

Special issue on data mining for decision making and risk management

Shusaku Tsumoto · Tzung-Pei Hong

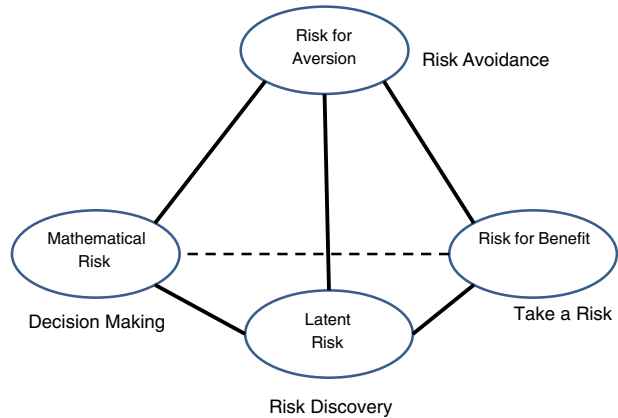
Received: 8 March 2011 / Accepted: 8 March 2011 /
Published online: 13 April 2011
© Springer Science+Business Media, LLC 2011

The organizations and systems in our modern society are becoming large and complex in order to provide more advanced services due to the growing variety of social demands. Such organizations and systems are efficient but highly complex, and may cause various unexpected situations. According to this observation, the importance of decision making and risk management in these organizations and systems has been strongly noticed in recent years. On the other hand, the accumulation of large amounts of data in the operations of the organizations and systems has become easier through the use of information technology. These data can be used to support decision making or risk management in the organizations and their systems.

“Risk” is both an old and a new problem. There are many ways to use this word: especially, in light of the rapid development and increase in complexity of social organizations, which gives an impression that risk management is strongly dependent on “application domains”. Our question is whether we can have a *generalized informatics approach* to risk by using data sciences techniques, which leads to a new field “risk sciences”, as proposed by Tsumoto et al. (2007). They classify the types of risk into four categories as shown in Fig. 1. The first category is “risk for aversion” such as medical risk, product safety, and disasters. The second type is “risk for benefit” or “take a risk”, used for future profit or innovation, such as the development of pharmaceutical, finance and insurance products. The third class is “latent risk”, which may be inferred by mining large databases. Finally, the fourth is “mathematical risk”, defined as a statistical model with weighted mean of loss functions. It is notable that these four classes are not mutually exclusive, rather, they overlap with some differences in the degree of membership. For example, risk in business may include the first and second type; if one aims at discovery of risk factors,

S. Tsumoto (✉)
Shimane University, Izumo, Shimane, Japan
e-mail: tsumoto@med.shimane-u.ac.jp

T.-P. Hong
National University of Kaohsiung, Kaohsiung, Taiwan
e-mail: tphong@nuk.edu.tw

Fig. 1 Components of risk

the third class is also very important. However, the contribution of the fourth class may be less than the others in this domain. Therefore, the classification is fuzzy.

This special issue focuses on both data mining and statistical techniques to detect and analyze the risks potentially existing in the organizations and systems and to utilize the risk information to improve their management and decision support, which can be viewed as a first step towards risk sciences. There are eight papers included, each of which has received recommendations from the reviewers and has been revised by the authors for this publication. Their contents are briefly introduced below.

The first paper presents a study of machine learning techniques used in computer-aided diagnosis systems. It proposes a new algorithm to derive well-performed hypotheses for diagnosis. The proposed algorithm not only exploits a specific data-editing technique to identify and discard possibly mislabeled examples, but also employs an adaptive strategy to decide on the editing occasions. Five pre-conditional theorems are derived to ensure the strategy with an iterative reduction of classification error and an increase in the scale of new training sets. This paper is primarily concerned with the second and third risk types.

The second paper discusses longitudinal consumer behavior from sequence analysis. It exploits typical acquisition patterns to predict a customer's next purchase. Different states represented by a set of variables are used to model complex, possibly coupled sequential phenomena. A dynamic Bayesian network based on the states is then used to represent longitudinal customer behavior by acquisition, product ownership and covariate variables. The authors show that the Bayesian network could exhibit adequate predictive performance to support a financial-services provider's cross-sell strategy. The paper also is concerned with the second and third risk types.

Prevention of drug dispensing errors is discussed in the third paper. The authors propose a risk management approach and implement a system based on it to prevent drug-dispensing errors. The proposed approach consists of two main procedures. The first one is to derive a decision tree and regression function from the given dispensing errors cases and drug databases, and the second procedure is to gather together similar drugs by clustering techniques. The drugs that may cause dispensing errors will then be alerted through the clustering results and the decision tree. The paper focuses on the first and third risk types.

The fourth paper proposes a simple but efficient probability-based model to evaluate information security investment on data centers. The authors present two

algorithms that calculate the probability of threat to each protected resource and find the optimal investment for data center security based on given data center network topology, respectively. The technique proposed in the paper can be used to facilitate the analysis and design of more secured data centers. The paper addresses the fourth type of risk.

The fifth paper examines the issue of business aviation by intelligent techniques. The authors hope to discover possible trends and needs of business aviation in Taiwan for supporting the government to make decisions in anticipation of eventual deregulation. They adopt a knowledge-discovery tool based on rough sets to analyze the potential for business aviation through an empirical study. Several interesting patterns are found and explained from their survey data. The paper can be viewed as research in the second and third types of risk.

The sixth paper investigates the usefulness of commercially available external databases for customer relationship management. The authors first present a methodology based on random forests predictive modeling techniques to create commercial variables for external data vendors. The methodology then utilizes the variables to predict spending pleasure, which is a composite measure of purchasing behavior and attitude, in 26 product categories for more than 3 million respondents. These predicted results can then be incorporated into a company's transactional database to strengthen the existing CRM models. The paper covers the second to fourth class of risk.

The seventh paper illustrates the usability and usefulness of character-string analysis techniques in shopping path analysis. It utilizes the technology of sensor networks to accurately track the in-store movements of customers and adopts a mining approach to analyze customers' purchasing behavior. The customer movements from store sections are first obtained by RFID and described by character strings. The character-string parsing techniques are then applied to find the purchasing patterns of customers who made a large quantity of purchases. The paper addresses the second and third risk types.

The final paper proposes a method for grouping trajectories represented as two-dimensional time-series data. It first compares two trajectories based on their structural similarity, and determines the best correspondence of partial trajectories. It then calculates the value-based dissimilarity for all pairs of matched segments and outputs their total sum as the dissimilarity of two trajectories. The method captures the structural similarity between trajectories even in the presence of noise and local differences, and provides better proximity for discriminating objects. This paper focuses on the first and third types of risk.

We hope that you will find the special issue is a valuable resource in the development of data mining techniques for decision making and risk management. Finally, we are grateful to all the authors for their contributions and the referees for their vision and efforts. We would also like to express our thanks to the editors-in-chief of the journal, Drs. Kerschberg, Ras and Zemankova, and to the JIIS staff for their great support to realize the special issue.

Reference

- Tsumoto, S., Washio, T., Miyamoto, S., Tsubaki, H., Kitagawa, G. (2007). Towards risk sciences (in Japanese). *Magazine for Japanese Society for Artificial Intelligence*, 22(5), 630–632.