

Editorial

Designing a market for data to enable chance discoveries

Yukio Ohsawa^{a,*} and Akinori Abe^b

^a*The University of Tokyo, Bunkyo-ku, Tokyo, Japan*

^b*Chiba University, Inage-ku, Chiba-shi, Chiba, Japan*

It is desired to create a social environment where active decision makers in businesses can be supported by data synthesis and analysis. However, the availability of data is still a critical bottle neck in the data-driven approach to businesses. In spite of the worldwide trend to disclose data owned by governments, potentially useful data owned by industrial organizations are still closed. The reasons for this difficulty of data disclosure include the expected merit of monopolizing the potential value of data, the privacy issue linked to the uncertainty about whom the data may be forwarded, and the uncertainty of purposes and processes of data usage.

In this special issue, we called for papers about techniques, methods, or theories about how we can design the market of data, that is a social environment where each person on the earth can obtain potentially useful data and combine them for efficient and effective decision making, without fearing loss of business opportunities. Participants of the market externalize and share the value of data, on which they come to buy and/or sell data in a reasonable condition, e.g., for a reasonable price or as open source if the data may give merits to people in general rather than to a particular segment in the market. Furthermore, data scientists are expected to import techniques from other experts dealing with different kinds of data. Here in the market of data, data

about use cases of data could be provided as metadata, to aid participants' learning from success/failure cases.

In MoDAT (Market of DATA) workshop series, participants exchanged thoughts and results relevant to the topics stated above. This series in the name of MoDAT started in the International Conference on Data Mining of 2013. However, the very origin of this activity is the session series on Chance Discovery – discovery of events significant for decision making – in International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES) since 2000, followed after by workshops in International Conferences on Artificial Intelligence (IJCAI), European Conference on AI (ECAI), etc. Methods and philosophies developed in chance discovery are now succeeded to MoDAT, because combination and reuse of data play essential roles in designing the market of data.

According to our finding in MoDAT workshops, experimental markets of data turned out to be innovative in that ideas born there tend to derive novel and productive proposals in real businesses. The market also turned out to be educative in that participants learn techniques for analyzing data in ways contributing to real businesses. Furthermore, data scientists came to learn and create techniques that accelerate discoveries from data. Areas relevant to the market of data turned out to be (not restricted to) electronic commerce (EC), mining data or text for externalizing potential values of data, analysis of relevance among metadata for externalizing and visualizing the possibility to combine

*Corresponding author: Yukio Ohsawa, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan. E-mail: ohsawa@sys.t.u-tokyo.ac.jp.

data, knowledge representation including ontologies of stakeholders (an their thoughts) of the market, creative data-based communication, etc. These all contribute to the discoveries of the value of an event or of data for decision making of any participant in the market of data. So far, MoDAT workshops have been filled with studies from domains relevant to chance discovery, the background of MoDAT, by computer scientists, sociologists, and knowledge workers in business.

Intelligent Decision Technologies is a journal that fits the concept of MoDAT, in that the market of data is a social system that enables people to have essential data and technologies for sharing and reusing data and knowledge. Papers on chance discovery – the origin of MoDAT for us – were really published in this journal, so we can say this IDT journal has been ready to make a special issue on MoDAT. We set the review process to highlight contributions to developing new method of research about designing the market of data. Topics expected for this special issue were as follows, but not restricted as far as the submission fits the viewpoint of MoDAT.

- (A) Data reuse with techniques of data/text mining and visualization, for externalizing the latent value of data;
- (B) Analysis of metadata, for estimating the possibility to combine data;
- (C) Knowledge representation, including ontologies of variables, for visualizing relevance of data;
- (D) Data-based communication, with discovering the value of data for decision making;
- (E) Any system for triggering, sustaining, and representing meaningful thoughts of stakeholders.

In each MoDAT workshop, 10 to 12 papers have been selected via the review process of 50% acceptance rate. And then, for this special issue, six among all papers including the significant extensions of papers presented in MoDAT workshops were selected, via blind review by two or more reviewers. Let us introduce below the finally accepted seven papers.

The first group of papers are about systems for triggering, sustaining, and representing toward innovations in the market of data. The first one is *Comparison of Conflict Resolution Behavior and Scenario Generating Process in Group and Individual* by Handwriting Process Analysis, by Hayashi and Ohsawa. Action Planning (AP) is introduced here, as a method for creating strategic scenarios of actions in data sharing and reusing. In AP, ideas for data utilization are grounded into feasible plans, i.e., scenarios, by adding

related elements (stakeholders, data, analysis methods, etc.). Participants notice and resolve conflicts and reduce risks in actions through the process of externalizing and serializing elements with conflict resolution behaviors. The results suggest an individual decision maker can generate a scenario of a similar quality as generated by a group, when one takes into account the cost due to inconsistency of scenarios. The work above is essentially relevant to the next paper *Abduction dealing with potential values and its datasets towards IMDJ*, by Akinori Abe. In abductive reasoning, Abe suggests, human considers values included in the knowledge. However, in some contexts, such values are unnecessary and sometimes harmful. Outside of the main abduction procedure, the proposed inference system is expected to deal with values to generate hypotheses considering the context (user's preference, situation, and the current trends, etc). In addition, Abe introduces the expiration of a hypothesis, where a hypothesis recently generated will not be generated during an abduction procedure, so that the system can sustainably generate novel hypotheses. The author points out the inference can be applied to recommendation in shops. Finally, the author briefly illustrates the process of Ohsawa's Innovators Marketplace on Data Jackets (IMDJ), as an extended abduction, projecting proposal production procedure in IMDJ to abduction.

The second group is about data reuse with analysis in the process of chance discoveries in the market of data. The first paper is *Tangled String for Sequence Visualization as Fruit of Ideas in Innovators Marketplace on Data Jackets*, by Ohsawa and Hayashi. This paper presents cases of IMDJ, that is, a market-based creativity support system for providers, users, and analysts of data. From these cases, a method for visualization was created as a result of IMDJ, responding to the requirements for data analysis. This created method called Tangled String (TS) visualizes sequential data, computing the importance of items in the overall string – here a sequence is compared to a string. The performance of TS is shown by experiments, showing TS meets requirements in IMDJ that reflects the social demands for reusing data. In other words, this paper showed the effect of the market of data to create a technique for data analysis.

On the other hand, the next paper in this group shows a case of analysis positioned in the process to the discovery of useful knowledge. In the paper *A study on the view of oral health and oral risk management in Japan: narrative analysis in combination with text-mining and KJ Method*, by Kobayashi and Nara, the

mutual awareness and understanding of oral risk management are supposed to be the basis for improving dental and oral health treatment quality. Based on interviews with ordinary people and dentists, authors analyzed narratives with KeyGraph, extracted the keywords, and enumerated key concepts. Then, by arranging the key concepts by affinity graph, the authors externalized similarities and differences between views of oral health and oral risk management of ordinary people and of dentists. These similarities and differences will contribute to better oral risk communication and management. Thus, data about viewpoints of different stakeholders, compared and/or combined, work to externalize latent viewpoints that urge innovations via mutual understanding of stakeholders in medical treatments.

In the next paper, *Application of Search Engine Focusing on Trend-related Queries to Market of Data*, Kori et al. examines the applicability of a search engine focusing on trend-related queries to the market of data. Authors regard the market of data as a promising target domain for the proposed search engine, where value creation from data fusion is crucial activity. Because data resources from various domains have temporal characteristics, the proposed search engine can be useful for finding connection between different data resources in terms of temporal trend. The result shows the participants could find items related to their interests from various viewpoints. According to guest editors' impression, this paper will provide a promising tool to aid data analysts in all domains of the market of data.

The last paper in this group focuses more strongly on a technique mining and/or visualizing text or data. *Matrix-Like Visualization Based on Topic Modeling for Discovering Connections Between Disjoint Dis-*

ciplines, by Qi and Ohsawa, pays attention to the difficulty of interdisciplinary research. It is difficult for researchers to discover connections between disciplines. Literature-based discovery shows the potentials of solving this problem by using information retrieval and natural language processing techniques. However, it still faces large amounts of manual works with prior knowledge, difficulty in understanding the discoveries, and limitation of the extension of the domain. Here, a matrix-like visualization approach is presented, based on topic modeling. With this approach, it is expected that interdisciplinary connections are to be discovered by detecting the topics we call mixed topics. For achieving this purpose, the authors visualize the document-phrase matrix generated by topic modeling and develop an original bi-clustering algorithm, which can help users without prior knowledge to detect connections between disciplines.

Thanks to these papers, the guest editors are finding that the topics (B), (C), (D) listed above can be regarded as elements for (A) or (E), because each submission included one of (B) (C) or (D) but aims to achieve (A) and/or (E), that position human(s) rather than machines at the center. We suggest this basic finding from these papers should be reflected to the real marketplace of data. For example, the method IMDJ above has been introduced in governmental activities by METI (the Ministry of Economics, Trade, and Industries) of Japan, DXC (Data Exchange Consortium) organized by the industrial section, and educational organizations including schools in Japan, China, Taiwan, USA, and India. Activities linked to MoDAT is being continued, welcoming submissions and audience. We express many thanks to the editorial committee of this journal, including Professor Lakhmi Jain and Professor Gloria Phillips-Wren as well as to reviewers.