

Web intelligence and chance discovery

Ajith Abraham · Yukio Ohsawa · Yasuhiko Dote

Published online: 25 May 2006
© Springer-Verlag 2006

Web Intelligence deals with the scientific research and development to explore the fundamental roles as well as practical impacts of soft computing and advanced intelligent paradigms for the next generation of Web-based systems, services, and environments. Chance discovery means discovering chances — the breaking points in systems, the marketing windows in business, etc. It involves determining the significance of some piece of information about an event and then using this new knowledge in decision making. The techniques developed combine data mining methods for finding rare but important events with knowledge management, groupware, and social psychology.

Soft computing has evolved not only from a theoretical point of view but also with a large variety of realistic applications related to the Internet, business intelligence, consumer products and industrial systems. Applications of soft computing have provided the opportunity to integrate human-like vagueness and real-life uncertainty into an otherwise hard computer programs.

The Fourth IEEE International Workshop on Soft Computing as Transdisciplinary Science and Technology

A. Abraham (✉)
School of Computer Science and Engineering,
Chung-Ang University, 221, Heukseok-dong,
Dongjak-gu Seoul 156-756, Republic of Korea
e-mail: ajith.abraham@ieee.org

Y. Ohsawa
Department of Quantum Engineering and Systems Science,
School of Engineering, The University of Tokyo, Tokyo, Japan
e-mail: ohsawa@q.t.u-tokyo.ac.jp

Y. Dote
School of Computer Science and Engineering,
Muroran Institute of Technology, Muroran, Japan
e-mail: dote@csse.muroran-it.ac.jp

(WSTST'05) gathered soft computing researchers working in the area of Web intelligence and chance discovery besides many other disciplines. This special issue comprising of ten papers is focused on using advanced soft computing tools for Web intelligence (first five papers) and chance discovery (last five papers). Papers were selected on the basis of fundamental ideas/concepts rather than the thoroughness of techniques deployed. The papers are organized as follows.

In the first paper *Walker* developed Web explorers to forage a simulated information ecosystem, the Internet, for useful information. Each explorer is designed to detect and report dynamic changes within the infrastructure of the Internet to its Web explorer dispatcher, which is responsible for coordinating thousands of explorers. The foraging behavior of active honeybee colonies serves as a model for Web explorers that are reactive, proactive, and robust.

Borges and Levene in the second paper presented two methods for testing the predictive power of a variable length Markov chain induced from a collection of user web navigation sessions. The first method used a statistical test to measure the significance of the distance between the distribution of the probabilities assigned to the test trails by a Markov model build from the full collection of sessions and a model built from the training set. The second method evaluates the model ability to predict the last page of a navigation session based on the preceding pages viewed by recording the mean absolute error of the rank of the last occurring page among the predictions provided by the model. Experimental results conducted on both real and random data sets are reported and the results show that in most cases a second-order model is able to capture sufficient history to predict the next link choice with high accuracy.

In the third paper, Yue et al. proposed a novel behavior-based anti-spam technology for email service based on an artificial immune-inspired clustering algorithm. The suggested method is capable of continuously delivering the most relevant spam emails from the collection of all spam emails that are reported by the members of the network. Two main concepts are introduced, which defines the behavior-based characteristics of spam and to continuously identify the similar groups of spam when processing the spam streams.

Finding a product with high quality and reasonable price online is a difficult task due to uncertainty of Web data and queries. Gu and Zhang in the fourth paper developed a new type-2 fuzzy online decision support system to handle the uncertainty problem. In the developed Web shopping expert, a fast interval type-2 fuzzy method is used to directly use all rules with type-1 fuzzy sets to perform type-2 fuzzy reasoning efficiently. The parameters of type-2 fuzzy sets are optimized by a least square method. The Web Shopping Expert based on the interval type-2 fuzzy inference system provides reasonable decisions for online users.

In the fifth paper, Badica et al. proposed logic wrappers inspired by the logic programming paradigm. The developed Logic wrappers (L-wrappers) have declarative semantics and therefore their specification is decoupled from their implementation and can be generated using inductive logic programming. Authors have also defined a convenient way for mapping L-wrappers to XSLT for efficient processing using available XSLT processing engines. The developed framework is illustrated for Web information retrieval purposes.

On the sixth to the tenth papers, let us give a short introduction to chance discovery, which is a research domain started from 2000. A chance here means an event or a situation, which is significant for the decision of human(s) — providing information about opportunity/risk. Because a chance can be a significant seed of benefits/loss for human, the process of chance discovery has been characterized by three steps: (1) human's awareness of a chance based on real data, (2) explaining its significance, and (3) decide and do real actions. Data mining and visualization techniques may contribute to step (1), the expertise and communication of humans are relevant to step (2), and the real data and human expertise are obtained from his/her actions in (3) and then reflected to (1) and (2). The accepted five papers on chance discovery can be summarized as follows, according to the steps (1), (2), and (3):

In the sixth paper by Iwase, Seo, and Takama, a system is presented for reflecting users' communication to their annotation in a 2D visualization of the target data. This annotation is reflected to an XML based description of

users' interest to the dataset the users share. For example, if the users are members of a marketing team, by this system they can insert their creative opinions to the data on customers' claims and integrate their thoughts to a business decision. This paper is on linking steps (1) and (2) above.

The seventh by Sakaki and Ohsawa shows a method for understanding scenarios of event-occurrences underlying a given dataset. The system gradually changes the granularity of the dataset to be visualized with a 2D data visualization tool. This corresponds to rearranging the dataset to fit to the user's understanding, in the process to chance discovery. Taking a dataset on questionnaire results from potential consumers of a product, visualized using the visualization tool KeyGraph, the authors show this accelerates step (1) and step (2).

In the eighth by Matsumura, Goldberg, and Xavier, a method is proposed for detecting signs of conflicts and contextual gaps among participants of human-human communication, and for taking advantage of this detection for improving the teamwork performance. For the manager of a team, for example, this is a promising method for the decision in team organization. Their new method is based on original metrics for the contextual distance among participants. This shows a cutting edge for step (2) above.

The ninth paper, Abe et al, introduces a behavioral model for nursing risk management in a hospital. They review previous nursing incidents and protections, and introduce a new model for explaining them based on abductive reasoning, that is a non-monotonic inference framework. Their models is a combination of abductive reasoning and scenario-based decision, where nursing activities are regarded as a scenario and an incident is regarded as a scenario violation which should be explained and is a critical chance for nurses. This paper presents a new vision in step (3) and step (1).

The tenth is by Yada, presenting a new platform of data mining. The interface of CODIRO system, designed by the author's special expertise on data-based marketing, allows user to collect and input data relevant to his/her concern with the market. In that the user(s) can reflect experiences in the real-world actions to acquisition of data, we can trust this system as the powerful engine for running steps from step (3) to step (1), and step (1) to step (2).

The editors wish to thank the referees who have critically evaluated the papers within the short stipulated time. We would like to take this opportunity to thank Professor Vincenzo Loia, Managing Editor, *Soft Computing Journal* for all the timely advices and help and also for the opportunity for editing this important scientific work. Finally we hope the reader will share our joy and find this special issue very useful.