First International Workshop on Chance Discovery

June 21, 2001

The first International Workshop on Chance Discovery was held in the hot-springs resort of Matsue, Shimani Prefecture, Japan, on 22 May 2001, as part of the Fifteenth Annual Conference of the Japanese Society for Artificial Intelligence (JSAI-2001). Thirteen presentations were made to the workshop [12], with 25 people attending. The majority of presenters and attendees were from Japan. An edited selection of the papers presented will be included in a volume of papers from the various International Workshops of JSAI-2001, to be published in the Springer series, *Advanced Information Processing*. A forthcoming Special Issue of the journal, *New Generation Computing*, will also address this topic.

What exactly is Chance Discovery? This was a question discussed with vigor at the workshop, and not everyone agreed that the subject was new, or worthy of special attention. The Workshop Chairman, Yukio Ohsawa, who had originally coined the name, opened the workshop with his definition. In his formulation, a chance is defined as some event which is significant for decision-making in a specified domain. Thus, a chance event may be either positive — an opportunity, say — or negative — a potential risk, say. Chance Discovery (CD) he defined as learning and/or explaining a chance event. The aims of CD as a discipline within Artificial Intelligence are the determination of methods for discovering such chance events, which are typically rare or hard to find.

Ohsawa distinguished CD from previous methods of knowledge discovery, such as time series analysis, statistical pattern matching or association rule detection. All these methods rely on past data or on cases similar to the data/case in question, whereas chance discovery applications typically have no such past data or similar cases upon which to draw. How do we forecast demand, for example, for an innovative new product? We can ask potential consumers what they think of it, and whether or not they will buy it, but if they have no prior experience of the product or anything related their answers are likely to be misleading. Likewise, looking at demand for similar or substitute products may also mislead: it makes no sense to forecast the demand for a new bridge, for example, on the basis of the numbers of people who currently swim across the river it will span. This forecasting problem is accentuated if one person's utility from the innovation depends on its prior adoption by others, as is the case with many telecommunications products. Forecasts for terrestrial mobile telecommunications services, for instance, have consistently and greatly under-estimated the actual demand [4]. The task of demand forecasting for really-new products is an important one in marketing, and some attempts have been made to tackle it, such as the simulation of purchase situations [19], or the formal incorporation of minority viewpoints into the forecast [8]. But these methods are a long way from maturity.

The same is true of the identification of risks, another application of Chance Discovery. If we wish to make a decision under uncertainty, classical decision theory provides us with a sound methodology provided we have first identified all the outcomes possible. What if we know or suspect that our list of possible outcomes is incomplete? What decision methodology is appropriate then? If the possible outcomes excluded are rare but catastrophic, the methods of maximum expected utility decision theory may lead us very astray. The chemical thalidomide, for example, was tested on human and animal subjects before its commercial release, but none of these test subjects were pregnant [17]; presumably the possibility that pregnant women could be affected differently to other women had not occurred to the drug's testers. The problem of the identification of all possible outcomes thus presents itself to those engaged in risk assessment of novel systems, technologies or substances, and to those building decision support systems for risk and safety assessment tasks. Although not called Chance Discovery, the problem has therefore received some attention from researchers in AI, for example [1].

What may be a risk in one decision context may not be so in another, and Ohsawa's definition of a chance event emphasized the decision aspect. He drew a distinction between Chance Discovery and Chance Management, a distinction taken up particularly in the talk by Helmut Prendinger [13]. Prendinger had sought to provide a philosophical grounding for the development of chance discovery and chance management methods, and for this, he drew on general systems theory. His work opened discussion on the relationship between a system and its environment, and between predictions of future environment states and actions which may impact such states. For example, in the case where actions and states have an intimate connection, predictions may be self-fulfilling or self-denying. Someone seeking to make accurate predictions or, alternatively, seeking to achieve or avoid certain world states would need to be fully aware of such effects. Forecasting the weather is thus an inherently different undertaking to forecasting demand in some market where the company doing the forecasting has a large market share. In my opinion these are very important issues for the field of Chance Discovery and Management, and considerable philosophical and modeling work along the lines begun by Prendinger is needed before we could claim to fully understand them. Use of complex and dynamic systems theory seems a promising approach to pursue here.

Three other presentations discussed formal and methodological issues. Peter McBurney [7] presented a formal model for a dialogue aimed at discovering chance events. This model, drawing on the philosophy of dialogue games, was based on a view that Chance Discovery dialogues incorporate both a search for truth in some situation, along with the imposition of values over the search space, for example, the value that certain search outcomes such as risks are undesirable. Thus, the search for truth in Chance Discovery is not a disinterested one, which makes it different in principle to the common understanding of scientific endeavor. With this view, the authors presented a set of dialogue locutions which they argued would enable a set of autonomous agents, possibly with differing knowledge of the situation, to undertake Chance Discovery dialogues. Much work will be needed to prove these proposals in a realistic application. In a more traditional vein, Takuichi Nishimura [11] presented a methodology based on time-series analysis for the identification of patterns in non-traditional data, such as video images or audio messages. Other applications of this method may be to the detection of patterns in genomic data or to the identification of potential typhoons from satellite-image data. However, these problem domains and the methodology proposed for them appear to be standard data mining applications, and I would not classify them under Chance Discovery. Another formal approach was that of Vicenc Torra [18], who proposed an architecture for CD in fuzzy logic-based systems. This architecture drew on a hierarchical conceptual model of the ways in which a system may interact with its environment, and so potentially provides the basis of a deeper understanding and modeling of chance management. It would be interesting to link Torra's work to that of Prendinger, and so potentially take both lines forward.

The other papers presented at the workshop discussed various aspects of chance discovery, often with a specific application in mind. Yoshikiyo Kato [3] discussed the detection of potential risks in complex systems, such as those used in space missions. He proposed a framework for identifying undetected patterns in operating data originating from space vehicles following deployment, a framework which combined data mining techniques with management of design information. Many accidents in complex engineering systems may ultimately be traced to organizational causes, as for example has been argued was the case with the Challenger Space Shuttle disaster [20]; it will therefore be interesting to see how well Kato's methods work when deployed in specific organizational contexts.

Several papers discussed systems to support creativity and learning. Hirohito Shibata [14] presented a system called *IdeaManager*, which was developed to support researchers with the recording and management of original ideas. The system was developed following a study in which 12 academic researchers in psychology and software engineering kept a diary of their research ideas for an extended period. The findings of this study informed the design of *IdeaManager*, which combines both an idea-storage module with an idea-management module. At present, this system only supports textual, and not graphical, representations of ideas, a feature seen as a severe limitation by some in the workshop audience. Wataru Sunayama [16] also proposed a framework to support creativity. In this framework, users are able to access a data mining engine in some web-domain by means of a search system, which links user-defined search terms to web-information. The system is still under development and so further details about its operation were not yet available. Tomoya Horiguchi [2] discussed chance discovery in a learning environment, and argued that learning often occurs following misunderstandings and mistakes. Seeking to systematize such events for computer-assisted learning applications, he proposed a means to identify effective counter-examples, i.e. those which have the potential to lead to learning. Also related to creativity, Hiroko Shoji [15] discussed a framework for on-line support for potential consumers which would draw the customer's attention to information relevant to their intended purchase. Such support aimed to imitate the ability of good human sales assistants to suggest products or accessories which lead to mental leaps in the minds of customers.

Naohiro Matsumura [5] and Yukio Ohsawa [10] both spoke of chance discovery as the finding of connections between previously disconnected communities. Here a community is defined as a group of people who share a context, and a connection between two separate communities effectively provides a new context which envelopes or links both the previous contexts. Matsumura's applications were web-based communities, for example customers of different electronic book outlets. Ohsawa's application was to the perception by ordinary people of their risk from natural disasters, such as earthquakes. Both these applications presented connections between communities in a graphical format and graph-theoretic techniques may eventually prove useful in analysing such data. Such techniques were adopted by Yutaka Matsuo [6] in his analysis of the relationships between the words *within* a single document. Identifying the most important words in a document, even when they may appear only rarely, is one goal of such analysis, and Matsuo wittily applied his methods to his very own paper.

Donald McCloskey has argued that an academic discipline is a 2-tuple consisting of an agreed scope of enquiry and an agreed set of methods of enquiry [9]. Discussion at this first international workshop focused on the scope of Chance Discovery — exactly what is the discipline, and is it really just data mining or AI search or statistical pattern analysis by another name. In my opinion, the case is clear that Chance Discovery is not any of these, but is something new. Indeed, by giving this particular scope of enquiry a name, we have brought into the one tent a diverse range of problems that have already arisen in different fields: risk assessment and management in systems engineering, in environmental management and in medicine; natural disaster prediction and management in public safety; identification of emergent trends and behaviors in complex systems; opportunity identification in business and marketing; demand forecasting for innovative products in marketing; etc. The key benefit of inclusion under a single tent is that common features to these problems can be recognized, which may facilitate their solution, and methods which have shown promise in one domain can be more readily applied to others. Having identified, however inchoately, the scope of this new discipline, our next and greater challenge will be to articulate and formalize its methods of enquiry.

Workshop reviewed by: Peter McBurney, Department of Computer Science, University of Liverpool, U. K.

References

[1] J. Fox. Will it happen? Can it happen? A new approach to formal risk analysis. *Risk, Decision and Policy*, 4 (2):117–128, 1999.

[2] T. Horiguchi and T. Hirashima. The role of counterexamples in discovery learning environments: awareness of the chance for learning. pages 5–10 of [12].

[3] Y. Kato, T. Yairi, and K. Hori. Integrating data mining techniques and design information management for failure prevention. pages 11–15 of [12].

[4] T. Kelly. Forecasting the mobile communications market: A finger in the airwaves? *Paper presented to Conference on Market Forecasting in the Telecoms Industry, Hong Kong, March 1996*, International Telecommunications Union, Geneva, Switzerland, 1996.

[5] N. Matsumura, Y. Ohsawa, and M. Ishizuka. Future directions of communities on the web. pages 17–20 of [12].

[6] Y. Matsuo, Y. Ohsawa, and M. Ishizuka. A document as a small world. pages 21–25 of [12].

[7] P. McBurney and S. Parsons. Chance discovery using dialectical argumentation. pages 37–45 of [12].

[8] P. McBurney and S. Parsons. Using Belief Functions to forecast demand for Mobile Satellite Services. In R. P. Srivastava and T. Mock, editors, *Belief Functions for Business Decisions*. Physica Verlag, New York, NY, USA, 2001 (In press).

[9] D. N. McCloskey. *The Rhetoric of Economics*. University of Wisconsin Press, Madison, WI, USA, 1985.

[10] Y. Nara and Y. Ohsawa. A method for discovering seeds of consensus applied to family risk perceptions. pages 91–94 of [12].

[11] T. Nishimura and R. Oka. Retrieval of similar time-series patterns for chance discovery. pages 79–86 of [12].

[12] Y. Ohsawa, editor. *Proceedings of the First International Workshop on Chance Discovery*. Japanese Society for Artificial Intelligence, Matsue, Japan, 2001.

[13] H. Prendinger and M. Ishizuka. Some methodological considerations on chance discovery. pages 1–4 of [12].

[14] H. Shibata and K. Hori. A creativity support system for long-term use in everyday life – Idea Manager: its concept and utilization. pages 47–56 of [12].

[15] H. Shoji and K. Hori. Chance discovery by creative communicators observed in real shopping behavior. pages 57–65 of [12].

[16] W. Sunayama and M. Yachida. Support system for creative activity by information acquirement through Internet. pages 67–70 of [12].

[17] H. Teff and C. R. Munro. *Thalidomide: The Legal Aftermath*. Saxon House, Westmead, Farnborough, Hampshire, UK, 1976.

[18] V. Torra. Fuzzy knowledge based system and chance discovery. pages 71–78 of [12].

[19] G. L. Urban, B. D. Weinberg, and J. R. Hauser. Premarket forecasting of really-new products. *Journal of Marketing*, 60:47–60, 1996.

[20] D. Vaughan. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. University of Chicago Press, Chicago, IL, USA, 1996.