

Multi-aspect Data Analysis in Brain Informatics

Ning Zhong

The International WIC Institute & Department of Information Engineering,
Maebashi Institute of Technology,
460-1 Kamisadori-Cho, Maebashi-City 371-0816, Japan
zhong@maebashi-it.ac.jp

Abstract. In order to investigate human information processing mechanism systematically, various methods of brain data measurement and analysis are required. It has been observed that multiple brain data such as fMRI brain images and EEG brain waves extracted from human multi-perception mechanism involved in a particular task are *peculiar* ones with respect to a specific state or the related part of a stimulus. Based on this point of view, we propose a way of *peculiarity oriented mining* for multi-aspect analysis in multiple human brain data, without using conventional image processing to fMRI brain images and frequency analysis to brain waves. The proposed approach provides a new way in Brain Informatics for automatic analysis and understanding of human brain data to replace human-expert centric visualization. We attempt to change the perspective of cognitive scientists from a single type of experimental data analysis towards a holistic view.

1 Introduction

Brain Informatics (BI) is a new interdisciplinary field to study human information processing mechanism systematically from both macro and micro points of view by cooperatively using experimental brain/cognitive technology and Web Intelligence (WI) and Data Mining centric advanced information technology. In particular, it attempts to understand human intelligence in depth, towards a holistic view at a long-term, global field of vision to understand the principle, models and mechanisms of human multi-perception, language, memory, reasoning and inference, problem solving, learning, discovery and creativity [17].

Although brain sciences have been studied from different disciplines such as cognitive science and neuroscience, Brain Informatics (BI) represents a potentially revolutionary shift in the way that research is undertaken. As a crucial step in understanding human intelligence in depth, we must first fully master the mechanisms in which human brain operates. These results reported, over the last decade, about studying human information processing mechanism, are greatly related to progress of measurement and analysis technologies. Various noninvasive brain functional measurements are possible recently, such as fMRI and EEG. If these measurement data are analyzed systematically, the relationship between a state and an activity part will become clear. Furthermore, it is useful

to discover more advanced human cognitive models based on such measurement and analysis. Hence, new instrumentation and new data analysis methods are causing a revolution in both AI and Brain Sciences. The synergy between AI and Brain Sciences will yield profound advances in our understanding of intelligence over the coming decade [9,11].

In recent papers [8,14,15,16], we have reported an approach for modeling, transforming, and mining multiple human brain data obtained from visual and auditory psychological experiments by using fMRI and EEG. We observed that each method (fMRI and EEG) has its own strength and weakness from the aspects of time and space resolution. fMRI provides images of functional brain activity to observe dynamic activity patterns within different parts of the brain for a given task. It is excellent in the space resolution, but inferior time resolution. On the other hand, EEG provides information about the electrical fluctuations between neurons that also characterize brain activity, and measurements of brain activity at resolutions approaching real time. Hence, in order to discover new knowledge and models of human multi-perception activities, not only individual data source obtained from only single measuring method, but multiple data sources from various practical measuring methods are required.

It is also clear that the future of Brain Informatics will be affected by the ability to do large-scale mining of fMRI and EEG brain activations. The key issues are how to design the psychological and physiological experiments for obtaining various data from human information processing mechanism, as well as how to analyze such data from multiple aspects for discovering new models of human information processing. Although several human-expert centric tools such as SPM (MEDx) have been developed for cleaning, normalizing and visualizing the fMRI images, researchers have also been studying how the fMRI images can be automatically analyzed and understood by using data mining and statistical learning techniques [7,9,12]. Furthermore, spectral analysis [1] and wavelet analysis [5] are the main stream as the frequency analysis methods of EEG brain waves.

We are concerned to extract significant features from multiple brain data measured by using fMRI and EEG in preparation for multi-aspect data mining that uses various data mining techniques for analyzing multiple data sources. Our purpose is to understand activities of human information processing by

- investigating the features of fMRI brain images and EEG brain waves for every state or part;
- studying the neural structures of the activated areas to understand how a peculiar part of the brain operates and how it is linked functionally to individual differences in performance.

As a step in this direction, we observe that fMRI brain imaging data and EEG brain wave data extracted from human information processing mechanism are *peculiar* ones with respect to a specific state or the related part of a stimulus. Based on this point of view, we propose a way of *peculiarity oriented mining* for knowledge discovery in multiple human brain data, without using conventional imaging processing to fMRI brain images and frequency analysis to EEG

brain waves [8,14,16]. The proposed approach provides a new way for automatic analysis and understanding of fMRI brain images and EEG brain waves to replace human-expert centric visualization. The mining process is a multi-step one, in which various psychological experiments, physiological measurements, data cleaning, modeling, transforming, and mining techniques are cooperatively employed to investigate human information processing mechanism.

The rest of the paper is organized as follows. Section 2 describes the design of visual and auditory calculation experiments as a case study. Section 3 introduces our peculiarity oriented mining approach and its application in multiple brain data analysis is discussed in Section 4. Finally, Section 5 gives concluding remarks.

2 Visual and Auditory Calculation Experiments

In order to study the relevance between auditory and visual information processing in a more advanced information processing activity, we deal with fMRI brain images and EEG brain waves measured by the visual and auditory stimuli about human's calculation activities (2-figures addition). Each experiment took about 7 or 8 minutes, and we execute each of visual and auditory experiments 3 times, respectively. Thus, in the experiments, five states (tasks), namely, *auditory on-task*, *auditory off-task*, *visual on-task*, *visual off-task*, and *no-task*, exist by the difference in the stimulus given to a subject.

We try to compare and analyze how brain waves change along with the different tasks stated above. And in the experiments, we defined the position of electrode as shown in Fig. 1, which is an extension of the international 10-20 system. Moreover, the sampling frequency is determined as 500Hz in consideration of the ingredient of brain waves. On the other hand, in the fMRI experiments, the presenting speed of the visual and auditory stimuli are decided to identify the rate of correct answers for each subject by using the results of the rate of correct answers measured by traditional psychological experiments. For our experiments, the rate of correct answers both of the auditory and visual stimuli sets is 85-90%. Numbers of stimuli and control periods are decided by the presenting speed of the auditory and visual stimuli. Thus, we got 30 fMRI images in one experiment from one subject.

3 Peculiarity Oriented Mining (POM)

The main task of peculiarity oriented mining is the identification of peculiar data. An attribute-oriented method, which analyzes data from a new view and is different from traditional statistical methods, is recently proposed by Zhong *et al.* and applied in various real-world problems [10,13,14].

Peculiar data are a subset of objects in the database and are characterized by two features: (1) very different from other objects in a dataset, and (2) consisting of a relatively low number of objects. The first property is related to the notion of distance or dissimilarity of objects. Intuitively speaking, an object is different

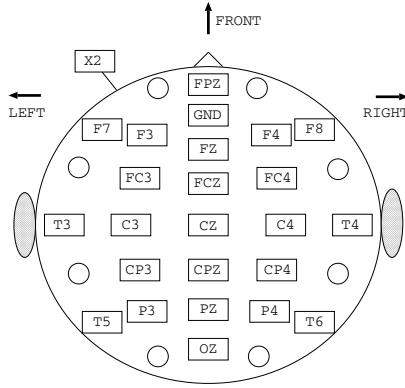


Fig. 1. Position of electrode

from other objects if it is far away from other objects based on certain distance functions. Its attribute values must be different from the values of other objects. One can define distance between objects based on the distance between their values. The second property is related to the notion of support. Peculiar data must have a low support.

At attribute level, the identification of peculiar data can be done by finding attribute values having properties (1) and (2). Let x_{ij} be the value of attribute A_j of the i -th tuple in a relation, and n the number of tuples. Zhong *et al* [13] suggested that the peculiarity of x_{ij} can be evaluated by a *Peculiarity Factor*, $PF(x_{ij})$,

$$PF(x_{ij}) = \sum_{k=1}^n N(x_{ij}, x_{kj})^\alpha \tag{1}$$

where N denotes the conceptual distance, α is a parameter to denote the importance of the distance between x_{ij} and x_{kj} , which can be adjusted by a user, and $\alpha = 0.5$ as default.

Based on peculiarity factor, the selection of peculiar data is simply carried out by using a threshold value. More specifically, an attribute value is peculiar if its peculiarity factor is above minimum peculiarity p , namely, $PF(x_{ij}) \geq p$. The threshold value p may be computed by the distribution of PF as follows:

$$\begin{aligned} threshold = & \text{mean of } PF(x_{ij}) + \\ & \beta \times \text{standard deviation of } PF(x_{ij}) \end{aligned} \tag{2}$$

where β can be adjusted by a user, and $\beta = 1$ is used as default. The threshold indicates that a data is a peculiar one if its PF value is much larger than the mean of the PF set. In other words, if $PF(x_{ij})$ is over the threshold value, x_{ij} is a peculiar data. By adjusting parameter β , a user can control and adjust threshold value.

4 Application in Multiple Human Brain Data Analysis

4.1 Data Modeling and Transformation

The existing multimedia data such as fMRI brain images and EEG brain waves might not be suitable for data mining. Hence, a key issue is how to transform such data into a unique representation format (i.e. table). For such transformation, we develop a system and cooperatively utilize a software tool called MEDx (SPM) to formalize, clean and conceptualize fMRI brain images, so that such images can be represented in a relational data model and stored in a relational database. Figure 2 shows examples of brain images transformed by using the MEDx. Furthermore, the ER (Entity-Relationship) model conceptualized fMRI brain images [14].

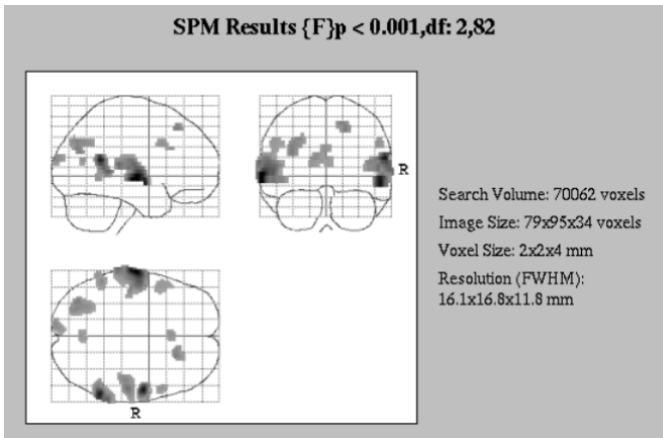


Fig. 2. Examples of brain images transformed by MEDx

Although the brain images transformed by using the MEDx can be represented in a relational data format, a problem is that the number of such data is too big. For instance, the number of the data for each subject is 122880 (i.e. 64 (pixels) \times 64 (pixels) \times 30 (images)). A way to reduce the number of data is to use a software tool called Talairach Daemon that is based on the Brodmann map as prior knowledge.

Brodmann assigned numbers to various brain regions by analyzing each area's cellular structure starting from the central sulcus (the boundary between the frontal and parietal lobes). Table 1 provides a general view of brain functions that refers to the Brodmann map as shown in Fig. 3. In our experiments, the Brodmann map is used as prior knowledge to obtain the Brodmann area values from the Talairach Daemon, before using our peculiarity oriented mining system to analyze the visual and auditory calculation related databases.

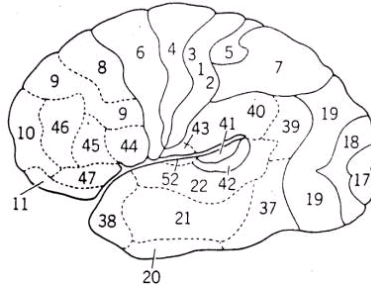


Fig. 3. Brodmann human brain map

Table 1. Functional areas with respect to the Brodmann human brain map

Function	Brodman Area
Vision	17, 18, 19
Audition	22, 41, 42
Body Sensation	1, 2, 3, 5, 7
Sensation, tertiary	7, 22, 37, 39, 40
Motor	4, 6, 8, 44
Motor, tertiary	9, 10, 11, 45, 46, 47

4.2 The Mining Process of fMRI Imaging Data

Two peculiarity oriented mining processes, namely mining (1) and mining (2), are carried out on the fMRI data, respectively. In the process of mining (1), the prior knowledge of Brodmann areas is used before peculiarity oriented analysis. This process can be divided into the following steps [14]:

- Step 1.* Formalize/transform fMRI data by using the MEDx system.
- Step 2.* Obtain the Talairach Daemon coordinate and active values from the MEDx data.
- Step 3.* Get Brodmann area values from the Talairach Daemon.
- Step 4.* Create the visual and auditory calculation related databases, respectively, by using the Brodman area values. If there are multiple values in a Brodman area, use the sum of the values.
- Step 5.* Carry out peculiarity-oriented mining in the visual and auditory calculation related databases, respectively.
- Step 6.* Compare/evaluate the results.

On the other hand, in the process of mining (2), our peculiarity oriented mining is carried out on the fMRI data transformed in MEDx, directly (i.e. without prior knowledge of Brodmann areas is used before data mining), so that we will be able to analyze and compare the results of using the Brodmann area or not.

4.3 The Mining Process of Brain Waves Data

The mining process of brain-wave data as shown in Fig. 4 is a multi-step one with data modeling and transformation. Brain wave data are typically a kind of time series data with noises. Hence, the removal of noises and data modeling are required. Usually, the low pass filter (LPF) and the band pass filter (BPF) are employed to remove noises by frequency analysis. However, we applied the FIR band pass filter that does not need to take into account the gap of a phase, because we paid our attention to the form of brain waves.

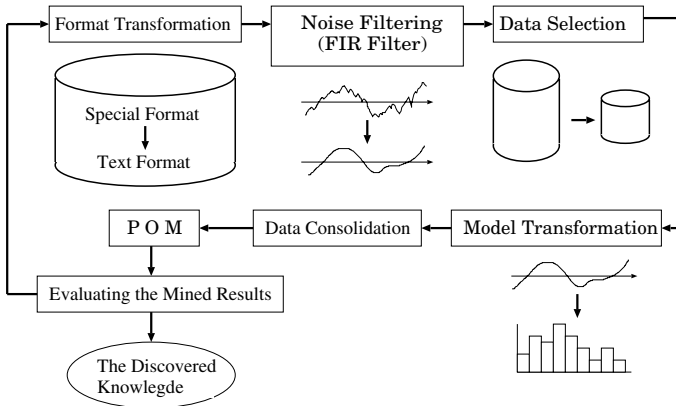


Fig. 4. The mining process of brain wave data

Brain wave data are clustered for every task after noise removal. Furthermore, it is necessary to differentiate valid brain waves from invalid ones, because the eye movement like a blink has a remarkable influence on brain waves. The 512 data per sample were extracted from the effective data. Thus we were able to obtain about 120 samples for each task.

The next important work is the model transformation of brain wave data. A problem of using peculiarity oriented mining in time series data is that it is influenced by a phase. That is, it is necessary to shift the cut area of each data so that the correlation may become strong. Hence, we change the wave data into 2-variate histogram that makes slope of a line and a potential in order to leverage the feature of raw data. Such 2-variate histogram data with respect to a channel can be collected in a two-dimension table as shown in Fig. 5, where v_i denotes the class value of a potential, and d_j denotes the class value of a slope. Furthermore, the frequency of appearance is denoted in a rate a_{ij} , so that they can be compared equally since the number of each task is different.

In order to investigate how subject A’s brain waves are different from subject B’s brain waves, it is necessary to calculate the difference between histograms. For example, let p be the number of subject A’s sample, q be the number of subject B’s sample, and x_{ij} be the difference of 2 histograms (histograms A

potential [μV]	v_1	a_{11}	a_{12}	a_{13}	\dots	a_{1m}
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	v_3	a_{31}	a_{32}	a_{33}	\dots	a_{3m}
	v_2	a_{21}	a_{22}	a_{23}	\dots	a_{2m}
	v_1	a_{11}	a_{12}	a_{13}	\dots	a_{1m}
		d_1	d_2	d_3	\dots	d_m
		slope [$\mu V/2msec$]				

Fig. 5. The sample of 2-variate histogram

and B). It is not possible to compare each of them by the frequency equally, because the number of samples of two tasks is different. Hence, frequency a_{ij} and frequency b_{ij} are converted into percentage $\overline{a_{ij}}$ and percentage $\overline{b_{ij}}$, respectively, such as $\overline{a_{ij}} = (100 \times a_{ij}) / (512 \times p)$ and $\overline{b_{ij}} = (100 \times b_{ij}) / (512 \times q)$. Thus, the difference can be computed as $x_{ij} = \overline{a_{ij}} - \overline{b_{ij}}$. As a result, the new histograms are able to be utilized as the data chunks for peculiarity oriented mining. Such histogram data with respect to a channel can be collected in a two-dimension table as shown in Fig. 6, where z is the maximum number of channels and is set to 24 for our experiments.

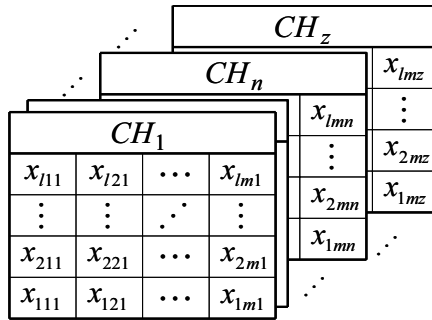


Fig. 6. Structure of mining data

Since our purpose is to find the *peculiar channel* by finding the difference of the states in the transformed datasets, the formula of calculating peculiarity factor needs to be extended from Eq. (1) into Eq. (3).

$$PF(CH_n) = \sum_{k=1}^z \sum_{j=1}^m \sum_{i=1}^l N(x_{ijn}, x_{ijk})^\alpha \tag{3}$$

which means calculating the distances among channels after the calculation of distances for histogram data in each channel. After the calculation of peculiarity

factors of all 24-channels, the threshold value is computed for judging whether some channel is peculiar one or not by Eq. (2) in which the β is set to 0.1 for our experiments.

5 Concluding Remarks

We presented a new methodology for investigating human multi-perception mechanism by combining cognitive technology and peculiarity oriented mining. We showed that multi-aspect analysis in multiple data sources is an important data mining methodology in Brain Informatics. The proposed methodology attempts to change the perspective of cognitive scientists from a single type of experimental data analysis towards a holistic view.

Since this project is very new, we just had preliminary results [8,14,15,16]. Our future work includes studying the neural structures of the activated areas and trying to understand how a peculiar part of the brain operates and how it is linked functionally to individual differences in performance by combining various mining methods with reasoning. Some of lessons in cognitive science and neuroscience are applicable to novel technological developments in Brain Informatics, yet others may need to be enhanced or transformed in order to manage and account for the complex and possibly more innovative practices of sharing, analyzing and creating data/knowledge that are made technically possible by the Wisdom Web and Knowledge Grids [6,15].

Acknowledgments

I am grateful to all my research collaborators, assistants, and students who have, over the years, together contributed to the development of Web Intelligence (WI) and Brain Informatics (BI). I would like to express my gratitude to Jiming Liu, Yiyu Yao, and Jinglong Wu for our joint projects and discussions. Finally, I would like to acknowledge the support of the following research grants: (1) the grant-in-aid for scientific research on priority area “Active Mining” from the Japanese Ministry of Education, Culture, Sports, Science and Technology, (2) Open Foundation of Beijing Municipal Key Laboratory for Multimedia and Intelligent Software Technology (KP0705200379), (3) Maebashi Institute of Technology Faculty Research Grants.

References

1. J.P. Banquet, “Spectral Analysis of the EEG in Meditation”, *Electroencephalography and Clinical Neurophysiology*, 35, (1973) 143-151.
2. J. Banfield and A. Raftery, “Model-based Gaussian and Non-Gaussian Clustering”, *Biometrics*, 49 (1993) 803-821.
3. L.A. Consularo, R.A. Lotufo, and L.F. Costa, “Data Mining Based Modeling of Human Visual Perception”, K.J. Cios (ed) *Medical Data Mining and Knowledge Discovery*, Physica-Verlag (2001) 403-431.

4. S. Cerutti, G. Chiarenza, D. Liberati, P. Mascellani, and G. Pavesi, "A Parametric Method of Identification of Single Trial Event-related Potentials in the Brain", *IEEE Trans. Biomed. Eng.*, 35(9) (1988) 701-711.
5. R. Hornero, M. Martin-Fernandez, A. Alonso, A. Izquierdo, and M. Lopez, "A DSP Implementation of Wavelet Transform to Detect Epileptiform Activity in the EEG", *Proc. 8th Annual International Conference on Signal Processing Applications and Technology, ICAPAT'97* (1997) 692-696.
6. J. Liu, "Web Intelligence (WI): What Makes Wisdom Web?", *Proc. 18th International Joint Conference on Artificial Intelligence (IJCAI'03)* (2003) 1596-1601.
7. V. Megalooikonomou and E.H. Herskovits, "Mining Structure-Function Associations in a Brain Image Database", K.J. Cios (ed.) *Medical Data Mining and Knowledge Discovery*, Physica-Verlag (2001) 153-179.
8. S. Motomura, N. Zhong, and J.L. Wu, "Brain Waves Data Mining for Human Multi-perception Activity Analysis", *Proc. Inter. Workshop on Advanced Technologies for e-Learning and e-Science (ATELS'04)* (2004) 65-72.
9. T.M. Mitchell, R. Hutchinson, M. Just, and R.S. Niculescu, F. Pereira, and X. Wang, "Classifying Instantaneous Cognitive States from fMRI Data", *Proc. American Medical Informatics Association Annual Symposium* (2003) 465-469.
10. M. Ohshima, N. Zhong, Y.Y. Yao, and S. Murata, "Peculiarity Oriented Analysis in Multi-people Tracking Images", H. Dai et al (eds) *Advances in Knowledge Discovery and Data Mining (PAKDD'04)*, LNAI 3056, Springer (2004) 508-518.
11. R.J. Sternberg, J. Lautrey, and T.I. Lubart, *Models of Intelligence*, American Psychological Association (2003).
12. H. Tsukimoto and C. Morita, "The Discovery of Rules from Brain Images", *Proc. First Inter. Conf. on Discovery Science*, LNAI 1532, Springer (1998) 198-209.
13. N. Zhong, Y.Y. Yao, and M. Ohshima, "Peculiarity Oriented Multi-Database Mining", *IEEE Transaction on Knowledge and Data Engineering*, 15(4) (2003) 952-960.
14. N. Zhong, J.L. Wu, A. Nakamaru, M. Ohshima, and H. Mizuhara, "Peculiarity Oriented fMRI Brain Data Analysis for Studying Human Multi-Perception Mechanism", *Cognitive Systems Research*, 5(3), Elsevier (2004) 241-256.
15. N. Zhong, J. Hu, S. Motomura, J.L. Wu, and C. Liu, "Building a Data Mining Grid for Multiple Human Brain Data Analysis", *Computational Intelligence*, 21(2), Blackwell Publishing (2005) 177-196.
16. N. Zhong, S. Motomura, and J.L. Wu, "Peculiarity Oriented Multi-Aspect Brain Data Analysis for Studying Human Multi-Perception Mechanism", *Proc. SAINT 2005 Workshops (Workshop 8: Computer Intelligence for Exabyte Scale Data Explosion)*, IEEE Computer Society Press, (2005) 306-309.
17. N. Zhong, J. Liu, Y.Y. Yao, and J. Wu, "Web Intelligence (WI) Meets Brain Informatics (BI)", *Proc. CME'05* (in press).