ISSN:0975-9646

Knowledge Discovery – Techniques and Application

Ajay Kumar
Department of Computer Science
Shivaji College,
University of Delhi

Indranath Chatterjee
Department of Computer Science
University of Delhi,
Delhi - 7

Abstract: Knowledge Discovery in Database (KDD) brings the latest research in statistics, machine learning, databases and AI. These are the part of the rapidly growing field of Data Mining or Knowledge Discovery. Topics covered here included fundamental issues, classification, clustering and application. Different phases of data collection and research issues have been focused.

Keywords: KDD, Machine learning, Data Mining.

1. INTRODUCTION:

Knowledge Discovery in Databases or Data Mining is an interdisciplinary area for extracting useful information from raw data. It is the method of exploring data in order to discover previously unknown patterns. The rapid growth of online data due to Internet and wide use of databases have created an immediate need of Knowledge Discovery methodologies. The challenges of extracting information from raw data brings upon research in pattern recognition, statistics, machine learning, databases, data visualization, optimization and high-performance computing.

Within overall processes of Knowledge Discovery in databases, data mining is an important part. The accessibility and abundance of information today makes data mining a matter of importance and necessity.

Several literatures suggests, such as in Brachmana et al. (1994), Fayyad et al**.** (1996), Maimon et al. (2000) and Reinartz et al. (2002) [1-3], different ways has been proposed by us to divide the KDD process into several phases. We suggests few the KDD process into the following several phases. Note that the process is iterative.

a. We need relevant previous knowledge and the goals for developing an understanding of the application domain.

b. Select a data set.

a. Data Pre-processing: Dimension Reduction is done in this stage (like Feature Selection and Sampling), Data Cleansing (like Removal of Noise or Outliers and Handling Missing Values) and Data Transformation (like Attribute Extraction).

b. Choose the proper Data Mining technique such as: regression, clustering, classification, summarization, etc.

c. Choose the algorithm: In this stage, selection of the specific method to be used for searching patterns is done

d. Employ the selected algorithm.

e. Evaluate and interpret the selected patterns.

f. Deployment: Putting the knowledge into another using the knowledge directly or simply documenting the discovered knowledge.

2. TECHNIQUES:

Data mining techniques can be classified, discovering the knowledge and utilizing the techniques.

Different features of Knowledge discovery: Accuracy, Automated learning, large amount of data, High Level Language, Interesting Results and E
fficiency.

2.1 Association rules:

It Detects sets of attributes that recurrently co-occur and rules. Among them, e.g. 80% of the people_who buy biscuits, also buy sugar (70% of all shoppers buy both).

2.2 Sequence mining (categorical):

Find sequences of events that usually occur together e.g. In a DNA set's sequences ACGTC is followed by GTCA after gap of 9, with probability of 30%.

CBR or Similarity search:

The objects that are within a defined distance of the queried object otherwise it will find all pairs that are within some distance of each other

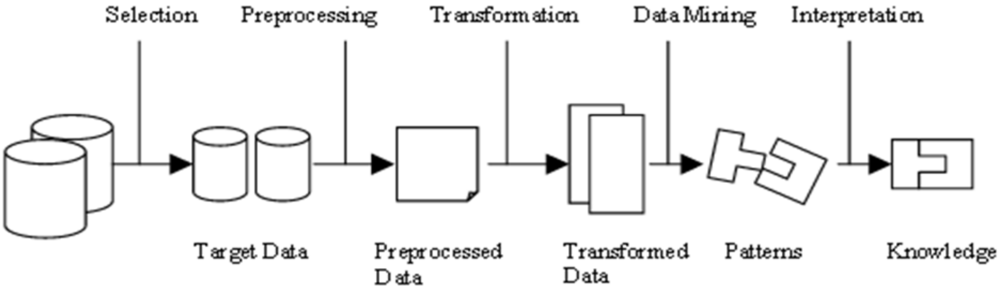

Fig1: KDP

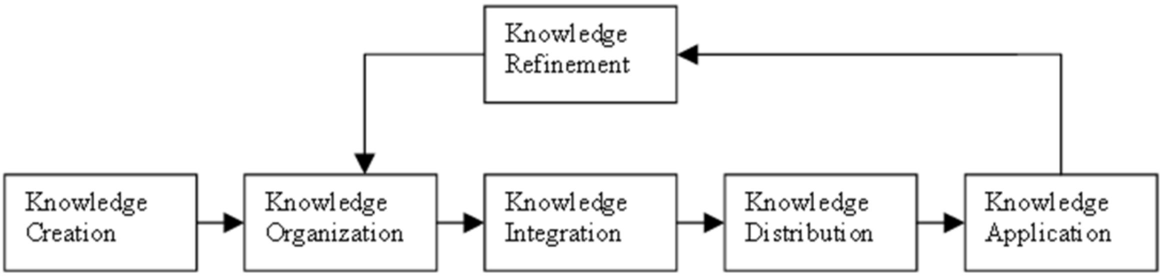

Fig 2: Knowledge discovery in database

Deviation detection:
Discover the records that are the different from the other records, i.e., finding all outliers.

Classification and Regression:
Assigning a new data record to one of the several pre-defined categories or classes. Real-valued fields can be predicted through regression, may called as supervised learning.

Clustering:
Partitioning the dataset into subsets such as the elements of a subset share a common set of properties, with more within group similarity and less inter-group similarity which may called as unsupervised learning.

Many other methods, such as Decision trees, soft computing: rough and fuzzy sets, Hidden Markov models, Time series, neural networks, Genetic algorithms, Bayesian networks.

3. APPLICATION:

- Health care application
- Business application
- Association mining
- Mobile marketing
- Micro-array data mining
- Internet mining
- Application of biological data mining

4. FUTURE SCOPE:

- Scalability:
  Scalability can be tracked into control through various ways such as, Efficient and sufficient sampling, High performance computing, in-memory vs. disk-based processing.

5. CONCLUSION:

We conclude that we presented some definitions of basic terms in the Knowledge discovery. Our primary focus was to clarify the relation between data mining and knowledge discovery. Overview of the KDD process and basic data mining methods have been provided. There are several data-mining techniques, specialized methods for particular kind of data and field. Understanding knowledge discovery and model induction clarifies the task of any KD algorithm. This paper represents a step toward a common idea that we hope will finally provide a unified vision of the overall goals and methods used in KDD. We hope that it will lead to a better understanding of the variety of approaches in the multidisciplinary field of KD.

REFERENCES:

1. Fayyad, Usama M., et al. "Advances in knowledge discovery and data mining." (1996).
2. Frawley, William J., Gregory Piatetsky-Shapiro, and Christopher J. Matheus. "Knowledge discovery in databases: An overview." *AI magazine* 13.3 (1992): 57.
3. Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." *AI magazine* 17.3 (1996): 37.
4. Liu, Huan, and Hiroshi Motoda. *Feature selection for knowledge discovery and data mining*. Vol. 454. Springer Science & Business Media, 2012.
5. Cios, Krzysztof J., Witold Pedrycz, and Roman W. Swiniarski. *Data Mining and Knowledge Discovery*. Springer US, 1998.
6. Bankier, John Duncan, et al. "Method and apparatus for knowledge discovery in databases." U.S. Patent No. 6,567,814. 20 May 2003.
7. Valtchev, Petko, Rokia Missaoui, and Robert Godin. "Formal concept analysis for knowledge discovery and data mining: The new challenges." *Concept lattices*. Springer Berlin Heidelberg, 2004. 352-371.
8. Shahabi, Cyrus, et al. "Knowledge discovery from users web-page navigation." *Research Issues in Data Engineering, 1997. Proceedings. Seventh International Workshop on*. IEEE, 1997