Paper 1545-2014

Association Mining of the Brain Data: An EEG Study

Pankush Kalgotra; Ramesh Sharda; Goutam Chakraborty, Oklahoma State University, Stillwater, OK, US

ABSTRACT

Many neuroscience researchers have explored how various parts of the brain are connected. No one has performed association mining using the brain data. In this study, we used SAS® Enterprise Miner™ 7.1 for association mining of the brain data collected by a 14-channel EEG device. An application of the association mining technique is presented in this novel context of brain activities and by linking our results to theories of cognitive neuroscience. The brain waves are collected while a user processes the information about Facebook, the most well known social networking site. The data is cleaned using Independent Component Analysis via an open source MATLAB package. Next, by applying the LORETA algorithm, activations at a fraction of a second are recorded. The data is codified into transactions to perform association mining. Results showing how various parts of brain that get excited while processing the information are reported. This study provides preliminary insights into how brain wave data may be analyzed by commonly available data mining techniques to enhance a researcher's understanding of brain activation patterns.

INTRODUCTION

There is a significant amount of research to discover the functional connectivity of human brain. It focuses on how different parts of the brain share information among each other on any stimulus. Several neuroscience researchers have tried to find how various parts of the brain are connected and respond to a stimulus. As found in the neuroscience literature, there is a strong functional connectivity among different regions of the brain. There is a network of regions in human brain that share information among each other during a specific activity (Heuvel and Pol, 2010; Corbetta and Shulman, 2002; Bassett and Bullmore, 2006). Different networks of brain regions are found by resting state fMRI (Bassett & Bullmore, 2006). Human experts then analyze the images to identify relations among various regions. Given that such analyses are time consuming, we explore applications of data mining techniques for identifying patterns of related images.

Association mining is a data mining technique typically used to analyze the transactional data. Association rules are generated from the data based on minimum support and confidence. The rules generated give information about the correlation of the items in the data set (Agrawal et al. 1993). This technique is helpful in analyzing large market-basket type databases. In addition to market-basket analysis, this technique has been used in other areas such as web usage mining, intrusion detection, etc. In this project, we extend the application of the association mining to analyze the data collected from the human brain. In our study, we used this technique to generate association rules of the brain activations. From the rules with high support and high confidence uncovered by the Enterprise Miner, we concluded that the various parts of brain activated together during the same course of time while the user participated in the experiment. Results of this study also provide insights into how a consumer processes information from a popular social networking site.

Internet has linked every part of this world and made it easy for people to connect with each other across the globe. A lot of the credit for global connection goes to the numerous social networking websites those have made this possible. We know that people spend a lot of time on social networking sites and many studies in marketing have shown that the consumers rate their experience with social networking sites as quite different from their experience on other sites on the Internet such as while shopping online via ecommerce sites. What is relatively unknown is how a subject's brain wave pattern may be different based on what types of sites on the Internet he/she is visiting.

We performed a single subject EEG (Electroencephalogram) while a user recalled the properties of the most popular social networking website, Facebook. An EEG device measures voltage fluctuations on the scalp that result from changes in membrane conductivity elicited by synaptic activity and intrinsic membrane processes. The electrodes on the skin capture the summed postsynaptic potentials generated by a large number of neurons. It has a high temporal resolution in milliseconds.

We used a 14-channel EEG device named EMOTIV EPOC (Emotiv Systems Inc., San Francisco, CA, USA), which is used in the gaming industries. It records the brain waves using a dry electrode system based on saline sensors located on the scalp using International 10/20 system (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) with CMS/DRL references at P3/P4 locations. It uses sequential sampling at 2048Hz and down sample to 128Hz. The device resembles the Bluetooth headset that is commonly used to listen to the music in daily lives. The data collected by the EEG were cleaned using Independent Component Analysis with EEGLAB, an open source MATLAB package. Next, by applying the LORETA algorithm, activations at a fraction of a second were recorded. Further, we reorganized the data into transactions to perform association mining using Enterprise Miner 7.1.

EEG EXPERIMENT

The primary goal of the experiment was to record the task related activities of the various parts of the human brain. One subject was selected who was a regular Facebook user. The participant was a 21 years old male. He was selected randomly from a convenient population.

PRIMARY TASK

The purpose was to analyze the features of Facebook. The brain waves of the participant were collected while he recalled the features of the Facebook to answer the questions listed below. The questions were designed to make the subject think about Facebook. Five questions were designed related to the mission statement by Facebook: “Facebook's mission is to give people the power to share and make the world more open and connected”. The questions asked were:

1. Facebook makes the world more open and connected?
2. Facebook gives people the power to share?
3. Facebook is easy to use?
4. Facebook is the best social networking website?
5. Facebook is a clean social networking website?

The experiment was completed in two sessions. The first session was an orientation session in which the subject was got familiar with the EEG device. In this session, he was also given 10 minutes to browse the Facebook website that could help him answer the questions easily. The subject was made to sit in an armrest chair with EEG device on his headset. He was requested to move as little as possible during the experiment to minimize the noise in the data.

In the second session, the real data was collected. Five questions were asked to the participant one after the other on the computer screen. Each question lasted for six seconds on the screen and three more seconds were given to mark a choice on a piece of paper after each question. During the orientation, he was instructed to answer the questions on a 5-points Likert scale with 1 as completely disagree and 5 as completely agree. We assume the subject processed the information in the first six seconds and the other three seconds were used to mark the answer on the screen. The experiment lasted up to 45 seconds among which 30 seconds (5 questions * 6 seconds) were spent for processing the information and 15 seconds for marking the answers. Although the brain waves were collected for all 45 seconds, only 30 seconds were extracted from the data for the analysis reported in this paper.

DATA COLLECTION AND ANALYSIS

To process the EEG recording collected at the sampling rate of 128Hz, a MATLAB package named EEGLAB and sLORETA software were used. The EEGLAB was used to filter the data (between 1Hz to 50Hz), create average and reference the data, and to compute the independent components. The Independent Component Analysis was done to detect and remove the common artifacts like eye-blinks and muscle contraction. Some of the noise that was clearly visible in the continuous data was removed manually. Then, this cleaned data was used to find the exact location of the activations in the brain.

Some researchers have argued that an EEG method is not a brain imaging technique because it just measures electrical signals resulting from ionic current flows within the neurons of the brain. Mathematically, the current produced in the brain have both negative and positive potentials that may cancel out. Therefore, EEG measurement may not result in the exact origin of the signals. This problem is referred to as the inverse problem. To solve this problem, many algorithms have been developed. We use a low-resolution brain electromagnetic tomography (LORETA) algorithm developed by Pascual-Marqui (2007) to find the exact location of the activity into the brain. By applying this algorithm using software sLORETA, we locate the sources of the signals recorded by the EEG.

The sLORETA software allowed us to save brain activities at a fraction of a second that were codified into the transactions for the association mining. The data collected was for 30 seconds with each second having 128 frames. So, the numbers of observations were 3,840 (30*128).

Another thing to be taken care of was the length of one transaction. According to the literature of cognitive neuroscience, it takes approximately 200-300 ms for processing one stimulus (Averbach & Coriell, 1961; Eriksen & Collins, 1969; Muller & Rabbitt, 1989). The activities of the excited regions start decreasing after 300ms of the stimulus. So, following the established theories, we selected 300 ms as one transaction time. We were able to extract

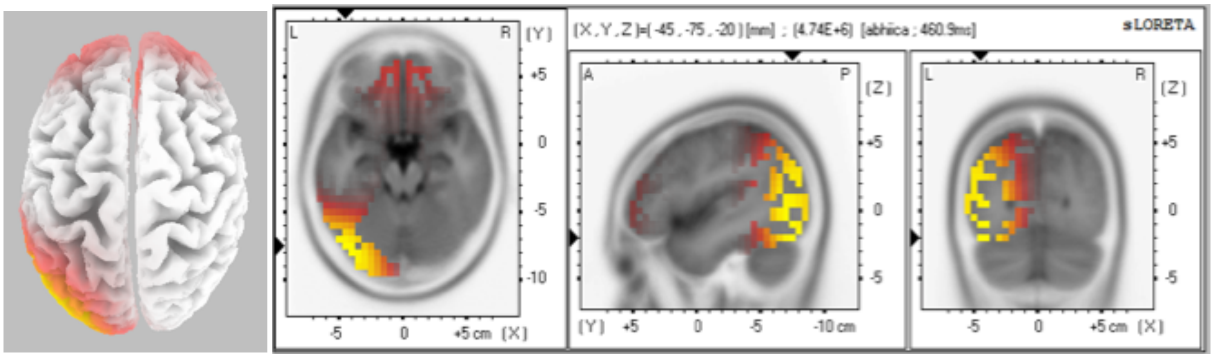

Figure 1. Brain Activations

about 100 transactions from the 30 seconds data (30,000ms/300ms). We assumed that one transaction contained only those regions that were activated by one stimulus. These transactional data was finally used for performing association mining.

RESULTS

Independent components computed from the EEGLAB were imported to sLORETA to find the exact location of the activities in the brain of the subject. Figure 1 shows the brain activity at a random point of time. The figure on the leftmost side is the 3D image of the cortex. The other three images are three orthogonal slices through the location of maximal increase or decrease. The increased activities are labeled by yellow. In the images, the Talairach coordinates are indicated by black triangles on the coordinate axes. The Talairach x coordinate (middle-left image) shows right (R) and left sides (L), y coordinates (middle-right image) as posterior (P) and anterior (A) and z-coordinates (rightmost image) as inferior (I) and superior (S). In Figure 1, the activated area indicated by yellow is Fusiform Gyrus that is present in the Temporal lobe. Similarly, brain activations at each time frame during the experiment were recorded to perform association mining.

ASSOCIATION MINING RESULTS

The data recorded from the sLORETA software was imported into Enterprise Miner for generating meaningful rules. Based on a minimum confidence of 10% and a minimum support of 4%, 199 association rules were generated. The top fifteen rules generated by the Enterprise Miner are presented in Figure 2. These rules state which parts of the brain activated together during the whole course of time. The top rule has the confidence of 62.5%. This suggests that when Precentral Gyrus & Fusiform Gyrus in the human brain were activated together, Middle Occipital Gyrus & Cuneus were also activated 62.5% of the time. This rule had a support of 4.67%, which means that in the data collected, all the parts of the brain stated in the association rule activated together 4.67% times. From the neuroscience literature, the functions of these parts provide interesting insights. Precentral Gyrus is responsible for motor control (Yousry et al. 1997), Fusiform Gyrus is associated with imagining of faces (Ishai et al. 2002), Middle Occipital Gyrus is responsible for the spatial attention (Renier et al. 2010) and Cuneus is associated with basic visual information processing (Collignon et al. 2011). It is interesting to note that all these brain regions are associated with imagining of visuals. The other areas that were activated in the next association rules were Lingual Gyrus and Superior Temporal Gyrus. Lingual Gyrus play a significant role in vision (Bogousslavsky et al. 1987) and Superior Temporal Gyrus is associated with emotions (Narumoto et al. 2001) and trustworthiness (Winston et al. 2002). In addition to imagining of visuals, the brain regions related to emotions and trust were also excited while processing the information about the questions asked.

From the rules generated, we found many intriguing results. First, we observe that the different parts of brain activated while processing the information about the social networking website. The parts activated are mostly related to imagining faces and visualizations. One of the most popular features of Facebook is sharing photos, thus the response appears to be relevant. The most interesting results were the activities in the brain related to emotions and trustworthiness. The subject was emotionally connected to the website and perhaps would therefore trust its content.

Second, the results show that when some parts are always activated together, it means they have a functional connectivity among each other. These parts are connected and form a network in the human brain. The graph shown in Figure 3 presents the functional connectivity of the brain while processing information about a social networking website.

Relations	Confidence(%) ▼	Support(%)	Rule
4	62.50	4.67	Precentral Gyrus & Fusiform Gyrus ==> Middle Occipital Gyrus & Cuneus
4	50.00	4.67	Superior Temporal Gyrus & Middle Occipital Gyrus & Lingual Gyrus ==> Fusiform Gyrus
4	50.00	4.67	Lingual Gyrus & Cuneus ==> Precentral Gyrus & Middle Occipital Gyrus
4	45.45	4.67	Superior Temporal Gyrus & Medial Frontal Gyrus & Lingual Gyrus ==> Fusiform Gyrus
4	40.00	5.61	Middle Occipital Gyrus & Lingual Gyrus ==> Medial Frontal Gyrus & Fusiform Gyrus
4	40.00	5.61	Middle Occipital Gyrus & Lingual Gyrus ==> Inferior Frontal Gyrus & Fusiform Gyrus
3	40.00	5.61	Middle Occipital Gyrus & Lingual Gyrus ==> Fusiform Gyrus
4	40.00	5.61	Superior Frontal Gyrus & Middle Occipital Gyrus & Lingual Gyrus ==> Fusiform Gyrus
4	40.00	5.61	Middle Temporal Gyrus & Middle Occipital Gyrus & Lingual Gyrus ==> Fusiform Gyrus
4	40.00	5.61	Middle Occipital Gyrus & Middle Frontal Gyrus & Lingual Gyrus ==> Fusiform Gyrus
4	40.00	5.61	Middle Occipital Gyrus & Medial Frontal Gyrus & Lingual Gyrus ==> Fusiform Gyrus
4	40.00	5.61	Middle Occipital Gyrus & Lingual Gyrus & Inferior Frontal Gyrus ==> Fusiform Gyrus
4	40.00	5.61	Middle Occipital Gyrus & Lingual Gyrus ==> Middle Frontal Gyrus & Fusiform Gyrus
4	40.00	5.61	Middle Occipital Gyrus & Lingual Gyrus ==> Middle Temporal Gyrus & Fusiform Gyrus
4	40.00	5.61	Middle Occipital Gyrus & Lingual Gyrus ==> Superior Frontal Gyrus & Fusiform Gyrus

Figure 2. Association Rules

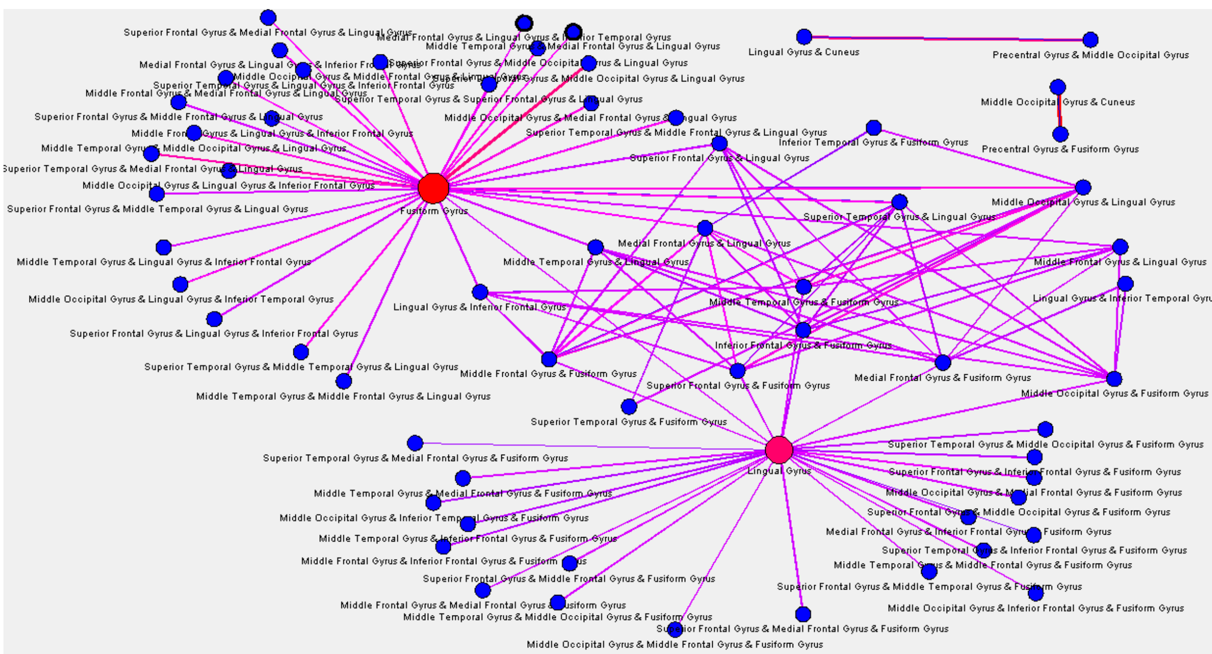

Figure 3. Link Graph

DISCUSSION

This is an exploratory study to present an application of association mining. Neuroscientists and other researchers use different kinds of network analysis techniques to understand the functional connectivity of the different parts of the brain. We have used the commonly available data mining technique to understand the activation patterns of the human brain. The association rules generated clearly show which parts of the brain are connected with each other. If they are connected, it means they share common information on the stimulus. This study gives researchers a new method to analyze the brain imaging data.

Specifically, the rules generated in this study state, which parts of the brain worked together while processing Facebook information. The particular activations of the brain parts provide interesting insights about how users process information from Facebook. This method can be used to help developers to design website that successfully activates the required parts of human brain.

Though the study has many useful implications for the researchers and application developers, it has many limitations. First, only one subject was used to collect the data. So, the sample size in our study is extremely small. Most of the EEG studies use more than 5 subjects to generalize the results. Second, the EEG data contains different kinds of noise artifacts. We tried to remove some of them using Independent Component Analysis and manually but still there might be some others that were not detected e.g., artifact due to environment. Third, the analysis is very exploratory in nature. More research is needed in this domain, as this is a very promising area.

Notwithstanding all the limitations, this study provides many insights to the research community and application developers and designers. This study gives researchers a new method to study the brain imaging data collected by any tool such as EEG, fMRI, etc. Different companies evaluate their performances based on the user ratings and self-reports but these methods have many unknown biases. However, the technique such as braining imaging is perhaps less biased because we are directly measuring the response in a user's brain.

Future work includes comparing different websites using the same technique and analyzing users' behavior. For instance, we can compare the activation patterns of the human brain while a user uses a networking site like Facebook, which is used for entertainment to another networking site LinkedIn, which is primarily used for professional purposes. It would be very interesting to see what parts of the brain get excited while using these similar networking sites that have extremely different mission statements.

By introducing the brain imaging technique in our field, we can strengthen our existing theories and find many hidden facts that are not possible to observe with the traditional approaches.

REFERENCES

Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. *In ACM SIGMOD Record* (Vol. 22, No. 2, pp. 207-216). ACM.

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal*.

Bassett, D. S., & Bullmore, E. D. (2006). Small-world brain networks. *The neuroscientist*, 12(6), 512-523.

Bogousslavsky, J., Miklossy, J., Deruaz, J. P., Assal, G., & Regli, F. (1987). Lingual and fusiform gyri in visual processing: a clinico-pathologic study of superior altitudinal hemianopia. *Journal of Neurology, Neurosurgery & Psychiatry*, 50(5), 607-614.

Collignon, O., Vandewalle, G., Voss, P., Albouy, G., Charbonneau, G., Lassonde, M., & Lepore, F. (2011). Functional specialization for auditory–spatial processing in the occipital cortex of congenitally blind humans. *Proceedings of the National Academy of Sciences*, 108(11), 4435-4440.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, 3(3), 201-215.

Eriksen, C. W., & Collins, J. F. (1969). Temporal course of selective attention. *Journal of experimental psychology*, 80(2p1), 254.

Ishai, A., Haxby, J. V., & Ungerleider, L. G. (2002). Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *Neuroimage*, 17(4), 1729-1741.

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2), 315.

Narumoto, J., Okada, T., Sadato, N., Fukui, K., & Yonekura, Y. (2001). Attention to emotion modulates fMRI activity in human right superior temporal sulcus. *Brain Research. Cognitive Brain Research*, 12, 225–231.

Pascual-Marqui, R. D. (2007). Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization*. arXiv preprint arXiv:0710.3341.*

Renier, L. A., Anurova, I., De Volder, A. G., Carlson, S., VanMeter, J., & Rauschecker, J. P. (2010). Preserved functional specialization for spatial processing in the middle occipital gyrus of the early blind. *Neuron*, 68(1), 138-148.

Van Den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8), 519-534.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, 5, 277–283.

Yousry, T. A., Schmid, U. D., Alkadhi, H., Schmidt, D., Peraud, A., Buettner, A., & Winkler, P. (1997). Localization of the motor hand area to a knob on the precentral gyrus*. A new landmark. Brain*, 120(1), 141-157.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Pankush Kalgotra, Oklahoma State University, Stillwater OK

Email: pankush@okstate.edu

Pankush Kalgotra is a doctoral student majoring in Management Science and Information Systems at Oklahoma State University. He has three-years experience of using SAS® tools for Data Mining, Texting Mining, and Sentiment Analysis projects. He is a SAS® certified Predictive Modeler using SAS® Enterprise Miner 6.1. In December 2012, he received his SAS® and OSU Data Mining Certificate. He has published in SAS Global Forum, 2013 Proceedings.

Ramesh Sharda, Oklahoma State University, Stillwater OK,

Email: ramesh.sharda@okstate.edu

Ramesh Sharda is Director of the PhD in Business for Executives Program, the Institute for Research in Information Systems (IRIS), ConocoPhillips Chair and a Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University. He has coauthored two textbooks (Decision Support and Business Intelligence Systems, 9th edition, Prentice Hall and Business Intelligence: A Managerial Approach, 2nd Edition, Prentice Hall). His research has been published in major journals in management science and information systems including Management Science, Operations Research, Information Systems Research, Decision Support Systems, Interfaces, INFORMS Journal on Computing, and many others. He is a member of the editorial boards of journals such as the Decision Support Systems, ACM Transactions of MIS, and Information Systems Frontiers. He is currently serving as the Executive Director of Teradata University Network. He recently received the 2013 INFORMS HG Computing Society Lifetime Service Award.

Goutam Chakraborty, Oklahoma State University, Stillwater OK

Email: goutam.chakraborty@okstate.edu

Dr. Goutam Chakraborty is a professor of marketing and founder of SAS® and OSU data mining certificate and SAS® and OSU business analytics certificate at Oklahoma State University. He has published in many journals such as Journal of Interactive Marketing, Journal of Advertising Research, Journal of Advertising, Journal of Business Research, etc. He has chaired the national conference for direct marketing educators for 2004 and 2005 and co-chaired M2007 data mining conference. He has over 25 years of experience in using SAS® for data analysis. He is also a Business Knowledge Series instructor for SAS®.