International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.9756/INTJECSE/V14I5.729 ISSN: 1308-5581 Vol 14, Issue 05 2022

"A Study of Algorithms, Systems, and Applications of Multi-Agent Systems for Distributed Data Mining"

Ph.D Schoolar: - Pooja
Subject:- Computer Science

UNDER THE SUPERVISION OF
Dr. Manish Varshney
Professor
Dept. of Computer Science,
Maharishi School of Engg. & Tech.,
MUIT University, Lucknow, U.P.

Abstract

Data mining is a powerful technology that converts data into competitive intelligence which businesses can use to proactively predict future trends, uncover meanings to historical happenings and discover business imperatives which was hitherto unknown to business. Once the task have been completed, the DM agents have been sending the results back to the task agent. When the task agent has all the results, it has been processed the results (for example it may combine them), and notifies the user agent (in some cases there may of course only be one set off results). The user agent then displays the combined results to the user.

Keyword: - Data Mining, Intelligence, Historical, Business.

Introduction

Data mining is a "process of discovering and interpreting previously unknown patterns in data to aid business in better decision-making". Data mining is a set of automated techniques used to extract buried or previously unknown pieces of information from large databases.

Data mining by nature is an iterative process and gets refined for further probing into data in a continuous manner. From a data management point of view, the data mining process requires exploration of data, creating the analytic data sets for evaluation, generating patterns and forecasting models.

Data Mining

Data mining is the process of discovering interesting patterns and knowledge from large amounts of data. The data sources can include databases, data warehouses, the web, other information repositories or data that are streamed into the system dynamically. Data mining which is sometimes also called as Knowledge Discovery in Database (KDD) is the process of analyzing data from different perspectives and summarizing it into useful information.

Today, data mining is used by many organizations with a strong consumer focus such as retail, financial, communication, and marketing organizations. Extraction of hidden predictive information from large databases is a powerful new technology with great potential to help organizations focus on the most important information in their data warehouses. For knowledge discovery from databases numerous techniques such as Genetic Algorithm, Clustering, Association Rules, Classification, Neural Networks, Decision Trees, and Regression are used. In recent years, data mining has been used widely in the areas of science and engineering, such as bioinformatics, genetics, medicine, education and electrical power engineering. It has been said that knowledge is power, and this is exactly what data mining is about. It is the acquisition of relevant knowledge that can allow making strategic decisions.

DISTRIBUTED DATA MINING

Distributed data mining in distributed environments like virtual organization networks, the Internet, corporate intranets, sensor networks, and other decentralized infrastructures questions the suitability of centralized KDD architectures for large-scale knowledge in a networked environment.

Distributed data mining works by analyzing data m a distributed fashion and pays a careful attention to the trade-off between centralized collection and distributed analysis of data. When the data sets are large scaling up the speed of the data mining task is crucial. Parallel knowledge discovery techniques address this problem by using high performance multi-computer machines.

Review of Literature

Distributed data mining (DDM) is an important research area. One of the approaches suitable for the DDM is to select relevant local patterns from the distributed databases. Such patterns, often called prototypes,

are subsequently merged to create a compact representation of the distributed data repositories. To assure obtaining homogenous prototypes the feature selection requires collaboration of agents [2].

Ashadevi (2011) have proposed a cost effective approach for materialized views selection in data warehousing environment. For efficient design of data warehousing system needs better materialized view selection. Query execution frequencies, Query access costs, base relation update frequencies, system storage space constraints are the performance metrics. The framework designed is cost effective with optimized maintenance that leads to efficient data warehousing system. In this research work focus is given to the essential constraints maintenance cost and storage space.

Astha and Manish (2012) has presented a review of data mining techniques in cloud computing database. Data mining is the extraction of information from large data set with the help of different data mining techniques. The existing data mining techniques like clustering, classification, genetic algorithms help in finding unknown information from the database. Cloud Computing is the technology where the resources are the shared services. A bulk of business data is stored in Cloud Data centers. In this work data mining technique in cloud computing using Google App Engine and Cloud SQL is implemented.

Research Objectives

The in-depth study of literature and the challenges thus presented laid the foundation of current research work. Following is the list of objectives that are being achieved during this research work.

(i) To investigate and analyze the literature thoroughly and to investigate some of the MADM issues: scalability, efficiency, and portability, extendibility and privacy protection.

(ii) To study the under-pinning philosophy of MADM frameworks and proposed the EMADS framework

(iii) To study the MADM, and consequently the EMADS framework, requirements, architecture, design and implementation.

(iv) To experiment at least with respect to meta ARM, that MADM and by extension EMADS, offers positive advantages in that all the meta ARM algorithms has been more computationally efficient than the bench mark algorithm.

(v) Literature survey

(vi) Conceptual framework development

Research Methodology

In order to achieve the above stated objectives, the instruments used in the study are:

(vii) Development of Computer Algorithms/ Programs

(viii) Extensibility of MADS requirements, architecture, design and implementation

(ix) Interpretation of Results

(x) Publication of technical papers

(xi) Improvement by the way of feedback of survey and technical publications, models studied in the field.

(xii) Compilation of the final results, the theories developed/ evolved & final preparation of thesis.

Research Procedure

A number of different categories of task agent have been identified in the context of EMADS; these have been discussed in different chapters. However, in general terms, the task agent have been sending a request to each identified DM agent in the list (there may only be one) to begin DM together with appropriate references to the identified data agents. Each DM agent has been accepted the request and begins mining their applicable data source. Once the task have been completed, the DM agents have been sending the results back to the task agent. When the task agent has all the results, it has been processed the results (for example it may combine them), and notifies the user agent (in some cases there may of course only be one set off results). The user agent then displays the combined results to the user.

CONCLUSION

MADM, as proposed here, seeks to benefit from the possibilities offered by MASs to improve overall DM performance; or in other words, the management of intelligent agents with the capacity to perform data mining. The term “management” denotes the ability to dispatch such agents across participant sites, and also the potential to support the communications between these agents. MADM allows each participant to utilise their own local data and algorithms and, at the same time, benefit from the data and algorithms that are made available by the other participants worldwide. The principal MADM advantages envisaged are those of experience and resource sharing, flexibility and extendibility, and (to an extent) protection of privacy and intellectual property rights

References

1. R. Agrawai, M. Mehta, J. Shafer, R. Srikant, A. Arning, and T. Bollinger. The Quest Data Mining System, In Proceedings of the 2nd international Conference Knowledge Discovery and Data Mining, (KDD), 1996.
2. R, Agrawai and G. Psaila. Active Data Mining. In Proceedings of the 1st International Conference Knowledge Discovery in Data Mining, AAAI, pages (3-8), 1995.

3. R, Agrawai and R. Srikant. Fast algorithm for mining association rules. In Proceedings of the 20th Internationa! Conference on Very Large Data Bases, Santiago de Chile, Chile, pages (487-499), 1994.
4. K. Ali, S. Manganaris, and R. Srikant. Partial classification using association rules. In Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD, AAAI Press), Newport Beach, CA, USA, pages (115-118), 1997.