Agent-Enriched Data Mining: A Case Study in Brain Informatics

Ning Zhong and Shinichi Motomura, *Maebashi Institute of Technology*

An agent-enriched peculiarity-oriented mining approach demonstrates the brain informatics methodology by transforming and mining human-brain data obtained from cognitive event-related potential experiments.

To understand human intelligence in depth, we must first master the brain's operation mechanisms. Ignoring the brain's activity and focusing instead on behavior has seriously impeded our ability to understand how human beings accomplish complex adaptive and distributed problem-solving and reasoning.[1,2] Research from the last decade on how humans process information[3–5] has led to advances in measurement and analysis technologies. Recently, researchers have introduced various noninvasive brain functional measurements, including event-related potential/electroencephalography (ERP/EEG) and functional magnetic resonance imaging (fMRI). Systematically analyzing this measurement data lets us clarify the relationship between a state and an activity. We can also use such measurement and analysis to develop more advanced human cognitive models. Hence, new instrumentation and data analysis methods are creating a revolution in both AI and brain sciences. The synergy between the two fields promises to yield profound advances in our understanding of intelligence over the coming decade.[3,6,7]

Brain informatics (BI) is a new interdisciplinary field that systematically studies the human information-processing mechanism from macro and micro viewpoints. It does this using experimental, computational cognitive neuroscience technologies and Web-intelligence-centric advanced information technology. In particular, BI attempts to understand human intelligence in depth to support a long-term, holistic vision to uncover the principle and mechanisms underlying human information-processing systems (HIPS).

The ability to perform large-scale analysis and simulation of brain data will shape BI's future. Current research focuses on two key questions:

- How can we design psychological and physiological experiments to systematically obtain various data from HIPS?
- How can we manage and analyze such data from multiple viewpoints to discover new models of HIPS?

Researchers have developed expert tools—such as the Brain Vision Analyzer and MEDx with statistical parametric mapping—for cleaning, normalizing, and visualizing ERP and fMRI data, respectively. They've also studied how to analyze and understand ERP and fMRI data using data mining and statistical learning techniques.[3-5,8]

To understand human information processing (IP) principles and mechanisms relating to higher cognitive functions—such as problem solving, reasoning, and learning—we must develop new brain data-mining

1541-1672/09/$25.00 © 2009 IEEE
Published by the IEEE Computer Society

techniques based on the BI methodology. The human brain is too complex for a single data mining algorithm. Agent-enriched brain data mining for multi-aspect data analysis is thus a key BI methodology for analyzing all available cognitive experimental data. We've developed an agent-enriched peculiarity-oriented mining (A-POM) approach for multi-aspect ERP data analysis. We present our approach here, along with a case study that demonstrates the BI methodology.

Overview: Brain Informatics Methodology

Researchers have traditionally studied brain sciences using various disciplines, including cognitive science and neuroscience. BI, however, represents a shift in brain research. We can regard BI as brain sciences in the Web-intelligence-centric IT age,[2,6,7] studying the human brain from the informatics viewpoint—that is, studying the brain as a HIPS.

BI researchers use informatics to support brain science studies and attempt to capture new forms of collaborative and interdisciplinary work. Thus, new kinds of BI methods and global research communities will emerge through the wisdom Web, an enormous, intelligent organism that will use data, information, knowledge, and a wisdom hierarchy to move toward human-level Web intelligence reality.[9] These new BI methods will incorporate knowledge grids, which will enable high-speed, large-scale, distributed agent-based analysis and computations, as well as radical new ways of sharing data and knowledge. Despite these changes, lessons from both cognitive science and neuroscience remain applicable to BI's novel technological developments.[2,6,10]

BI emphasizes a *systematic* approach to investigating human IP mechanisms, including measuring, collecting, modeling, transforming, managing, and mining brain data obtained from various cognitive experiments. Such systematic study currently focuses on four main research questions:

- How do thinking-centric brain mechanisms work?
- How can we best design cognitive experiments?
- How can we manage brain data in an integrated way?
- How can we analyze brain data deeply and systematically?

Brain informatics emphasizes a systematic approach to investigating human information-processing mechanisms.

In the first case, we can broadly divide the capabilities of human intelligence into two main aspects: perception and thinking. Cognitive neuroscience researchers have achieved advanced results in perception-oriented studies, but have reported only a few separate, preliminary studies that were thinking-oriented or focused on the overall human IP process.[11] Systematic investigation of thinking-centric mechanisms is therefore based on both Web intelligence research and state-of-the-art cognitive neuroscience.[1,2,7,11,12]

Second, to systematically design cognitive experiments, we must design tasks for both psychological and physiological experiments. From these, we can systematically obtain HIPS data for use in multipurpose investigations of human thinking- and perception-centric cognitive functions. To discover new knowledge and models of human IP activities, we must use multiple data sources and practical measuring methods, such as ERP and fMRI. Furthermore, we must systematically design cognitive experiments so the resulting data is useful for multiple purposes.

Third, we can investigate how to manage the brain data by using a conceptual brain data model. This model represents functional relationships among multiple brain data sources with respect to all major HIPS aspects and capabilities. Such data representation offers multilevel modeling, abstraction, and transformation for multi-aspect analysis and simulation.

Finally, to systematically examine how to analyze the brain data deeply, we can extract significant patterns and features from multiple brain data sources obtained by using powerful tools, such as ERP and fMRI, and then engage in multi-aspect data analysis by combining various data mining and reasoning methods.[3,5,6,12] We can also deploy agents for data preprocessing, mining, reasoning, and simulation in a multiphase process to achieve multi-aspect analysis and multilevel conceptual abstraction and learning.

By addressing each of these key research areas, the BI framework combines analysis and simulation to understand human intelligence in depth; agent-enriched data mining will play a central role in its multiphase process.

Case Study: Agent-Enriched Peculiarity-Oriented Mining

Our work focuses on human IP activities at two levels: spatiotemporal features and flow based on functional relationships among activated brain areas for various tasks; and the neural structures and neurobiological processes related to those activated areas. More specifically, we're trying to understand how neurobiological processes support a cognitive process based on BI methodology. We're thus investigating how a specific part of the brain operates at a specific time, how those operations change over time, and how the activated areas work cooperatively to implement an overall IP system.

As a step in this direction, we're studying ERP data. Such data is *peculiar* with respect to a specific state or the related part of a stimulus. To automate ERP data analysis and understanding, we propose the A-POM knowledge-discovery approach. A-POM doesn't use conventional ERP analysis and doesn't require human-expert-centric visualization.[6] Instead, it investigates the human IP mechanism through a multistep mining process that cooperatively employs various psychological experiments, physiological measurements, data cleaning, modeling, transforming, managing, and mining techniques.

A-POM has two main benefits for addressing the complexity and diversity of brain data and applications:

- A-POM agents cooperate in a multiphase process and support multilevel conceptual abstraction and learning.
- This agent-based approach supports task decomposition for distributed data mining.

Researchers can apply our methodology to interpret a HIPS's spatiotemporal features and flow. In the cognitive process from perception (in our case, a cognitive task stimulated by vision) to thinking (computation), the A-POM system collects data from several event-related points in time and transforms them into various forms suitable for multi-aspect data analysis. The system then explains the results of the separate analyses and synthesizes them into an overall flow.

Multi-Aspect ERP Data Analysis

We identified the best use of each feature using two aspects of ERP data analysis—the potential change and the frequency element—and experiments with multiple difficulty levels. A-POM can also find interesting temporal and spatial features in ERP data using the potential change and frequency aspects.

It's clear that a specific brain section operates in a specific time and those operations change over time. Although it's easy to detect ERP data's concavity and convexity (P300 and so on) using an existing tool, it's difficult to find peculiar data when there are multiple channels with concavity and convexity.[13] Thus, to gain new knowledge and develop new models of human IP activities, we must attend to both the peculiar channel and peculiar time of ERPs to investigate the HIPS's spatiotemporal features and flow.

Agent-enriched peculiarity-oriented mining can find interesting temporal and spatial features in event-related potential data.

There are many ways to find data peculiarities.[14–18] Researchers have applied POM's attribute-oriented method, which analyzes data from a new viewpoint and differs from traditional statistical methods, in various real-world problems.[8,18]

Defining peculiar data. "Peculiar" data is a subset of database objects with two characteristics:

1. They're clearly different from the other data set objects.
2. They constitute a relatively low percentage of the total objects.

The first property relates to the objects' distance or dissimilarity. Intuitively speaking, an object is different from other objects if it's regarded as far away from them on the basis of certain distance functions. The peculiar object's attribute values must differ from those of other objects. We can therefore define distance between objects based on the distance between their values.

The second property relates to the notion of support. Peculiar data's support must be low frequency. The brain doesn't directly compare one object to another; it first recognizes objects by comparing them to stored representations.[19] However, as we describe later, we use a simplified method of comparison.

Identifying peculiar data. At the attribute level, the system can identify peculiar data by finding attribute values with properties (1) and (2) above. Let x_{ij} be the value of attribute A_j of the i-th tuple in a relation, and n be the number of tuples. We can evaluate x_{ij}'s peculiarity using a peculiarity factor, $PF(x_{ij})$:

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

where N denotes the conceptual distance, α is a parameter to denote the importance of the distance between x_{ij} and x_{kj} (which users can adjust), and $\alpha = 0.5$ is the default.

On the basis of the PF, we simply use a threshold value to select peculiar data. More specifically, an attribute value is peculiar if its peculiarity factor is over a minimum peculiarity p—namely, $PF(x_{it}) \geq p$. We can compute the threshold value p by the distribution of PF as follows:

$$\begin{aligned} p = {} & \text{mean of } PF\left(x_{it}\right) + \\ & \beta \times \text{standard deviation of } PF\left(x_{it}\right) \end{aligned} \tag{2}$$

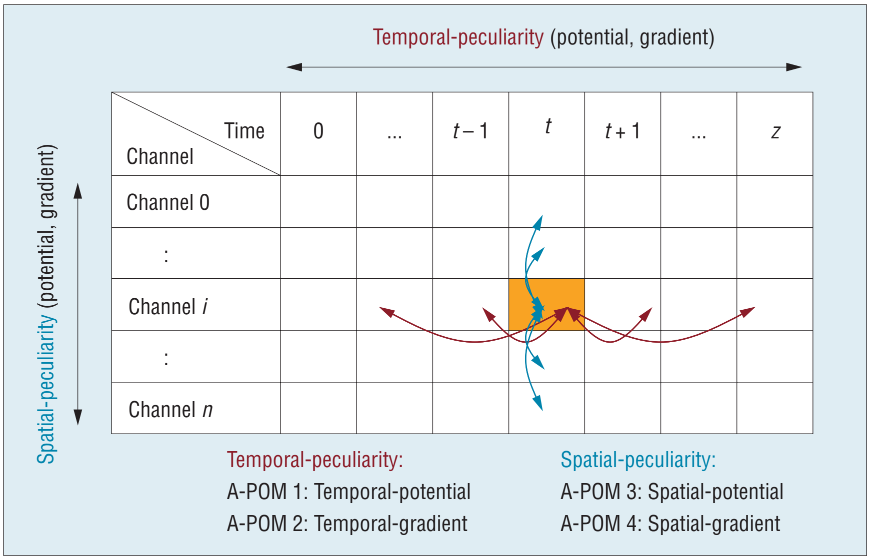

Figure 1. Event-related potential (ERP) peculiarities and multi-peculiarity-oriented mining (POM) agents. Four kinds of POM agents (A-POM 1 to 4) are used to find peculiar patterns, which are classified into two types of peculiarities with respect to the temporal (time) and spatial (channel) axes.

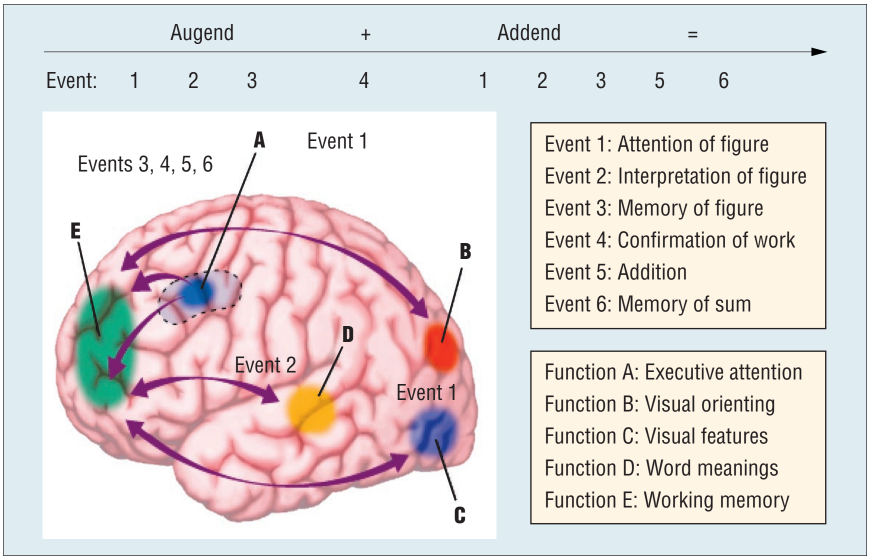

Figure 2. Human computation processing. Such processing consists of several component functions, including attention, interpretation, short-term memory, understanding of work, computation, and checking. The top-most legend in the figure denotes the cognitive task of addition (`augend + addend =`) and gives related descriptions of various cognitive events and functions corresponding to the task.

where β can be adjusted by a user, and $\beta = 1$ is the default. By adjusting the parameter β, users can control and adjust the threshold value.

Unfortunately, such a POM is not totally fit for ERP data analysis. The reason is that, for ERP data analysis, the useful aspect is latent time, not amplitude. To solve this problem, in addition to using POM for ordinary ERP potential data analysis, we also analyze the gradient data derived by difference of potential. However, when the raw data's sampling frequency is high or noise removal is insufficient, noise contaminates the gradient data. To solve this, we first block potential data each 50 milliseconds and then use the average potential data in the block for POM analysis.

The role of POM agents. Having thus prepared, we look for peculiar patterns using four kinds of POM agents (see Figure 1). We classify these patterns according to two types of peculiarities: temporal (time) and spatial (channel).

- A-POM 1 (time/potential mining) examines whether the potential at a specific time is peculiar when compared with other times in channel i,
- A-POM 2 (time/gradient mining) examines whether the gradient at a specific time is peculiar when compared with other times in channel i,
- A-POM 3 (spatial/potential mining) examines whether the potential of channel i is peculiar when compared with the potential on other channels in a specific time t,
- A-POM 4 (spatial/gradient mining) examines whether the gradient of channel i is peculiar when compared with the gradient on other channels in a specific time t.

We can use these four agent types on the spatiotemporal data in all ERP channels in a distributed cooperative mode.

Experimental Results

Explaining and integrating our A-POM-based multi-aspect analysis results is a key issue. We do this in four distinct steps. First, we examine an integrated model of the results in relation to spatiotemporal features. As Figure 2 shows, we use an example of computation processing from the macro viewpoint, which consists of several component functions of the human computation mechanism, including attention, interpretation, short-term memory, understanding of work, computation, and checking.

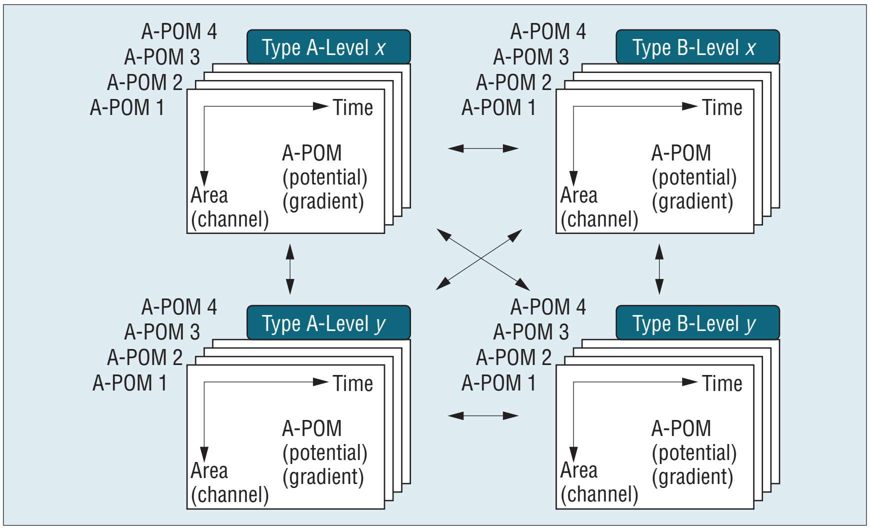

Figure 3. The multi-agent peculiarity-oriented mining (POM) analysis integration model. These POM agents (A-POM 1–4) can act on multiple event-related potential (ERP) data sources with different difficulty levels and task types in a distributed cooperative mode.

We use each cognitive experiment's resulting data for multiple purposes, in keeping with the BI methodology. Our experiments might, for example, satisfy the requirements for investigating the mechanisms of human visual and auditory systems, computation, problem solving (that is, we regard the computation processing as an example of problem-solving processes), and HIPS general spatiotemporal features and flow.

Furthermore, we set up two types of cognitive experiments with respect to a series of computation tasks. The two types differ in terms of visual attention:

- Type A: numbers remain on the screen.
- Type B: the subjects must continue to remember numbers while the numbers on the screen change.

For the spatiotemporal viewpoint, we collect ERP data of 128 EEG channels from six event-related points in time with respect to six computational processing stages: `Augend 1`, `Augend 2`, "+", `Addend 1`, `Addend 2`, and "=".

Figure 3 provides a global view of our proposed model for integrating and explaining our results. The horizontal axis denotes time and the vertical axis denotes area (channel). Because we manage our results in a layered structure, it's easy to discover a unique phenomenon and new knowledge. On the other hand, given the difference in spatiotemporal resolution in multi-aspect data analysis, it's important that we adequately collate the representation of results.

We've applied the A-POM-based approach to all ERP channels with multiple difficulty levels for finding peculiar channels and peculiar time bands. We obtained some remarkable results by comparing results from Type A and Type B experiments.

As an example of human-expert-centric visualization, Figure 4 shows a computation process's spatiotemporal feature, represented by Type A and Type B topographies. We obtain these topographies by adding the average of seven subjects using an ordinary EEG analysis tool. When human experts read such figures, they naturally note the points of the noticeable positive and negative potentials in a spatiotemporal mode. Our A-POM-based peculiarity analysis supports automatic judgment on the location of the more remarkable points—that is, the peculiar ones—to enrich the expert-centric visualization.

Figure 5 shows the *PF* values of two channels, F8 and P6, in a computation process for Type A and Type B, respectively. The reason we selected F8 and P6 as examples is that their peculiarity tendencies are the most different among the four kinds of mining agents (A-POM 1 to 4). In our analyses, we normalized the peculiarity threshold to 100—that is, we judge a data as peculiar if its peculiarity factor is more than 100. Although both channels' *PF* values increase when a stimulus is presented, our experiments show that the two channels' peculiarities differ among the agent types and when comparing Type A and Type B experiments. As for P6, the electric potential is almost flat, except when a stimulus is presented. Hence, the *PF* values of A-POM 1 and 2, which detect temporal peculiarity, suddenly increase when a stimulus is presented, while the *PF* values of A-POM 3 and 4 are almost low because they detect spatial peculiarity. Furthermore, F8 isn't affected by a stimulus presentation, and it has long time bands for both positive and negative potentials. Hence, we can regard F8 as a channel with a higher spatial peculiarity.

We can obtain remarkable results when we compare Type A and Type B experiments. Figure 6 shows results of A-POM integration for ERP differential data (Type A and Type B) in two-digit mental arithmetic tasks. The top of the figure shows the ERP differential data's topography. Furthermore, we transformed the ERP data into various forms for four kinds of A-POM-based multi-aspect analysis corresponding to five brain areas: frontal lobe, left temporal lobe, parietal lobe, right temporal lobe, and occipital lobe. Each block in the time axis denotes 200 milliseconds; the plot density shows the average *PF* values during that time. Thus, if more blocks have deep color in one of the brain areas, that area has more peculiar data. Furthermore, orange denotes that Type

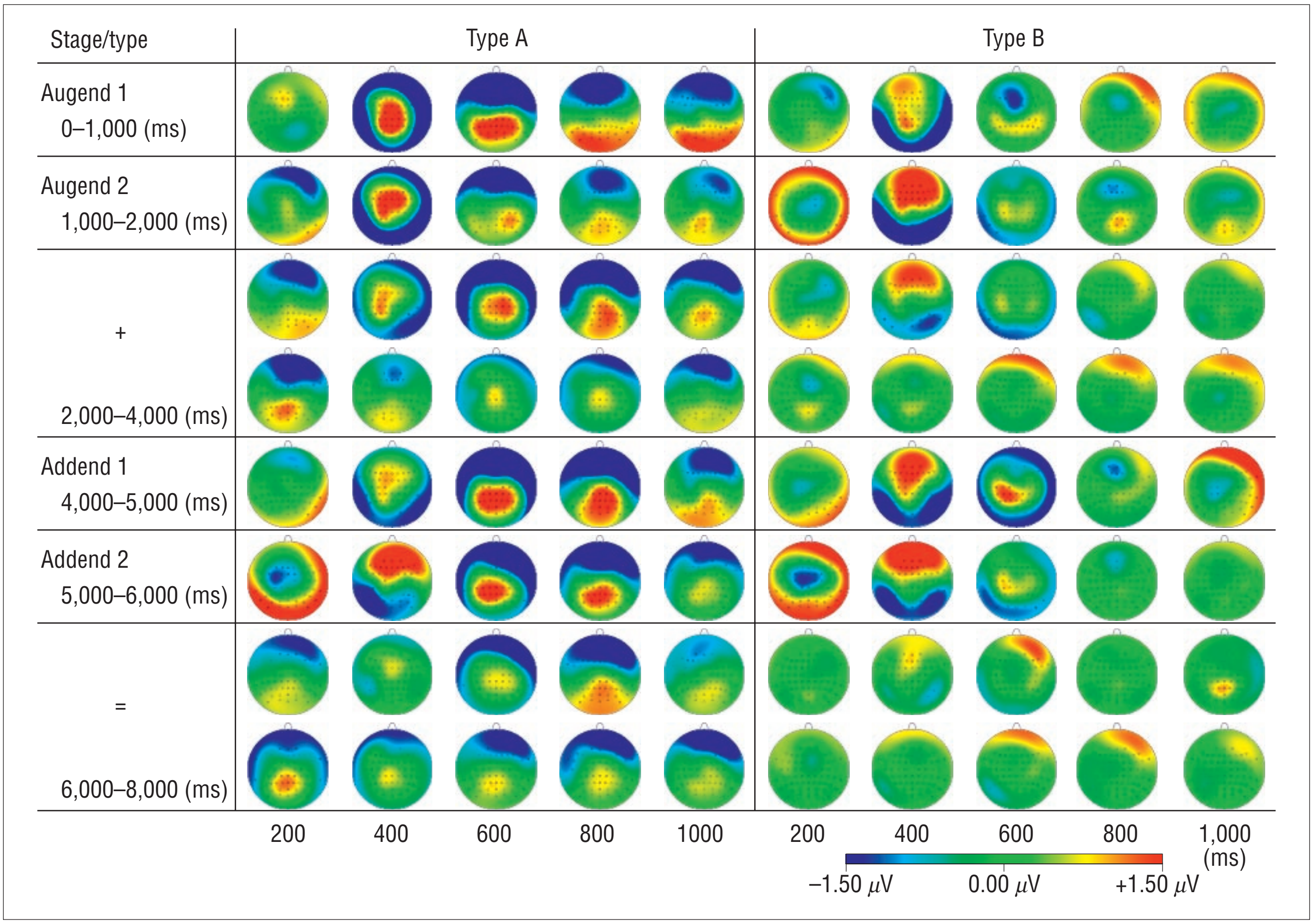

Figure 4. The spatiotemporal feature of a computation process represented by Type A and Type B topographies. Red denotes noticeable positive potentials and blue denotes noticeable negative potentials with respect to a specific time at the computation stage for a Type A or B task.

A has a positive value, while blue denotes that Type B has a positive value when we derive the difference between the two experiment types. In other words, a white block is uninteresting from a spatiotemporal viewpoint.

As Figure 6 shows, many parietal and occipital lobe blocks are deep orange with respect to the timing of presenting stimuli. This means that stronger potential appeared in Type A over Type B. In contrast, the same measure shows many deep blue blocks in the frontal and right temporal lobes. Thus, stronger potential appeared in Type B in these regions. We can obtain the same results by observing Figure 4's topographies.

When we analyze the results after the process has been running for 4,600 ms and 5,000 ms, we see that A-POM 3 (spatial potential) has high peculiarity in both the frontal lobe and the right temporal lobe between the two time measures. The areas' potentials differ greatly in Type A and Type B experiments, which we can see from a spatial viewpoint in Figure 4. On the other hand, both A-POM 1 and A-POM 2 (temporal peculiarity) have high peculiarity between 5,000 and 5,400 ms. By referring to Figure 4, we see that the Type A and Type B topographies are quite similar. Hence, this time zone's feature is apparently distinct from that of other time zones.

Our A-POM-based ERP data mining case study demonstrates the BI methodology's usefulness. By using the A-POM-based multi-aspect analysis, we can evaluate data from the peculiarity viewpoint, and thus offer a new way to investigate HIPS's spatiotemporal features and flow. Our methodology attempts to broaden cognitive and brain scientists' perspective from a single type of experimental data analysis toward a long-term, holistic vision that can reveal the underlying HIPS's principles and mechanisms.

Acknowledgments

We measured the EEG data with support from the International WIC Institute (Beijing University of Technology). We thank Shengfu Lu, Peipeng Liang, and other colleagues for their support and assistance. Our work is partially supported by a Japanese Society for the Promotion of Science grant-in-aid for scientific research

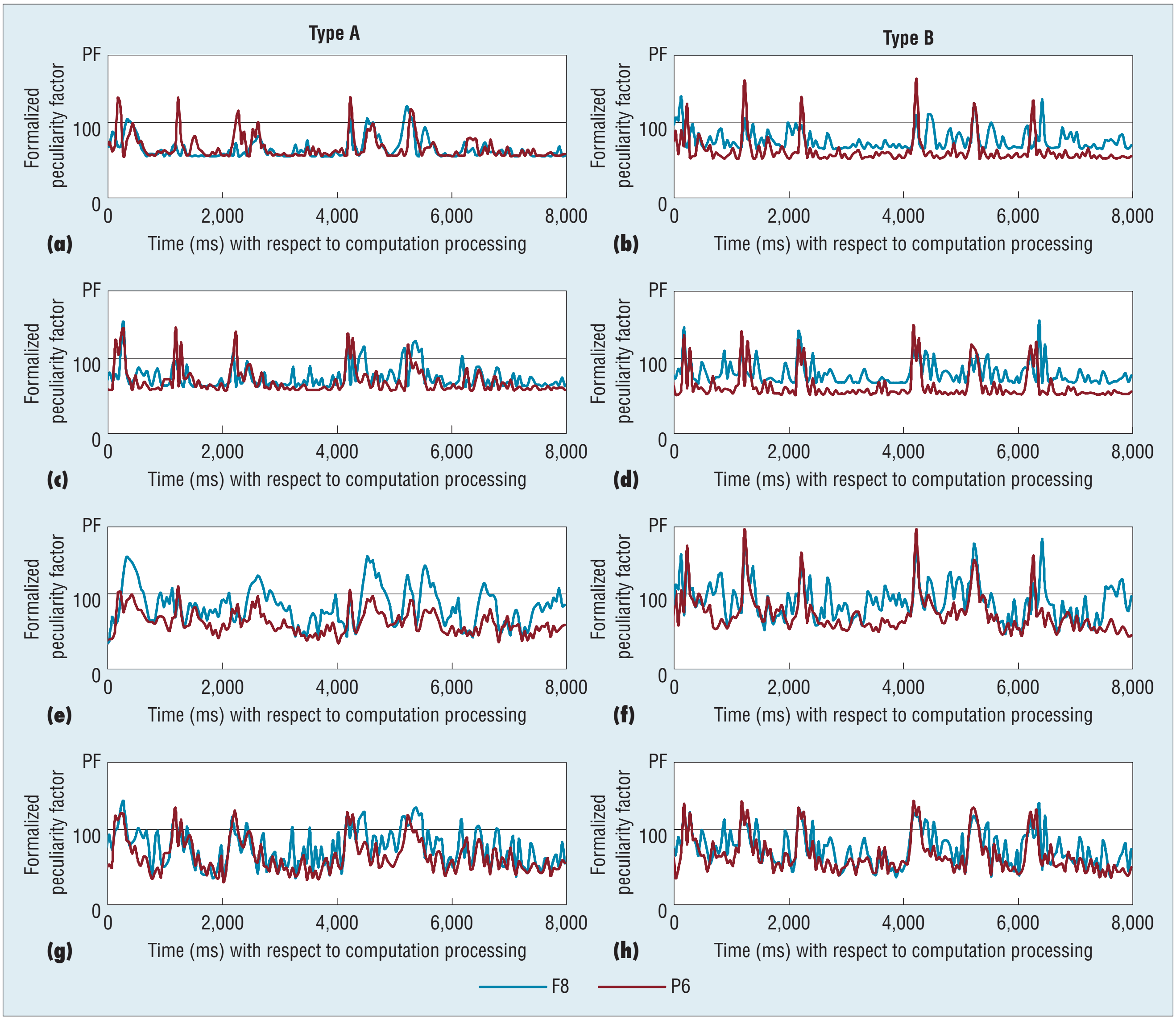

Figure 5. Examples of agent-enriched peculiarity-oriented mining (A-POM) analysis. Two example channels, F8 and P6, show the results of mining agents (a, b) A-POM 1, (c, d) A-POM 2, (e, f) A-POM 3, and (g, h) A-POM 4 for Type A and Type B experiments.

(18300053) and by the Support Center for Advanced Telecommunications Technology Research Foundation.

References

1. R. Morris, L. Tarassenko, and M. Kenward, eds., *Cognitive Systems: Information Processing Meets Brain Science*, Elsevier, 2006.
2. N. Zhong, "Impending Brain Informatics (BI) Research from Web Intelligence (WI) Perspective," *Int'l J. Information Technology and Decision Making*, vol. 5, no. 4, 2006, pp. 713–727.
3. T.M. Mitchell et al., "Predicting Human Brain Activity Associated with the Meanings of Nouns," *Science*, vol. 320, 2008, pp. 1191–1195.
4. T.C. Handy, *Event-Related Potentials, A Methods Handbook*, MIT Press, 2004.
5. F.T. Sommer and A. Wichert, eds., *Exploratory Analysis and Data Modeling in Functional Neuroimaging*, MIT Press, 2003.
6. N. Zhong and S. Motomura, "WI Based Multi-Aspect Data Analysis in a Brain Informatics Portal," *Autonomous Intelligent Systems: Agents and Data Mining*, V. Gorodetsky et al., eds., LNAI 4476, Springer, 2007, pp. 46–59.
7. N. Zhong et al., "Web Intelligence Meets Brain Informatics," N. Zhong et al., eds., *Web Intelligence Meets Brain Informatics*, LNAI 4845, Springer, 2007, pp. 1–31.
8. N. Zhong et al., "Peculiarity Oriented fMRI Brain Data Analysis for Studying Human Multi-Perception Mechanism," *Cognitive Systems Research*, vol. 5, no. 3, Elsevier, 2004, pp. 241–256.
9. N. Zhong, J. Liu, and Y.Y. Yao, "Envisioning Intelligent Information Technologies through the Prism of Web Intelligence," *Comm. ACM*, vol. 50, no. 3, 2007, pp. 89–94.
10. N. Zhong et al., "Building a Data Mining Grid for Multiple Human Brain Data Analysis," *Computational Intelligence*, vol. 21, no. 2, Blackwell, 2005, pp. 177–196.
11. M.S. Gazzaniga, ed., *The Cognitive Neurosciences III*, MIT Press, 2004.
12. J.R. Anderson, *How Can the Human Mind Occur in the Physical Universe?*, Oxford Univ. Press, 2007.

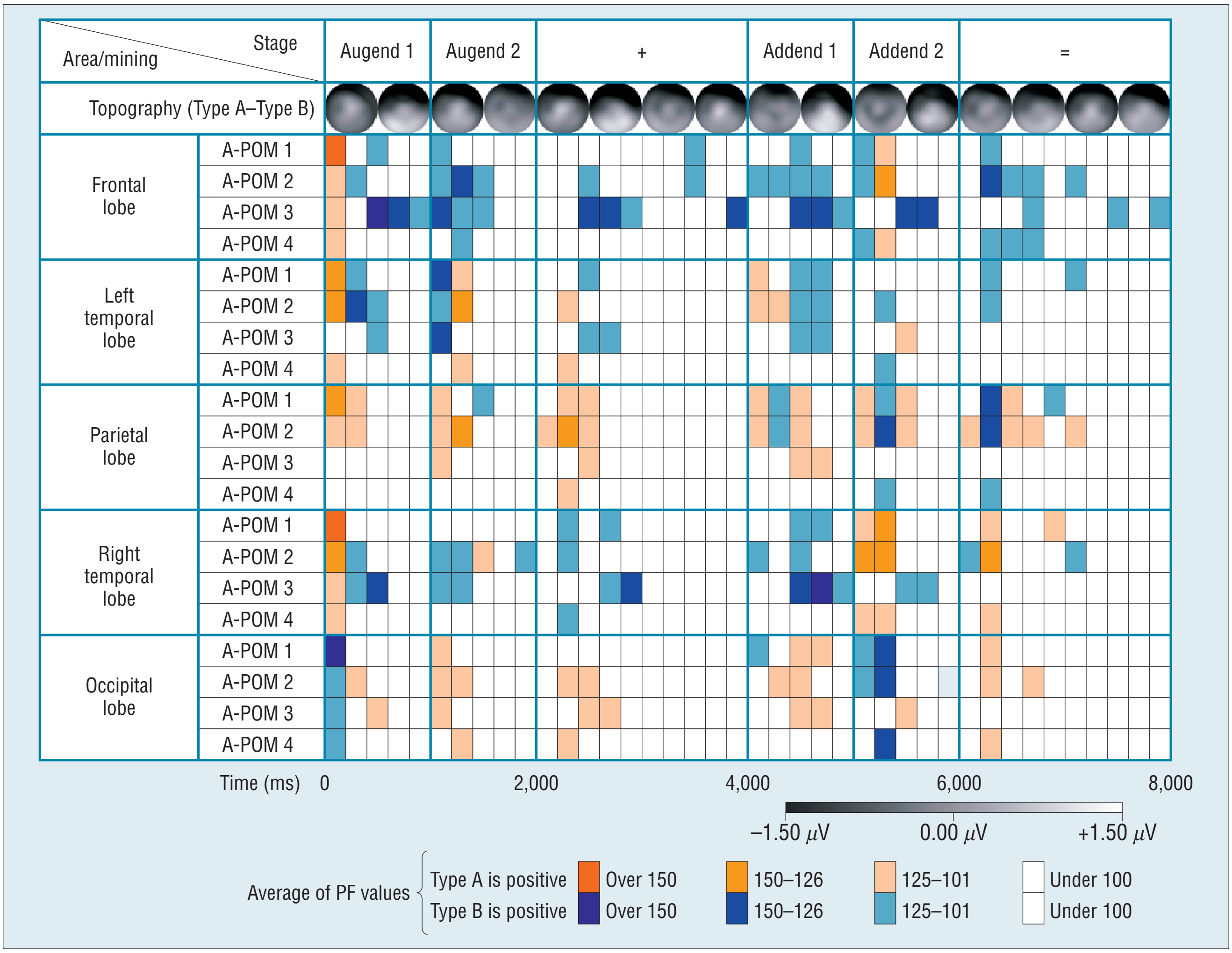

Figure 6. Results of agency-enriched peculiarity mining (A-POM) analysis for event-related potential (ERP) differential data in Type A and Type B experiments. Each time-axis block is equal to 200 milliseconds; plot density shows the average PF values during that time. Orange denotes a positive value for Type A, while blue denotes a positive value for Type B.

13. H. Nittono et al., "Event-Related Potential Correlates of Individual Differences in Working Memory Capacity," *Psychophysiology*, vol. 36, no. 6, 1999, pp. 745–754.
14. G. Dong and J. Li, "Efficient Mining of Emerging Patterns: Discovering Trends and Differences," *Proc. 5th Int'l Conf. Knowledge Discovery and Data Mining* (KDD 99), AAAI Press, 1999, pp. 43–52.
15. A.A. Freitas, "On Objective Measures of Rule Surprisingness," J. Zytkow and M. Quafafou, eds., *Principles of Data Mining and Knowledge Discovery*, LNAI 1510, Springer, 1998, pp. 1–9.
16. B. Liu et al., "Analyzing the Subjective Interestingness of Association Rules," *IEEE Intelligent Systems*, vol. 15, no. 5, 2000, pp. 47–55.
17. E. Suzuki, "Autonomous Discovery of Reliable Exception Rules," *Proc. 3rd Int'l Conf. Knowledge Discovery and Data Mining* (KDD 97), AAAI Press, 1997, pp. 259–262.
18. N. Zhong, Y.Y. Yao, and M. Ohshima, "Peculiarity Oriented Multidatabase Mining," *IEEE Trans. Knowledge Data Eng.*, vol. 15, no. 4, 2003, pp. 952–960.
19. L.I. Perlovsky, "Toward Physics of the Mind: Concepts, Emotions, Consciousness, and Symbols," *Phys. Life Rev.*, vol. 3, no. 1, 2006, pp. 23–55.

THE AUTHORS

Ning Zhong is a professor in the Department of Life Science and Informatics at the Maebashi Institute of Technology, Japan, and an adjunct professor in the International Web Intelligence Consortium (WIC) Institute at the Beijing University of Technology, China. His research interests include data mining, Web intelligence, intelligent agents, and brain informatics. He received his PhD in information engineering from the University of Tokyo. Contact him at zhong@maebashi-it.ac.jp.

Shinichi Motomura is a doctoral candidate at the Maebashi Institute of Technology, Japan. His research focuses on human brain data mining and brain informatics. He received his MS in information engineering from the Maebashi Institute of Technology. Contact him at motomura@maebashi-it.org.