Data Mining for Knowledge Management in Technology Enhanced Learning

JELENA MAMCENKO, INGA BELEVICIUTE
Department of Information Technologies
Vilnius Gediminas Technical University
Saulėtekio al. 11, Vilnius
LITHUANIA

Abstract: A huge amount of data, relevant information must be identified, structured, and directed toward work processes to effectively facilitate or enable the achievement of goals and end products. Knowledge management structures, maintains, and provides access to knowledge within information systems. Data mining extracts relevant tacit knowledge from information system assets through data correlation. The paper describes how data mining techniques can be used for knowledge management in technology enhanced learning.

Key-words: data mining, knowledge management, knowledge discovery, e-learning

1 Introduction

Both organizational learning and knowledge management contribute towards enhancing competitiveness through the active involvement of people. It follows therefore that a complementary link between organizational learning and knowledge management should exist since organizational learning should engage everyone in the exploration [1]. Data mining is intended to provide support in the complex data rich but information poor situations. Organizations often attempt to transform raw data into usable knowledge as a part of their knowledge management initiatives [2]. Data mining is then used to translate structured data into knowledge.

In the paper we argue that data mining can make an important contribution to a knowledge management effort. Our purpose is to show how data mining techniques can be used for knowledge management in technology enhanced learning, which would lead to a better performance and understanding of e-learning participants' behavior.

Knowledge Discovery and Data Mining is an interdisciplinary area focusing upon methodologies for extracting useful knowledge from data. The ongoing rapid growth of online data due to the Internet and the widespread use of databases have created an immense need for KDD methodologies. The challenge of extracting knowledge from data draws upon research in statistics, databases, pattern recognition, machine learning, data visualization, optimization, and high-performance computing, to deliver advanced business intelligence and web discovery solutions.

The paper is organized as follows: the next section presents a review of knowledge management and knowledge management technologies. The following sections discusses about data mining technologies and application of data mining techniques for knowledge management. The final part provides a summary and presents the discussion about future works.

2 Knowledge management and knowledge management technologies

To understand knowledge management is better to distinguish data, information and knowledge. Most notably in IT literature, data is raw numbers and facts, information are these numbers and facts patterned in a certain way. Knowledge is information possessed in the mind of individuals: it is personalized information (which may or may not be new, unique, useful, or accurate) related to facts, procedures, concepts, interpretations, ideas, observations, and judgments [3]. Knowledge can be classified into two types: explicit knowledge, the knowledge included in documents or books and tacit knowledge, the knowledge that can be acquired by experience,

communication. Tacit and explicit forms of knowledge can converse [4].

Knowledge management considers mainly the four basic processes: knowledge creating, storing, transferring and applying. Knowledge management is the name given to the set of systematic and disciplined actions that an organization can take to obtain the greatest value from the knowledge available to it [5]. Information technologies do not have single role or single technology comprising knowledge management. Knowledge management technologies are developed to support and enchase the organizational of knowledge creation, storage, transfer, and application [6]. We will discus each of these processes.

Creation of knowledge takes place through conversion of tacit and explicit knowledge [7] and shown in Table 1 together with examples of applied technologies.

Sharing of tacit knowledge is socialization. It is connected to ideas of communication and collaboration in which information technology plays minimal role [8].

Converting tacit knowledge to explicit is externalization. Examples of information technology applications could be newsgroups and forums. Internalization refers to creation of new tacit knowledge from explicit knowledge and could be promoted by e-learning technologies.

The combination approach refers to the creation of new explicit knowledge by merging, categorizing and synthesizing existing explicit knowledge. This mode is the most flexible of applying information technologies. There are could be used capturing, data mining, search and other techniques. On the explicit side, data mining reflects the highest level of knowledge attainment that requires skills in data domain, data querying and presentation and artificial intelligence/machine [9].

Knowledge storage. Advanced computer storage technology and sophisticated retrieval techniques, such as query language, multimedia data bases, document management technology and database management systems, can be effective tools in enhancing organizational memory [3].

Knowledge transfer is the process through which one unit (e.g., group, department, or division) is affected by the experience of another. They further point out the transfer of organizational knowledge can be observed through changes in the knowledge or performance of recipient units.

Table 1. Conversion of tacit and explicit knowledge

Type of knowledge	**Tacit**	**Explicit**
Tacit to	Socialization *Video conferencing, Instant messaging*	Externalization *Forums, Newsgroups*
Explicit to	Internalization *On-line education*	Combination *Search, Data mining*

Technology can support knowledge application by embedding knowledge into organizational routines. It can enhance knowledge integration and application by facilitating the capture, updating and accessibility of organization directives [3].

3 Data Mining

As independently developed subjects, knowledge management and data mining have not been analyzed as complementary, much less exploited. Both knowledge management and data mining grow and thrive at the confluence of information technologies (machine learning, knowledge-based systems, and databases), statistics and data analysis, as well as the business and management sciences.

One of the successive data analyses in the recent 15 years became data mining technology. There are a lot of data mining definitions such as process of discovering novel, previously unknown, potentially useful, understandable and valid patterns from a large amount of data [10], but in common all of them have the same meaning and is called data mining or knowledge discovery in databases. This discipline brings together databases, decision support systems, machine learning, artificial intelligence, statistics, data visualization, and several other disciplines. Some analytics describes data mining as a complimentary sub discipline in Knowledge Discovery in Databases. Actually, it is enough difficult to distinguish these two technologies. We are considering that data mining application without results interpretation is nothing to mean. So, data mining is a technology, which is used

in the pre last step of knowledge discovery. Data mining has four levels (

Fig. 1) [11]:

- Data selection
- Data transformation
- Data Mining
- Results interpretation

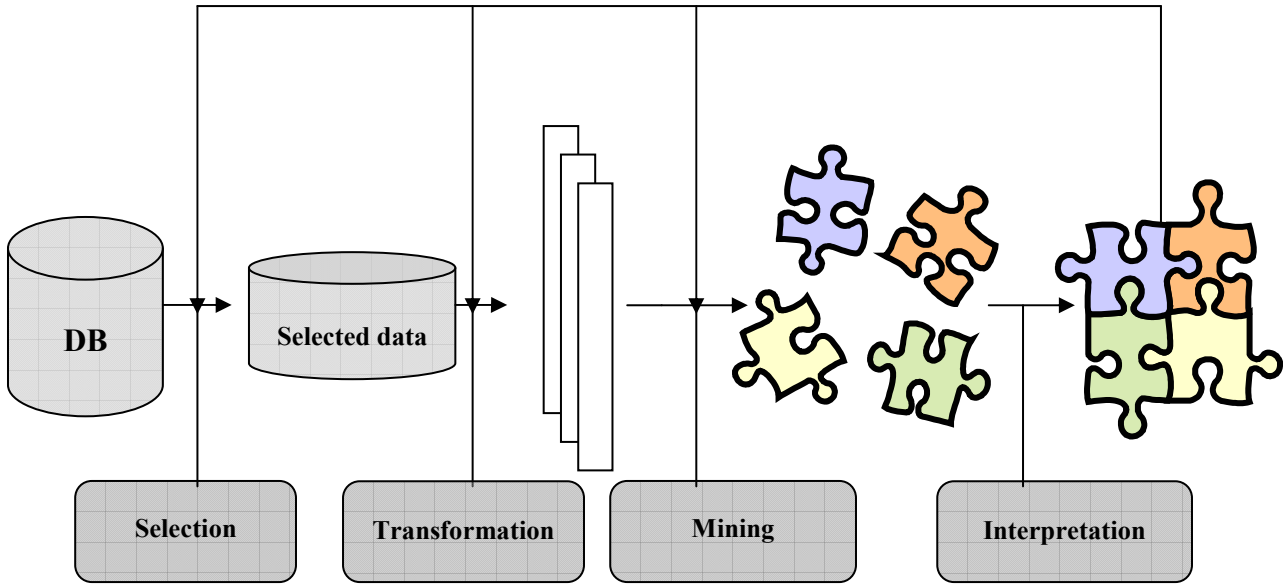

Fig. 1. The data mining process scheme

Difficult level in this big process takes data selection - you need to decide which fields are needed. If you analyze telecoms data, it is enough problematic to choose from 50 or more attributes necessaries fields, that can influence or discover the true value of subscribers. Another heavy step is data transformation. Usually it takes much more time comparing with data analysis. Due to various data saving formats some software is needed their own special format. That is a disadvantage. But almost all of them accept data stored in relational databases. During mining function the needed model is built and after that model should be tested on the real data. The last step is results interpretation. In this level decision should be taken by domain specialist where technology was applied.

Actually, data mining is more suitable for organizations which store and save a lot of data, such as banks transactions or similar. In this case applied technology to huge amount of data can bring the biggest advantages.

4 Data mining for knowledge discovery in e-learning environment

The implementation of data mining techniques for knowledge discovery in the university domain, especially to log files, can help to understand visitors needs, for instance, modification of web page to

better fit for the user, Web page creation that are unique per user or using the desires of a user to determine what documents to retrieve. These pieces of knowledge could lead to better course location on web site and education strategies. Likewise, an analyst could perform OLAP on the data warehouse to determine what kind of courses are the most popular and what students attend the most. For instance as a dimension could be student, event, date or teacher. As a measure we can choose a degree of learning, satisfaction, complementary or other. But there is one requirement for analyzed data: data should be organized and stored so that it can be analyzed from any dimension or at any level of the hierarchy.

According to the results which were received in previous works [12], [13] data mining would enable to help the learners who are interested in certain areas by suggesting relevant or complimentary courses of which they might not be aware, providing with a personalized registration Web page [14]. For the learning provider, they will have the chance to view data of learners and courses from different angles in order to have full picture, enabling them to make the most profitable decision via targeting the class of users of interest to them and investing more in courses that are highly required by their targeted classes of students [15].

There are four most common tasks of data mining [16], [17], which can be applied to e-learning data:

- *Association rules*. These are descriptive patterns of the form A implies B, where A and B are statements about the values of attributes of an instance in a database. In the e-learning database a data mining system may find associations rules such as: learners from 25 to 30 years old with an income of 200€ to 300€ registered for Distant Learning in Information Technologies course.
- *Clustering*. The clustering model also known as segmentation model. Clustering analyses data objects without consulting a known class label. In general, the class labels are not present in the training data simply because they are not known to begin with. Clustering can be used to generate such labels. The objects are clustered or grouped based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity. That is, clusters of objects are formed so that objects within a cluster have high similarity in comparison with one another, but are very dissimilar to objects in other clusters. Each cluster can be viewed as a class of objects from which rules can be derived [17].
- *Classification and prediction*. This refers to discovering predictive patterns where a predicted attribute is nominal or categorical. The predicted attribute is called the class. The derived model is based on the analysis of a set of training data and can be represented in the form of classification rules, decision trees, mathematical formula, neural network and etc [9]. Universities which are providing e-education may be interested, for instance, in classifying courses to expensive, cheap or, maybe, more attractive, normal, etc.
- *Forecast*. This term refers to using the content of the database, which reflects historical data, to automatically generate a model that can predict a future learner's behaviour. Neural networks are the most common forecasting technology in data mining.

The key feature of an e-learning system is to be able to offer an opportunity for getting knowledge from the best teachers and experts in that domain to everybody of high motivation whenever and wherever they are. In order to realize an e-learning society, it is necessary to apply knowledge management applications and data mining techniques for cumulated data about users, courses and etc.

5 Conclusions and future work

The knowledge management has extensive capabilities in many facets of knowledge management, including highly secure data warehousing, expert system information-of-value determination, advanced document and content indexing and access systems, and knowledge engineering. Data mining can be performed in highly automated fashion on single documents or archives of related documents, using both expert systems and conceptual clustering.

A knowledge management system can, for instance, allow curriculum planners, instructors, or learners to search for subject matter experts in the corporation, or find existing relevant materials within a company's intranet.

References:

[1] Sharma R. K., Understanding organizational learning through knowledge management. *Journal of Information & Knowledge Management*, Vol. 2, No. 4, pp. 343-352, 2003.

[2] Davenport T., Grover V., Knowledge management. *Journal of Management Information Systems*, Vol. 18, No 1, 2001.

[3] Alavi M., Leidner D. E., *Review: Knowledge management and knowledge management systems: conceptual foundations and research issues*. MIS Quarterly Vol. 25 No 1, pp. 107-136, March 2001.

[4] Nonaka, I., Takeuchi, H., *The knowledge-creating company*. New York: Oxford University Press, 1995.

[5] Davenport T. H., Prusak L., *Working knowledge: How organizations manage what they know.* Harvard Business School Press, Boston, 1998.

[6] Argote, L., Ingram P., Knowledge transfer A Basis for Competitive Advantage in Firms. *Organizational Behavior and Human Decision Processes* 82(1): 150-169, 2000.

[7] Nonaka I., *The knowledge creating company.* Harvard Business Review 69, pp 96-104, 1991.

[8] Marwick A., Knowledge management technology. *IBM systems journal*, Vol. 40, No 4, pp. 814-830, 2001.

[9] Luan, J., Data Mining and Knowledge Management in Higher Education-Potential Applications. *Presentation at AIR Forum*, Toronto, Canada, 2002.

[10] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R., *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, 1996.

[11] Han, J., Kamber, M., *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publisher, 2001.

[12] Kulvietis G., Mamcenko J., Sileikiene I., Data Mining Application for Distance education Information System. *WSEAS Transactions on Information Science and Applications*, Issue 8, Vol. 3, p.1482-1488, August 2006.

[13] Belevičiūtė I., Šileikienė I., Knowledge Management Applications in a Virtual University Infrastructure. *WSEAS Transactions on Advances in Engineering Education*, Issue 5, Vol. 3, May 2006, p. 354-361.

[14] Monk D., Using Data Mining for E-Learning Decision Making. *The Electronic Journal of e-Learning*, Vol. 2, Issue 1, pp. 41-54, 2005. Is available online at www.ejel.org.

[15] Hanna M., Data Mining in the e-learning domain. *Campus-Wide Information Systems*, Vol. 21 Nr.1, 2004, pp. 29-34.

[16] Simoudis E., Reality check for data mining, *IEEE expert*, Vol. 11 Nr. 5, 1996, pp. 28-30.

[17] Ha S. H., Park S. C., Application of data mining tools to hotel mart of the Internet for database marketing. *Expert Systems with Applications*, Vol. 15, pp.1-31.