

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/269484827>

An Introduction to Data Mining Technique

Article · August 2014

CITATIONS

0

READS

2,348

2 authors, including:



[Agha Salman Haider](#)

Jazan University

5 PUBLICATIONS 1 CITATION

SEE PROFILE



An Introduction to Data Mining Technique

Mohd. Furkan¹, Agha Salman Haider²

¹Sr Lecturer, Jazan University, Saudi Arabia

² Sr Lecturer, Jazan University, Saudi Arabia

Abstract—

Data mining is the process of extracting out valid and unknown information from large databases and use it to make difficult decisions in business (Gregory, 2000). Data mining or data analysis with complex and large datasets brings the wealth of research and knowledge in machine learning and statistics for the task of discovering new sets of knowledge in large databases. Over the past three decades, large amounts of difficult data's of business are stored electronically and this volume will continue to increase in future. In order to manage huge volumes of data, the techniques of data mining are also becoming sophisticated and advanced, day by day.

Keywords—Data Mining ,DDM, Data Miners,

INTRODUCTION

Smyth, Mannila and Hand (2001) have defined that “progress in digital acquisition storage technology has resulted in the development of vast database. This has happened in all areas of human attempt from the mundane to the exotic. Little wonder then that attention has development in the possibility of tapping these data, of demanding from them information that might be of value to the proprietor of the database. The regulation concerned with this assignment has become recognized as data mining. Defining a scientific discipline is always a contentious task; researchers often disagree about the exact range and limits of their field of study. Bearing this in and tolerant that others might oppose about the details, they shall accept as their functioning definition of data mining”.

Data mining is the analysis of observational data sets to get unsuspected relationships and to sum up the data in novel ways that are both understandable and useful to the data proprietor.

The relationships and summaries derived throughout a data mining exercise are habitually referred to as patterns or models. Illustrations include linear equations, graphs, tree structures, clusters, rules and recurrent patterns in time series. The description over refers to observational data as opposed to investigational data. Data mining classically deals with data that have already been composed for various reasons other than the data mining analysis. This means that the objectives of the data mining implement play on role in the data collection plan. This is single way in which data mining is at variance from much of statistics in which data are frequently composed by using well-organized strategies to reply particular questions (Tan, Kumar and Steinbach, 2006).

Data mining techniques:

1. Web data mining:

The last decade has witnessed the web revolution which has ushered a new information retrieval age. The revolution has had a profound impact on the way they search and find information at home and at work. Searching the web has become an everyday experience for millions of people from all over the world. From its beginning in the early 1990s the web had grow to more than four billion pages in 2004 and perhaps would grow to more than eight billion pages by the end of 2006.



Figure 2.9: Web data mining

Source: articlesweb.org

2. Multi-Relational Data mining:

Chong, Feng and Cao (2010) have described that “Multi-relational classification is an important data mining task, since much real world data is organized in multiple relations. The major challenges come from firstly, the large high dimensional search spaces due to many attributes in multiple relations and secondly, the high computational cost in feature selection and classifier construction due to high complexity in the structure of multiple relations. Mining multi-relational data repositories is an essential task in many applications such as business intelligence. Multi-relational classification is arguably one of the fundamental problems in multi-relational data mining. Multi-relational classification is challenging. First, there may be a large number of attributes in a multi-relational database where classification is conducted. Since, relations are often connected in one way or another; virtually multi-relational classification has to deal with a very high dimensional search space.

3. Distributed data mining:

Pralhad, Ramachandrarao and Adhikari (2010) have explained that “Distributed Data Mining (DDM) algorithms deals with mining multiple databases distributed over different geographical regions. In the last few years, researchers have started addressing problems where databases stored at different places cannot be moved to a central storage area for variety of reasons. In multi-database mining there are no such restrictions. Thus distributed data mining could be considered as a special type of multiple database mining. Distributed data mining environment often comes with different distributed sources of computation.

4. Graph mining:

Kantardzic (2011) has described that “Graph mining applications are far more challenging to implement because of the additional constraints that arise from the structural nature of the underlying graph. The problem of frequent pattern mining has been widely studied in the context of mining transactional data. Recently, the techniques for frequent pattern mining have also been extended to the case of graph data”.

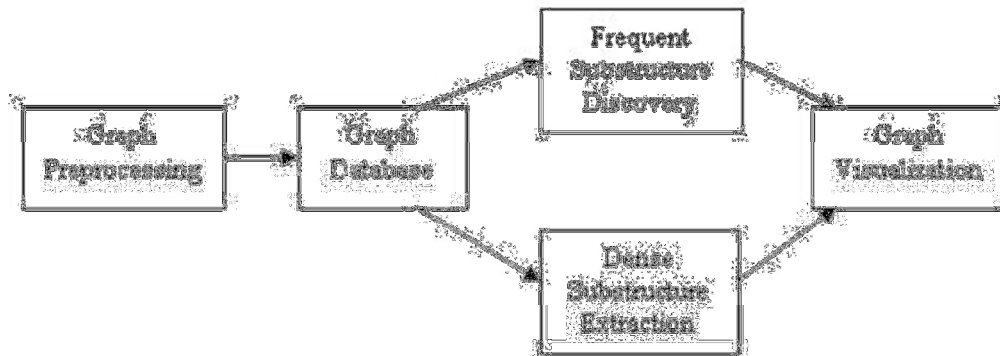


Figure 19: Graph mining

Source: cdacmumbai.in

5. Visual data mining:

Kimani, Dix and Catarci (2010) have explained that” visual data mining is the use of visualization technique to allow data miners and analysts to evaluate, monitor and guide the inputs, products and process of data mining”. The field of visual data mining primarily around the exploitation of the human visual system in mining knowledge. In essence, this can be realized by placing the user at a strategic place in the system framework while the same time exploitation effective visual strategies. Visual data mining may therefore, be fined as the exploitation of appropriate visual strategies in order to allow enable or empower the data mining user to process data and also to drive guide or direct entire process of data mining.

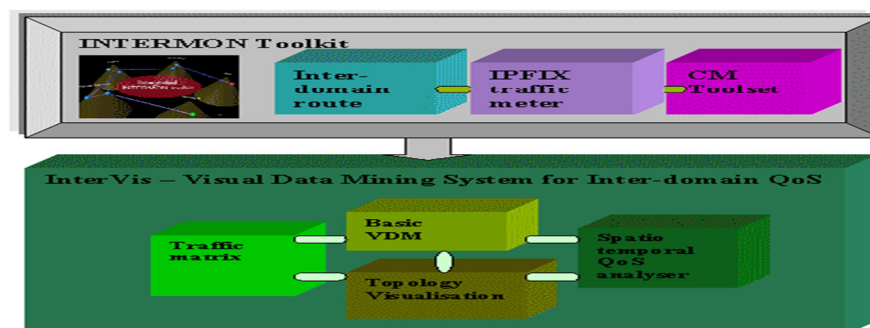


Figure 4.10: visual data mining



CONCLUSION

This study explores the evolution of data mining, we have discussed different data mining techniques used for data mining, we will discuss the best data mining technique in our next paper.

REFERENCES

- Smyth, Mannila and Hand (2001), Principles of data mining, MIT Press, p 1.
- Tan (2007), Introduction to Data Mining, Pearson Education India, p 1.
- Hand (1998). "Data Mining Statistics and More?", The American Statistician, USA, p 112–118
- Cerrito (2006), Introduction to data mining using SAS Enterprise Miner, SAS Publishing, p 1.
- Larose (2005), Discovering knowledge in data: an introduction to data mining, John Wiley and Sons, p 2.
- Sullivan (2011), Introduction to Data Mining for the Life Sciences, Springer, p 2.
- Pei, Kamber and Han (2011), Data Mining: Concepts and Techniques, Elsevier, p 1.
- Sivanandam and Sumathi (2006), Introduction to data mining and its applications, Springer, p 2.