ISSN 2348 – 8034
Impact Factor- 3.155

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

AN AGENT BASED DISTRIBUTED DATA MINING USED IN E-COMMERCE ORGANIZATION AND ITS EXPLICIT SPECIFICATION OF CONCEPTUALIZATION

Miss Rupali Hinglaspure[*1] & Prof. S. W. Ahmad [2]
[*1,2]Computer science and Engineering Department, PRMIT&R Badnera, India

ABSTRACT

Distributed data mining technology has emerged as a means of identifying patterns and trends from large quantities of data which located onto the different sites. Distributed data mining has used a data warehousing model of gathering all data into a central site, then running an algorithm against that data and extract the useful knowledge, pattern from data. With the increased complexity in number of applications and due to large volume of availability of data from heterogeneous sources, there is a need for the development of suitable a set of types, properties, and relationship which can handle the large data set and present the mined outcomes for evaluation intelligently. In the era of intensive data driven applications distributed data mining can meet the challenges with the support of agents. This system is used in large organization like e-commerce, supermarket, and banks etc. System developed for the e-commerce organization for maintain and extract the distributed database.

Keywords: *Introduction, Literature Survey, System Architecture, System Design, System Implementation*

I. INTRODUCTION

Distributed data mining or knowledge discovery is concerned with extracting knowledge from databases and/or knowledge bases using machine learning techniques. Knowledge discovery from the databases presents another approach where pattern from the large repository is extracted and analyzed. Data is almost growing in exponential terms and most of the relevant applications are driven by it. One simple data mining technique cannot provide the required solution for it. Main challenges of data mining includes, existence of huge data set located at different sites, intensive computation process with large set and rapid change in input data[1]. The distributed data mining (DDM) mines data sources irrespective of their physical locations. The first order learning is to enables us to explore the aspects of knowledge integration and theory refinement which do not appear in propositional systems. Software has the response to the problem of using the vast amounts of information stored on networked systems. There are many types of software agent, however agents are typically thought of as being 'intelligent' programs which have some degree of self-sufficiency. We intend to design an open, flexible data mining agent. A group of these agents will be able to cooperate to discover knowledge from distributed sources [3].

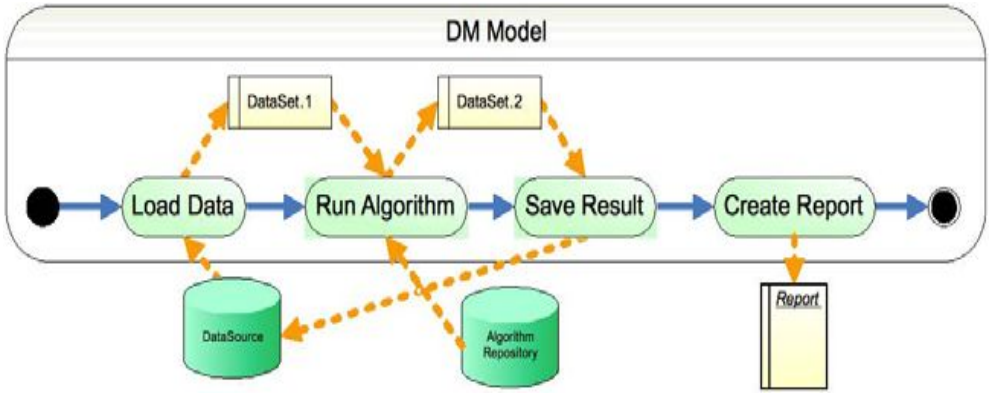

Fig.1. A simple distributed data mining model

Figure 1 shows a simple data mining model defines a series of activities: Load data, run algorithm, save result, and create report. Each activity may produce any arbitrary data sets and store them in the model memory for other activities to consume. Sharing inputs and outputs does not occur in a single memory unit, due to distributed

ISSN 2348 – 8034
Impact Factor- 3.155

environment of the system. Inputs and outputs are determined to be transferred from one agent to another agent as needed.

Distributed data mining

Distributed computing plays an important role in the data mining process for several reasons. First data mining often requires huge amounts of resources in storage space and computation time. To make systems scalable, it is important to develop mechanisms that distribute the work load among several sites in a flexible way. Second data is often inherently distributed into several databases, making a centralized processing of this data very inefficient and prone to security risks. Distributed data mining explores techniques of how to apply Data Mining in a non-centralized way.

What is an agent?

Agents are defined as software or hardware entities that perform some set of tasks on behalf of users with some degree of autonomy. In order to work for somebody as an assistant, an agent has to include a certain amount of intelligence, which is the ability to choose among various courses of action, plan, communicate, adapt to changes in the environment, and learn from experience. In general an intelligent agent can be described as consisting of a sensing element that can receive events, a recognizer or classifier that determines which event occurred, a set of logic ranging from hard coded programs to rule based inference, and a mechanism for taking action [4].

Agent based distributed data mining

Distributed data mining (sometimes referred by the acronym DDM) considers data mining in this broader context. Distributed data mining may also be useful in environments with multiple compute nodes connected over high speed networks. Even if the data can be quickly centralized using the relatively fast network, proper balancing of computational load among a cluster of nodes may require a distributed approach. The privacy issue is playing an increasingly important role in the emerging data mining applications [3].

A huge amount of data is stored in databases. Within these databases there is the potential to discover new knowledge about the world. The distributed data mining applications can be further enhanced with agents. Agent based distributed data mining takes data mining as a basis foundation and is enhanced with agents, therefore this novel data mining technique inherits all powerful properties of agents and as a result yields desirable characteristics [7].

Data mining agents seek data and information based on the profile of the user and the instructions she/he gives. A group of flexible data mining agents can cooperate to discover knowledge from distributed sources. They are responsible for accessing data and extracting higher level useful information from the data. A data mining agent specializes in performing some activity in the domain of interest. Agents can work in parallel and share the information they have gathered so far [24].

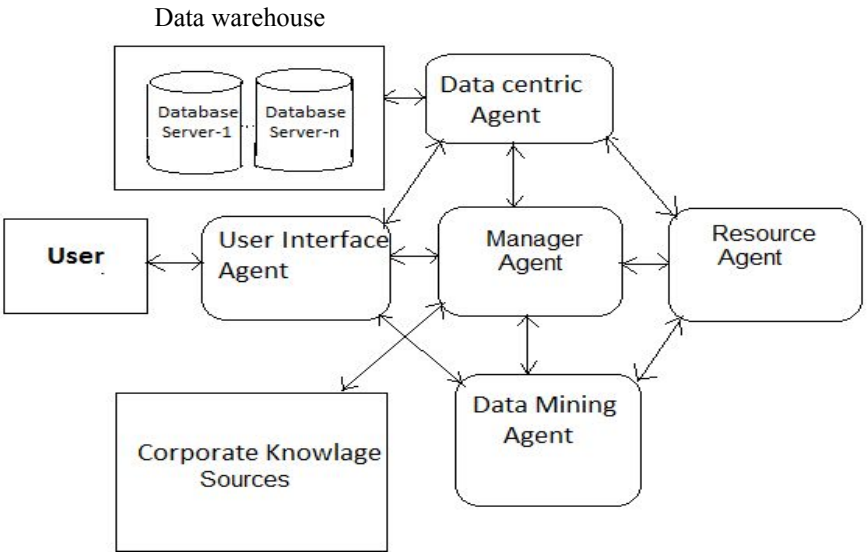

Figure 2.Block diagram of agents based distributed data mining system

ISSN 2348 – 8034
Impact Factor- 3.155

Role of agents in distributed data mining

Agents are defined as software or hardware entities that perform some set of tasks on behalf of users with some degree of autonomy. In order to work for somebody as an assistant an agent has to include a certain amount of intelligence which is the ability to choose among various courses of action, plan, communicate, adapt to changes in the environment, and learn from experience. In general, an intelligent agent can be described as consisting of a sensing element that can receive events, a recognizer or classifier that determines which event occurred a set of logic ranging from hard coded programs to rule based inference, and a mechanism for taking action. According to Wooldridge intelligent agents are defined as agents capable of flexible autonomous action to meet their design objectives. They must involve:

• **Reactivity:** to perceive and respond in a timely fashion to changes occurring in their environment in order to satisfy their design objectives. The agent's goals and/or assumptions that form the basis for a procedure that is currently executed may be affected by a changed environment and a different set of actions may have to be performed.

• **Pro-activeness:** ability to exhibit goal directed behavior by taking the initiative, responding to changes in their environment in order to satisfy their design objectives.

• **Sociability:** capability of interacting with other agents through negotiation and/or cooperation to satisfy their design objectives.

II. LITRATURE SURVEY

In **[May 1999 Ayse Sehdim]** has explained more on agents the special types of software applications, has become a very popular paradigm in computing in recent years. The author states that, the agent based studies can be implemented for clustering, classification, and summarization. Recent increase in agent-based applications is also because of the technological developments in distributed computing, robotics and the emergence of object-oriented programming paradigms. Advances in distributed computing technologies have given rise to use of agents that can model distributed problem solving [30].

In **[1999 Berry and Linoff]** has proposed automated data mining by taking a picture with an automatic camera. The quality of results is not always as good as what can be achieved by an expert however the ease of use empowers non- expert users to achieve reasonable results with minimum effort. This method is considered in this research work to automate the selection process of user required attributes and mining algorithms for better decision making process. By considering this approach, the time taken to select the attributes and mining algorithm is reduced while compared with manually process is being considered [33].

In **[March 2000 Badrul Sarwar et al.]** has described the recordation algorithm for e-commerce. The largest e-commerce sites offer the million of product for sale. Choosing among so many options is challenge for customer. recommdation system has emerge for this problem [29].

In **[June 2003 Matthias Klusch and Stefano Lodi et al**.] has proposed the increasing demand to scale up to massive data sets inherently distributed over a network with limited bandwidth and computational resources available motivated the development of distributed data mining (DDM). Distributed data mining is expected to perform partial analysis of data at individual sites and then to send the outcome as partial result to other sites where it is sometimes required to be aggregated to the global result. Quite a number of distributed data mining solutions are available using various techniques such as distributed association rules, distributed clustering, Bayesian learning, classification (regression), and compression, but only a few of them make use of intelligent agents at all. Many large companies, public institutions and non-profit organizations have resources dedicated to internal knowledge management efforts, often as a part of their business strategy, information technology, or human resource management departments. Several consulting companies provide strategy and advice regarding agent base distributed data mining to this organization [28].

In [**March 2004 Saravanan and Vivekanandan]** has proposed an automated data mining system which encompasses familiar data mining algorithms. According to author the system will automatically select the appropriate data mining technique and select the necessary field needed from the database at the appropriate time without expecting the users to specify the specific techniques and the parameters. Association and Classification rule mining is incorporated in this approach with the help of software agents. The system also has multiple association and classification techniques and selects the appropriate techniques based on user interest / data type / data size. Automatic detection of clusters and multi-dimensional visualization is not being considered by the author [26].

ISSN 2348 – 8034
Impact Factor- 3.155

In **[2005 Eleni Mangina]** has discusses real world monitoring engineering applications and specifies the role of knowledge engineers in complexity and diversity of tasks associated with specific problem domain. This paper also deals about tackling simultaneously different types of knowledge (inaccurate, incorrect or redundant) from different data sources that require being processed using different reasoning mechanisms. Within this paper, an intelligent agent-based platform is being considered for implementation, where the approach of integrating the use of two or more techniques is taken, in order to combine their different strengths and overcome each other's weaknesses and generate hybrid solutions. Different types of knowledge are discovered using different data mining techniques based on the user is considered to implement in this research [24].

In **[April 2007 Marcos Campos and Peter Stengard et al.]** proposes a new approach to the design of data mining applications platform to targeted user communities. This approach uses a data centric focus where information are stored in a location and by implementing automated methodologies to make data mining process more accessible to non experts. The automated methodologies are exposed through high-level interfaces. This frame work hides the data mining concepts away from the users, helping to bridge the conceptual gap generally associated with data mining. Automated mining algorithm used in this approach is classification and regression techniques. Clustering techniques is being used in this research work with the help of automated data mining system to create new cluster for user community [25].

In **[April 2009 Dr. Sujni Paul]** has proposed an optimized distributed association rule mining algorithms in parallel and distributed data mining. Many current data mining tasks can be accomplished successfully only in a distributed setting. The field of distributed data mining has therefore gained increasing importance in the last decade. The Apriori algorithm by Rakesh Agarwal has emerged as one of the best Association Rule mining algorithms. Ii also serves as the base algorithm for most parallel algorithms. The enormity and high dimensionality of datasets typically available as input to problem of association rule discovery, makes it an ideal problem for solving on multiple processors in parallel. The primary reasons are the memory and CPU speed limitations faced by single processors. In this paper an Optimized Distributed Association Rule mining algorithm for geographically distributed data is used in parallel and distributed environment so that it reduces communication costs [31].

In **[June 2009 Wout Dullaert and Tijs Neutens et al.]** implements an intelligent agent based communication for particular platform. Intelligent agents are used in the form of high potential output such as increase cost efficiency, better service and safety communication among the agents. They are also autonomous, communicative and intelligent. Author also proposed real-time decision is also possible with the presence of intelligent agent. Agents are used to overcome the quality, reliable service, trust concerns and confidentiality during the exchange process. Agent technology is used for automated transport process. In this research work, intelligent agent-based concept is considered for cluster formation and cluster visualizations [20].

In **[May 2010 Chayapol Moemeng and Xinhua Zhu et al**.] has describe agent based workflow has been proven its potential in overcoming issues in traditional workflow-based systems, such as decentralization, organizational issues, etc. The existing data mining tools provide workflow metaphor for data mining process visualization, audition and monitoring; these are particularly useful for distributed environments. In agent-based distributed data mining (ADDM), agents are an integral part of the system and can seamlessly incorporate with workflows. Author describes a mechanism to use workflow in descriptive and executable styles to incorporate between workflow generators and executors. This shows that agent-based workflows can improve agent based distributed data mining interoperability and flexibility and also demonstrates the concepts and implementation with a supporting the argument, a multi-agent architecture and an agent based workflow model are demonstrated [19].

In **[September 2010 Vuda Rao]** has explained communications among the agents with in multi-agent system. According to the author, multi agent system often deals with complex applications that required to solve the existing problem during data mining process in distributed system with individual and collective behaviors of the agents depends on the observed data from distributed system. Based on this concept, an integration of multi-agent system with data mining is incorporated and it also defines how multiple agents are communicated with respect to specific applications. Declaration of different agents with respect to specific task and communication behavior among agents is considered in this research work to meet the user requirements [18].

In **[Year 2011 A. Vasudeva Rao]** has described extracting or mining useful data from large amounts of data. Due to the wide availability of huge amount of useful information and knowledge, data mining can be used for market analysis, fraud detection and customer retention. Network intrusion detection system is used to detect any intruder which might have entered into the computer system. A multi agent based approach is used for network intrusion detection. The method proposed is the enhancement of the previous methods. Here more numbers of agents are used which will be continuously monitoring the data to check for any intruder which might have entered in the system. Each agent is trained accordingly so that it can check for any type of intruder entering into the system.

ISSN 2348 – 8034
Impact Factor- 3.155

Verification through many agents will ensure the safety of the data. Multi agent systems are often distributed and agents have proactive and reactive features which are very useful for Knowledge Management Systems, combining distributed data mining with multi agent for data intensive applications is appealing [4].

In **[March 2012 Trilok Pandey and Niranjan Panda et al.]** has proposed autonomous agents and multi-agents and knowledge discovery (or data mining) are two of the most active areas in information technology. Ongoing research has revealed a number of intrinsic challenges and problems facing each area, which can't be addressed solely within the confines of the respective discipline. A profound insight of bringing these two communities together has unveiled a tremendous potential for new opportunities and wider applications through the synergy of agents and data mining. With increasing interest in this synergy, agent mining is emerging as a new research field studying the interaction and integration of agents and data mining. In this paper, it gives an overall perspective of the driving forces, theoretical underpinnings, main research issues, and application domains of this field, while addressing the state-of-the-art of agent mining research and development. Our review is divided into three key research topics: agent-driven data mining, data mining-driven agents, and joint issues in the synergy of agents and data mining. This new and promising field exhibits a great potential for groundbreaking work from foundational, technological and practical perspectives [10].

In **[May 2012 Philippe Viger and Cheng Wei Wu et al.]** has proposed the mining Top-K association rules. Association rule mining consists of discovering associations between items in transactions. It is one of the most important data mining tasks. It has been integrated in many commercial data mining software and has wide applications in several domains. Mining association rules is a fundamental data mining task. This is a serious problem because in practice users have limited resources for analyzing the results and thus are often only interested in discovering a certain amount of results, and fine tuning the parameters is time-consuming. To address this problem, author propose an algorithm to mine the top-*k* association rules, where *k* is the number of association rules to be found and is set by the user. The algorithm utilizes a new approach for generating association rules named rule expansions and includes several optimizations. Experimental results show that the algorithm has excellent performance and scalability, and that it is an advantageous alternative to classical association rule mining algorithms when the user want to control the number of rules generated. This idea of mining top-k association rules presented in this paper is analogous to the idea of mining top-k itemsets and top-k sequential patterns in the field of frequent pattern mining [13].

In **[December 2012 Reecha Prajapati and Sumitra Menaria]** introduces the data mining technology normally adopts data integration method to generate data warehouse on which to gather all data into a central site, and then run an algorithm against that data to extract the useful Module Prediction and knowledge evaluation's. A single data mining technique has not been proven appropriate for every domain and data set. Data mining techniques involving in such complex environment must encounter great dynamics due to changes in the system can affect the overall performance of the system. Agent computing whose aim is to deal with complex systems has revealed opportunities to improve distributed data mining systems in a number of ways. Multi-agent systems often deal with complex applications that require distributed problem solving. The field of distributed data mining deals with these challenges in analyzing distributed data and offers many algorithmic solutions to perform different data analysis and mining operations in a fundamentally distributed manner that pays careful attention to the resource constraints. Since multi-agent systems are often distributed and agents have proactive and reactive features which are very useful for knowledge [11].

In **[2012 Thornton k. and T. Berman]** has described that how a novel application of data mining techniques can be used to provide the engine for a tool which can be used to identify email correspondence which may be an early indication of virtual bullying or harassment. The approach that makes use of linear discriminated approaches to classify normal and non-normal style of email correspondence for each sender has been taken. The change in email style could be used to provide an early indication of virtual harassment/bullying. This approach has great potential for use in large organization where it often appears to be hard to identify unacceptable information transmission between two colleagues. By identifying indicative behavior it has been made possible to start company anti bullying processes in a timelier manner [7].

In **[January 2012 R. Jayabrabu and Dr. V. Saravanan et al.]** has describe data mining techniques plays a vital role like extraction of required knowledge, finding unsuspected information to make strategic decision in a novel way which in term understandable by domain experts. To develop an approach which performs the process autonomously that needs more user interaction by way of selecting the appropriate user specifies objectives are the primary goal of this research. The developed system also chooses the best data mining technique for knowledge extraction and also for better understanding. Thus, level of automation incorporated in this system is an important issue. The automated system performs the process based on user interface agent, data mining agent and

ISSN 2348 – 8034
Impact Factor- 3.155

visualization agent. User interface agent is used to navigate the history of the user profile among the frequent user to mine the related data. Data mining agent is used in this system to perform different types of data analysis by selecting [26].

In **[June 2013 Harshit Srivastava and Virendra Kumar et al.]** has Data mining includes the use of data analysis tools to find out hidden data, unknown, valid patterns and relationships in large databases. These tools contain mathematical algorithms, statistical models, and machine learning methods like decision trees or neural networks. Resultantly, data mining comprise of more than managing and collecting data, it also consists of analysis and prediction. Data mining can be used on data represented in textual, multimedia or quantitative forms. Data mining applications consist of sequence or path analysis, association, classification, clustering. Association rule learning in data mining is a good method for finding interesting relations between huge data sets. It is used to recognize strong rules founded in databases using distinct measures of interestingness [7].

In **[May 2013 Fan Min and William Zhu]** has proposed to mine the top-k granular association rules for each user. Recommender systems are important for e-commerce companies as well as researchers. Recently, granular association rules have been proposed for cold-start recommendation. However, existing approaches reserve only globally strong rules; therefore some users may receive no recommendation at all. In this paper author propose to mine the top-k granular association rules for each user. Author defines three measures of granular association rules. These are the source coverage which measures the user granule size, the target coverage which measures the item granule size, and the confidence which measures the strength of the association. With the confidence measure, rules can be ranked according to their strength. Then we propose algorithms for training the recommender and suggesting items to each user. An example of such a rule might be \young women rate adventure movies released in 1990s with a probability of 35%; 21% users are young women and 15% movies are adventure ones released in 1990s." Here 35% is the confidence of the rule. With the confidence measure, the strength of any two rules can be compared. , we propose an algorithm for training the recommender. This is done through building connections between source and target granules that satisfy the coverage thresholds. Author proposes an algorithm to obtain top-k rules which in turn suggest k types of items to each user [8].

In **[2013 María del Pilar Angeles and Jonathan Córdoba Luna]** has described a multi agent distributed data mining framework as an approach to performance and data security issues. It has been implemented by ontology's in order to incorporate semantic content to improve the intelligence and efficiency of data mining agents. Each agent is only responsible for specific duties. Agents communicate and coordinate with each other to enhance data mining and keep privacy and confidentiality of data. The developed prototype shows a parallel, distributed data mining process, and a real-world use case, which integrates birth rate data registered during 2011-2012 in México by the official censuses [9].

In [**September 2013 Vinaya Sawant and Dr.Ketan Shaha]** has describe the distributed data mining using agent, Data mining had played an important role in analyzing large set of data and understanding the complex systems in almost all areas. The data mining process becomes more challenging in a heterogeneous distributed environment. The various algorithms in distributed data mining are proposed that aims at integrating the knowledge from the data that are geographically distributed. Mobile agents in distributed data mining prove to be one of the best and robust methods to handle distributed data in a faster way. Advances in computing and communication over wired and wireless networks have resulted in many pervasive distributed computing environments. The internet, intranets, LANs, and wireless networks are some examples. Many of these environments have different distributed sources of voluminous data and multiple computer nodes. Analyzing and monitoring these distributed data sources require data mining technology designed for distributed applications. Author needs data mining architectures that pay careful attention to the distributed resources of data, computing and communication in order to consume them in a near optimal fashion [22].

In **[March 2014 Monika Khatri and S.Dhande]** has described history and current and future trends of data mining techniques. From human life birth people have been seeking patterns in data. In all sectors such as government agencies, scientific institutions, and business have all dedicated enormous resources to collecting and storing data. From this large data only a small amount of these data will ever be used while others will be not useful for the task. There is need to understand these large data. The ability to extract useful knowledge hidden in these data is called as data mining and to act on that knowledge is becoming increasingly important in today's competitive world. There are several data mining techniques that have been developed such as association, classification, clustering, prediction and sequential patterns, etc., are used for knowledge discovery from databases. Association is a technique which is used to find a pattern that is based on a relationship of a particular item on other items in the same operation [5].

ISSN 2348 – 8034
Impact Factor- 3.155

In **[June 2014 R. Hemamalini** and **Dr. L. Josphin Mary]** has describe the distributed data mining (DDM) is a branch of the field of data mining that offers a framework to mine distributed data paying careful attention to the distributed data and computing resources. Usually, data-mining systems are designed to work on a single dataset. On the other hand with the growth of networks, data is increasingly dispersed over many machines in many different geographical locations. Also, even as most practical data-mining algorithms operate over propositional representations are known as first order learning. In existing system, the concept of knowledge is very important in data mining. In order to get the correct knowledge from the data mining system, the user must define the objective and specify the algorithms and its parameters exactly with minimum effort. If the data mining system produces large number of meaningful information by using a specialized data mining algorithm like association, clustering, decision trees etc. It will take more time for the end users to choose the appropriate knowledge for the problem discussed. Developing a data mining system that uses specialized agents with the ability to communicate with multiple information sources, as well as with other agents requires a great deal of flexibility [3].

In **[2014 Drasti Patel and Reema Patel]** has present the t techniques for mining top-k ranked association rules from datasets. Association rule mining greatly helps to identify trends and patterns from huge datasets. Author describes the different technique to mine top-k ranked association rules and closed frequent item sets. In this we study the mining of top-K frequent closed item sets, where the support threshold is chosen as the maximum value sufficient to guarantee that the item sets returned in output is at least K. Another algorithm is to mine the top-k association rules, where k is the number of association rules to be found and is set by the user. Rule expansion method is used for mining top-k association rules [6].

In **[year 2014 Bhimalendu Pathak and Madhavi Sinha**] has described agent based distributed data mining and its ontology. This proposed the agent base distributed data mining. With the increased complexity in number of applications and due to large volume of availability of data from heterogeneous sources, there is a need for the development of suitable ontology, which can handle the large data set and present the mined outcomes for evaluation intelligently. In the era of intensive data driven applications distributed data mining can meet the challenges with the support of agents. This paper discusses the underlying principles for effectiveness of modern agent-based systems for distributed data mining. Relational database systems lead the way of efficient storage and retrieval of data from the organized set of information long back. Knowledge discovery from the databases presents another approach where pattern from the large repository is extracted and analyzed. Data is almost growing in exponential terms and most of the relevant applications are driven by it. One simple data mining technique cannot provide the required solution for it. Main challenges of data mining includes, existence of huge data set located at different sites, intensive computation process with large set and rapid change in input data. The distributed data mining (DDM) mines data sources irrespective of their physical locations. An intelligent agent is an autonomous entity or software program that automatically performs tasks on behalf of the user. An intelligent agent or group of such agents act on data from heterogeneous sources and distributed in proactive manner [1].

In **[January 2015 G. S. Bhamra and A. K. Verma et al.]** describe distributed association rule mining (DARM) is the task for generating the globally strong association rules from the global frequent itemsets in a distributed environment. The intelligent agent based model, to address scalable mining over large scale distributed data, is a popular approach to constructing distributed data mining (DDM) systems and is characterized by a variety of agents coordinating and communicating with each other to perform the various tasks of the data mining process [2].

Summary

The above paper's aims to survey the role of agent based distributed data mining in large organization by conducting literature review and classification of articles from 1999 to 2015 in order to explore how agent based distributed data mining and applications have been developed in this period. Agent based distributed data mining plays the important role in e-commerce, banks etc. This paper provide the different algorithms which is used for the distributed data mining and shows the different technique of data mining and types of agents which helps the user to extract the data from huge amount of database which located into the different sites.

From this review of literature idea of "An agent base distributed data mining and its explicit specification of conceptualization" arose. Likewise this implement and explore the agent base data mining which used to extract the knowledge from huge amount of database located onto the different sites.

ISSN 2348 – 8034
Impact Factor- 3.155

III. SYSTEM ANALYSIS

Analysis has been done to identify the problem area in distributed data mining. Many large organizations like e-commerce, bank, need to a distributed data mining.

Analysis

Traditional agent technology has been challenged in many aspects such as developing organizational and social intelligence. In this system analysis, concern ourselves with the challenges that may benefit from the involvement of data mining. Following aspects agent awareness, agent learning, agent action ability, and agent distributed processing, agent in-depth services, and agent constraint processing.

One or more agents per network node are responsible for examining and analyzing a local data source. In addition, an agent may query a knowledge source for existing knowledge. The agents communicate with each other during the discovery process .This allows the agents to integrate the new knowledge they produce into a globally coherent theory. In addition, a supervisory agent, responsible for coordinating the discovery agents may exist. A graphical interface allows the user to assign agents to data sources, and to allocate high level discovery goals. It allows the user to critique new knowledge discovered by the agents, and to direct the agents to new discovery goals, including ones that might make use of the newly discovered knowledge.

An agent based distributed data mining (ADDM) system can be generalized into a set of components and viewed as depicted in figure-1 below. Author may generalize activities of the system into request and response, each of which involves a different set of components [1].

Component Diagram: It shows the components of distributed data mining.

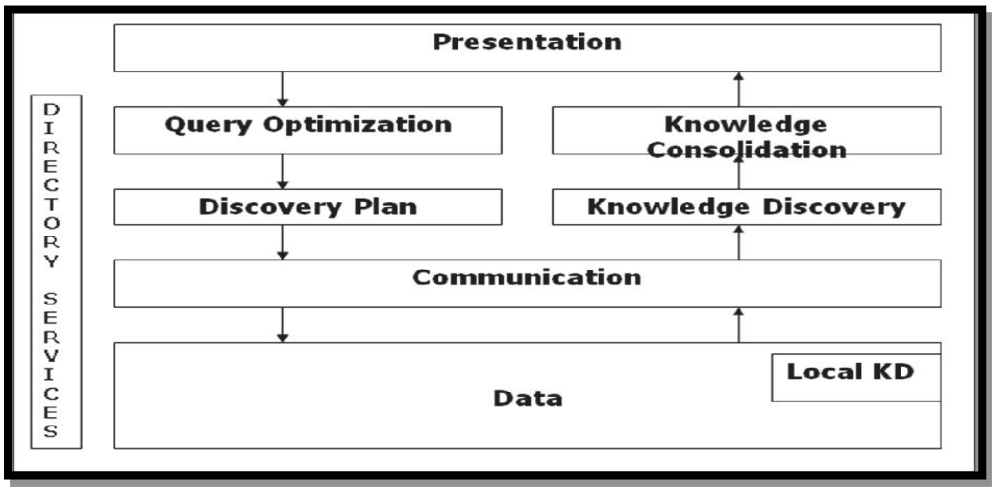

Fig. 3. Basic components of agent based distributed data mining system

Basic components of an agent based distributed data mining system:

- Data: Data is the foundation layer of our interest. In distributed environment, data can be hosted in various forms, such as online relational databases, data stream, web pages, etc., in which purpose of the data may be varied.
- Communication: The system chooses the related resources from the directory service, which maintains a list of data sources, mining algorithms, data schemas, data types, etc. The communication protocols may vary depending on implementation of the system, such as client-server, peer-to peer,etc.
- Presentation: The User Interface (UI) interacts with the user as to receive and respond to the user. The interface simplifies complex distributed systems into user-friendly message such as network diagrams, visual reporting tools, etc. On the other hand, when a user requests for data mining through the UI, the following components are involved. A query optimizer analyses the request as to determine type of mining tasks and chooses proper resources for the

ISSN 2348 – 8034
Impact Factor- 3.155

request. It also determines whether it is possible to parallelize the tasks, since the data is distributed and can be mined in parallel.

- Query optimization: A query optimizer analyses the request as to determine type of mining tasks and chooses proper resources for the request. It also determines whether it is possible to parallelize the tasks, since the data is distributed and can be mined in parallel.
- Discovery Plan: A planner allocates sub-tasks with related resources. At this stage, mediating agents play important roles as to coordinate multiple computing units since mining sub-tasks performed asynchronously as well as results from those tasks. On the other hand, when a mining task is done, the following components are taken place:
- Local Knowledge Discovery: In order to transform data into patterns which adequately represent the data and reasonable to be transferred over the network, at each data site, mining process may take place locally depending on the individual implementation.
- Knowledge Discovery: It may also be considered as mining. It executes the algorithm as required by the task to obtain knowledge from the specified data source.
- Knowledge Consolidation: In order to present the user with a compact and meaningful mining result, it is necessary to normalize the knowledge obtained from various sources. The component involves a complex methodology to combine patterns from distributed sites [1].

IV. SYSTEM ARCHITECTURE

This distributed data mining architecture is an agent based system developed for performing knowledge discovery from large distributed sources of data. Due to the diversity of mining algorithms and the diversity of data sources, it is difficult to generate a mining model by combining mining rules on different sites. In this system works independently to combine result from different sites.

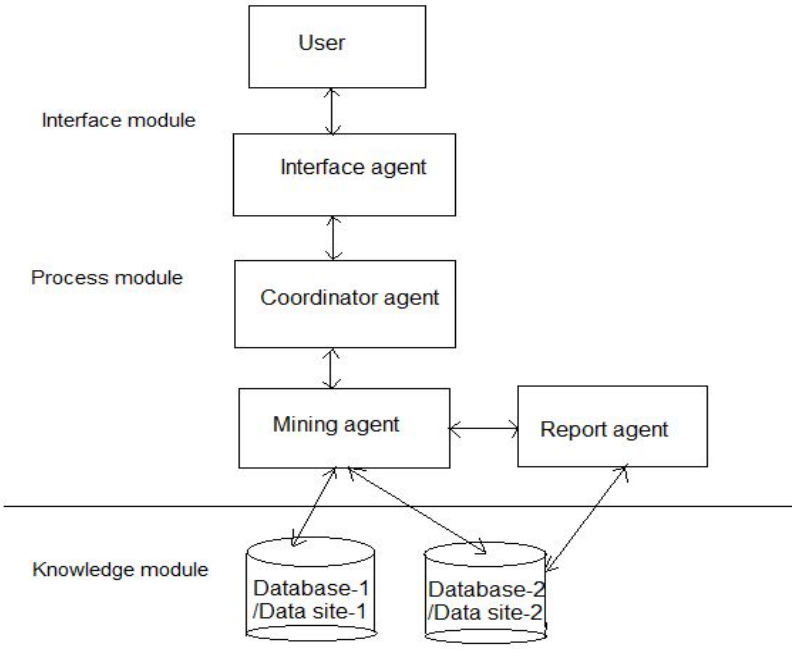

Fig. 4. Architecture of agent based distributed data mining

The following is a definition for the most common agents that are used in such systems. The names might be different but they share the same functionalities in most cases.

Interface Agent (or User Agent): It interacts with the user (or user agent). It asks the user to provide his requirements, and provides the user with mined results (may be visualized). Its interface module contains methods for inter agent communication and getting input from the user. The process module contains methods for capturing the user input and communicating it to the facilitator agent. In the knowledge module, the agent stores the history of user interaction, and user profiles with their specific preferences.

ISSN 2348 – 8034
Impact Factor- 3.155

Facilitator Agent (or Manager Agent): The facilitator agent is responsible of the activation and synchronization of different agents. It elaborates a work plan and is in charge of ensuring that such a work plan is fulfilled. It receives the assignments from the interface agent and may seek the services of a group of agents and synthesize the final result and present it to the interface agent. The interface module is responsible for inter-agent communication.

Mining Agent: The data mining agent implements specific data mining techniques and algorithms. The interface module supports inter-agent communication. The process module contains methods for initiating and carrying out the data mining activity, capturing the results of data mining, and communicating it to result agent or the facilitator agent. The knowledge module contains meta knowledge about data mining methods, i.e. what method is suitable for what type of problem, input requirements for each of the mining methods, format of input data, etc. This knowledge is used by the process module in initiating and executing a particular data mining algorithm for the problem at hand.

Result Agent: Result agent observes a movement of mining agents, and obtains result from mining agents. When result agent obtains all results, it arrangement/integrates with the facilitator agent to show the result to the user. The interface module may provide access to other visualization software that may be available within the organization.

Mobile Agent: A mobile agent travels around the network. On each site, it processes the data and sends the results back to the main host, instead of expensive transferring large amount of data across the network. This has the advantage of low network traffic because the agents do data processing locally.

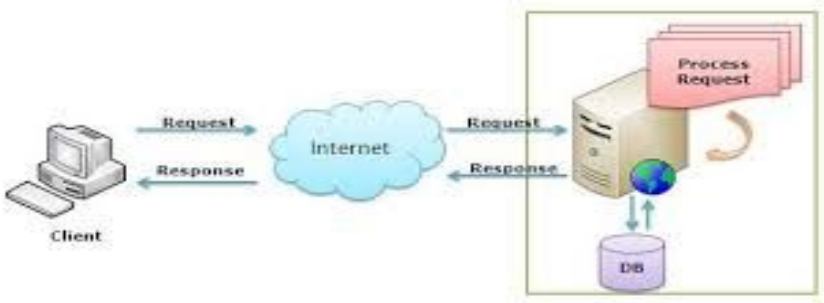

Fig.5. Three tier architecture of distributed data mining

It is a three-tier distributed data mining architecture wherein the three tiers are referred to as:

- Client
- Middle tier
- Distributed database

Client

It has a user interface, which allows the user to select and submit the datasets on which data mining needs to be performed. The pre and post-processing of data is taken care of by two modules namely, association rule module and rule induction module. Mining for association rules basically involves, finding item sets that occur with high frequency and then generating rules based on these results.

Middle tier

The frequent item sets and the contingency tables required for the processes described above are generated in this tier. Thus the pre and post processing of data involves both the client tier and the middle tier.

General services of middle tier which can be classified as:

- Connection and access services
- Remote administration services
- Work management

The connection and access services allow the client to connect to the middle tier. Remote administration allows the middle tier software to be configured from a computer that is physically separate from the middle tier. The work manager is responsible for the execution of data mining tasks.

ISSN 2348 – 8034
Impact Factor- 3.155

Distributed database

The distributed database tier consists of a File system and a Database Management System.

Advantages of distributed data mining

- Reflects organizational structure
  Many organizations are naturally distributed over several locations. For example, a bank has many offices in different cities. It is natural for databases used in such an application to be distributed over these locations. A bank may keep a database at each branch office containing details such things as the staff that work at that location, the account information of customers etc. The staff at a branch office will make local inquiries of the database. The company headquarters may wish to make global inquiries involving the access of data at all or a number of branches.
- Improved share ability and local autonomy
  The geographical distribution of an organization can be reflected in the distribution of the data; users at one site can access data stored at other sites. Data can be placed at the site close to the users who normally use that data. In this way, users have local control of the data, and they can consequently establish and enforce local policies regarding the use of this data. A global database administrator (DBA) is responsible for the entire system. Generally, part of this responsibility is assigned the local level, so that the local DBA can manage the local DBMS.
- Improved availability
  In a centralized DBMS, a computer failure terminates the applications of the DBMS. However, a failure at one site of a DDBMS, or a failure of a communication link making\ some sites inaccessible, does not make the entire system in opera bite. Distributed DBMSs are designed to continue to function despite such failures. If a single node fails, the system may be able to reroute the failed node's requests to another site.
- Improved reliability
  As data may be replicated so that it exists at more than one site, the failure of a node or a communication link does not necessarily make the data inaccessible.
- Improved Performance
  As the data is located near the site of 'greatest demand', and given the inherent parallelism of distributed DBMSs, speed of database access may be better than that achievable from a remote centralized database.
- Economics
  It is now generally accepted that it costs much less to create a system of smaller computers with the equivalent power of a single large computer. This makes it more cost effective for corporate divisions and departments to obtain separate computers. It is also much more cost-effective to add workstations· to a network than to update a mainframe system.
- Modular growth
  In a distributed environment, it is much easier to handle expansion. New sites can be added to the network without affecting the operations of other sites. This flexibility allows an organization to expand relatively easily. Adding processing and storage power to the network can usually handle the increase in database size. In a centralized DBMS, growth may entail changes to both hardware (the procurement of a more powerful system) and software (the procurement of a more powerful or more configurable DBMS).

V. SYSTEM DESIGN

The results and discussion may be combined into a common section or obtainable separately. They may also be broken into subsets with short, revealing captions.

Algorithms Used

In this dissertation there is emerging TopKRules algorithm which provides the rule based mining for extracting the knowledge. Algorithm is a set of rules that precisely defines a sequence of operations an algorithm is an effective method that can be expressed within a finite amount of space and time and in a well-defined formal language for calculating a function.

The step by step implementation of TopKRules algorithm as follows.

The TopKRules algorithm

TOPKRULES(T, k, minconf) R := Ø. L := Ø. minsup := 0.

1. Scan the database T once to record the tidset of each item.
2. FOR each pairs of items i, j such that |tids(i)| ×|T| ≥ minsup and |tids(j)| ×|T| ≥ minsup:

ISSN 2348 – 8034
Impact Factor- 3.155

3. sup({i}→{j}) := |tids(i) ∩ tids(j)| / |T |.
4. sup({j}→{i}) := |tids(i) ∩ tids(j)| / |T|.
5. conf({i}→{j}) := |tids(i) ∩ tids(j)| / |tids(i)|.
6. conf({j}→{i}) := |tids(i) ∩ tids(j)| / |tids(j)|.
7. IF sup({i}→{j}) ≥ minsup THEN
8. IF conf({i}→{j}) ≥ minconf THEN SAVE({i}→{j}, L, k, minsup).
9. IF conf({j}→{i}) ≥ minconf THEN SAVE({j}→{i}, L, k, minsup).
10. Set flag expandLR of {i}→{j}to true.
11. Set flag expandLR of {j}→{i}to true.

12. R := R∪{{i}→{j}, {j}→{i}}.

13. END IF
14. END FOR

15. WHILE ∃r ∈ R AND sup(r) ≥ minsup DO

16. Select the rule rule having the highest support in R
17. IF rule.expandLR = true THEN
18. EXPAND-L(rule, L, R, k, minsup, minconf).
19. EXPAND-R(rule, L, R, k, minsup, minconf).
20. ELSE EXPAND-R(rule, L, R, k, minsup, minconf).
21. REMOVE rule from R.

22. REMOVE from R all rules r ∈ R | sup(r) <minsup.

23. END WHILE

Algorithm description

First scans the database once to calculate tids({c}) for each single item c in the database (line 1). Then, the algorithm generates all valid rules of size 1*1 by considering each pair of items i, j, where i and j each have at least minsup×|T| tids (if this condition is not met, clearly, no rule having at least the minimum support can be created with i, j) (line 2). The supports of the rules {i}→{j} and {j}→{i} are simply obtained by dividing |tids(i→ j)| by |T| and |tids(j→ i)| by |T| (line 3 and 4). The confidence of the rules {i}→{j} and {j}→{i} is obtained by dividing |tids(i→ j)| by |tids(i)| and |tids(j→ i)| by | tids(j)| (line 5 and 6). Then, for each rule {i}→{j} or {j}→{i} that is valid, the procedure SAVE is called with the rule and L as parameters so that the rule is recorded in the set L of the current top-k rules found (line 7 to 9). Also, each rule {i}→{j} or {j}→{i} that is frequent is added to the set R, to be later considered for expansion and a special flag named expandLR is set to true for each such rule (line 10 to 12).

After that, a loop is performed to recursively select the rule r with the highest support in R such that sup(r) ≥ minsup and expand it (line 15 to 23). The idea is to always expand the rule having the highest support because it is more likely to generate rules having a high support and thus to allow to raise minsup more quickly for pruning the search space. The loop terminates when there is no more rule in R with a support higher than minsup. For each rule, a flag expandLR indicates if the rule should be left and right expanded by calling the procedure EXPAND-L and EXPAND-R or just left expanded by calling EXPAND-L. For all rules of size 1*1, this flag is set to true.

Advantage of TopKRules algorithm

1) Redundancy: This new approach reduces the redundant rules so it will find accurate result.
2) Time Efficient: An approach to add new mine rule to LIST give more control on program. It can be suitable for user who wants to satisfy on first result.

ISSN 2348 – 8034
Impact Factor- 3.155

3) Reduce Memory Usage: As with new approach with reduce the redundant data and mine only give rule less memory will capture by approach as previous methods can display output when all rules will mine.
4) Sorting of Data: When output store in LIST it will be arrange in descending order by SUP value of new mine rule thus not required user to sort data again. 5) Reduce computation: It also save computation time because computation is needed to be done for only top rules.

Feasibility study

Economical Feasibility:

A research paper depicts some of the module and their applications are developed in ASP.NET. This systems front end is PHP and back end processing in Java. PHP code can be simply mixed with HTML code, or it can be used in combination with various tinplating engines and web frameworks.

Java is open source language. Java is a general-purpose computer programming language that is concurrent, class-based, object-oriented, and specifically designed to have as few implementation dependencies as possible This means there are certain costs associated with Java development.

Technical Feasibility:

Java is used in this system because of its feature. Java is object oriented. Java can run on many different operating systems. This makes Java platform independent. Java does this by making the Java compiler turn code into Java bytecode instead of machine code. This means that when the program is executed, the Java Virtual Machine interprets the bytecode and translates it into machine code. This provides a lot of power and flexibility to the web pages.

Operational Feasibility:

With Java this is attempt to provide an environment for extract the data /knowledge from huge amount of data sites. In this system agents are developed in java by using the agent, it extracts the data/knowledge from different sites. In this environment there are different users which are going to access our agent based distributed mining system. There are different stages of the execution. The starting point is that the user has to register himself or herself into to agent based distributed data mining system, by giving the basic information user id, password and other information like email id, date of birth, mobile no, address etc. when the registration form is completely filled up that means user is a valid member of agent based distributed data mining system.
Inputs and output specification of the system is, user extracts the knowledge/data from different sites and maintain the database. In user detail the login form plays as an input specification and the output specification is account created and the data/knowledge is extract from different sites, extracted knowledge visible to the user as well as to the administrator of ecommerce organization like Amazon, flip cart. Once the user is the part of the organization user will be the member of the agent based distributed data mining system. In this system agent are works between user and data sites and it helps to extract the data from different sites.

Administrator adds other product in the database and maintains the user detail. The output specification of this input will be account creation and the extraction of data/knowledge by using the different agents.

VI. IMPLIMENTATION DETAIL

Agent based distributed data mining system implements different modules. It is design for the e-commerce like flip cart, amazon etc. In this type of organizations, owner wants to see the witch product is mostly seen by customer and which product has a good support of customer, this type of knowledge extraction is perform by using the top k rule algorithm. System modules are as follows:

1. **User login and signup module**
2. **Add entry and performs ADDM (Agent based Distributed Data Mining) module**
3. **Admin module**

By using these modules the execution of the project is completed.

User login and signup module:

User login and sign is module in which the user is able to login the main application. User has to fill the user-id and password and sign up by using the other information. After registration the user will login to application with their ID and password. In this module, user enters his/her other information like Name, Mobile no, mail ID etc. All this facility is provided for user.

ISSN 2348 – 8034
Impact Factor- 3.155

Add entry and performs ADDM module:
After user signup user login onto this system and extract the useful knowledge/data which is located in different sites. If user wants to see the number of items as per there quantity then user submit the add entry or user wants to see which product has good support and their confidence factor then user submit the ADDM (Agent based Distributed Data Mining).

Admin module:
In this module the administrator administers user's web usage record. It maintains product record, add product in record, change some product of the record and maintain the databases.

VII. CONCLUSION

Thus the dissertation provides an environment where owner of any large organization like e-commerce extract the knowledge of the product which is located into the different sites and provide the environment where knowledge/data extracted by using the agents. This system extracts the data by using the top k rule algorithm. System is design to increase the efficiency and productivity of the e-commerce organization.

The system provides the concept of distributed data mining and describes the various agents and their role. Agent systems are fundamentally designed for collaborative problem solving in distributed environments. Many of these application environments deal with empirical analysis and mining of data. This Agent based distributed data mining system in order to improve data mining performance and data security considering negotiation and a metadata for further information and better decision regarding how many and agents and where they are required. It focuses on the existing approach of agent data mining system in distributed environment. The refinement in existing approach observed with the development of suitable ontology, communication interface, and deployment of agent systems on the right track.

Distributed data mining enables learning over huge volumes of data that are situated at different geographical locations. It supports several interesting applications ranging from fraud and intrusion detection, e-commerce, to knowledge discovery from remote sensing data around the globe.

VIII. ACKNOWLEDGEMENTS

I wish to acknowledge with deep gratitude the valuable guidance received from my guide respected Prof. S.W.Ahmad , who has not only encouraged us through this venture but also took great pain in letting us understand the concepts.

REFERENCES

1. *Bimalendu Pathak and Dr. Madhavi Sinha.(BIT'S): "An agent base distributed data mining and its ontology ", International Conference on Computing for Sustainable Global Development (INDIACom). IEEE 2014 International Conference on , vol., no vol., no.pp.400,404, 5-7 March 2014 .*
2. *G. S. Bhamra, A. K. Verma and R. B. Patel, "Agent Based Frameworks for distributed association Rule mining", International journal in foundations of computer science & technology, pp 85-96, 2015.*
3. *R.Hemamalini, Dr.L.Josephini Mary,"An Analysis on Multi-Agent Based Distributed Data Mining System", International Journal of Scientific and Research Publications, Volume 4, Issue 6,pp 1-6, June 2014.*
4. *Vasudeva Rao , "Agent Based Approach to Knowledge Discovery in Data mining", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 6, pp 180-185,June 2014.*
5. *Monika khatri and S.Dhande "History and Current and Future trends of Data mining Techniques", International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 3, pp. 311-315 , March 2014.*
6. *Drashti Patel, Reema Patel, "Mining top-k association rules: A Survey", IJIRT, Volume 1 Issue 6, ISSN: 2349-6002, 2014*
7. *Fan Min and William Zhu, "Mining top-k granular association rules for recommendation", cs-IR, Vol.1, pp-1-5, May 2013.*
8. *María Del Pilar Angeles and Jonathan Córdoba-Luna (UNAM Mexico): "Multi-Agent Distributed Data Mining by Ontologies", IEEE International Journal on Advances in Software, vol 6 no 3 & 4, year 2013.*
9. *Trilok Nath Pandey, Niranjan Panda and Pravat Kumar Sahu, "Improving performance of distributed data mining (DDM) with multi-agent system", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, pp 73-82, March 2012.*

10. *Reecha B. Prajapati, Sumitra Menaria, "Multi Agent Based Distributed Data Mining", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 10, pp 76-80, December 2012.*
11. *Liu, Guoxiang, and Fengxia Yang. "The application of data mining in the classification of spam messages", In Computer Science and Information Processing (CSIP), 2012 International Conference on, pp. 1315-1317 IEEE, 2012.*
12. *Philippe Fournier-Viger, Cheng-Wei Wu and Vincent S. Tseng " Mining Top-K Association Rules",25th Canadian Conference on Artificial Intelligence, Canadian AI 2012, Canada, Series ISSN 0302-9743,pp 61-73, May 2012,*
13. *Burn-Thornton, K., and T. Burman. "The Use of Data Mining to Indicate Virtual (Email) Bullying", In Intelligent Systems (GCIS), Third Global Congress on, pp. 253-256. IEEE, 2012*
14. *R. Jayabrabu, Dr. V. Saravanan, Prof. K. Vivekanandan, "Cluster detection and multidimensional visualization of automated data mining using intelligent agent", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.1, pp 125-138,January 2012 .*
15. *N.P. Trilok, P. Niranjan, and K.S.Pravat, "Improving performance of distributed data mining (DDM) with multiagent system", International Journal of Computer Science, vol. 9, no. 2& 3, ISSN: 1694-0814, pp. 74-82, 2011.*
16. *J. Han, M. Kamber, and J. Pei, "Data mining: concepts and Techniques", 3rd ed., Elsevier, pp.744, 2011*
17. *Vuda Sreenivasa Rao, "Multi Agent-Based Distributed Data Mining:An Overvie", International Journal of Reviews in Computing, Vol3/11Vol3, pp 83-92, 2010.*
18. *Chayapol Moemeng, Xinhua Zhu, and Longbing Cao, "Integrating Workflow into Agent-Based Distributed Data Mining Systems", ADMI 2010, LNCS 5980, pp. 4–15, 2010.*
19. *Wout Dullaert , Tijs Neutens, Greet Vanden Berghe ,Tijs Vermeulen , Bert Vernimmen , Frank Witlox, MamMoeT, "An intelligent agent based communication support platform for multimodal transport", Expert Systems with Applications 36 ,Pp10280–10287, june 2009.*
20. *S. Sumathi and S.N. Sivavavdam, "Introduction to data mining and its applications", Studies in Computational Intelligence, Springer Verlag, 2006, p. 828.*
21. *Datta, S, Bhaduri, K., Giannella, C., Wolff, R. & Kargupta, H "Distributed Data Mining in Peer-to-Peer Networks", IEEE Internet Computing 10(4), 18-26, 2006.*
22. *L. Panait and S. Luke. "Cooperative Multi- Agent Learning", The State of the Art. Autonomous Agents and Multi-Agent Systems, 11(3):387–434, 2005.*
23. *Eleni Mangina, "Intelligent Agent-Based Monitoring Platform for Applications in Engineering", International Journal of Computer Science & applications Vol.2, No.1, pp. 38-48, 2005.*
24. *Marcos M. Camos, Peter J.Stengard, Boriana L Milenova, "Data- Centric Automated Data Mining", Proceedings of the Fourth International Conference on Machine Learning and Applications (ICMLA'05), Pages 97-104, 2005*
25. *V. Saravanan,K. Vivekanandan: "Design and Implementation of Automated Data Mining Using Intelligent Agents in Object Oriented Databases0", Intelligent Information Processing, pages 221- 226, 2004.*
26. *Mafruz Zaman Ashrafi "ODAM: An Optimized Distributed Association Rule Mining Algorithm", IEEE DISTRIBUTED SYSTEMS ONLINE 1541-4922 © 2004 Published by the IEEE Computer Society Vol. 5, No. 3; March 2004.*
27. *Matthias Klusch, Stefano Lodi, Gianluca Moro, "Issues of Agent-Based Distributed Data Mining", AAMAS'03, pp 1034-1035, July 2003.*
28. *Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl "Analysis of Recommendation Algorithms for E-Commerce", 2nd ACM conference of electronic commerce, USA, pp 158-167, 2000.*
29. *Ayse Yasemin Seydim, "Intelligent Agent: Data mining perspective", Diparment of Computer Science and Engineering,Southan Mathodis Univercity,Dillas, pp 20-28,May 1999.*
30. *R. Bose and V. Sugumaran. IDM: "An intelligent software agent based data mining environment", IEEE International Conference on Systems, Man, and Cybernetics, 3, 1999*
31. *S. R. Vuda, "Multi agent-based distributed data mining, an overview," International Journal of Reviews in Computing, pp. 83-92, ISSN: 2076-3328, E-ISSN: 2076-33.*
32. *Berry, M. and Linoff, G" Mastering Data Mining: The Art and Science of customer Relationship Management", Wiley, 1999.*