



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

Optimization Based Data Mining in Business Research

Praveen Gujjar J¹, Dr. Nagaraja R²

Lecturer, Department of ISE, PESITM, Shivamogga, Karnataka, India^{1,2}

ABSTRACT: Business research is a process of collecting, classifying, analyzing and interpreting the data for making business decisions. Optimization in data mining and business research has found classical improvements in making decisions. Optimization in data mining provides a classical tool set to generate several data-driven classification systems, which helps in taking business decisions. Optimization approach in data mining strengthens the validity of organizing results and the improved collection, classification, analysis and interpretation of primary data. This paper focuses on the commonly used techniques for data mining viz., graph analysis, combinatorial optimization, Mathematical programming methods, Simulated and Genetic Method, Market Basket Analysis, APRIORI algorithms and randomized algorithms, Decision Tree induction method, K- means clustering algorithms, Drill down analysis. The optimization in data mining techniques helps decision makers to get precise and accurate information for making business decisions. Optimization techniques deal with data separation, classification. This paper also focuses on issues and Problems associated with data mining which involves optimal attribute subset, optimal number of clusters, missing values and incomplete data.

KEYWORDS: Data Mining, Business Decision, Optimization Techniques, Business Research.

I. INTRODUCTION

Data mining is becoming a buzzword for the today's world. Data mining is a process of looking a record available in the database for the given patterns and keywords, this process is performed by an experienced analyst to take business decision in the ongoing task in the organization. Data mining can show imperative and perceptive data that has been sprinkled all through systems and departments. Data mining looks directly into the database, which is the key for all multifaceted system. It helps in managing sales and marketing information.

Optimized Data mining enables us to pull out necessary information that is needed to take business decision. Optimization in data mining helps us to forecasting efforts to a novel point, allowing us to predict how the cost of goods and service might be fluctuate over a period of time with respect to material costs change.

Research is the process of finding solutions to a problem after a thorough study and analysis of the situational factors. Business Research provides the needed information that guides managers to make informed decisions successfully to deal with problems. The information provided it could be a result of cautious study and analysis of gathered data.

Business research comes with a problem or a query. It requires a clear articulation of a goal. Business research procedure it involves making principal problem statement into more manageable sub units applying hypothesis and interpreting of data in attempting to resolve the problem. The given solution must bring profit to the organization and that solution must be treated as commendable decision to bring profit.

Leaders who willing to make data driven decision he/she must undergone careful monitoring the key metrics of their organization so they can tune their strategy for better performance. It clearly indicate who like to purchase computer they will also purchase a mouse pad these insights made possible by optimized data mining approach.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

II. GENERAL PROCEDURE FOR DATA MINING

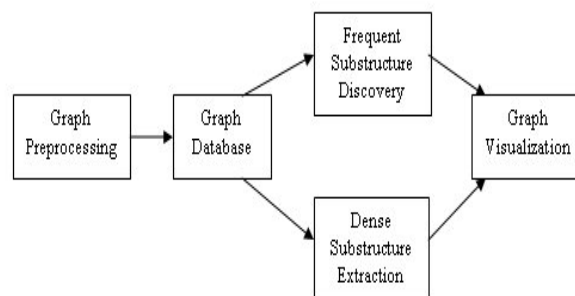
Data mining as an investigative Procedure it involves careful exploring data, typically business or market related data; in search for dependable patterns and/or methodical relationship between variables, and then to validate the result by applying the detected patterns to new subsets of data. General procedure for data mining it involves following steps those are 1) The initial exploration 2) Model building 3) Deployment.

Stage 1: Exploration – This stage involves data preparation which involves cleaning data, transformation of data, selecting subsets of records in case of data sets with large fields. Stage2: Model Building – It involves considering several models and choosing the best one based on their analytical performance. Stage 3: Deployment – Model selected as best in model building stage and applying it to new data in order to generate expected outcome.

III. DATA MINING TECHNIQUES

3.1 Graph Mining Tool Architecture

Graph mining tool that provides facilities for input data preprocessing for upload of source data into graph representations, frequent substructure discovery, dense substructure extraction and visualization techniques on the graph representation of data. The Graph mining tool architecture is shown in fig 1.



a. Combinatorial Optimization

Combinatorial optimization it involves obtaining the feasible solution in two ways. Those are selecting the best rule from a finite set of rules and another one is selecting the best subset of attributes. Numerous different subset solutions may have to consider. Hence branch and bound and Random search may used for the selecting the feasible solution for the business problem. Branch and bound technique is greater than the weka exhaustive search.

b. Random Search

Random search it involves initial solution $x^{(0)}$ and let $k=0$.

Loop going to work as shown below

Loop:

Consider the neighbors $N(x^{(k)})$ of $x^{(k)}$

Select a candidate x' from $N(x^{(k)})$

Check the acceptance criterion

If accepted then let $x^{(k+1)} = x'$ and otherwise let $x^{(k+1)} = x^{(k)}$

Until stopping criterion is satisfied

c. Simulated and Genetic Algorithms

Simulated Annealing (SA)

Common idea here is accept inferior solutions with a given probability that decreases as time goes on

Tabu Search (TS)

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

Common idea here is restrict the neighborhood with a list of solutions that are tabu (that is, cannot be visited) because they were visited recently

Genetic Algorithm (GA)

Common Idea is neighborhoods based on 'genetic similarity'

d. K-Mean Method

K-mean algorithm creates clusters by determining a central mean for each cluster. The algorithm starts by randomly select K entities as the means of K clusters and randomly adds entities to each cluster. Then, it re-computes cluster mean and re-assigns entities to clusters to which it is most similar, based on the distance between entity and the cluster mean. It describes as shown in fig 2

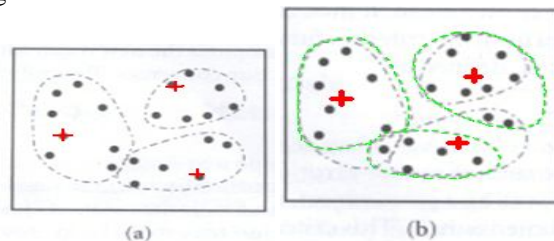


Fig 2: Showing K-mean algorithms procedure

e. Market Basket Analysis

Retailer gives the idea about each customer purchases different set of products, different quantities, different times which in-turn may use in business decision. Market basket analysis uses following information. 1) Identify who customers are (not by name). 2) understand why they make certain purchases. 3) Gain insight about its merchandise (products).

Market Business analysis helps in taking necessary action to take up the business decision those are 1) Store layouts : Which products to put on specials, promote, coupon etc. 2)Combining all of this with a customer loyalty card it becomes even more valuable.

f. APRIORI algorithm -Frequent item sets

A frequent (used to be called large) itemset is an itemset whose support (S) is $\geq \text{minSup}$. Apriori property (downward closure): any subsets of a frequent itemset are also frequent itemsets.

Using the downward closure, we can prune unnecessary branches for further consideration.

APRIORI

k = 1

Find frequent set L_k from C_k of all candidate itemsets

Form C_{k+1} from L_k ; k = k + 1

Repeat 2-3 until C_k is empty

Apriori's Candidate Generation

For k=1, C_1 = all 1-itemsets.

For k>1, generate C_k from L_{k-1} as follows:

The join step

C_k = k-2 way join of L_{k-1} with itself

If both $\{a_1, \dots, a_{k-2}, a_{k-1}\}$ & $\{a_1, \dots, a_{k-2}, a_k\}$ are in L_{k-1} , then add $\{a_1, \dots, a_{k-2}, a_{k-1}, a_k\}$ to C_k

The prune step- Remove $\{a_1, \dots, a_{k-2}, a_{k-1}, a_k\}$ if it contains a non-frequent (k-1) subset.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

g. Drill-Down Analysis

The Drill- down process analysis works by considering break down in the data of some variable of concern. Several Statistics, tabular representation of data and histograms can be grouped for the sake of computation. This technique is having its own kind of advantage, consider one simple example if we have to check the male customer of the particular region we can easily get our desired data by using the drill down analysis technique.

h. Decision Tree Induction method

Basic algorithm (a greedy algorithm)

1. Tree is constructed in a top-down recursive divide-and-conquer manner

2. At start, all the training examples are at the root

Attributes are categorical (if continuous-valued, they are discretized in advance) Examples are partitioned recursively based on selected attributes

3. Test attributes are selected on the basis of a heuristic or statistical measure (e.g., information gain)

i. Mathematical Programming

The Mathematical programming of data mining it involves Continuous decision variables. Constrained versus non constrained from of objective function those are linear programming, quadratic programming and General Mathematical Programming

IV. ISSUES ASSOCIATED IN DATA MINING FOR BUSINESS DECISION

Above mentioned Data mining Technique are having certain issue that is to be resolved in quick span of time. Some of the issues associated in data mining for taking up the business decision are pin pointed in this section. Major issues arises in the data mining is that developing a unifying method of data mining for scaling up for high dimensional data. Further there is needed to be having dealing mechanism for non static cost sensitive data. Some issues raised by data mining technology are individual privacy. Technique used here is possibly analyzing routine truncations in the business and it may disclose a significant amount of information about individual buying practice. Another issues roused by a data mining is that data integrity.

Some of the business issues in data mining are listed in the following points

- 1) Methodological issues in applying the data mining technique
- 2) Cost of data mining
- 3) Distributed Data Mining and Mining Multi-agent Data
- 4) Missing value
- 5) Incomplete Data

V. MERITS OF DATA MINING IN BUSINESS RESEARCH

Business point of view optimized data mining helps the decision makers to improve Branding and marketing. Data can disclose much information like direction of marketing and finance department. Using of data mining help in predict the future market in the society. Hence it helps in creating good will in the mind of people who were all leaving in that region.

Further Optimized data mining can help us to gain more profit if we predict our future. Knowing the customer wants is so important in the production side. Fulfilling customer need gives more satisfaction to both producer and consumer. Moreover decision makers can tap new regions for making their product to be introduced in new geographical area. Data mining is potential for pervasive use in a wide range of sectors, which stems from their capacity to fulfill a generic solution for a given set of problem in business.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

VI. CONCLUSION

Having optimized data mining technique it helps the decision makers to take precise decision about the organization to gain more incremental profit. There is security issues upset the rhythm of using the tools for data mining. Scientific science and scientific computation increase the demand of using optimized techniques to mine the data for making most effective decision from business point of view.

REFERENCES

- [1] Cristianini, N., and Shawe-Taylor, J., An introduction to Support Vector Machines and other Kernel-Based Learning Methods, Cambridge University Press (2000).
- [2] E.J. Anderson and P. Nash, John Wiley and Sons Ltd, Linear Programming in Infinite-Dimensional Spaces, 1987.
- [3] Han, J., and Kamber, M., Data Mining: Concepts and Techniques, The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers (2000).
- [4] Hastie, T., Tibshirani, R., and Freedman, J., The Elements of Statistical Learning - Data Mining, Inference and Prediction, Springer Series in Statistics, 2001.
- [5] Osmar R. Zaiane, Mohammad El-Hajj, and Paul Lu. Fast parallel association rule mining without candidacy generation. In ICDM, 2001.