ISSN No : 2454-4221 (Print)
ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

Regression Technique using Data mining with Frequent Data Item Sets

S.V.Subramanyam[1]
Professor, Department of Artificial Intelligence and Machine Learning,
School of Engineering, Malla Reddy University
Hyderabad, Telagana State, India.
svsubramanyam@mallareddyuniveristy.ac.in

H.Packiaraj[2]
Assistant.Professor, Department of Artificial Intelligence and Machine Learning,
School of Engineering, Malla Reddy University
Hyderabad, Telagana State, India.
packiaraj@mallareddyuniveristy.ac.in

Abstract

Data mining is considered to deal with huge amounts of data which are kept in the database, to locate required information and facts. Innovation of association rules among the huge number of item sets is observed as a significant feature of data mining. The always growing demand of finding pattern from huge data improves the association rule mining. The main purpose of data mining provides superior result for using knowledge base system. Researchers presented a lot of approaches and algorithms for determining association rules. This paper discusses few approaches for mining association rules.

Frequent pattern mining approach is the most efficient data mining method to find out hidden or required pattern among the large volume of data. It is responsible to find correlation relationships among various data attributes in a huge set of items in a database. Studying Proposed algorithm, Apriori that is used to extract frequent item sets from large item sets. The Proposed algorithm has a limitation of wasting time for scanning the whole database searching on the frequent itemsets. Here we are using an improved proposed algorithm to reduce the time consumed in transaction scanning for candidate itemsets by reducing the number of transactions to be scanned by using smallest minimum support.Another approach discussed is regression technique for pairing the unpaired itemsets also reducing the time consuming of the item sets. Frequent pattern generation in data mining using Regression Technique.

Keywords :Data mining, regression techniue algorithm, minimum support threshold, multiple scan, frequent item sets regression.model, regression technique.

Introduction

Currently the world has a wealth of data, stored all over the planet (the Internet and Web are prime examples), but we need to understand that data. It has been stated that the amount of data doubles approximately every twenty months. This is especially true since the use of computers and electronic database packages. The amount or quantity of data easily exceeds what a human can comprehend on their own and thus if we wish

Cite this article as: S.V.Subramanyam & H.Packiaraj "Regression Technique using Data mining with Frequent Data Item Sets", International Journal of Research in Advanced Computer Science Engineering, (IJRACSE), Volume 8 Issue 5, October 2022, Page 1-6.

ISSN No : 2454-4221 (Print)
ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

to use and understand as much data as possible we need tools to help us. From this overwhelming state, the field of data mining has taken off and become hotly utilized.

The role of data mining is simple and has been described as "extracting knowledge from large amounts of data". Frequent pattern miningis one of the dominating data mining technologies. Frequent pattern miningis a process for finding associations or relations between data items or attributes in large datasets. It allows popular patterns and associations, correlations, or relationships among patterns to be found with minimal human effort, bringing important information to the surface for use. Frequent pattern mininghas been proven to be a successful technique for extracting useful information from large datasets. Various algorithms or models were developed many of which have been applied in various application domains that include telecommunication networks, market analysis, risk management, inventory control and many others. The success of applying the extracted rules to solving real world problems is very often restricted by the quality of the rules. However, the quality of the extracted rules has not drawn adequate attention. Measuring the quality of association rules is also difficult and current methods appear to be unsuitable, especially .

Literature Survey

1. In 2008, He Jiang et al. [1] suggest the weighted association rules (WARs) mining are made because importance of the items is different. Negative association rules (NARs) play important roles in decision-making. But according to the authors the misleading rules occur and some rules.

2. In 2009, Yuanyuan Zhao et al. [2] suggest that the Negative association rules become a focus in the field of data mining. Negative association rules are useful in market-basket analysis to In 2012, Yihua Zhong et al. [3] suggest that association rule is an important model in data mining. However, traditional association rules are mostly based on the support and confidence metrics, and most algorithms and researches assumed that each attribute in the database is equal. In fact,.

3. may have different frequency. So he puts forward a discovery algorithm for mining positive and negative fuzzy association rules to resolve these three limitations.

4. In 2013, Luca Cagliero et al. [10] tackle the issue of discovering rare and weighted itemsets, i.e., the Infrequent Weighted Itemset (IWI) mining problem. They proposed two novel quality measures to drive the IWI mining process.

Problem Domain

A regression algorithm estimates the value of the target (response) as a function of the predictors for each case in the build data. These relationships between predictors and target are summarized in a model, which can then be applied to a different data set in which the target values are unknown.

ISSN No : 2454-4221 (Print)
ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

The traditional Apriori-based implementations are efficient but cannot generate all valid positive and negative ARs. So we try to solve that problem without paying too high a price in terms of computational costs and reducing space with less time.

The simple linear regression model is represented by the equation:

$$y = \alpha + \beta X$$

By mathematical convention, the two factors that are involved in a simple linear regression analysis are designated *X* and *y*. The equation that describes how *y* is related to *x* is known as the **regression model** [2, 3]. Here in the equation α is the *y* intercept of the regression line and *β* is the slope.

Limitations of Proposed algorithm:

Proposed algorithm suffers from some weakness in spite of being clear and simple. The main limitation is costly wasting of time to hold a vast number of candidate sets with much frequent itemsets, low minimum support or large itemsets. For example, if there are 10^4 from frequent 1- itemsets, it need to generate more than 10^7 candidates into 2-length which in turn they will be tested and accumulate. Furthermore, to detect frequent pattern in size 100 (e.g.) v1, v2… v100, it have to generate 2^{100} candidate itemsets that yield on costly and wasting of time of candidate generation. So, it will check for many sets from candidate itemsets, also it will scan database many times repeatedly for finding candidate itemsets. Apriori will be very low and inefficiency when memory capacity is limited with large number of transactions.

Proposed work

There is much scope in he domain of frequent pattern rule mining, as in today era huge amount of data is generating day by day. Also synthesis of various techniques together to solve complex problems is proved very efficient. Hence province of frequent pattern mining can also be made more efficient with the help of various optimization techniques for instance genetic algorithm, fuzzy logics, rough set, soft set etc.

The computational cost of association rules mining can be reduced in the following ways:

- It can be reduced by reducing the passes.
- Need sampling.
- Using regression technique
- Reduce time consuming with large data scan.
- Reduce again and again large data scan problem.

Methodology

it's one in every of the info mining techniques. The aim is to search out that things are oftentimes purchased along in order that they're organized consequently on the shelves of the shop. This data also can be utilized in cross commercialism. It consists of if/then statements that are wont to verify the connection between the info, hold on in warehouses or alternative repositories, which can otherwise appear unrelated. As an example if an individual buys a replacement automotive he's presumably to urge its

ISSN No : 2454-4221 (Print)
ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

insurance done. Information is analyzed for locating out frequent patterns to make association rules. Association rules are wanted to predict the behavior of the purchasers. These are used for market basket analysis. Additionally alternative areas of association rules embody web usage mining, Intrusion detection, Continuous production, and Bioinformatics. There square measure several algorithms for association rules mining.

Steps to perform algorithm: Regression technique

C_k denotes the set of candidate k-itemsets and F_k denotes the set of frequent k-itemsets

Step 1: Scan the database D, generating candidate 1- setC1;

Step 2: According to the min_sup, frequent item 1- setL1 is generated from the candidate 1- setC1;

Step-3: According to the min_sup, frequent item (k+1) – set Lk+1 is generated from the candidate (k+1) - set CK+1;

Step-4: Get frequent itemsetsLk;

Step-5: Generate candidates from frequent items CK.;

Step-6: Prune the results to find the frequent itemsets using Linear Regression. Lk, and candidate (k+1) – set Ck+1 are generated.

Step-7: Goto Step-4 till outliers are removed.

Step-8: Generate strong association rules from frequent itemsets. According to the min_sup, frequent item (k+1) - set Lk+1 is generated from the candidate (k+1) - set CK+1;

Step-9: A Rule which satisfy the min. support and min. confidence threshold.

Implementation

For this problem, we can use weka tool as well as mat lab for pattern matching.

Through tables; we find out paired data set.

General steps:

1. In the first pass, the support of each individual item is counted, and the large ones are determined

2. In each subsequent pass, the large item sets determined in the previous pass is used to generate new item sets called candidate item sets.

3. The support of each candidate item set is counted, and the large ones are determined.

4. This process continues until no new large item sets are found.

Conclusion

The Linear Regression technique predicts a numerical value. Regression performs operations on a dataset where the target values have been defined already. And the result can be extended by adding new information. The relations which regression establishes between predictor and target values can make a pattern. This pattern can be used on other datasets where the target values are not known. In this paper we have formulate a linear regression technique,

ISSN No : 2454-4221 (Print)
ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

further we have designed the linear regression algorithm. The test data are taken to prove the relationship between predictor and target variable which is being represented by the linear regression equation

$$Y = \alpha + \beta X$$

Frequent pattern mining rules are very useful in applications going beyond the standard market basket alanysis. We have shown here various Proposed algorithms used to find frequent items in a given transaction of database. Since Proposed algorithm was first introduced and as experience was pile up, there have been many attempts to devise more efficient algorithms of frequent itemset mining. Many algorithms share the same idea with Apriori in that they generate candidates.

References

1. Manisha rathi Regression modeling technique on data mining for prediction of CRM CCIS 101, pp.195-200,2010Springer–Verlag Heidelberg 2010.

2. Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman, 2nd ed.

3. K.Kotaiah Swamy & E.Ravi Kumar, "Knowledge Mining With Big Data Handling with Data Driven Models", International Journal & Magazine of Engineering, Technology, Management and Research (IJMETMR), ISSN 2348-4845, Volume 3 Issue 5, May 2016, Page 52-56.

4. Breese, J. S., Heckerman, D., and Kadie, C. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Proceedings of UAI-1998: The Fourteenth Conference on Uncertainty in Artificial Intelligence.

5. Burges, C. (1998). A tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2(2):955–974Mohammad Al.Maolegi,Bassam Arkok,An improved proposed algorithm for association rules.

6. Charu C.Aggarwal,Philip S.Yu.Anew framework for itemset generation in IBM T J Watson Research Centre,1998.

7. Shweta Sharma and Ritika Pandhi,Proposed algorithm for frequent itemsets patterns,vol.3,no.6,june2012

About the Authors:

1.S.V.Subramanyam
Professor,
Department of Artificial Intelligence and Machine Learning, School of Engineering,
MallaReddy University
svsubramanyam@mallareddyuniversity.ac.in

S.V.Subramanyam he is working in Dept of Artificial intelligence and Machine Learning at Malla Reddy University . He did B.Tech In computer science and information technology

ISSN No : 2454-4221 (Print)
ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

from J.B.I.E.T ,and M.Tech in Computer Science and Engineering from Indur Instituteof Engineering and Technology , He Published papers in many national and international conferences and journals . his area of interest is Data Base,Data mining and Data Warehouseand computer networking

2.H.Packiaraj
Assistant. Professor
Department of AI & ML,
School of Engineering
Malla Reddy University,
packiaraj@mallareddyuniversity.ac.in

Mr.H.Packiaraj, well known author he received B.Tech from CSI Institute of Technology,Thovalai. Affliated by Anna University, Chennai and M.E (NIE) from Karunya University, Coimbatore. He worked for 3 years in PSN Engineering College, Tirunelveli Dist. He worked for 10 years as Asst. Professor in Dept. of CSE in Loyola Institute of Technology and science,Thovalai, Kanyakumari Dist. Presently he is working in Dept of AI & ML in Malla Reddy University, Hyderabad. He Published papers in many national and international conferences and journals. His areas of Interest include Data Warehouse and Data Mining, Networking information security, and Machine Learning.