Impact of Sampling Features on EEG Classification

Ala Alam Falaki[†] **and Seyyedeh Hoora Fakhrmoosavy**[††]

[†]Department of Computer Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran
[††]Department of Computer Engineering, Islamic Azad University, Lahijan Branch, Lahijan, Iran

Summary
In this paper, a sampling method is going to be introduce for classifying Electroencephalogram (as known as EEG) signals. This method consists of three steps. (i) Reduce EOG artifacts. (ii) Calculating the bandpower of the signal. (iii) Finding the best time segment of features with highest classification accuracy in the range of bandpower peaks. The Butterworth algorithm used for feature extraction and the classification accuracy measured by Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Naïve-Bayes (NB) and Random Forest (RF) algorithms. For the experiment, dataset 2b from BCI competition IV that recorded in 3 channels for motor imagery tasks were studied, two different mental tasks are examined for each subject in two class labels for right- and left- hand movement.

Key words:
EEG, Motor Imagery, Hand Movement, band power, Sampling

1. Introduction

The EEG signal (also known as brain waves) is a technique that captures brain's electrical activity. In more details, multiple channels of EEG are recorded simultaneously from various locations on the scalp for comparative analysis of activities in different regions of the brain [1]. EEG signal is viable for detecting motor imagery tasks e.g. hand movement.

Based on world health organization (WHO) report on 2016, about 15% of the world population deal with some sort of disability [2], whilst there are some forms of disabilities that have nothing to do with how brain functions e.g. spinal cord injuries and amyotrophic lateral sclerosis (ALS). Developing brain computer interface (BCI) devices like artificial hand [3] and mind controlled wheelchair [4] will have a huge impact on lots of people living, especially in less developed countries.

There are massive numbers of researchers available that exploring different aspects of EEG signal processing, Siuly et al. presented a sampling technique that extracts features from signals using statistical concepts like mean, max, min, etc. [5, 6]. Robinson et al. used common spatial pattern (CSP) to provide spatial spectral information. Combination of CSP and linear discriminant analysis (LDA) classifier led to identifying hand movement speed in two fast/slow classes [7].

In this article will discuss if the proposed method for sampling EEG signals going to have a notable effect on classification accuracy compares to apply classification on raw features vector (Features without sampling). EEG signal can be described as time-varying spectral analysis, The Butterworth IIR filter is suitable for translating signals to a selected range of flatted frequency and calculate each channel's bandpower. Also, the extracted features can be divided into segments based on recorded signal bandpower peaks and sample rate in each trial length. The results will be compared amongst four different classification algorithms, Linear Discriminant Analysis, Support Vector Machine, Naive Bayes and Random Forest.

The rest of paper is organized as follows: in the next section will discuss about the dataset in more details, in section 3, going to describe what algorithms were used in the experiment, after that in section 4 specify how these algorithms used and describe the proposed method in more details, section 5 will share the results and discuss what the results mean. Finally, conclusions are drawn in Section 6.

2. Data Acquisition

The dataset is taken from BCI competition IV, provided by Institute for Knowledge Discovery and the Graz University of Technology (available in [8]). The dataset 2b consists of 3 bipolar EEG channels (0.5-100Hz; notch filtered) and 3 EOG channels from 9 subjects in two classes for the left-hand and right-hand motor imagery tasks, for each subject, 5 sessions are provided and each session consists of several runs. For data recording, three electrodes in the position of C3, Cz and C4 (as illustrated in Figure 1) were recorded with the sample rate of 250Hz. For more information about the dataset visit [9].

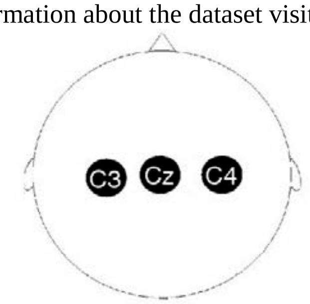

Fig. 1. Position of EEG electrodes on the scalp

Manuscript received November 5, 2017
Manuscript revised November 20, 2017

In this article, the accuracy will be measured for each specific subject and different trials. There are also 3 EOG channels provided in the datasets for noise reduction that will be discussed in "Preprocessing" section.

3. Algorithms

3.1 Preprocessing

Considerable origin of EEG signal's noise comes from Electrooculographic (EOG) artifacts which generate from eye movement and its electrical activity. So an EOG artifact reduction method is going to apply on 3 EEG channels to prepare noise-free data, A. Schlogl et al. presented a fully automated noise reduction method in [10] that can calculate correction coefficients with the following equation:

$$S(ch,t) = Y(ch,t) - U(ch,t) * b(ch,t)$$

Where S is the signal without EOG artifact, Y is the recorded signal, U denotes the noise source that recorded from three electrodes shown in Figure 2, and b indicates the weights of the EOG artifacts at the EEG channel ch at time t. The result of this procedure is shown in Figure 3.

3.2 Bandpower

Butterworth IIR filter applied on data to take signal into the frequency domain, this filter aims to generate as flat as possible signal in a selected pass-band, in another word this algorithm will remove any information that is outside of the defined frequencies [11]. After that, it is possible to determine the energy of each channel in the specified band with squaring the time series magnitude. $y(t)$ in the following equation is the output of Butterworth filter.

$$E = \sum_t |y(t)|^2$$

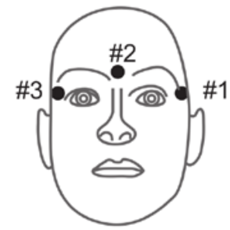

Fig. 2 Position of EOG electrodes [8]

3.3 Linear Discriminant Analysis (LDA)

LDA is based on Fisher's Linear Discriminant, is a statistical model that separates data using hyper-planes [12]. LDA can find a vector in data, in a way that minimizes the space between mean of data and each class. In more details, LDA tries to minimize the following equation.

$$\frac{(m_\phi - m_\psi)^2}{s_\phi^2 + s_\psi^2}$$

m_ψ and m_ϕ stand for mean in data, s_ψ and s_ϕ means standard derivation where two classes are labeled as ψ and ϕ.

The advantage of LDA algorithm is low computational complexity which makes it suitable for online Brain Computer Interface (BCI) applications, and simplicity of design.

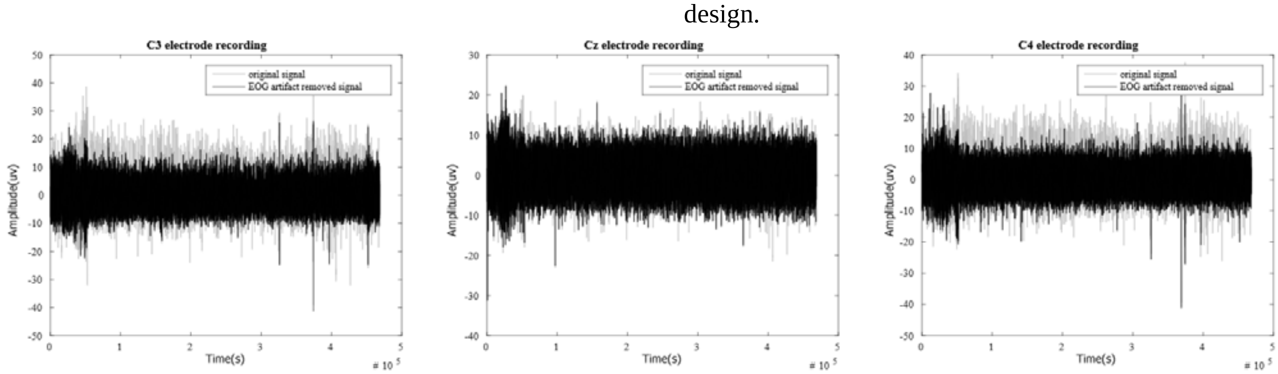

Fig. 3. Compare noise removed (black color) signal with original signal (gray color) – first person of the examination, on 3 recorded channels. X axis is amplitude (micro-volt) and Y axis is time (samples)

3.4 Support Vector Machine (SVM)

SVM classifier also uses discriminant hyper-planes for detecting each class by trying to maximize the distance between hyper-plane and the nearest training data point. With the available parameters for SVM, it is immune to over-fitting and curse of dimension but these parameters cause low execution speed [13].

Mentioned hyper-plane can be defined by the equation:

$$y = w^T x + b$$

3.5 Random Forest

RF classification algorithm generates a group of decision trees, each group uses a random subset of data for training, all the trees grow fully and independently, so the training/testing process can be done in parallel which can save execution time. The algorithm uses voting mechanism to predict output of the forest, simply the most popular class will be selected as the output [14].

3.6 Naïve Bayes

NB is based on Baye's Theorem (following equation), the key idea of this method is looking at the variable independently, but this algorithm also performs well when the variables are dependent. This algorithm performs very well for high dimensional data space. One of the applications of this algorithm is real-time predictor which like LDA makes it suitable for online BCI [15].

$$P(c|D) = \frac{P(D|c)\,P(c)}{P(D)}$$

It tries to find the posterior probability of P(c | D) (means, c is the best class label according to training data D) from P(D | c) (the probability of belonging data D to class label c), P(c) (the initial probability) and P(D) (the prior probability).

4. Proposed Method

As illustrated in figure 4, in this work, the process starts with noise reduction, after that the signal's bandpower will be measured and continues to the novel sampling technique, in the last phase, final feature matrix will be segmented and passed to the classifiers, to achieve the best classification result. Each of these steps is described in sub-sections with more details.

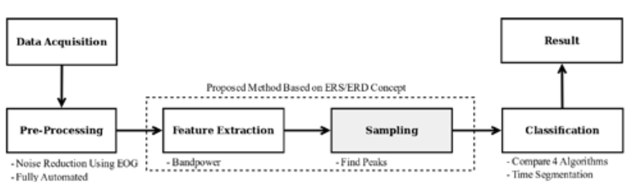

Fig. 4. Proposed method diagram

4.1 Feature Extraction

As discussed, the Bandpower is used as feature extraction algorithm, for band-pass filtering two frequency bands are selected, alpha rhythm in range of 10 Hz to 16 Hz (that is responsible for peace mental state) and beta rhythm in range of 20 Hz to 30 Hz (that is responsible for conscious thought and logical thinking mental state) are selected [16], simply because the mentioned rhythms generate features for two different mental states, thereby it is easier to distinguish features.
Based on Event Related Synchronization (ERS) and Event Related Desynchronization (ERD) methods [17], the rise and fall in signal's bandpower have a direct relation with subject's brain activity. This logic will be used in the proposed sampling technique.

4.2 Sampling

In statistics, Sampling can be defined as the process of selecting a subgroup of data in a way that the trimmed data could represent the main features of the whole population. For sampling data, first signals divided by each trial, and a feature matrix creates in the time range of Local Maxima of each segment, the idea is taken from "Maximal Variation Sampling" in statistics and also it is safe to say that with this method, all the sudden changes in signals, which means different mental activities are captured. This sampling method can reduce the features matrix dimension by 75%. As shown in figure 5, the proposed sampling method will keep overall characteristics of the signal with fewer data points.

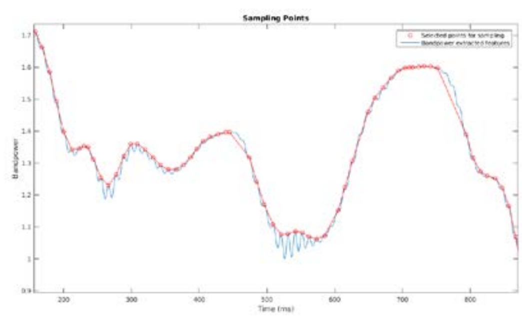

Fig. 5. Selected data points from feature matrix

qEEG (also known as EEG brain mapping) illustrated in figure 6 and 7, random trials were selected in 30Hz frequency for left- (A) and right- (B) hand motor imagery cerebral activity. The pattern shown, the proposed sampling method will have same characteristics as original signal in 30Hz frequency band, in which produces the main features in this study.

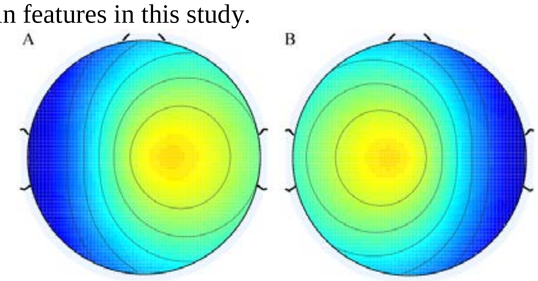

Fig.6 EEG brain mapping for original signal, band-passed on 30Hz frequency, (A) for left hand and (B) for right hand motor imagery

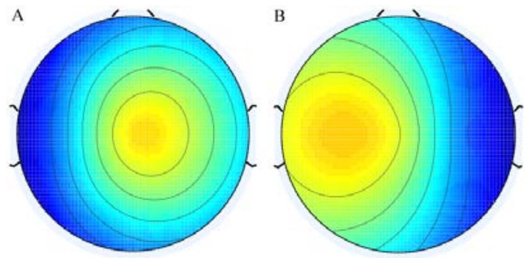

Fig. 7. EEG brain mapping for signal with proposed sampling method, band-passed on 30Hz frequency, (A) for left hand and (B) for right hand motor imagery.

4.3 Classification

After preparing features, classification algorithms used to find a relation between features and classes (left-/right-hand movement). In this stage, the feature matrix will be segmented [18], The classification accuracy measured with four algorithms in each segment. There is a brief discussion about this procedure in sub-sections below.

```
1  SegmentLen;
2  NoS = ceil(SampleRate/SegmentLen);
3  Segments = reshape((1:NoS*SegmentLen),SegmentLen,NoS);
4  [CL, ~, cl] = unique(Classes);
5  for k = 1:size(Segments,1),
6    c = []; d = []; t = [];
7    for k1 = 1:length(CL),
8          t = allPossibleSumOfTwoVectors(Trigers(cl==k1),Segments(k,:));
9          tempFeatures = [tempFeatures; Features(t(:),:)];
10         tempClasses = [tempClasses; repmat(k1,numel(t),1)];
11   end;
12   Classifier(k) = train(tempFeatures, tempClasses, classifierAlg);
13   accuracy (k) = Classifier(k).ACC;
14 end;
15 BestClassifier = Classifier( max(accuracy) );
```

The segmentation process will be used for every classifier. In the mentioned algorithm, The Segments variable contains the length of each segment based on signal's sample rate (line 1 to 4), after that the time vector will be calculated, based on all possible sum of segments vector and subject's classes trigger position (line 8). The computed time vector will be used to find the relevant features in each segment and the class label in connection with that trial (line 9 to 10). This process will be repeated for each time segment and at the end, best classification model will be the output of the algorithm. As shown in line 12, the classification algorithm is a variable and can be one of the described algorithms.

5. Result and Discussion

Accuracy is the measurement of true classification results compare to observations. As ISO discussed in [19], higher accuracy has direct relation with trueness and precision. The accuracy can be measured by the following equation:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Where the variables stand for True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

The accuracy of each classification method is shown in table 1, for each algorithm and trial, the accuracy measured with/without the proposed sampling method. As it is obvious, the proposed method has a huge impact on classification accuracy, mainly because the discussed method chooses valuable features, therefor with lower feature matrix dimension, the accuracy is getting better.

Among the algorithms, RF has the lowest accuracy, but the other three have kind of close scores. The reason is that the random forest algorithm has a built-in sampling technique to distribute features into different trees, but with the proposed method, all the extracted features are important and can't miss any of them randomly. Although it is safe to say that with knowing how RF works, still, the proposed sampling method worked very well.

The Receiver Operating Characteristic (ROC) curve shows the performance of the generated models. Figure 6 shows ROC plot for each algorithm. It draws False Negative Rate (1-specificity) against True Positive Rate (sensitivity) which are calculated as shown in following:

$$False\ Negative\ Rate\ (FNR) = 1 - Specifity = \frac{TP}{TP + FN}$$

$$True\ Positive\ Rate\ (TPR) = Sensitvity = \frac{TN}{TN + FP}$$

The ideal ROC plot happens when TPR grows rapidly whilst the FNR hardly increases.

A better metric to translate the meaning of the ROC curve is Area Under Curve (AUC) that is presented in Table 2, keep in mind that the maximum number for the area under a ROC curve is 1 [20].

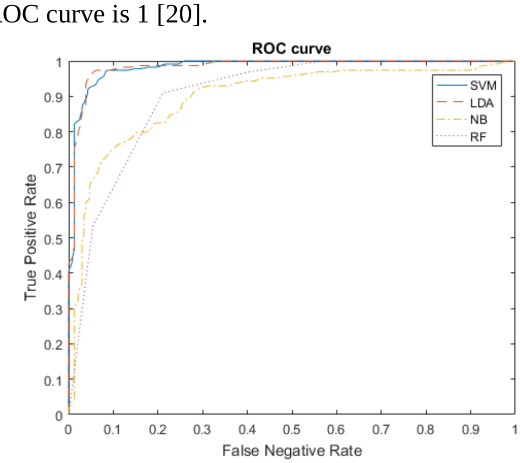

Fig. 8. ROC curve plot for LDA, SVM, NB and RF classifiers

In the second part of this section, the Kappa value measured for two top classification algorithms based on Table 1 results (LDA, NB). BCI competition IV used Kappa value as a measurement to identify the winner, which is the algorithm with the highest Kappa value, this variable gives a number from 0 to 1. If Random Accuracy declares as follows:

$$Random\ Accuracy = \frac{(TN + FP) \times (TN + FN) + (FN + TP) \times (FP + TP)}{Total\ \times Total}$$

Then Kappa value will be:

$$Kappa = \frac{Accuracy - Random\ Accuracy}{1 - Random\ Accuracy}$$

As compared in Table 3, the proposed method with LDA and NB classification algorithms have the greater Kappa value compare to the competition winner scores.

The winner algorithm by Zheng Yang Chin used Filter Bank Common Spatial Pattern (FBCSP) as feature extraction algorithm and used Naïve Bayes as classification method [21]. Jaime Fernando Delgado Saa and Yang Ping used LDA as the classification method. As it is obvious the proposed method with the same classification algorithm achieves better Kappa values.

Table 1: Accuracy for each classification algorithm and trial, compared with (W) and without (W/O) the proposed method, all the numbers are in percent.

Algorithm		#1	#2	#3	#4	#5	#6	#7	#8	#9	MEAN
LDA	W/O	55	59	54	54	48	51	55	59	55	54.4
	W	75	57	61	93	84	86	84	93	86	79.9
SVM	W/O	54	58	56	53	56	50	54	51	56	54.2
	W	76	57	64	95	58	85	81	93	84	77
RF	W/O	59	60	59	55	54	59	58	58	61	58.1
	W	66	55	58	93	80	74	55	85	75	71.2
NB	W/O	54	57	54	53	54	53	57	50	58	54.4
	W	68	62	63	95	86	82	80	93	86	79.4

Table 2: The area under curve from ROC plot for each algorithm

Algorithms	LDA	SVM	NB	RF
AUC	0.983	0.982	0.895	0.831

Table 3: Compare BCI competition Kappa scores with proposed method two best algorithms

Contributor	Kappa	1	2	3	4	5	6	7	8	9
Proposed method with NB	**0.61**	0.40	0.36	0.31	0.91	0.72	0.62	0.60	0.87	0.72
Proposed method with LDA	**0.60**	0.51	0.14	0.22	0.87	0.68	0.71	0.67	0.87	0.72
Zheng Yang Chin	**0.60**	0.40	0.21	0.22	0.95	0.86	0.61	0.56	0.85	0.74
Huang Gan	**0.58**	0.42	0.21	0.14	0.94	0.71	0.62	0.61	0.84	0.78
Damien Coyle	**0.46**	0.19	0.12	0.12	0.77	0.57	0.49	0.38	0.85	0.61
Shaun Lodder	**0.43**	0.23	0.31	0.07	0.91	0.24	0.42	0.41	0.74	0.53
Jaime Fernando Delgado Saa	**0.37**	0.20	0.16	0.16	0.73	0.21	0.19	0.39	0.86	0.44
Yang Ping	**0.25**	0.02	0.09	0.07	0.43	0.25	0.00	0.14	0.76	0.47

6. Conclusion

The process of translating EEG signals is complicated, various combination of techniques are available when it comes to noise reduction, feature extraction, feature selection and classification. The main view of this article is the usage of this technique on online BCI, to propose a superior algorithm for this purpose, a sampling method was introduced in a way to decrease the computation time and power with reducing data points. The proposed method's accuracy was tested among different classifiers and also compared with the BCI competition IV results, it shows that with the same classifiers used in the competition, the proposed method performs better.

References

[1] R. M. Rangayyan, "Introduction to Biomedical Signals," in Biomedical Signal Analysis: John Wiley & Sons, Inc., 2012, pp. 1-59.

[2] W. H. Organisation. Disability and health. Available: http://www.who.int/mediacentre/factsheets/fs352/en/

[3] V. Klig, "Biomedical applications of microprocessors," Proceedings of the IEEE, vol. 66, no. 2, pp. 151-161, 1978.

[4] I. A. Mirza et al., "Mind-controlled wheelchair using an EEG headset and arduino microcontroller," in 2015 International Conference on Technologies for Sustainable Development (ICTSD), 2015, pp. 1-5.

[5] Siuly, Y. Li, and P. Wen, "Classification of EEG signals using sampling techniques and least square support vector machines," in Rough Sets and Knowledge Technology, 2009, pp. 375-382: Springer Berlin Heidelberg.

[6] S. Siuly, E. Kabir, H. Wang, and Y. Zhang, "Exploring Sampling in the Detection of Multicategory EEG Signals,"

Computational and Mathematical Methods in Medicine, vol. 2015, p. 12, 2015, Art. no. 576437.
[7] N. Robinson, A. P. Vinod, K. K. Ang, K. P. Tee, and C. T. Guan, "EEG-Based Classification of Fast and Slow Hand Movements Using Wavelet-CSP Algorithm," IEEE Transactions on Biomedical Engineering, vol. 60, no. 8, pp. 2123-2132, 2013.
[8] G. U. O. Technology. (2008). Description of dataset 2B. Available: http://www.bbci.de/competition/iv/desc_2b.pdf
[9] G. U. O. Technology. (2008). BCI Competition IV. Available: http://bbci.de/competition/iv/index.html
[10] A. Schlogl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," (in eng), Clin Neurophysiol, vol. 118, no. 1, pp. 98-104, Jan 2007.
[11] R. M. Rangayyan, "Filtering for Removal of Artifacts," in Biomedical Signal Analysis: John Wiley & Sons, Inc., 2012, pp. 73-176.
[12] E. Alpaydin, Introduction to Machine Learning. The MIT Press, 2010, p. 584.
[13] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," Journal of Neural Engineering, vol. 4, no. 2, p. R1, 2007.
[14] L. Breiman, "Random Forests," Machine Learning, journal article vol. 45, no. 1, pp. 5-32, 2001.
[15] I. Rish, "An empirical study of the naive Bayes classifier," in IJCAI 2001 workshop on empirical methods in artificial intelligence, 2001, vol. 3, no. 22, pp. 41-46: IBM New York.
[16] A. Baskaran, R. Milev, and R. S. McIntyre, "The neurobiology of the EEG biomarker as a predictor of treatment response in depression," (in eng), Neuropharmacology, vol. 63, no. 4, pp. 507-13, Sep 2012.
[17] G. Pfurtscheller, "Functional brain imaging based on ERD/ERS," Vision Research, vol. 41, no. 10–11, pp. 1257-1260, 5// 2001.
[18] M. Lovrić, M. Milanović, and M. Stamenković, "Algoritmic methods for segmentation of time series: An overview," Journal of Contemporary Economic and Business Issues, vol. 1, no. 1, pp. 31-53, 2014.
[19] "BS ISO 5725-1: Accuracy (trueness and precision) of measurement methods and results -- Part 1: General principles and definitions," no. 1, p. 17, 1994.
[20] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," (in eng), Radiology, vol. 143, no. 1, pp. 29-36, Apr 1982.
[21] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter Bank Common Spatial Pattern Algorithm on BCI Competition IV Datasets 2a and 2b," (in English), Frontiers in Neuroscience, Methods vol. 6, no. 39, 2012-March-29 2012.