

# Operations Research and Knowledge Modeling in Data Mining

Masato KODA

Graduate School of Systems and Information Engineering  
University of Tsukuba, Tsukuba Science City, Japan 305-8573  
[koda@sk.tsukuba.ac.jp](mailto:koda@sk.tsukuba.ac.jp)

## ABSTRACT.

*Data mining has been a subject of substantial interest both in academia and industry. Covering machine learning, statistics, and operations research (OR), the technology of knowledge discovery now represents a new indispensable tool to assist in intelligent decision making in the highly complex business environment. Recently, OR techniques, e.g., quadratic programming, have extensively applied to data mining problems, such as Vapnik's Support Vector Machine (SVM). In order to deal with complex data analyses, however, knowledge modeling methodology known as Nonaka's SECI Model may help yield a basis for the development of knowledge-rich data mining technology. Since 1998, the author and colleagues extended such a research idea into complex data classification problems via boosting or ensemble learning techniques. The methods differ from single SVM, artificial neural network, and decision tree. The purpose of this paper is to promote the research interests in the connection of OR, knowledge modeling, and data mining as well as real-life applications among the growing computing and informatics communities.*

**Keywords:** *Knowledge-rich data mining, Knowledge modeling, Operations research.*

## 1.0 INTRODUCTION

Today, through Internet channels, businesses are facing competition from international sources, communicating electronically with suppliers, and interacting in real-time with customers. The growth of Internet connectivity, the increased availability of data warehouses, and the imperative of ever-swifter responses underscore the urgency of developing new advanced technologies of knowledge discovery in such areas as ERP (Enterprise Resource Planning), SCM

(Supply Chain Management), CRM (Customer Relationship Management), KM (Knowledge Management), and even in a newly emerging discipline of Service Science to build knowledge-based economy and society.

Data mining has been a subject of substantial interest both in academia and industry. This is due to an increased recognition of significant advantages of data mining (i.e., knowledge discovery) compared to hypothesis-based conventional approaches in diverse domains ranging from business and economics to engineering and sciences. Covering machine learning (AI), statistical analyses, operations research (OR), and database research, the technology of data mining now represents a new indispensable tool to assist in intelligent decision making for the highly complex business environment under various business models with different service design contexts.

In terms of the information required to perform the CRM applications via “service interface,” the Internet brings fundamental changes both in communication and knowledge acquisition; people no longer have to meet face-to-face to exchange business ideas and/or transfer knowledge, either tacit or explicit, since all information can be delivered through the service interface online and on demand. The new Web2.0 architecture comprises innovative technology features: it offers a platform that is ubiquitous to access customers; it supplies an interface with services for easier availability; it promotes an interoperable synchronization mechanism among business partners including customers and enables how the responsibility to provide the information is divided among the partners and consumers.

As the Web matured as a platform for online commerce and information services, a service provider becomes multi-channel by adding an Internet channel to its existing operations. Thus, in the rest of this paper, we

will mainly focus on information-intensive business models in which the information systems including Internet are responsible for the greatest proportion of value creation. Examples typically include data mining, data entry and transcription, customer support, computer programming, etc. In these domains, databases, software applications, or other explicit repositories or sources of information are ubiquitous and essential to meeting the goals of the customer for value creation.

This paper argues that for the substantial subset of data mining models which can be described as a collective intelligence, i.e., boosting, bagging, and ensemble learning, it is possible to take a more abstract view in terms of knowledge modeling contexts that highlights knowledge-rich data mining. Section 2 gives a relationship between OR and data mining. Knowledge modeling framework, defined in Section 3, is based on the well-known Nonaka's SECI model (Nonaka and Takeuchi, 1995) for spiral process of knowledge creation and evolution. Section 4 shows how this abstract description of knowledge modeling makes the different data mining models into combinable building blocks of knowledge-rich data mining, and suggests some unifying design concepts that apply to realize collective intelligence. In Section 5, we conclude that the premise of this paper is building a unified conceptual architecture of knowledge management model for the knowledge-rich data mining.

## **2.0 OPERATIONS RESEARCH AND DATA MINING**

In our approach, knowledge discovery using data mining tools plays a key role. Data mining is heavily used in many CRM and service applications. In data mining arena, Support Vector Machine and artificial neural network may be the most popular methods to estimate non-linear patterns and complicated non-explicit rules that are hidden in CRM transaction data.

Support Vector Machine (SVM) is invented by Vapnik (Vapnik, 1998) and known to be a powerful learning method that uses optimization algorithms based on OR, e.g., quadratic programming, Karush-Kuhn-Tucker conditions, etc., in order to achieve high generalization capability. SVM is essentially a binary classification algorithm in which the maximal margin hyperplane that separates two (separable) sets of data in a high

dimensional feature space is derived from statistical learning theory.

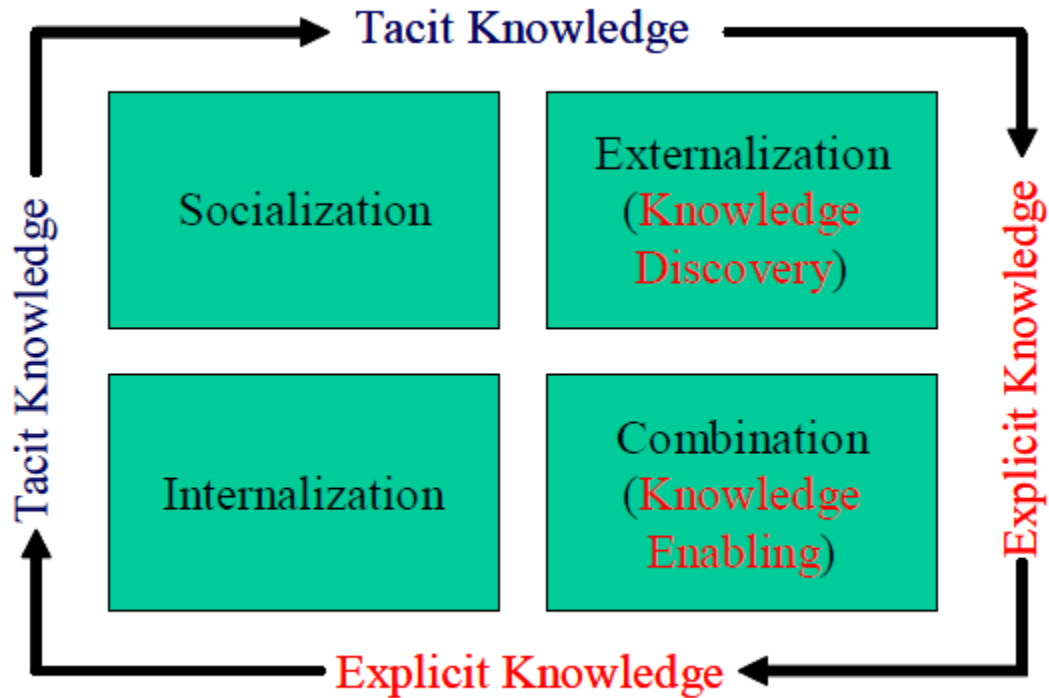
Artificial neural network (ANN) consists of nodes and connections between the nodes. Nodes process information and perceive data features. Connections and the weights generate and store knowledge in the network. The back-propagation algorithm of ANN is based on the steepest descent algorithm of OR technique, and noise-driven mechanisms provide a new approach to designing robust adaptation mechanisms, e.g., subconscious noise reaction (Koda, 1997a, 1997b; Koda and Okano, 2000), in neural network learning.

Thus, an effective combination of data mining technology and OR is expected to provide efficient knowledge discovery and knowledge enabling tools (see, e.g., Dupret and Koda, 2001; Okano and Koda, 2003).

## **3.0 SECI MODEL AND KNOWLEDGE-RICH DATA MINING**

Knowledge-based modeling can provide a unified approach to CRM and service applications in developing information-intensive business models (e.g., service marketing, medical care, travel, etc.) that incorporate the advanced flexibility of Web 2.0 architecture. For example, CRM system must allow multiple agents and clients to participate interactively together and, hence, it requires high level of messaging linkage model that unifies support of desktop applications with interactivity of rich Web contents for diverse clients over heterogeneous network environments.

In this section, we will illustrate a constructive knowledge modeling approach to CRM and service applications based on knowledge discovery and knowledge enabling to establish a knowledge-rich data mining technology. The proposed approach is essentially derived from the well-known SECI model (Nonaka and Takeuchi, 1995) for spiral process of knowledge creation and evolution. It should be noted that our final goal is to bring the concept of knowledge modeling into the so-called collective intelligence to build the knowledge-rich data mining process and accordingly increase the usefulness of data mining results in CRM and service applications. The SECI model for spiral process of knowledge creation and evolution, is shown in Figure 1.



Nonaka et. al

Figure 1. SECI Knowledge Model

Socialization-Externalization-Combination-Internalization (SECI) is a continuous process of dynamic interactions between tacit and explicit knowledge. Here, we will focus on Socialization, Externalization, and Combination processes. Socialization implies sharing tacit knowledge (e.g., master's expert knowledge or wisdom of craftsmanship, etc.) through face-to-face or shared communication to transfer it. Utilizing the Internet, however, people no longer have to meet face-to-face to exchange business ideas and/or transfer knowledge since all necessary information can be delivered through the service interface online and on demand. Externalization is a developing process that supports external communication of knowledge according to protocols and templates. In data mining terminology, we may interpret that Externalization is essentially equivalent to knowledge discovery process. Combination is the process that can be described as

knowledge enabling which combines discovered knowledge and leads to value creation.

The SECI model can be equivalently transformed to and represented by the diagram depicted in Figure 2. The model accommodates the nature of interaction between tacit and explicit knowledge, where the processes of knowledge transfer or messaging are mainly focused. The transformed SECI model provides a unified view which incorporates the SECI model in four messaging schemas, i.e., Socialization, Externalization, Combination, and Internalization. In this framework, however, care should be taken to treat the tacit knowledge for Socialization and Internalization. Because of the relatively complicated character of tacit knowledge with only small amount of samples, it may bring difficulties to the knowledge-rich data mining processes.

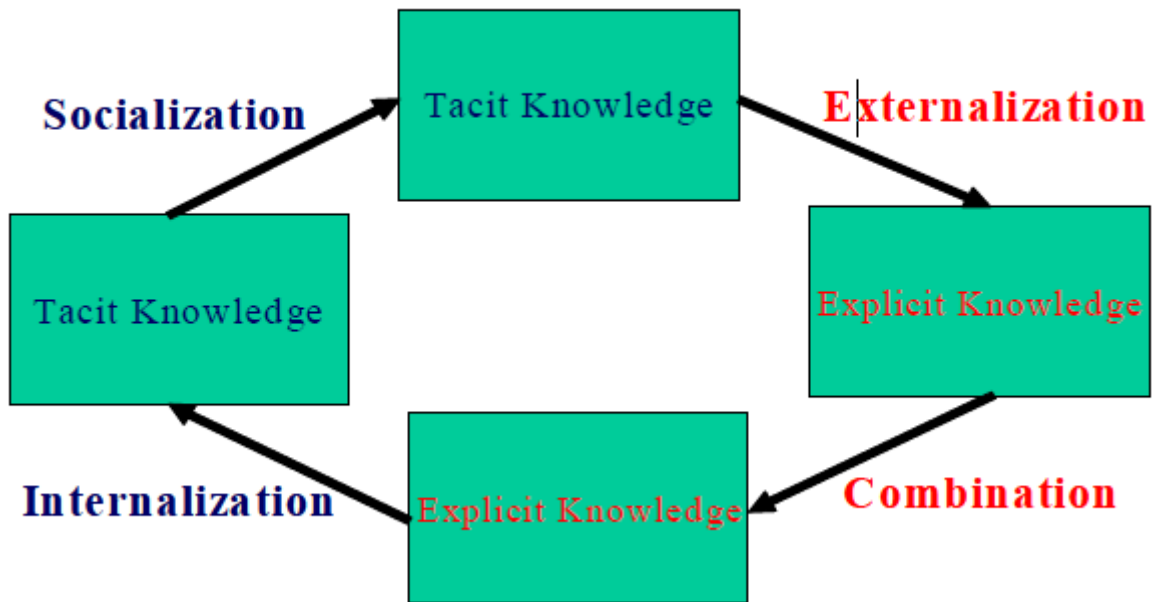


Figure 2. Transformed SECI Model

#### 4.0 COLLECTIVE INTELLIGENCE AND KNOWLEDGE-RICH DATA MINING

Collective intelligence, also known as a committee model, refers to a collection of models that learn a target function asymptotically by training a number of individual learners and combining their decisions. It is a technique for dynamical integration of learning machines and the most widely used methods include AdaBoost (Freund and Schapire, 1997). Since 1998, the author and colleagues extended the knowledge modeling methodology into complex data classification problems via boosting or ensemble learning techniques. The methods differ from single SVM and artificial neural network. For technical details, the readers are referred to the references (Sano, Suzuki, and Koda; 2004, 2008).

Based on the knowledge-rich data mining and the transformed SECI model which are described in Section 3, the proposed unifying design concepts of collective intelligence is shown in Figure 3. This maps the original SECI model onto spiral messaging infrastructure using the knowledge-rich data mining

tools as combinable building blocks to realize information-intensive business systems. Potentially, it can greatly improve CRM and service applications portability by reducing their dependency on underlying knowledge transfer schemas and middleware platforms. However, it requires a powerful messaging infrastructure for knowledge discovery and knowledge enabling, which reconcile the differences between underlying knowledge transfer schemas and deployment of collective intelligence platforms.

The proposed design concept exemplifies the spiral evolution of business processes and the cyclic nature of knowledge-rich data mining processes (see Figure 3). The sharing of business data can be done asynchronously and synchronously in knowledge discovery and enabling processes. It should be noted that knowledge-rich data mining technologies are embedded in each of the knowledge discovery and knowledge enabling processes. Also, we may note that the sharing can be organized through group (intranet) communication by designing and using suitable middleware (we may call it as “know-ware”).

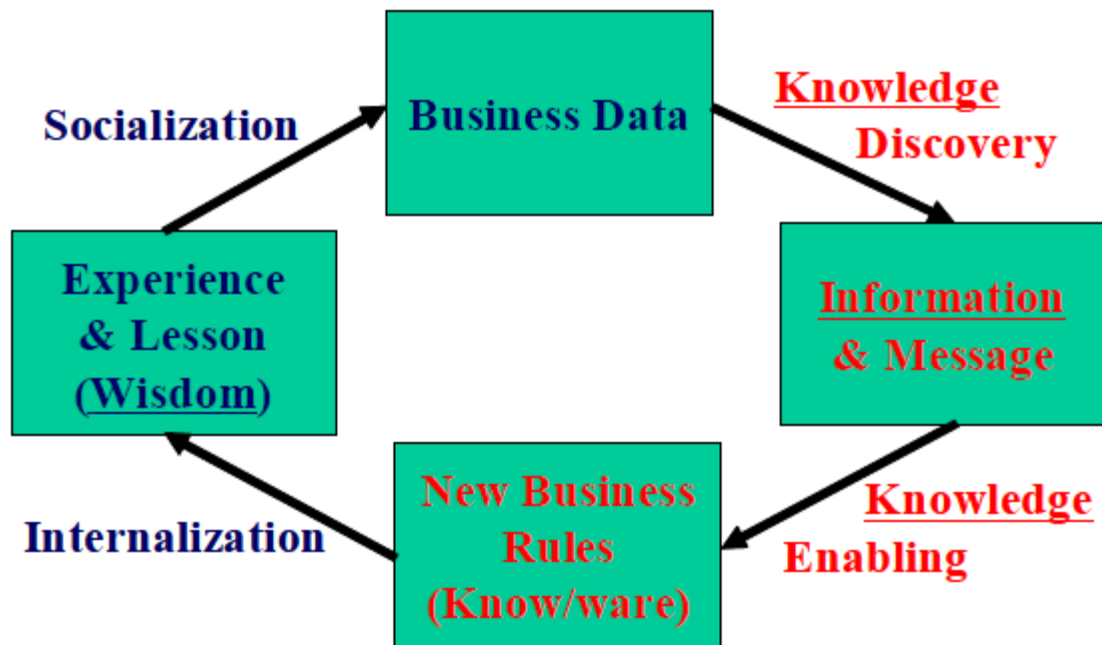


Figure 3. Proposed Design Concepts

In general, collaboration in collective intelligence may bring the core problem for CRM and service applications. For the objective of our approach, design and implementation of a uniform architecture for collective intelligence with automatic collaboration capability has fundamental importance. In the proposed unifying design concepts, we may develop a business “know-ware” platform that is a middleware necessary for the implementation of knowledge-rich data mining.

## 5.0 CONCLUSIONS

Most complex CRM and information-intensive service systems should combine one-to-one encounters through multi-channel interactions based on business “know-ware” platforms utilizing service interface. This paper examined the knowledge-rich data mining to propose a unifying concept that realizes advanced collective intelligence under knowledge modeling framework, especially when the system is information-intensive.

The knowledge messaging, sharing, and collaboration models can be used for advanced design of knowledge-rich data mining in the Web2.0 environment. We have proposed a new knowledge modeling architecture that incorporates the SECI model in knowledge discovery

and knowledge enabling processes. The proposed approach allows maximum use of knowledge transfer or messaging schema with shared context and memory. It suggests that our architecture enables innovative design methodology for collective intelligence. Based on our conceptual design, we may need to develop a business know-ware platform which is a middleware for efficient CRM collaboration.

More thorough analysis of existing collective intelligence technology will identify design patterns necessary for knowledge-rich data mining. In addition, it should be possible to extend the unifying ideas about service interfaces and information exchange to better understand the nature of collective intelligence based on the knowledge-rich data mining.

## Acknowledgement

This work was supported in part by a Grant-in-Aid for Scientific Research of Japan Society for the Promotion of Science.

## REFERENCES

- Dupret, G., and Koda, M. (2001), "Bootstrap re-sampling for unbalanced data in supervised learning," *European Journal of Operational Research*, vol. 134, pp. 141-156.
- Freund, Y., and Schapire, R.E. (1997), "A decision-theoretic generalization of online learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, pp. 119-139.
- Koda, M. (1997a), "Stochastic sensitivity analysis and Langevin simulation for neural network learning (Invited Paper)," *Reliability Engineering and System Safety*, vol. 57, pp. 71-78.
- Koda, M. (1997b) "Neural network learning based on stochastic sensitivity analysis," *IEEE Trans. System, Man, and Cybernetics, Part B: Cybernetics*, vol. 27, pp. 132-135.
- Koda, M., and Okano, H. (2000), "A new stochastic learning algorithm for neural networks," *Journal the Operations Research Society of Japan*, vol. 43, pp. 469-485, 2000.
- Nonaka, I. and Takeuchi, H. (1995), *The Knowledge-Creating Company*. New York, Oxford: Oxford University Press.
- Okano, H., and Koda, M. (2003), "An optimization algorithm based on stochastic sensitivity analysis for noisy objective landscapes," *Reliability Engineering and System Safety*, vol. 79, pp. 245-252.
- Sano, N., Suzuki, H., and Koda, M. (2004), "A robust boosting method for mislabeled data," *Journal of the Operations Research Society of Japan*, vol. 47, pp. 182-196.
- Sano, N., Suzuki, H., and Koda, M. (2008), "A robust ensemble learning using zero-one loss function," *Journal of the Operations Research Society of Japan*, vol. 51, pp. 95-110.
- Vapnik, V.N. (1998), *Statistical Learning Theory*. New York: John Wiley & Sons.