

Role of Taxonomy in Knowledge Discovery Patterns

Subashka Ramesh and Dr.A.Chandrasekar

Assistant Professor, Saveetha University, Chennai
subashka@gmail.com
Professor, St.Joseph College of Engineering, Chennai
drchandrucse@gmail.com

Abstract

Data Mining is a key in acquaintance discovery process. It is an extraction of erudite data. Knowledge search is done by many methods like statistical algorithms are used to discover patterns and relations in large preexisting databases. It is found that knowledge taxonomy is growing faster and helps in developing many software applications based on different patterns. Knowledge Taxonomy emphasizes on facilitating the efficient recovery and sharing of evidence's and data's for building the taxonomy. Study on building Taxonomy incompetent pattern has engrossed much responsiveness in these days. This paper studies the role of Taxonomy about knowledge innovation process which identifies some directions that offer weighty potential in refining the effectiveness of data by excavating application using Knowledge Nomenclature.

Keywords- Data mining, Knowledge Taxonomy, Information Retrieval, Innovative Pattern.

INTRODUCTION

In the 21st century the human beings are using huge statistics and these statistics are in the different arenas. It may be in the form of **text**, **audio**, still images, **animation**, **video**, or interactivity at ease forms. As the statistics is available in the dissimilar layouts proper stroke to be occupied to retrieve the statistics from data base. The technique used to make the better decision is called as data mining or knowledge Discovery Process (KDD). The essential object the information technology attracted with data Mining is due to the perception of “we are data rich but Information poor”. Data mining, widely identified as Knowledge Discovery in Databases (KDD), it is the nontrivial pulling out of implied, previously unidentified and potentially useful

statistics from data in databases. It is actually the practice of verdict the unknown statistics/form of the repositories.

Taxonomy is a designed fixed of labels and explanations used to unify statistics and papers in a unswerving approach. A acquaintance taxonomy, concentrations on supporting the proficient repossession and partaking of acquaintance, information and facts across an officialdom by building the taxonomy everywhere workflows and facts needs in an innate construction.

Taxonomy is a hierarchical classification of items, which can be used to organize all kinds of knowledge. A classification contains classes of similar items. Taxonomies are critical for the administration of organizations. Pincher claims that, devoid of a taxonomy intended for stowing and organization, or one that cares well searching, all types of management systems in an organization are approximately unworkable. Nonetheless, there is still hesitancy among organizations to promise the essential resources to the proposal and maintenance of taxonomies. There is hesitancy may be interconnected to an scarce understanding of what taxonomies are and what can they do for an associations. Taxonomies can donate to creating explicit knowledge drive in the available point of need. They also help the plotting and classification of tacit knowledge drive in staff expertise. They support group effort and sharing among units and constituent part of an organizations by mapping and orchestrating the sharing. They also help planting knowledge into practice by making sense of the information of the officialdom and creating a common terms and a common way of working. Taxonomies have therefore to be canned as an central part of the knowledge management strategy of the officialdom and when the strategy is implemented as a project, taxonomies are a key task that needs to be scheduled and executed by teams fitted out with the necessary acquaintance and assistances.

This paper describes 5 sections. Section 1 is completely literature survey based on taxonomies for n u m e r o u s tenacities. Section 2 describes the data mining classification tasks and Knowledge Taxonomy. Section 3 describes different taxonomies and there special purposes. Section 4 describes taxonomy development steps and the benefits of developing taxonomy Section 5 describes the application and future direction of data mining using KT so that the research area can pin pointed the following area.

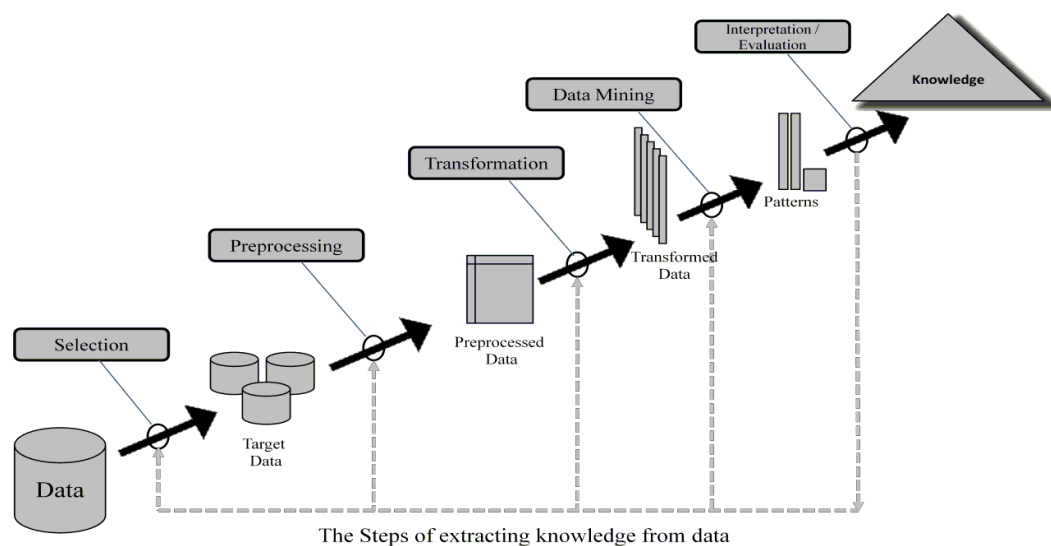
I. LITERATURE SURVEY

The scholars have developed taxonomies for n u m e r o u s tenacities; nonetheless they are not branded as m a n y. Nickerson(1997) developed taxonomy of collective solicitations grounded on the characteristic of communiqué in a group. He made a research archetype that aims to discourse new knowledge about artificial entities that are deliberate to encounter specific goals and provide effectiveness to the users. His contributions were constructs, models, methods, and instantiations. And Rossi, Chatterjee and Williams(2008) chosen dualistic meta-characteristic enterprises and targets in sprouting their taxonomy of numeral amenities. He probed and promoted methodologies that helps employees to choose how to implement and assimilate skills in their work practices and aims at system design that assimilates workers' hidden

awareness of work techniques. Kim, Leem and Suh(2004) develop a taxonomy scheme for mobile industry models based on preliminary with the meta-characteristic of business players. He made a study on how to calculate the taxonomy by measuring its efficacy to categorise objects in the problem domain of estimation. Alden et al. 2012, Jans et al. 2010 designed a standard that facilitate a ready extraction of entity-level transactional data for audit interrogation. In this arena, there has been a small but growing literature that applies data mining techniques to auditing in general and fraud detection in particular. Lewis 2012, Rohman and Berg 2013) in this scholar took a scatter shot approach, investigating patterns in financial statement disclosures, and the nature of journal entries without appropriate guidance drawn from lessons in known fraud patterns.

II. DATA MINING CLASSIFICATION

Data mining is a step in the Knowledge Discovery Process. The series of stages identified in mining knowledge from facts are shown in figure [1]. Data mining has been used in locale such as Statistics, information rescue, data warehousing, fact Visualization, machine erudition, database systems. Data mining method have been commence to fresh areas including spatial fact analysis, patterns detection, neural networks, image databases and numerous appliance fields such as immense business, bio-Informatics and finances. The Practice starts with establishing the Knowledge Discovery targets, and ends with the execution of the revealed information. The overall procedure of discovering and inferring patterns from facts involves the following steps:



Developing and Understanding of the application field

The people who work in Knowledge Discovery Projects should understand the goals of end user in many decisions in which the KDD process will take place.

Selection of a Data Set from which discovery will be achieved

This process includes obtaining additional necessary information, integrating data for the knowledge discovery and selecting a subset of large data discovery. The input selection of a data is large database and output is a target data. This method is very important because the information is extracted from the available data.

Preprocessing and Cleaning

It is the process of eliminating noise and unreliable data. There are many data mining algorithm and statistical methods to do the task. It also includes collecting required information from selected data fields, providing appropriate strategies for dealing with missing data and redundant data.

Transformation of data

It is a stage in which the certain data is changed into forms suitable for the mining process. The methods used in this process include dimension reduction and attribute transformation. The preprocessed data needs to be transformed into predefined format, depending on the data mining task.

Data Mining

It is a stage in which intellectual techniques are applied in order to extract data patterns. Description and Prediction are the major goals in data mining. Prediction is often referred to as supervised data mining and Description is referred unsupervised and conception aspects of data mining. Most data mining methods are based on inductive learning. Data mining methods includes Association rules, Regression, Clustering, Decision Tree and Bayesian.

Interpretation and Evaluation

It is important to gather information from the available methods. In this process the evaluation and interpretation is done with respect to the goals defined in the application field.

Data Repository

Knowledge Discovery in Databases (KDD) can be referred to as the term of data mining which aims for discovering interesting patterns from a database. Therefore, knowledge discovery and data mining should be relevant to any kind of data warehouse. There are many different data supplies where extraction can be applied, including relational databases, transactional databases, multimedia databases and text.

I. Knowledge Taxonomy

Knowledge taxonomy enables the efficient rescue and distribution of knowledge, data and information across an organization by constructing using workflow and information in an intuitive structure. Taxonomies should be defined more by their purpose and use, than by the structural form they happen to take. Different taxonomies are Lists, Tree structures, Hierarchies, Matrices, Facets, System maps.

Table 1: Knowledge domains and the benefits of taxonomies

Domain	Knowledge	Benefit of a taxonomy
Information	Right information is provided when it is obligatory.	Explicit knowledge is embedded in documents.
Expertise and learning	Some of the expertise and experience individuals are within the organization.	Mapping and classification of implied knowledge rooted in staff expertise.
Collaboration	Social aspects of knowledge and sharing within an organization.	Mapping and coordination of sharing.
Culture of the organization	The sharing of knowledge takes place in the organization.	Creates a common vocabulary and a common way of working.

III. TAXONOMY DEVELOPMENT STEPS

The Classification of Taxonomy is based on Meta-Characteristics and Comprehensive. The choice of meta-characteristics is constructed on the persistence of taxonomy. For example if the scholar is trying to classify computer platform in to a taxonomy and based on processing power, then the meta characteristics of the computer platform is the hardware and the software characteristics such as memory, CPU power and OS competence. On the other hand if the computer platform is classified based on how users use them, then the meta-characteristics of the platform is to interact with users, such as the number of concurrent users and user edge. The choice of the meta-characteristic must be done carefully which critically results taxonomy. To develop a beneficial Taxonomy it should have the following attributes[7]

It should be succinct.

It should be sufficiently inclusive. It should be comprehensive.

It should be extendible.

Figure-2 explains Bailey's model for the development of taxonomy. It begins by inspecting a subcategory of objects to organize and next finds general features of these objects.

The features are clustered into dimensions that form the original taxonomy. Each dimension contains characteristics that are equally exclusive and cooperatively exhaustive. The new characteristics and dimensions are finding out based on their effectiveness in categorizing objects. This process follows Bailey's deductive to practical method. The process is repeated until the scholars suitably satisfied that the taxonomy has the qualities of succinctness.

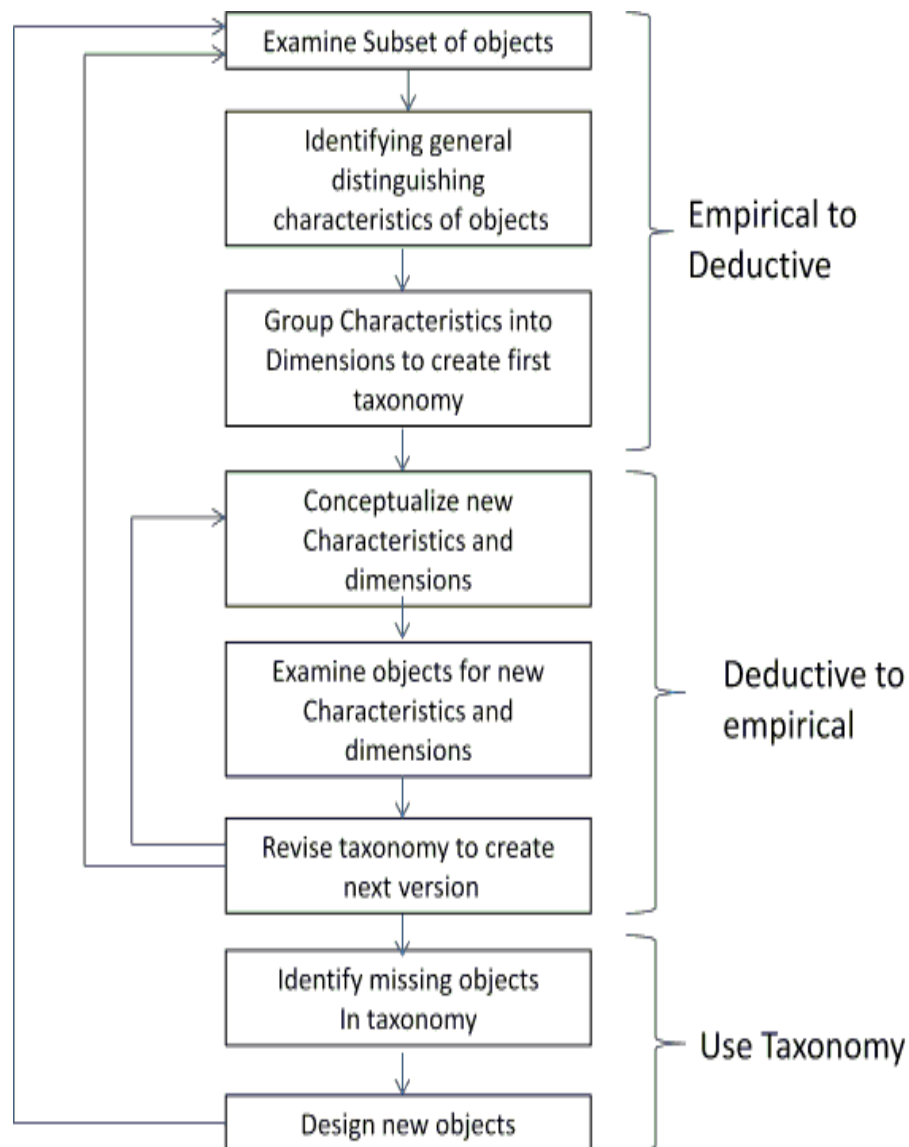


Fig.2 Taxonomy Development Method

The Benefits of developing a Taxonomy includes

- a. Improved quality
- b. Easier navigation
- c. More efficient search
- d. Improved information sharing
- e. Discovery
- e. A better user experience

IV. TASKS AND CHALLENGES USING KNOWLEDGE TAXONOMY FOR GETTING RELEVANT PATTERN

Knowledge discovery tasks depend on which type of functionalities the knowledge system performs and which kind of patterns the system looks for. Different functionalities are developed for achieving different tasks. However, some goals with particular results need to be reached by using the combination of several KDD methods. The main KDD tasks can be classified into the following

Classification

Classification is the process of assigning data objects to desired predefined categories or classes. It also can be viewed as the process of finding a proper method to distinguish data classes or concepts. Objects without a class label are then classified using this method. Generally training data is required for concept learning before classification can be preceded.

Clustering

Clustering is the task of dividing the set of items into a number of groups such that the items in the same group have related features. In other words, clustering aims for maximizing the intra-class resemblance and minimizing the inter- class comparison. The major difference between classification and clustering is that the latter analyses object without consulting class labels, whereas the former needs such information to begin with.

Change and Deviation Detection

Such a task involves the discovery of changes and deviation of specific values in data objects (e.g., the change in time-series data, protein sequencing in a genome and the difference between expected values in ordering data objects).

Mining Association Rules

Associations are rules that describe the frequency and certainty of two groups of data values. This task usually is applied to a transactional database. It discovers the implication between an decedent and consequent, both of which represent sets of items in the transactions. For example, an association rule can be “70% of customers who purchase bread also purchase milk.

V. THE FUTURE OF TAXONOMIES IN DATA MINING

The future lies not in the choice of taxonomies in Data Mining, but rather in their integration to maximize the opportunities they provide. Currently, the application of data mining is to auditing an early stage of development. The researchers create taxonomy for developing structures in data. Future Taxonomies in data mining focus on i. automatic extraction of meta data using Textual data mining on a massive scale. ii. Techniques to search large amount of data from unknown patterns or relationships.

VI. CONCLUSION

Knowledge Taxonomy has received a lot of focus in new emerging Data Mining research field and its significance has been highlighted in both industry and academia. A better understanding of Knowledge Taxonomy assurances to help people share and transfer knowledge within this domain, and thus speed up the advance of this study area. For this reason an important thing should be figure out, how this new research area is connected to present research topics. In this paper the study has been carried out regarding Knowledge Taxonomy by applying data mining analysis.

References

1. "Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999".
2. "Nagle, D., Serenyi, D., Matthews, A., "The Panasas Active Scale Storage Cluster: Delivering Scalable High Bandwidth Storage", Proceedings of the 2004 ACM/IEEE conference on Supercomputing, pp. 53-, 2004."
3. "Dunham, M. H., Sridhar S., "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition, 2006".
4. "Yu, W., Liang, Sh., Panda, D.K., "High performance support of parallel virtual file system (PVFS2) over Quadrics", Proceedings of the 19th annual international conference on Supercomputing, pp. 323-331, 2005".
5. "Fayyad, U., Piatetsky-Shapiro, G., and Smyth P., "From Data Mining to Knowledge".
6. "Discovery in Databases, " AI Magazine, American Association for Artificial Intelligence, 1996".
7. "An Enhanced Data Mining Life Cycle".
8. "Zupan, and J. Demsa, " Open-Source Tools for Data Mining." Clinics in Laboratory Medicine, vol.28, pp. 37-54, 2008".
9. <http://www.lac.uic.edu/~grossman/papers/esj-98.html>
10. <http://www.kdnuggets.com>
11. "Hodge, V., & Austin, J., "A survey of outlier detection methodologies, " Artificial Intelligence Review, 22(2), 2004, (pp. 85–126)".
12. "Hwang, G. J., Hsiao, C. L., & Tseng, C. R., "A computer-assisted approach to diagnosing student learning problems in science courses, " Journal of Information Science and Engineering, 19, 2003, (pp. 229–248)".
13. "Jain, A. K., Murty, M. N., & Flynn, P. J., "Data clustering: A review, " ACM Computing Surveys, 31(3), 1999, (pp. 264–323)".
14. "Merceron, A., & Yacef, K., "Mining student data captured from a web-based tutoring tool: Initial exploration and results, " Journal of Interactive Learning Research, 15(4), 2004, (pp. 319–346)".
15. "Muehlenbrock, M., "Automatic action analysis in an interactive learning environment, " In Proceedings of the workshop on usage analysis in learning

- systems at the 12th international conference on artificial intelligence in education, Amsterdam, The Netherlands, 2005, (pp.73–80)”.
16. “Pahl, C., & Donnellan, C., “Data mining technology for the evaluation of web-based teaching and learning systems, ” In Proceedings of the Congress E-learning. Montreal, Canada, 2003, (pp. 1–7)”.
 17. “Rallo, R. Gisbert, M., & Salinas, J., “Using data mining and social networks to analyze the structure and content of educative online communities, ” In International conference on multimedia and ICTs in education, Caceres, Spain, 2005, (pp. 1–10)”.
 18. “Reyes, P., & Tchounikine, P., “Mining learning groups’ activities in forum-type tools, ” In Proceedings of the 2005 conference on computer support for collaborative learning, Taiwan, 2005, (pp 509– 513)”.
 19. “Romero, C., & Ventura, S., “Educational data mining: A survey from 1995 to 2005, ” Expert Systems with Applications, 33(1), 2007, (pp. 135–146)”.
 20. “Romero, C., Ventura, S., Espejo, P.G. and Hervás, C., “Data Mining Algorithms to Classify Students, ” In Proceedings of the 1st International Conference on Educational Data Mining, 2008, (pp. 8-17)”.

