ISSN:2321-1156
International Journal of Innovative Research in Technology & Science(IJIRTS)

NEW TECHNIQUE OF DATA MINING USING ANT COLONY OPTIMIZATION

Sonika Rana
Department of Computer Science and Engineering
Sachdeva Engineering College for girls
Gharuan (Mohali) PB
ranasonika862@gmail.com

Rakesh Sharma
Department of Computer Science and Engineering
Sachdeva Engineering College for girls
Gharuan (Mohali) PB
rakesh77kumar@yahoo.om

Nidhi Bhatla
Department of Computer Science and Engineering
Sachdeva Engineering College for girls
Gharuan (Mohali) PB
engineernidhi@yahoo.com

Abstract— Companies store the data in a very large amount so that they can use data whenever they need. The historical data is used for the purpose of analysis and surveys. And stored data is also used in future for the purpose of decision making. In order to extract the relevant information from large amount of data the process of data mining is used. Data mining is used to search the relationships and global patterns which are hidden inside the large amount of data. The techniques of data mining are used with the neural networks. Ant Colony Optimization (ACO) is a technique which is used for data mining [5]. ACO is used to eliminate the rule based classifiers from data. Neural networks are widely used with data mining another technique of data mining is decision and regression tree. Many algorithms are used for data mining one of them is Ant miner algorithm or technique.

Keywords— Data storage; ANN; Data integration; Preprocessed Data; Clustering;

I. INTRODUCTION

Data mining is a technique to finding out the relevant information from a large stored data. Because large amount of data is store for the purpose of decision making and history review. But not all the stored data is helpful or useful in decision making and history review. Because stored data is a kind of raw data. To convert it into in meaningful data or information data mining technique is used. Data mining technique includes many steps to find out the relevant information from raw data. Data mining use many algorithms for implementation. Algorithm like ACO (Ant Colony Optimization) is widely used for the purpose of extracting the meaningful data from data stores or databases. The process of KDD is as follows:[11]

- Data Model
- Target Data
- Preprocessed Data
- Pattern/Model
- Knowledge

Data Cleaning
It is a first step in KDD while extracting the relevant information from large database. In this noise and irrelevant dat from the database is removed. In this phase the data cleaning is done

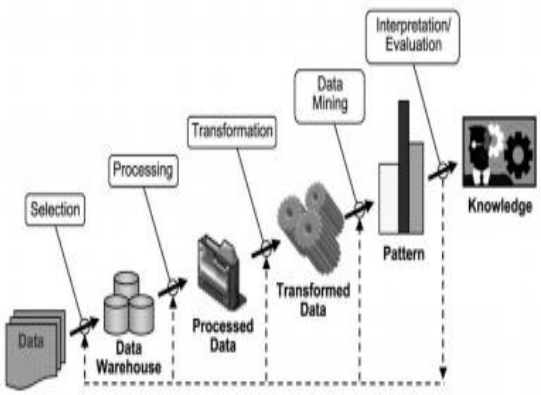

Figure 1 Steps included in the process of Data Mining

Data Integration
It is a phase or process in which the data is integrated. Data may be collected from more than one source. For example in case of Distributed Database, the data is coming from many databases located at various locations geographically. Therefore to use this scattered data for KDD it must be integrated firstly.

Data Selection
After gathering the data from heterogeneous sources, Data selection process is performed. In this step the dat which is important for decision making process is selected from integrated data.

Data Transformation
In this phase the selected data is transferred or converted into a proper meaning format. This step is must for implementing the data mining.

Data Mining
It is a process to extract the important or meaningful data from the large database. In this phase various techniques or algorithms are used for the purpose of extracting the data. Data mining is an important process to perform before the process of decision making on the basis of historical data.

Pattern Evaluation
Pattern evaluation is a step in which the extracted patterns are matched with already existing patterns. This is done for the purpose of decision making. Only relevant or necessary patterns are evaluated in this step [11].

Knowledge Representation

It is the last step in which the extracted data or patterns are visually represented to the end user. Various visual effects or techniques are used in this step. Visualization of data makes the data to be easily understood by the end users [11].

In data mining two learning approaches are used. These approaches are as follows:

- Supervised Learning
- Unsupervised Learning

Supervised Learning:

This technique of learning is also known as directed data mining. In this technique or approach the variable is divided into two categories. One variable is known as Explanatory variable and another is known as Dependent variable. The value of Dependent Variable must be known for data set.

Unsupervised Learning:

In this technique of learning the variables are not divide into parts. All the variables re treated as same. The only difference between supervised learning and unsupervised learning is that in case of unsupervised learning technique the target is to reduction of data in more generalized form like clusters.

II. Tasks performed by Data Mining

Various tasks performed in data mining is as follows:

- Classification
- Estimation
- Prediction
- Association Rule or Affinity grouping
- Clustering
- Description and Visualization

Classification

Classification is a process in which the whole database is monitored for the purpose of finding or extracting the relevant data or information. This phase is characterized by well define classes.

Estimation

Estimation is a technique which is performed to find out the output of some inputted data. This phase provides the output to corresponding input along with some unknown variables.

Prediction

In this phase the data is evaluated or monitored by considering some of the future predictions which can take place in near future and feasible. Normally prediction is perform or estimated on the basis of historical data. For example after evaluating the historical data it can be estimated or predicted that which customer will leave to buy the product or what can be reason behind this?

Association Rule

Association rule is used to define the relationship between various modules. Association tells that how modules or various objects are associated with each other.

Clustering

Clustering is a very strong technique used for data mining. It can only use as a data mining technique. Clustering used various algorithms to access the database deeply. Clustering is technique which divides the database into segments or sub groups. These segments are the collection of related variables. These sub divided segments are known as clusters. After dividing the database into clusters it is easy to extract the relevant information on the basis of divided clusters.

Description and Visualization

Data visualization is last step to perform. In this the mined data is represented to the end user. To make the data easily understandable to the user, various visualization techniques are applied to the mined data. [12]

III. Algorithm Used For Data Mining

ACO is an optimization algorithm that uses a graph for solving computational problems and finding good paths. ACO stands for Ant Colony Optimization, an algorithm that is based on the behavior of ants that finds path between their colony and the food source. ACO is the first algorithm of the Ant Colony algorithms family that aims to search optimal path in a graph. The ants when go in search of food, wanders and when finds a food source, lays pheromone trails in the path when returning from food source to their colony. The other ants follow the same path and find food source. The pheromone trail evaporates with time and hence loses its attractive strength. The pheromone trails stay for lesser time on longer paths because the more time an ant takes to travel down the path, the more time pheromone will evaporate. When the path is shorter, it is marched over frequently, pheromone trails are laid frequently which increases the density of pheromones and hence pheromone trails stay longer on shorter paths. The pheromone evaporation helps the follower ants to find shorter paths and when a shorter path is found the other ants are believed to follow that path instead of the earlier one. This avoids adherence to single path without searching other routes that are shorter. This behavior of ants helped in designing ACO that is an optimization algorithm for finding the path through a guided search in a graph. The application of this algorithm includes providing optimized solution to the travelling salesman problem fulfilling the conditions that every city be visited once. These algorithms can be applied for finding optimized solutions to many problems like vehicle routing problem, stochastic problems etc. The solutions that were earlier provided to these problems like approaches of simulated annealing and genetic algorithm were unable to adapt the abrupt changes in graph. ACO have advantage over these solutions that this algorithm runs uninterruptedly and readily adapt changes that happen in real time and thus can be applicable in network routing and urban transportation systems.

IV. Related Work

Rafael S. Parpinelli "Data Mining with an Ant Colony Optimization Algorithm":- In this author defines the Ant-Miner algorithm or technique for data mining. This algorithm is used to eliminate the classification from data. This algorithm is based on real ant colonies and other principles and concepts of data mining are also used. In this the comparison of Ant-Miner with CN2 is done with respect to the performance analysis. After simulation it is observed that the Ant-Miner is much reliable then CN2 with the point of simplicity and accuracy. Data mining is a process opf KDD. KDD stands for Knowledge Discovery in Database.

Bijaya Kumar Nanda "Class Based Rule Mining using Ant Colony Optimization" :- In this author explains that Ant Colony Optimization i.e. ACO is an algorithm which eliminates the process of classification from data i.e. rule based classifiers can be eliminated by using this algorithm. In this two data mining tasks are simulated. These tasks are ARM (Association Rule Mining) and another is classification. A classifier along with AOC is used which merge the association and classification.

DR. YASHPAL SINGH," NEURAL NETWORKS IN DATA MINING", In this paper author declares data mining is a technique which is used to fetch the relevant information from large data storage or data warehouse. All the work is being done online and lots of data is stored in order to save the history or to use it in near future. But to fetch the relevant information from huge amount of data is quite a difficult task. To fetch the important or meaningful data from large data storage, data mining is used. Data mining is a step-wise process. Many technologies are available for data mining like ANN, regression and decision tree. The ANN has a characteristic of black-box nature due to which researchers avoid this technology. In this paper author provides an overview to ANN and solve the problem of black- box to an extent.

Dr. T. Karthikeyan," A Study on Ant Colony Optimization with Association Rule", In this author defines the AOC is an algorithm used for data mining. ACO eliminates the process of classification from data. It is used as a discovery rule for optimization. AOC+ algorithm is uses MAX-MIN. It also uses system to develop rules in database. Soil classification is a technique which is used to differentiate between soils on the basis of their features and criteria. After simulation it is observed that the performance of Ant miner+ is better than the performance of Ant miner algorithm.

K. Amarendra," Research on Data Mining Using Neural Networks", In this paper it is defined by the author that ANN is widely used along with data mining. Data mining is a process to fetch the meaningful data from the large amount of data which may or may not important from some aspects. Example of this is relationship between patient's data and medical diagnosis. Data mining is a technique to extract important or meaningful or relevant data from huge data and then this extracted data is used for the purpose of decision making. Data mining search for the relationship and global patterns which are hidden in large amount of data. In this paper, author analysis the technology of data mining which is based on neural networks.

V. CONCLUSION

Data mining is defined as the process of detection and extraction of the hidden and the useful information from the data set .The information acquired through the process of data mining is initially unknown, useful, valid and of high quality. As studied form the literature it is concluded that data mining is the useful for the acquiring the information from the large set of data. An introduction to the data mining is given, its classification is done .Various techniques for data mining have been discussed. Various soft computing techniques have been used for the data mining process that shows that these are efficient method of data mining.

Traditionally ACO is considered as the best approach of data mining. In future this method can be further enhanced by using the trending soft computing algorithm. Along with this the hybrid approaches can be also used for improving data mining process.

References

[1] Rafael S. Parpinelli "Data Mining with an Ant Colony Optimization Algorithm",

[2] Bijaya Kumar Nanda "Class Based Rule Mining using Ant Colony Optimization", IJACTE, Vol 2, Issue 1, Pp 2319-2526, 2013

[3] DR. YASHPAL SINGH," NEURAL NETWORKS IN DATA MINING",JTAIT, Vol 5, Pp 37-42, 2005

[4] Dr. T. Karthikeyan ," A Study on Ant Colony Optimization with Association Rule", IJARCSSE, Vol 2, Issue 5,Pp 486-492, 2012

[5] K. Amarendra,"Research on Data Mining Using Neural Networks", IJCSI, Vol 2,Issue 1, Pp 1-8,

[6] Anand V. Saurkar," A Review Paper on Various Data Mining Techniques", IJARCSSE, Vol 4, Pp 98-101, 2014

[7] Sonalkadu," Effective Data Mining Through Neural Network", IJARCSSE, Vol 2, Issue 3, Pp 441-444, 2012

[8] Neelamadhab Padhy," The Survey of Data Mining Applications And Feature Scope", IJCSEIT, Vol 2, Isssue 3, Pp 43-58, 2012

[9] Kalyani M Raval," Data Mining Techniques", ijarcsse, Vol 2, Issue 10, Pp 439-442, 2012

[10] Nikita Jain," DATA MINING TECHNIQUES: A SURVEY PAPER", IJRET, Vol 2, Issue 11, Pp 116-119, 2013

[11] Ashish Kumar Dogra1,"A Review Paper on Data Mining Techniques and Algorithms", IJARCET, Vol 4, Pp 1976—979, 2015

[12] Anand V. Saurkar,"A Review Paper on Various Data Mining Techniques",IJARCSSE, Vol 4, Issue 4, Pp 98-101, 2014