Middle-East Journal of Scientific Research
24 (Recent Innovations in Engineering, Technology, Management & Applications): 160-163, 2016
ISSN 1990-9233; © IDOSI Publications, 2016
DOI: 10.5829/idosi.mejsr.2016.24.RIETMA125

Review on Extraction of Intelligence from Higher Order Mining

[1]*Mrs. M. Deepanayaki and* [2]*Vidyaathulasiraman*

[1]PhD Research Scholar, Department of Computer Science, Periyar University, Salem, India
[2]Assistant Professor & Head, Department of Computer Science,
Government Arts & Science College (W), Bargur, India

Abstract: The rapid growth of data sizes to terabytes in everyday life made the researchers to realize the need of efficient analysis of data to extract the intelligence from that enormous amount of complex data set. Intelligence is a term describing the capacities of mind, plays a key role in predicting our abilities and future. Most data mining algorithms and frameworks have concentrated on the extraction or mining of interesting rules and patterns directly from collected data. Investigating the mined interesting rules and the utility of mining rules from the results of other data mining routines still achieving the required intelligence is the process of Higher Order Mining (HOM). This process will greatly reduce the burden of programmers as it is a very high level as well as hierarchical programming scheme suitable for the development of knowledge-intensive tasks. This Paper discusses the process of analyzing the patterns and the utility of deriving rules from the results of data mining routines.

Key words: Data Mining and Algorithms・Artificial Intelligence・Big Data・High Dimensional Data

INTRODUCTION

Data Mining, an important term dominating the world by its widespread presence.It impacts our daily lives, whether we realize it or not such as shopping, searching, medical diagnosis and treatment, defense, law enforcement etc, [1]. Since everything and everywhere are computerized there occurs a collection of vast amount data with high dimensions. The important information hidden in these vast data is the whole sole attraction of the researchers of multiple disciplines to make and study in developing effective approaches to derive the hidden intelligence or knowledge within them.There are various different domains in data mining such as web mining, text mining, Spatial mining, customer relationship mining, Sequential mining etc. Among which Higher order mining is the combination of the various mining techniques such as Association rule mining, classification, prediction, cluster analysis, decision rules etc., Thus the process of mining the mined data or Non Primary data is higher order mining. The necessity of Higher Order Mining grows when the size and dimensionality of data grows.

Artificial Intelligence: The general problem of simulating (or inventing) intelligence has been broken down into a number of specific modules or sub-problems. Intelligence has been defined in many different ways such as in terms of one's capacity for logical thinking, abstract and innovative knowledge and level of understanding, self-awareness and communication skill, learning ability, emotional knowledge, memory power, proper and successful planning, creativity and problem solving. It can also be more generally described as the ability to grasp the information and modify it as knowledge for applying to itself or other instances of knowledge or information.

Although humans have been the primary focus of intelligence, researchers and scientists have also attempted to investigate and worked on Artificial intelligence that is intelligence in machines. Generally intelligence or a fulfilled Artificial Intelligence has not yet been achieved and is a long-term goal of AI research. Among the Researches so far, Scientists hope machines will exhibit are reasoning, knowledge, planning, learning, communication, perception and ability to move and to manipulate objects. In the field of artificial intelligence there is no assurance on how closely the brain should be simulated.

Corresponding Author: Mrs. M. Deepanayaki, PhD Research Scholar,
Department of Computer Science, Periyar University, Salem, India.

AI Techniques: Artificial Intelligence techniques are widely used in Data Mining Process. AI Techniques such as pattern recognition, machine learning and neural networks have received much attention towards Data mining. Other techniques in AI such as knowledge acquisition, knowledge representation and searching are relevant to the various process steps in DM. AI techniques that can be used for DM include case-based reasoning and intelligent agents. Case-based reasoning uses historical cases to recognize patterns and the intelligent agent approach employs a computer program (i.e. an agent) to shift through data.

Data Mining: Data mining is the process of applying various techniques and algorithms for analyzing and classifying with the intention of uncovering hidden patterns and knowledge in large data sets [2]. It bridges the gap from Applied Statistics and Artificial Intelligence to Database Management by exploiting the way data is stored and indexed in databases to execute the actual learning and discovery of knowledge more efficiently. The goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Not all patterns found by the data mining algorithms are necessarily valid. It is common for the data mining algorithms to find patterns in the training set which are not present in the general data set [3]. The necessity of mining the mined data arises here.

Research Areas of Data Mining: Research efforts of Data mining includes the integration of data mining with data warehouses or various database systems, need to mine patterns across distributed environments which leads to the development of Fast Distributed Mining (FAM), Optimized Distributed Association Mining (O-DAM) [4], new methods to handle and analyze the complex, huge, high dimensional data, mining software bugs, social network analysis, multi-relational and multi- database data mining etc.,

Higher Order Mining (HOM): The value of knowledge obtainable by analyzing large quantities of data is widely acknowledged. However, so-called *primary* or *raw data* may not always be available for knowledge discovery for several reasons.

First, cooperating institutions that are interested in sharing knowledge may not be willing (or allowed) to disclose their primary data.

Second, data in the form of streams are only temporarily available for processing. If stored at all, stream data are maintained in the form of synopses or derived, abstract representations of the original data.

Finally, even for non-stream data, there are limits on the computation speed to be achieved; such limits are set by hardware and firmware technologies. This problem can only be partially solved through parallelization and increased processing power. Ultimately, in many cases data must be summarized to be processed efficiently.

In the light of these observations, we anticipate the need for defining and practicing data mining without the luxury of primary data. To that end, the paradigm of *Higher Order Mining* [5] is introduced as a form of data mining that is applied over non-primary, derived data or patterns. Although Higher Order Mining is a new paradigm, there are already research advances on knowledge discovery methods from patterns rather than data. HOM is a routine mining process. Though the process is complex, its necessity is inevitable.

Higher order Mining are employed in a number of real world applications such as law enforcement, homeland defense, supercomputing applications. Some of the Higher order mining under research are Distributed Higher order Text mining (DHOTM), Higher order Association Rule Mining (HOARM), Latent (LHOIM) and Explicit Higher order Itemset Mining(EHOIM).

HOM Process: It is identified that the process of HOM proceeds with the following steps:

s

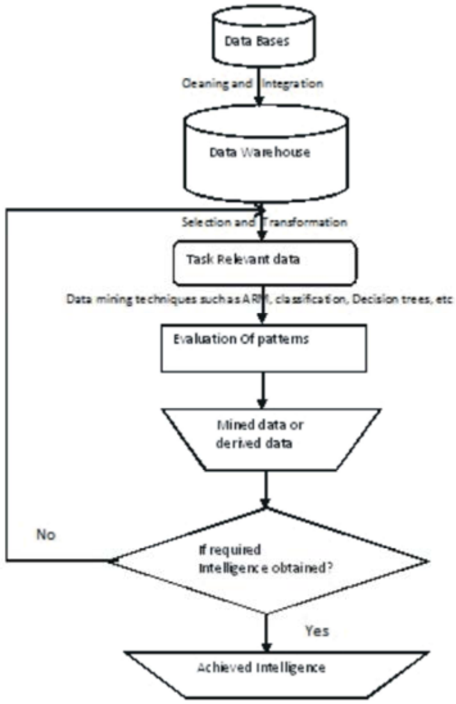
Data Bases
Cleaning and Integration
Data Warehouse
Selection and Transformation
Task Relevant data
Data mining techniques such as ARM, classification, Decision trees, etc
Evaluation Of patterns
Mined data or derived data
No
If required Intelligence obtained?
Yes
Achieved Intelligence

Fig. 2: Higher Order Mining Process: (HOM Process)

Step 1: The core databases such as Object-relational DB systems, Temporal and Time-series DB systems, Heterogeneous and Legacy DB systems, Data Stream Management System, Web-Based information systems are the data sets in which Cleaning and Integration [1] is performed and the results are stored in the respective Data Warehouses.

Step 2: After which Data selection and Data Transformation [1] is performed in the Warehouse data to transfer the data in the form suitable form data mining.

Step 3: Data obtained is applied to the desired Data mining Techniques [1] such as Frequent Item set Mining, Association Rule Mining, Classification and Prediction techniques, Cluster Analysis, Outlier Analysis and Evolution Analysis, Neural Networks, Genetic Algorithms etc., to obtain the interesting Patterns.

A Data Mining system can generate thousands or even millions of patterns or rules, but only few are interesting. A pattern is interesting if it is:

- Easily understood by humans
- Valid on a test data (with some degree of certainty)
- Useful & Novel
- Validates a user defined hypothesis

Step 4: Thus the Completeness of Data Mining Algorithms reaches only when it generates all of the interesting Patterns. When Data mining algorithm produces interesting patterns then it is highly desirable & efficient, but it is highly challengeable in Data Mining Domain. If the patterns are not realistic and efficient then the user can go for further search on interesting patterns. This Further Search can be done again with the similar data mining techniques which are used in the first iteration or it can be a different mining techniques. This loop can be continued till achieving the intelligence required. This Process is Called Higher Order Mining.

In particular, there is no reason to expect that one type of data mining approach will be suitable for all types of data, even for all high dimensional data. Statisticians and other data analysts are very cognizant of the need to apply different tools for different types of data.

Major Application Areas of HOM: Data mining has huge application areas such as[2]:

Banking: Loan/credit card approval predict good customers based on old customers

Customer Relationship Management: Identify those who are likely to leave for a competitor.

Targeted Marketing: Identify likely responders to promotions

Fraud Detection: Telecommunications, financial transactions from an online stream of event identify fraudulent events

Manufacturing and Production: Automatically adjust knobs when process parameter changes

Medicine: Disease outcome, effectiveness of treatments analyzes patient disease history and finds relationship between diseases

Molecular/Pharmaceutical: Identify new drugs

Scientific Data Analysis: Identify new galaxies by searching for sub clusters

Web Site/Store Design and Promotion: Find affinity of visitor to pages and modify layout

HOM dominantly work in all these areas including defense and Law enforcement, Astrology, Weather Forecasting, Space research, Financial data analysis, Biological data analysis, Scientific data analysis where lies the huge and high dimensional data with the scope of identifying intelligence.

Advantages of HOM: Some of the beneficiary tasks of High order mining are:

- With the ever-increasing size of data and the need for that data, mining of large and high dimensional databases poses a demanding task that should satisfy both the requirements of the computation efficiency and result quality.
- The usage of HOM is very much required in the field where there are Big Data, High Dimensional Datasets in all areas of science, Engineering and Businesses. These include genomics and proteomics, biomedical imaging, signal processing, astrophysics, finance, web and market basket analysis, among many others. The number of features in such data is often of the order of thousands or millions that is much larger than the available sample size. This renders classical data analysis methods inadequate, questionable, or inefficient at best and calls for new approaches.

- Some 15 years ago, Stanford statistician D. Donoho predicted that the 21st century will be the century of data.

"We can say with complete confidence that in the coming century, high-dimensional data analysis will be a very significant activity and completely new methods of high-dimensional data analysis will be developed; we just don't know what they are yet.", by D. Donoho,2000.

This paper is one of the suggestions of analysis of such data.

Thus, it is argued that this approach has three significant advantages.

Firstly, with the expansion of dataset size, the tractability of mining from the complete dataset may be difficult on a regular basis,

Secondly, changes in observations (and therefore in the observed system) can be more easily discovered by inspecting changes in extracted rules over time (or over any other sequential progression) and Finally, the natures of the rules extracted by this process are that they contain different higher order semantics from that exhibited by first order discovery process.

Limitations of HOM: The Limitations of HOM depends on the data it works that is data with high dimensions, complex data sets, statistical data etc., Some of the Limitations of this curse of dimensionality are the following [5]:

- High dimensional geometry defeats our intuition rooted in low dimensional experiences so that data presentation and visualization become particularly challenging.
- Distance concentration is the phenomenon of high dimensional probability spaces where the contrast between pair wise distances vanishes as the dimensionality increases - this makes distances meaningless and affects all methods that rely on a notion of distance.
- Bogus correlations and misleading estimates may result when trying to fit complex models for which the effective dimensionality is too large compared to the number of data points available.
- The accumulation of noise may confound our ability to find low dimensional intrinsic structure hidden in the high dimensional data.
- The computation cost of Higher order processing is often high.
- In general terms, problems with high dimensionality result from the fact that a fixed number of data points become increasingly "sparse" as the dimensionality increase.
- Privacy, security and misuse of information are the big problems if they are not addressed and resolved properly.

CONCLUSION

This review paper aims to promote new advances and research directions to address the curses and to uncover and exploit the blessings of high dimensionality in data mining with the help of HOM [5]. Research interest ranges from theoretical foundations, to algorithms and implementation, to applications and empirical studies of mining high dimensional, complex distributed data and big data with the influence of higher order mining.

REFERENCES

1. Jiawei Han and Micheline Kamber, 2008. Data Mining Concepts and Techniques" Second Edition, Elsevier, Reprinted 2008.
2. en.wikipedia.org/wiki/Data_mining
3. Multidimensional Data Analysis and data mining, Black Book, Arijay Chaudhry and Dr. P.S. Deshpande.
4. A Review on Data mining from Past to the Future. Venkatadri.M,Research Scholar, Dept. of Computer Science, Dravidian University, India. Dr. Lokanatha C. Reddy Professor,Dept. of Computer Science, Dravidian University, India. International Journal of Computer Applications (0975 – 8887) Volume 15– No.7, February 2011.
5. Proceedings of The 1st International Workshop on High Dimensional Data Mining (HDM) In conjunction with the IEEE International Conference on Data Mining (IEEE ICDM 2013) in Dallas, Texas.