

Best Feature Extraction and Classification Algorithms for EEG Signals in Neuromarketing

Jafar Zamani ¹, Ali Boniadi Naieni ^{2,*} 

¹ School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran

² School of Progress Engineering, Iran University of Science and Technology, Tehran, Iran

*Corresponding Author: Ali Boniadi Naieni
Email: bonyadi@iust.ac.ir

Received: 17 August 2020 / Accepted: 25 September 2020

Abstract

Purpose: There are many methods for advertisements of products and neuromarketing is new area in this field. In neuromarketing, we use neuroscience information for revealing Consumer behavior by extracting brain activity. Functional Magnetic Resonance Imaging (fMRI), Magnetoencephalography (MEG), and Electroencephalography (EEG) are high efficient tools for investigating the brain activity in neuromarketing. EEG signal is a high temporal resolution and a cheap method for examining the brain activity.

Materials and Methods: 32 subjects (16 males and 16 females) aging between 20-35 years old participated in this study. We proposed neuromarketing method exploit EEG system for predicting consumer preferences while they view E-commerce products. We apply some important preprocessing steps for noise and artifacts elimination of the EEG signal. In next step feature extraction methods are applied on the EEG data such as Discrete Wavelet Transform (DWT) and statistical features. The goal of this study is classification of analyzed EEG signal to likes and dislikes using supervised algorithms. We use Support Vector Machine (SVM), Artificial Neural Network (ANN) and Random Forest (RF) for data classification. The mentioned methods were used for whole and lobe brain data.

Results: The results show high efficacy for SVM algorithms than other methods. Accuracy, sensitivity, specificity and precision parameters were used for evaluation of the model performance. The results show high performance of SVM algorithms for classification of the data with accuracy more than 87% and 84% for whole and parietal lobe data.

Conclusion: We designed a tool with EEG signals for extraction brain activity of consumers using neuromarketing methods. We investigated the effects of advertising on brain activity of consumers by EEG signals measures.

Keywords: Neuromarketing; Electroencephalography Signal; Feature; Classification.

1. Introduction

Neuromarketing is referred to combination of two major fields' neuroscience and marketing. Neuroscience tools such as fMRI and EEG are used for analyzing the consumer brain activity in the neuromarketing to select products to improve the marketing [1-2]. There are many neurophysiological techniques for analyzing the consumer behavior and advertisement to reveal different aspects of marketing [3]. Functional Magnetic Resonance Imaging (fMRI), Magnetoencephalography (MEG), and Electroencephalography (EEG) were very useful methods used in neuromarketing [4]. FMRI is a predominant and noninvasive map systems-level activity in human brain. The information of FMRI obtained using Blood Oxygen Level-Dependent (BOLD) contrast for obtaining quantitative measures of oxidative metabolism, perfusion, and blood volume changes with activation. Many studies used fMRI in neuromarketing application, but the cost of these experiments is very expensive [5-6]. MEG is a method to obtain brain functions at the required high temporal resolution such as sensory percepts, cognitive processes, and motor activity [7]. EEG is one of the oldest methods for mapping brain activity in health and disease and one of the most widely used techniques in clinical practice because of its high temporal resolution and availability in many clinical environments. Brain-Computer Interfaces (BCIs) are efficient neuroimaging methods in neuromarketing. In this technology the users have effectively communicated with computer systems [8]. To overcome the higher cost, technical complexities and limited access limitations, in this study we use EEG system setup. EEG is very useful technique in Neuromarketing because of lower cost, easy handling, wireless connectivity and availability [9-10]. This study will improve the further promotions of the products for marketing.

In this paper we proposed neuromarketing method using EEG system for predicting consumer preferences while they view E-commerce products. A contrast of the activity was extracted by LIKE dataset and DISLIKE dataset. These two classes of datasets were classified using machine learning algorithms.

For classification we used total data and each of EEG frequency band data.

2. Materials and Methods

2.1. Dataset Description

32 subjects (16 males and 16 females) aging between 20-35 years old participated in this study. For our study, we considered 10 product categories which were chip, juice, lighter, gift bag, Kleenex, diary, ice cream, chocolate, tea, and mobile phones. For each category, 5 different images were shown to subjects. The image with same sizes presentation protocol is as follows. Each image was shown for 5 seconds and after that one white page was shown for 2 seconds. After finishing images of one category, one black page was shown for 5 seconds. After showing 25 images to subject, one black page was shown for 30 seconds. For each image, the participants were given a questionnaire and asked to fill in with rating scale of 1 to 5.

EEG data was recorded from subject while performing the protocol. The data was measured by 19 electrodes at locations FP1, FP2, F7, F8, F3, F4, T3, T4, C3, C4, T5, T6, P3, P4, O1, O2, FZ, CZ, PZ according to the International 10-20 system. The sampling frequency was 500Hz.

2.2. Preprocessing

Before feature extraction, it is necessary to apply some preprocess steps for obtaining as clean as possible EEG signal. We are interested in EEG signals existing in frequency range of 0.5Hz-100Hz. Thus, we apply low-pass filter with cut-off frequency of 100Hz and then high-pass filter with cut-off frequency of 0.5Hz. By doing so, the low frequency noises, including baselines and trends and also noises existing in frequencies higher than 100Hz are removed. The filtering process is continued by applying notch filter at 50Hz for removing line noise [11].

Some artifacts such as eye blink and Electromyogram (EMG) coexist with EEG signal in the same frequency ranges. Therefore, they cannot be eliminated from the EEG signal through conventional filtering. In this study, we consider the Independent

Component Analysis (ICA) technique for wiping these artifacts from EEG signal. ICA attempts to find a number of signals components which are temporally independent from each other, as possible. Some of these components are our desired EEG signals and some others are artifacts. By removing the artifact components from EEG signal, the resulted signal would be a clean one. For satisfactory finding of artifact components through ICA, it is needed for the EEG signal to become as clean as possible. For this reason, we apply the filtering process firstly. Then, for each subject, we implement the ICA analysis for computing 18 independent components. For one subject, the time series of 18 components are shown in Figure 1(A) and (B).

It is obvious that the first component is for eye blink and must be removed from EEG signal. The 2nd, 4th, 6th, 7th, 13th, and 17th components are suspect for EMG. The Power Spectrum Density (PSD) plot of these components can help for identifying the truly EMG components (Figure 1 (C)). The 4th, 7th, and 13th components show low power density at low frequency and their powers are increasing at 15Hz. Hence, we consider them as EMG artifacts and wipe them from EEG signal. The EEG signal before and after applying ICA technique are shown in Figure 2.

As it is evident, the ICA efficiently helps for removing eye blink and EMG artifacts from EEG signal.

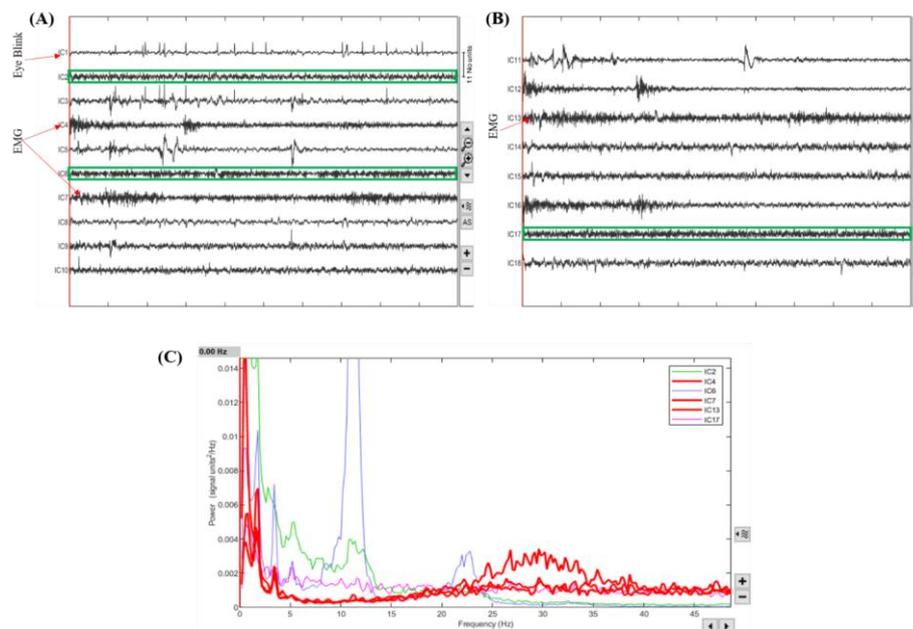


Figure 1. (A) and (B): The time series of 18 ICA components, the green squares indicate components which are suspicious of being EMG ones; the 1st component is for eye blink and the 4th, 7th and 17th components are for EMG. (C) The PSD of some components suspected for EMG; the selected components as EMG are drawn by red color

It is noteworthy that all the ICA analysis and also the conventional filtering are carried out in Brainstorm software [12].

In the next step, Savitzky-Golay (S-Golay) filter is employed for smoothing EEG signals through a frame span of size 5 and with a quadratic polynomial. S-Golay is a least square or polynomial based filter that smooths the noisy signal by fitting consecutive subsets of neighboring signal points using low degree polynomial and linear least squares [13-14].

2.3. Feature Extraction

We employ the Discrete Wavelet Transform (DWT) for feature extraction [15]. The DWT exerts a filter bank technique through which different low-pass and high-pass filters are applied to the signal and signals at different frequency bands are extracted. The outputs of low-pass and high-pass filter, respectively, offer approximate and detailed information of the signal. These outputs are called Approximate (A) and Detail (D) coefficients of DWT, respectively [16].

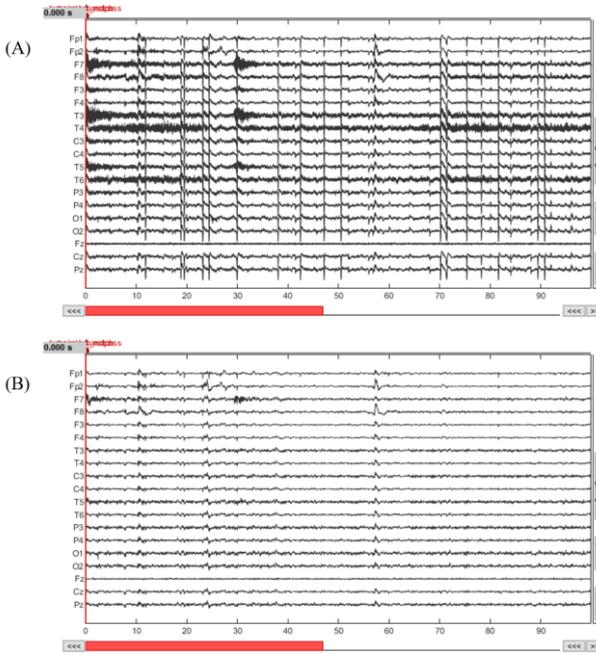


Figure 2. (A) and (B) show the EEG signal before and after applying ICA analysis. By using ICA technique, the EEG signal is cleaned from eye blink and EMG artifacts

In this work, we decompose the EEG signal at four levels by employing the Daubechies 4 (DB4) wavelet decomposition approach. This decomposition provides five group wavelet coefficients D1, D2, D3, D4, and A4. Each group represents one frequency band information of brain electrical activity. These groups correlate with the EEG spectrum having four different frequency bands, including Delta (1-4 Hz), Theta (4-8 Hz), Alpha (8-13 Hz), Beta (13-22 Hz), and Gamma (32-100Hz) [17]. The D1, D2, D3, D4, and A4 are corresponding to the Gamma, Beta, Alpha, Theta, and Delta bands, respectively.

For each frequency band, the wavelet coefficients of all 19 electrodes are computed. In addition, statistical features, including Mean (M), Standard Deviation (SD), Energy (EN) and Root-Mean-Square (RMS) are computed for each wavelet coefficient of each electrode. These features are calculated as follows:

$$M = \frac{1}{N} \sum_{j=1}^N x_j \tag{1}$$

Where x_j is the j^{th} sample of under study signal and N represents the total number of samples.

$$SD = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (x_j - \bar{x})^2} \tag{2}$$

Where \bar{x} is the mean of the under study signal.

$$EN = D \sum_{j=0}^{N-1} x_j^2 \tag{3}$$

Where D is duration of signal.

$$RMS = \sqrt{\frac{\sum_{j=1}^N x_j^2}{N}} \tag{4}$$

These four features are computed for each of 19 channels. Therefore, 78 features are attained. These statistical features along with the wavelet coefficients form the final feature vector.

We are also interested in knowing about brain lobes importance for consumer choice prediction. To know this, we investigate the classification performance for each brain lobe separately and based on their corresponding feature vectors. The 5 lobes include Frontal (Fp1, Fp2, F7, F8, F3, F4, Fz), Temporal (T3, T4, T5, T6), Central (C3, C4, Cz), Parietal (P3, P4, Pz), and Occipital (O1, O2).

3.3. Classification Methods

We exploit several algorithms with higher efficacy for classification of the data using extracted features. The machine learning based methods are popular algorithms for EEG classification, in this regard we use the Radial Basis Function (RBF) kernel Support Vector Machine (SVM) as a supervised learning algorithm for data classification. SVM use a kernel function to map input data into higher dimensional space using nonlinear methods, after which it separates the data via a hyper-plan with maximal margins [18]. SVM is very interested algorithms for binary classification of data [19]. These three algorithms were used in many application of the EEG signal classification. The algorithm implemented 100 iteration and meant of accuracy values were reported.

Artificial Neural Network (ANN) [20] and Random Forest (RF) are another algorithms that we use for EEG data classification [21]. RF uses multiple decision trees for training the data and outputs average prediction of each tree. This classifier generates random forests of tree [22].

As mentioned in Data Description section, the subject determined his/her opinion about product image by a discrete number from 1 to 5. In this study, we regarded the numbers 1, 2, and 3 as dislike class

(label=0) and the numbers 4 and 5 as like class (label=1).

3. Results

The proposed methods were implemented on Matlab for evaluation of the algorithms efficiency. In the first step we randomly split the data to test and train data by using the cross-validation. We select 70% of data as a training set. For evaluation of the model performance, we use accuracy, sensitivity, specificity and precision (Figure 3). These parameters are calculated as following formulas:

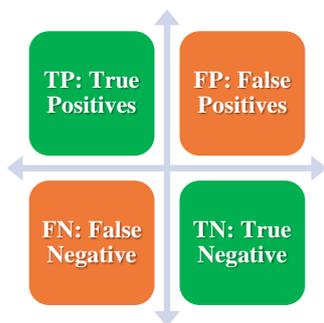


Figure 3. Four concepts accuracy, sensitivity, specificity and precision

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

$$Precision = \frac{TP}{TP + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

4. Conclusion

We design a setup with EEG signals for extraction brain activity of consumers using neuromarketing methods. The main goal of this study is to examine the effects of advertising on brain activity of consumers by EEG signals measures. The designed task is shown on the computer screen for all of the participants and recorded EEG data. We implement all of necessary of denoising, features selection and extraction methods of the EEG data. We classify the data to likes and dislikes using three high efficacy algorithms. The results show high performance of SVM algorithms for classification of the data with accuracy more than 87% and 84% for whole and parietal lobe data (Table1 , Table2).

Table 1. Numerical result of the classification algorithms for all of the brain

Methods	Accuracy	Sensitivity	Specificity	precision
SVM	87.87	71.43	98.59	97.05
RF	81.60	52.87	92.36	84.42
ANN	70.99	44.30	88.37	71.29

Table 2. Numerical result of the classification algorithms for all of the brain different brain lobes

Methods	Data	Accuracy	Sensitivity	Specificity	Precision
SVM	Frontal	73.51	37.31	97.05	89.17
	Central	81.94	56.18	98.63	96.36
	Occipital	68.33	27.37	95.17	78.79
	Parietal	84.31	61.62	99.08	97.77
	Temporal	80.00	52.91	97.59	93.46
RF	Frontal	68.93	28.75	89.18	60.31
	Central	78.06	41.25	92.86	84.56
	Occipital	65.63	25.89	84.95	61.32
	Parietal	79.58	48.13	94.67	92.35
	Temporal	74.48	43.87	91.48	91.29
ANN	Frontal	64.64	27.64	88.70	61.41
	Central	70.42	37.46	91.76	74.65
	Occipital	62.71	25.79	86.90	56.32
	Parietal	69.31	35.92	91.06	72.34
	Temporal	68.96	34.66	91.24	71.98

Acknowledgements

This work was supported by the Iran's National Elites Foundation.

References

- 1- Morin C. "Neuromarketing: the new science of consumer behavior". *Society*, vol.1;48(2), pp. 131-5, Mar 2011.
- 2- Nilashi M, Samad S, Ahmadi N, Ahani A, Abumalloh RA, Asadi S, Abdullah R, Ibrahim O, Yadegaridehkordi E. "Neuromarketing: A Review of Research and Implications for Marketing". *Journal of Soft Computing and Decision Support Systems*. vol. 4, no.7(2), pp. 23-31, Mar 2020.
- 3- Boksem MA, Smidts A. "Brain responses to movie trailers predict individual preferences for movies and their population-wide commercial success". *Journal of Marketing Research.*; vol. 52(4), no.4, pp. 82-92, Aug 2015.
- 4- Yadava M, Kumar P, Saini R, Roy PP, Dogra DP. "Analysis of EEG signals and its application to neuromarketing". *Multimedia Tools and Applications*. vol. 1, no. 76(18), pp. 19087-111, Sep 2017.
- 5- Berns GS, Moore SE. "A neural predictor of cultural popularity". *Journal of Consumer Psychology*. vol.1, no.22(1), pp. 154-60, Jan 2012.
- 6- Stanton SJ, Sinnott-Armstrong W, Huettel SA. "Neuromarketing: Ethical implications of its use and potential misuse". *Journal of Business Ethics*, vol.1, no. 144(4), pp. 799-811, Sep 2017.
- 7- Baillet S, Mosher JC, Leahy RM. "Electromagnetic brain mapping". *IEEE Signal processing magazine*, vol. 18, no. (6), pp. 14-30, Nov 2001.
- 8- Ramadan, R.A.; Refat, S.; Elshahed, M.A.; Ali, R.A. "Brain-Computer Interfaces. *In Intelligent Systems Reference Library; Springer International Publishing: Cham, Switzerland*, vol. 74, pp. 31–50, 2015.
- 9- Telpaz A, Webb R, Levy DJ. "Using EEG to predict consumers' future choices". *Journal of Marketing Research.*, vol. 52, no. (4), pp. 511-29, Aug 2015.
- 10- Golnar-Nik P, Farashi S, Safari MS. "The application of EEG power for the prediction and interpretation of consumer decision-making: A neuromarketing study". *Physiology & behavior*, vol. 1, no. 207, pp. 90-8, Aug 2019.
- 11- Cohen, Mike X. "Analyzing neural time series data: theory and practice". *MIT press*, 2014.
- 12- BrainStorm, Matlab Toolbox. (2019). [Online]. Available: <http://neuroimage.usc.edu/brainstorm>.
- 13- Hargittai S, Savitzky-Golay. "least-squares polynomial filters in ECG signal processing". *In Computers in Cardiology IEEE*, vol. 25 ,pp. 763-766, Sep 2005.
- 14- Awal MA, Mostafa SS, Ahmad M. "Performance analysis of Savitzky-Golay smoothing filter using ECG signal". *International Journal of Computer and Information Technology*, vol. 1, no. (02), pp. 24, Jan 2011.
- 15- URAL G, KAÇAR F, CANAN S. "Wavelet Phase Coherence Estimation of EEG Signals for Neuromarketing Studies". *NeuroQuantology*, vol. 17, no. (02), pp. 112-20, Jan 2019.
- 16- Amin HU, Malik AS, Ahmad RF, Badruddin N, Kamel N, Hussain M, Chooi WT. "Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques". *Australasian physical & engineering sciences in medicine*. Vol. 1, no. 38(1), pp. 139-49, Mar 2015.
- 17- M. Yadava, *et al.*, "Analysis of EEG signals and its application to neuromarketing, Multimed". *Tools Appl.*, vol. 76, no. 18, pp. 19087 –19111, 2017.
- 18- Cortes C, Vapnik V. "Support-vector networks. Machine learning". vol. 1, no.20(3), pp. 273-97, Sep 1995.
- 19- Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification.", pp. 1396-1400, 2003.
- 20- Sha'abani MN, Fuad N, Jamal N, Ismail MF. "kNN and SVM Classification for EEG: A Review". *In InECCE2019, Springer, Singapore*, pp. 555-565, 2020.
- 21- Subasi A, Ercelebi E. "Classification of EEG signals using neural network and logistic regression". *Computer methods and programs in biomedicine*. vol.1, no. 78(2), pp. 87-99, May 2005.
- 22- Fraiwan L, Lweesy K, Khasawneh N, Wenz H, Dickhaus H. "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier". *Computer methods and programs in biomedicine*, vol. 1, no.108(1), pp. 10-9, Oct 2012.