www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume – 5 Issue -04 April, 2016 Page No. 16172-16175

Performance Appraisal Of KDD Technique In Shopping Complex Dataset

Dr. K.Kavitha
Assistant Professor, Department of Computer Science
Mother Teresa Women's University, Kodaikanal
Kavitha.urc@gmail.com

Abstract

KDD field is concerned with the development of methods and techniques for making raw data into useful information. Real world practical application issues are also outlined. The basic assumption in predicting financial markets is that it is not possible. This is consistent with remarks many financial professionals have made. In this survey, lot of information from the customer and vendors of shopping mall is collected. Vendors are needed to extract the required combination of customers. To find out the solution of this automatic research application to improve the business strategy is undertaken. This paper highlights the concepts, review the status and limitations.

Keywords: KDD, Association Rule, Confidence, Support, Threshold

1. Introduction

Data mining and Knowledge Discovery is widely accepted as a key technology for enterprises. It is used to improve the decision support, data analysis of knowledge from data of their abilities. The data mining of various techniques have been proposed. Among those techniques, an association rule learning algorithm named Apriori is most famous. It depends on the generation of the candidate set and the test method. However, it is more efficient in mining and discovering knowledge.

The definitions of KDD and data mining and the general multistep process are provided. The application of multistep process is defined as the data mining algorithms as one particular step in the process. The KDD field is disturbed with the development of methods and techniques for making raw data into useful information. The algorithm and the application of data-mining step are discussed. The practical application issues are also outlined. The basic assumption in predicting financial markets is that it is not possible. This is consistent with of remarks many financial professionals. In particular, many trading professionals and vendors do sense that there is times certainly few where vendors can predict relatively well. It considers explicitly the trade-off between model coverage and model accuracy. Trying to give an accurate prediction for all data points is unlikely to succeed. On the other, it must be accurate enough and simultaneously general enough to allow sufficient high-probability opportunities to trade effectively.

The basic problem addressed by the KDD process is one of mapping low level into other forms that might be more compact more abstract or useful at the core of the process such as Selection through Interpretation. It is the application of specific data mining methods for pattern discovery and extraction.

2. Knowledge Discovery Approach in Shopping Complex Dataset

Information from the customer and vendors of shopping mall are collected. Vendors are needed to extract the required combination of customers. To find out the solution of this automatic research application to improve the business strategy is undertaken. From the driven rules to predict some solution based on that vendors can arrange the products and get maximum profit for their business and also it can get business enhancement if vendors are not having the items. The same rule can specify for specific products also like which brand customers, which combination customers. It can predict relatively in particular many trading professionals and vendors do feel that there are times. This attitude, "generally agnostic but irregularly making bets," has important implications. One of the major challenges is used to reduce the noisy periods by being more selective about the conditions under which to advance to find patterns that offer a reasonable number of opportunities to conduct high risk-adjusted-return trades.

A significant number of data mining techniques have been presented in order to perform different knowledge tasks. Association rule mining, frequent itemset mining, sequential pattern mining are some of these techniques. Association rules are used to extract interesting correlations, associations, frequent patterns between the item sets in the database or other repositories. Generally, the association rules can be expressed as$A \Rightarrow B$, where$A, B \subset I$ are called item set where A and B are referred as the antecedent and consequent respectively. Item set X can be mathematically denoted as, $X = X_1, X, \dots\dots, X$ be a set of *n* attributes. It produces dependency rules which will predict occurrence of an item based on occurrences of other items. The application of association rules discovery is supermarket shelf management. This rule is used to develop specifically for data mining. Then, it is used to find the correlations between items as rules. The examples for association rule discovery is supermarket, and attached mailing in direct marketing. It has confidence and support. Multiple level association rules items often form hierarchy. Itemset at the lower level are expected to have lower support, and transaction database can be encoded based on dimensions and levels. It can be used to relate pages that are most often referenced together in a single server session. Association rule discovery using the Apriori algorithm may reveal a correlation between users.

Classification is the most commonly applied data mining technique. It employs a set of pre-classified examples to develop a model that can classify the population of records at large. It frequently employs decision tree or neural network based classification algorithms. The data classification process includes learning and classification. The test data are used to estimate the accuracy of the rules of classification. The algorithm then encodes these parameters into a model called a classifier.

It is the task of mapping a data item into one of several predefined classes. It can be done by using supervised inductive learning algorithms such as decision tree classifier, and support vector machines. Support vector machine is defined as the Methods for analyzing and modeling data which can be divided into groups; supervised and unsupervised learning. Supervised learning requires is defined as the combination of both the value of independent variables and dependent variables. One of the most useful applications of statistical analysis is the development of a model to represent and explain the relationship between variables.

Classification rule mining aims to discover a small set of rules in the database to form a correct classifier. Association rule mining finds all rules in the database that satisfies some minimum support and minimum confidence constraints. Most of the research conducted on classification in data mining has been dedicated to single label problems. The problem can be defined as, let X denote the domain of possible training instances and Y be a list of class labels, let S denote the set of classifiers for X →Y, each instance $d \in X$ is assigned a single class y that belongs to Y.

Data mining adds to clustering the complications of very large datasets with very many attributes of different types. This forces unique computational requirements on relevant clustering algorithms. It is a division of data into groups of similar objects. It allows us to make associations between items in a column based not merely on examining items in adjacent rows. Clustering can be said as identification of similar classes of objects. This approach can also be used for effective means of distinguishing groups of objects. It can be used as preprocessing approach for attribute subset selection and classification. It is a common descriptive task where one seeks to identify a finite set of categories or clusters to describe the data.

The application of clustering in a knowledge discovery context includes discovering homogeneous subpopulations for consumers in marketing databases and identifying. It is used to find an effective dataset representation that supports counting while also automatically identifying transactions that contain the query itemset, a pattern-based flat clustering framework that uses association rules to generate a pattern. A given a set of data points, each having a set of attributes, and a comparison measure among them, find clusters such that the data points in one cluster are more similar to one another.

Clustering is useful in several exploratory pattern-analyses, decision-making, and machine-learning, including data mining, and pattern classification. It is a technique to group together a set of items having similar characteristics. Clustering of users tends to establish groups of users exhibiting similar browsing patterns. Sequential pattern discovery attempts to find inter-session patterns such that the presence of a set of items is followed by another item in a time-ordered set of sessions. Other types of temporal analysis that can be performed on sequential patterns include trend analysis, change point detection. The example of sequential pattern discovery is sequences in which customers purchase services.

It is the task of finding the complete set of frequent subsequences in given set of sequences. A huge number of likely sequential patterns are hidden in databases. It uses frequent items to recursively project sequence databases into a set of smaller projected databases and grows subsequences fragments in each projected database. A huge number of possible sequential patterns are hidden in databases. It is used to find the complete set of patterns, when promising, satisfying the minimum support. It is Apriori-based method. A sequence database is mapped to a large set of items. It is performed by growing the subsequences one item at a time by Apriori candidate generation.

Application of sequential pattern discovery is customer shopping sequences, stocks and markets. It is similar to the frequent itemsets mining, but with consideration of ordering. A database D of sequences called data sequences, in which:I = {$i_{1,}i_{2,...}i_n$} is the set of items. Each sequence is a list of transactions ordered by transaction-time. It consists of following fields such as sequence-id, transaction-id, and itemset. The technique of sequential pattern discovery attempts to find inter-session patterns like the presence of a set of items. It can be performed on trend analysis, and change point detection are included in the sequential patterns.

It is a set of objects, each associated with its own timeline of measures, find rules that predict strong progressive dependencies among different events. A sequential pattern mining is performed by growing the subsequences one item at a time by Apriori candidate generation. It is vertical format sequential pattern mining method. Regression is the easiest technique to use, but is also possibly the least powerful. Its analysis is designed to estimate the probability of observing a given data set given that a pre-determined hypothesis about the relationship between an outcome variable and a set of factors is assumed. It is a data mining function that predicts a number. Its tasks begins with a data set in which the target values satisfies.

This analysis can be used to model the relationship between one or more independent variables and dependent variables. It can be adapted for prediction. It consists of a family of techniques for prediction that fit linear and non-linear combinations of basic functions to combinations of the input variables. Nonlinear regression methods, although powerful in representational power, can be difficult to interpret. It is learning a function that maps a data item to a real-valued prediction variable. Applications for regression is predicting the amount, predicting consumer demand for a new product as a function of advertising expenditure, and predicting time series. It predicts a value of a given continuous valued variable based on values of other variables. For example, predicting sales volumes of new product based on advertising expenditure, and time series prediction of stock market indices. It is very difficult to predict because it may depend on complex interactions of multiple predictor variables. Neural networks too can create both classification and regression models.

Deviation detection can reveal surprising facts hidden inside data. In knowledge discovery, often such information is a vital part of important business decisions and scientific discovery. But majority of researcher are doing in first three methods only. The association rule discovery method is chosen in this research. It should work in conjunction with automatic selection of the appropriate function that promises the best results. It relatively is a new operation in terms of commercially available data mining tools. These applications include fraud detection in the use of quality control. Improving the shopping complex business strategy due to trends in fashion world is tried. The database of purchasing details, billing details from the point of sale is collected. It's contributed an adequate approach for developing the business planning. It is the combination approach of data mining and real time business scheme.

Some of the stores are using bar code reader system to recognize the products. In this case, the point of sale bar code details is collected. In this vendors have to apply the association rule discovery method to find out the similarity of the purchase. The importance of deviation detection in data has been recognized in the fields of databases and machine learning for a long time. It has been often viewed as outliers, or noise in data. It gives a new measure for deviation detection and an algorithm for detecting deviation. Then, it has linear density and also a necessary property for a data mining algorithm.

Sample data set and their association rule derived are given in Table 3.1. From the rules driven some solution based on that vendors need is predicted. The rule specifies answers to several topics like which brand the customers like more?, which combination customers like most ?, etc.,

Transaction ID	Items
MT1	Shampoo, Conditioner, Comb
MT2	Shampoo, Lotion
MT3	Lotion, conditioner, Soap
MT4	Soap, Oil, Lotion
MT5	Lotion, Oil, Detergent

Table 3.1 Sample Dataset from shopping Mall

MT – Member Transaction

Some of the Discovered rules are

Shampoo, Conditioner -> -> Lotion
Conditioner, Oil ->-> Shampoo
Shampoo -> -> Lotion

Limitations

- Absence of Decision Making Concept
- Increases Memory Consumption through Rule Length
- Increases Time Complexity
- Missing Classification and Clustering Techniques

DOI: 10.18535/ijecs/v5i4.20

3. Conclusion

The proposed work provides an efficient method for developing business strategy and helps to business firms such as the shopping mall to enhance the business strategy. This is the combined approach of data mining and real time business strategy. But still the research does not satisfy the Decision Making especially user’s Decision. so, the proposed work concentrated the user’s response. Based on that, transactions are limited and reduce the memory usage.

References

[1] P. Bollmann-Sdorra, A.M. Hafez and V.V. Raghavan,“A Theoretical Framework for Association Mining based on the Boolean Retrieval Model,” DaWaK 2001,September 2001.

[2] A.M. Hafez, “Association mining of dependency between time series,” Proceedings of SPIE Vol. 4384, SPIE AeroSense, April 2001.

[3] V.V. Raghavan and A.M. Hafez, “Dynamic Data Mining,“ IEA/AIE 2000, pp.220-229, 2000.

[4] A.K. Tung, J. Han, L.V. Lakshmanan and R.T. Ng, "Constraint-based Clustering in Large Databases," Proc. Int. Conf. on Database Theory, pp. 405-419, 2001.

[5] E Turban, 1997. Decision support systems and expertsystems (Fifth Edition), Prentice-Hall, London:UK.

[6] HL Viktor, 1999. Learning by Cooperation: An Approach to Rule Induction and Knowledge Fusion, PhD dissertation, Department of Computer Science, University of Stellenbosch, Stellenbosch: South Africa.

[7] Abraham Bernstein, Shawndra Hill, and Foster Provost. An Intelligent Assistant for the Knowledge Discovery Process. Technical Report IS02-02, New York University, Leonard Stern School of Business, 2002.

[8] Joerg-Uwe Kietz, Regina Zuecker, Anna Fiammengo, and Giuseppe Beccari. Data Sets, Meta-data and Preprocessing Operators at Swiss Life and CSELT. Deliverable D6.2, IST Project MiningMart, IST-11993, 2000.

[9] Pavel Brazdil. Data Transformation and Model Selection by Experimentation and Meta-Learning. In C.Giraud-Carrier and M. Hilario, editors, Workshop Notes – Upgrading Learning to the Meta-Level: Model Selection and Data Transformation, number CSR-98-02 in Technical Report, pages 11–17. Technical University Chemnitz, April 1998.