

Data Mining and Its Applications for Knowledge Management

Ms. Ekta Choudhari¹ Ms. Pooja Deshpande² Ms. Vrushali Uttawar³

^{1,2,3}Assistant Professor

^{1,2,3}DYPCOE, Akurdi, India

Abstract— Data mining or data processing is one of the most crucial steps in the knowledge discovery in databases process and is considered as subfield in knowledge management. Research in data mining continues growing in learning organization and in business over coming decades. This paper explores the applications of data mining techniques to support knowledge management process. The discussion is divided into 4 topics: (i) knowledge resource; (ii) knowledge types and/or knowledge datasets; and (iii) data mining techniques and applications used in knowledge management. The paper first briefly states the definition of data mining and data mining functionalities Then the knowledge management rationale and major knowledge management tools integrated in knowledge management cycle are described briefly. lastly, the applications of data mining techniques in the process of knowledge management are discussed.

Keywords: Data Mining, Knowledge Management, Knowledge Resources, Data Mining Techniques and Applications

I. INTRODUCTION

Many organizations have collected and stored vast amount of data. Managing knowledge resources can be a challenge. Knowledge management is nothing but process of data usage [6]. The basics of data mining is a process of using tools to withdraw or extract useful information from large datasets; data mining is an essential part of knowledge management [6]. Wang & Wang (2008) point that data mining can be useful for KM in two main manners: (i) to share common knowledge of business intelligence (BI) context among data miners and (ii) to use data mining as a tool to extend knowledge base of humans.

This paper is studies review literature related to application of data mining techniques for KM in academic journals between 2007 and 2012. This paper states data mining in following way: first, data mining definition and the data mining task primitively used in this study are described; second, the definition of knowledge management and the knowledge capture and creation tools are stated; third, articles about data mining in KM are studies and the results of the classification are reported; and last, the conclusions of the study are discussed in this paper.

II. DATA MINING

A. Definition of Data Mining

Data mining is an crucial step in the knowledge discovery in databases and datasets (KDD) process that produces useful patterns or models from data (Figure 1) [7]. The terms of KDD and data mining are totally different. KDD refers to the overall process of extracting useful knowledge from data. Data mining refers to find new patterns from a set of data in databases by focusing on the algorithms to withdraw useful knowledge [7].

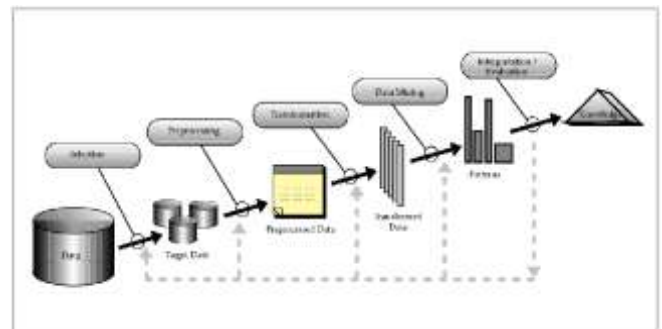


Fig. 1: KDD process consists of iterative methods

- 1) Selection: Selecting relevant data to the analysis task from the database
- 2) Preprocessing: Removing inconsistent and inappropriate data; combining multiple sources of data
- 3) Transformation: Transforming sources of data into appropriate forms to perform data mining
- 4) Data mining: Choosing a data mining algorithm which can be used to pattern in the data; Extracting data patterns
- 5) Interpretation/Evaluation : Interpreting the patterns into knowledge by removing duplicate or irrelevant patterns; Translating the useful patterns into human understandable format

B. Data Mining Tasks

- 1) Fayyad et.al. in 1996 defined six main tasks of data mining:
- 2) Classification is finding models that analyze and then classify data item into several already defined classes
- 3) Regression is mapping nothing but data item to a real valued prediction variables
- 4) Clustering is identifying a finite set of categories or clusters to describe the data
- 5) Dependency Modeling is obtaining model which describes states variable dependencies
- 6) Deviation Detection is discovering significant changes in the data
- 7) Summarization is obtaining a compact description for data

Data mining has two main features of prediction and description. Prediction involves using some variables in data sets in order to predict values of other relevant. Description involves finding human trends and understandable patterns in the datasets.

III. KNOWLEDGE MANAGEMENT

A. Definition of Knowledge Management

There are various knowledge management concepts. In this article we use the definition of knowledge management by McInerney (2002):

“Knowledge management (KM) in data mining is an effort to increase useful knowledge within any organization. Several ways to do this include encouraging communication,

offering opportunities, and promoting the sharing of appropriate knowledge artifacts”

This definition clearly states the interaction aspect of knowledge management and organizational learning.

Knowledge management process focuses on knowledge flows and the process of creation, sharing, and distributing knowledge (Figure 2) [5]. Each of knowledge units of capture and creation, sharing and dissemination, and acquisition and application can be done by information technology.



Fig. 2: KM Technologies Integrated KM cycle

B. Knowledge Management: Capture and Creation Tools

The classification of KM technologies and focuses on tools for capturing and creating of knowledge.

Classification of KM technologies:

- 1) KM Framework
- 2) Knowledge-Based Systems (KBS)
- 3) Information and Communication Technology
- 4) Data Mining
- 5) Database Technology
- 6) Artificial Intelligence (AI)/Expert Systems (ES)
- 7) Modeling

Ruggles et.al. 1997 classified KM technologies as tools that generate knowledge code, and transfer knowledge. Dalkir classifies KM tools according to the phase of the KM cycle. We can see that data mining involves in the part of knowledge creation and capture phase.

Knowledge Creation and Capture Phase	Knowledge Sharing and Dissemination Phase	Knowledge Acquisition and Application Phase
Content creation ■ Authoring tools ■ Templates ■ Annotations ■ Data mining ■ Expertise profiling ■ Blogs	Communication and collaboration technologies ■ Telephone ■ Fax ■ Videoconferencing ■ Chat rooms ■ Instant messaging ■ Internet telephony ■ E-mail ■ Discussion forums ■ Groupware ■ Wikis ■ Workflow management	E-learning technologies ■ CBT ■ WBT ■ EPSS
Content management ■ Metadata tagging ■ Classification ■ Archiving ■ Personal KM	Networking technologies ■ Intranets ■ Extranets ■ Web servers, browsers ■ Knowledge repository ■ Portal	Artificial intelligence technologies ■ Expert systems ■ DSS ■ Customization-personalization ■ Push/pull technologies ■ Recommender systems ■ Visualization ■ Knowledge maps ■ Intelligent Agents ■ Automated taxonomy systems ■ Text analysis—summarization

Fig. 3: Major KM Techniques, Tools, and Technologies

IV. THE APPLICATIONS OF DATA MINING IN KNOWLEDGE MANAGEMENT

A. Knowledge Resources

We divided knowledge resources into five 5 for knowledge resources.

- 1) Health Care Organizations: this domain was a use of the disease knowledge management system (KMS) of the hospital case study [10].
- 2) Financial/Banking: the domain knowledge covered financial and economic data; data mining can assist banking institutions making decision support and knowledge sharing processes to an enterprise bond classification [4].
- 3) Entrepreneurial Science: the knowledge resource was research assets in a knowledge institution [3]; there were three types of the research assets: research products, intellectual capital, and research programs.
- 4) Collaboration and Teamwork: Worker’s log and documents were analyzed each worker’s referencing behavior and construct worker’s knowledge flow.

B. Knowledge Types

Knowledge types in various sectors.

- Health-care System domain, the dataset composed of three databases: the health-care providers’ database; the out-patient health-care statistics database; and the medical status database [11]. Another data source was from hospital inpatient medical records [10].
- Construction Industry domain, sample data in the form of Post Project Reviews (PPRs) as defining good or bad information’s [24].

C. Data Mining DM Techniques and Applications Used in Knowledge Management

Applications of data mining have been used in various enterprises ranging from public health-care, construction industry, Food Company, retailing to finance.

- 1) Classification: Dividing or classifications is one of the most usual learning in data mining. This task aims to mapping of data item to predefined classes.
- 2) Clustering: This involved seeking to identify a finite set of categories and grouping together objects that are similar to each other and dissimilar to the objects belonging to other clusters. This technique has been applied in many fields, for example:
- 3) Common tools used for clustering include k-means, principal component analysis, the Kolmogorov-Smirnov test and the quintile range test and polar ordination.
- 4) Dependency Modeling: This is nothing but finding a model that describes relationships between attribute sets. In medical records management, it is helpful for clinical decision makings.

V. CONCLUSION

In organization, knowledge is an important resource. This aims for a research summary on the application of data mining DM in the KM domain. The results presented in this paper have some assumptions:

- However, we will see the hybridization techniques in order to solve different KM problems.

- KM is an interdisciplinary research area.
It is cleared from this is that data mining so important in data extractions, we can use it to solve several problems in KM.

REFERENCES

- [1] An, X. & Wang, W. (2010). Knowledge management technologies and applications: A literature review. *IEEE*, 138-141. doi:10.1109/ICAMS.2010.5553046
- [2] Berson, A., Smith, S.J. & Thearling, K. (1999). *Building Data Mining Applications for CRM*. New York: McGraw-Hill.
- [3] Cantú, F.J. & Ceballos, H.G. (2010). A multiagent knowledge and information network approach for managing research assets. *Expert Systems with Applications*, 37(7), 5272-5284. doi:10.1016/j.eswa.2010.01.012
- [4] Cheng, H., Lu, Y. & Sheu, C. (2009). An ontology-based business intelligence application in a financial knowledge management system. *Expert Systems with Applications*, 36, 3614-3622. Doi:10.1016/j.eswa.2008.02.047
- [5] Dalkir, K. (2005). *Knowledge Management in Theory and Practice*. Boston: Butterworth-Heinemann.
- [6] Dawei, J. (2011). The Application of Date Mining in Knowledge Management.2011 International Conference on Management of e-Commerce and e-Government, IEEE Computer Society, 7-9. doi: 10.1109/ICMeCG.2011.58
- [7] Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37-54.
- [8] Gorunescu, F. (2011). *Data Mining: Concepts, Models, and Techniques*. India: Springer.
- [9] Han, J. & Kamber, M. (2012). *Data Mining: Concepts and Techniques*. 3rd.ed. Boston: Morgan Kaufmann Publishers.
- [10] Hwang, H.G., Chang, I.C., Chen, F.J. & Wu, S.Y. (2008). Investigation of the application of KMS for diseases classifications: A study in a Taiwanese hospital. *Expert Systems with Applications*, 34(1), 725-733. doi:10.1016/j.eswa.2006.10.018
- [11] Lavrac, N., Bohanec, M., Pur, A., Cestnik, B., Debeljak, M. & Kobler, A. (2007). Data mining and visualization for decision support and modeling of public health-care resources. *Journal of Biomedical Informatics*, 40, 438-447. doi:10.1016/j.jbi.2006.10.003
- [12] Li, X., Zhu, Z. & Pan, X. (2010). Knowledge cultivating for intelligent decision making in small & middle businesses. *Procedia Computer Science*, 1(1), 2479-2488. doi:10.1016/j.procs.2010.04.280
- [13] Li, Y., Kramer, M.R., Beulens, A.J.M., Van Der Vorst, J.G.A.J. (2010). A framework for early warning and proactive control systems in food supply chain networks. *Computers in Industry*, 61, 852-862. Doi:10.1016/j.compind.2010.07.010
- [14] Liao, S.H., Chen, C.M., Wu, C.H. (2008). Mining customer knowledge for product line and brand extension in retailing. *Expert Systems with Applications*, 34(3), 1763-1776. doi:10.1016/j.eswa.2007.01.036
- [15] Liao, S. (2003). Knowledge management technologies and applications-literature review from 1995 to 2002. *Expert Systems with Applications*, 25, 155-164. doi:10.1016/S0957-4174(03)00043-5
- [16] Liu, D.R. & Lai, C.H. (2011). Mining group-based knowledge flows for sharing task knowledge. *Decision Support Systems*, 50(2), 370-386. doi:10.1016/j.dss.2010.09.004
- [17] Lee, M.R. & Chen, T.T. (2011). Revealing research themes and trends in knowledge management: From 1995 to 2010. *Knowledge-Based Systems*. doi:10.1016/j.knosys.2011.11.016
- [18] McInerney, C.R. & Koenig, M.E. (2011). *Knowledge Management (KM) Processes in Organizations: Theoretical Foundations and Practice*. USA: Morgan & Claypool Publishers. doi:10.2200/S00323ED1V01Y201012ICR018
- [19] McInerney, C. (2002). Knowledge Management and the Dynamic Nature of Knowledge. *Journal of the American Society for Information Science and Technology*, 53(12), 1009-1018. doi:10.1002/asi.10109
- [20] Ngai, E., Xiu, L. & Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36, 2592- 2602. doi:10.1016/j.eswa.2008.02.021
- [21] Ruggles, R.L. (ed.). (1997). *Knowledge Management Tools*. Boston: Butterworth-Heinemann.
- [22] Sher, P.J. & Lee, V.C. (2004). Information technology as a facilitator for enhancing dynamic capabilities through knowledge management. *Information & Management*, 41, 933-945. doi:10.1016/j.im.2003.06.004
- [23] Tseng, S.M. (2008). The effects of information technology on knowledge management systems. *Expert Systems with Applications*, 35, 150-160. doi:10.1016/j.eswa.2007.06.011
- [24] Ur-Rahman, N. & Harding, J.A. (2012). Textual data mining for industrial knowledge management and text classification: A business oriented approach. *Expert Systems with Applications*, 39, 4729- 4739. doi:10.1016/j.eswa.2011.09.124
- [25] Wang, F. & Fan, H. (2008). Investigation on Technology Systems for Knowledge Management. *IEEE*, 1-4. doi:10.1109/WiCom.2008.2716
- [26] Wang, H. & Wang, S. (2008). A knowledge management approach to data mining process for business intelligence. *Industrial Management & Data Systems*, 108(5), 622-634.
- [27] Wu, W., Lee, Y.T., Tseng, M.L. & Chiang, Y.H. (2010). Data mining for exploring hidden patterns between KM and its performance. *Knowledge-Based Systems*, 23, 397-401. doi:10.1016/j.knosys.2010.01.014