIPS, 1999, pp 754−760

[4] Cox, D.D., Savoy, R.L.: "Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex." NeuroImage 19, 2003, pp 261−270

[5] Wagner, A. D. et al. "Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity." Science, 281, 1998, pp 1188−1191.

[6] T.M. Mitchell, R. Hutchinson, R.S. Niculescu, F.Pereira, X. Wang, M. Just, and S. Newman "Learning to Decode Cognitive States from Brain Images" Machine Learning, Vol. 57, Issue 1−2, October 2004. pp. 145−175.

[7] M.T.T.Hoang, Y.G.Won and H.J.Yang, "Cognitive States Detection in fMRI Data Analysis using incremental PCA", ICCSA, 2007. pp.335-341.
[8] Luu-Ngoc Do, Hyung-Jeong Yang, "Classification of Cognitive States from fMRI data using Fisher Discriminant Ratio and Regions of Interest" International Journal of Contents, Vol.8, No.4, pp.55-62, 2012.12
[9] Michael Collins, Robert E. Schapire, Yoram Singer. "Logistic regression, AdaBoost and Bregman distances" Machine Learning, Vol.48 July 2002 Issue 1, pp 253-265
[10] Freund, Y. and R. E. Schapire. "A Decision -Theoretic Generalization of On-Line Learning and an Application to Boosting." Journal of Computer and System Sciences, Vol. 55, pp. 119–139, 1997.
[11] Rademacher, J., Galaburda, A.M., Kennedy, D.N., Filipek, P.A., Caviness, V.S.: Human celebral cortex: localization, parcellation, and morphometry with magnetic resonance imaging. J. Cogn. Neurosci. 4, 352–374 (1992)
[12] Etzel, J.A., Gazzola, V., Keysers, C.: An introduction to anatomical ROI-based fMRI classification analysis. Brain Res. 1282, 114–125 (2009)
[13] Sergios Theodoridis, Konstantinos Koutroumbas: Pattern Recognition, 273-274 (2009)

Authors

Duong-Tien Vu

He received his B.E. degree in Electrical & Electronics Engineering from Ho Chi Minh University of Technology (HCMUT) in 2014. He is currently a M.S student at Dept. of Electronics and Computer Engineering, Chonnam National University, South Korea. His main research interests include Bioinformatics and Pattern Recognition.

Hyung-Jeong Yang

She received her B.S., M.S. and Ph. D. from Chonbuk National University, Korea. She is currently an associate professor at Dept. of Electronics and Computer Engineering, Chonnam National University, Gwangju, Korea. Her main research interests include multimedia data mining, pattern recognition, artificial intelligence, e-Learning, and e-Design.

Luu-Ngoc Do

He received the B.S from Chonnam National University, South Korea in 2011. He is currently a PhD student at Dept. of Electronics and Computer Engineering, Chonnam National University, South Korea. His main research interests include Data Mining, Pattern Recognition, Machine Learning and Bioinformatics.

Thao-Nguyen Thieu

She received her B.E. degree in Faculty of Math and Computer Science, Ho Chi Minh City University of Science (HCMUS) in 2012. She is currently a M.S student at Dept. of Electronics and Computer Engineering, Chonnam National University, South Korea. Her main research interests include Bioinformatics, Data Mining and Pattern Recogniti