ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

FINDING OF REPEATED ITEMSETS USING GENETIC ALGORITHM

Kavya.T*, Sasikumar.R
* Student, Department of Computer Science and Engineering, R.M.D. Engineering College, Thiruvallur Dist -601206,T.N.,INDIA.
Associate Professor, Department of Computer Science and Engineering, R.M.D. Engineering College, Thiruvallur Dist-601206, T.N.,INDIA.

ABSTRACT

Extracting high profit itemsets from a deed database allude to the finding of itemsets with high profits. In existing algorithms they acquire the difficulty of producing huge numbers of candidate itemsets reduces the extracting performance in conditions of time and memory requirement. This difficulty may become poorer when the database contains long transactions. Therefore utility pattern growth (UP-Growth) and UP-Growth+ method gives out compact frequent item sets that is comparatively a good result when compared to already existing method. Even though it gives out compact frequent itemsets, UP-Growth generates minimum utility itemsets and UP-Growth+ generates maximum utility itemsets. But there is a difficulty for the provider to find frequent itemsets based on profit from the entire dataset . By considering this problem, an innovative solution has been proposed for finding frequent itemsets from the entire database according to the customer purchase details. Our main aim is to view the frequent itemsets from the combination of minimum and maximum utility itemsets using GENETIC algorithm.

KEYWORDS: high profit itemset, candidate pruning, profit mining.

INTRODUCTION

The rapid e-commerce growth has made both business community and customers face an innovative circumstance. Due to penetrating competition on one hand and the customer's option to choose from several alternative businesses community has realized the necessity of intelligent marketing strategies and relationship management. Business Intelligence is the gathering, management, and analysis of large amounts of data on the company's customers, products, services, operations, suppliers, and partners and all the transactions in between. Business intelligence applications include target marketing, market basket analysis, customer profiling and fraud detection in e-business. Intelligent techniques in web mining aid in managerial decision making, policy establishment, predicting customer's action etc. Existing methods, potential high profit itemsets are initiated and then extra database scan is performed for finding their profits. However, existing system often produces an enormous set of potential high profit itemsets and their extracting performance is reduced. The enormous amount of potential high profit itemsets forms a challenging problem to the extracting performance, the more turnaround time it consumes[1].

The UP-Growth is one of the efficient algorithms to generate high utility itemsets depending on construction of a global UP-Tree. In this phase, the framework of UP-Tree follows three steps: (i). Construction of UP-Tree. (ii). Generate PHUIs from UP-Tree. (iii). Identify high utility itemsets using PHUI. The construction of global UP-Tree is follows, (i). Discarding global unpromising items (i.e., DGU strategy) is to eliminate the low utility items and their utilities from the transaction utilities. (ii). Discarding global node utilities (i.e., DGN strategy) during global UP-Tree construction. By DGN strategy, node utilities which are nearer to UP-Tree root node are effectively reduced. Two algorithms utility pattern growth (UP-Growth) and UP-Growth+, and a reduced data structure, called a utility pattern tree (UP-Tree) are suggested for finding high profit items from a deed database efficiently.

A pre-processing technique extraction transformation loading is used for secure processing and maintaining the datasets efficiently. Here first we have login page for security purposes. Customer has to sign in using their corresponding login and from the item list user can purchase their products according to the available count of the product. Fig. 1 shows System architecture for finding frequent itemsets. Customer registration details are maintained

by the admin. Customer can give their reviews in the message box and it can be seen by the admin. Multiple providers can add their item details with their item profit in database and they can update and delete their item details after add the item. While adding the item provider will enter the profit and count of the product. Here admin can see registration details of customer and available product list. Then entire data set can also seen by the admin. Admin login option is used to show the available total item, current registered user. Multiple providers can view their frequent itemsets using their customer purchased details Provider can see the frequent itemsets from the entire data set using genetic algorithm in their login page according to the customer purchase from the minimum and maximum utility.

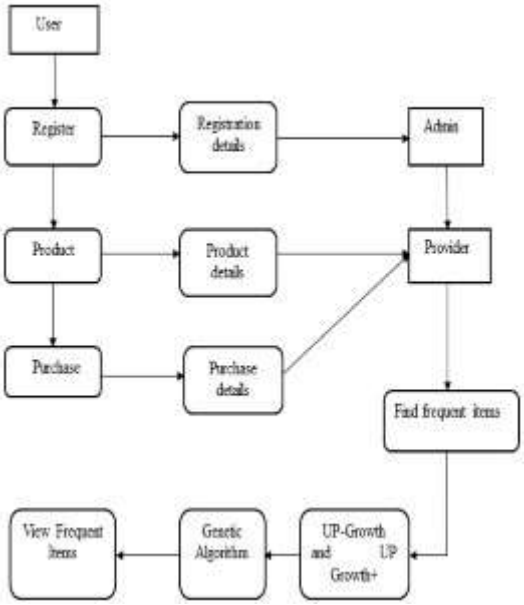

Fig. 1 System architecture

Providers can view their frequent itemsets using their customer purchased details and view the frequent itemsets in either monthly or daily basis. Thus genetic algorithm which helps to view frequent itemsets from the large data streams . The rest of the paper is organized as follows. Section II reviews the related work on finding high profit and frequent itemsets. Section III describes the solution to the problem of finding frequent itemsets and section IV related to the result and the future works.

RELATED WORKS

There are many related papers for finding high profit and frequent itemsets. R. Agrawal et al has proposed to solve the difficulty of mining sequential patterns and practically evaluate their performance using artificial data. The limitations are scanning the database many times; the number of candidate sequences is generated so there is a need for more efficient mining methods and execution time, memory space will be increased in existing approaches[2]. J. Han et al presents an approach frequent pattern tree (FP –Tree) for storing vital information about frequent patterns. FP growth is used for extracting a complete set of frequent patterns by frequent pattern growth.

Here the main drawback of FP –tree is expensive to build and needs more memory for storing the transactions and it is very tedious to implement because of its complex data structure[3].

Mahesh S. Raisinghani proposed a method for conceptualizing and identifying a combination of qualitative and quantitative variables for evaluating e-business strategies comprising e-commerce level and business-level strategies. The ANP research methodology process is employed since it is well suited for the study of a multi-attribute decision problem. The limitations used are Subjective bias because there were no tests concerning inter and intra consistency tests. This test was conducted as a cross sectional design. A longitudinal study may give different result[4].

Tamir Tassa presents an approach based on the Fast Distributed Mining (FDM), it is simpler and more efficient in terms of communication rounds, communication cost. The methodologies used are Fast distributed Mining (FDM), Apriori algorithm. There is a problem of mining generalized association rules and subgroup discovery in horizontally

partition data[5]. Jaejoo Lim have proposed this approach provides a point of reference for organizations using interconnected IT systems for e-business, and also serves to guide the standards work to ensure the coherence and integration of related techniques, services and protocols.

The methodologies are XML, simple object access protocol (SOAP) and business process modelling. The limitations are greater uncertainty in defining requirements for standard specification and high careful evaluation is needed[6]. Yu-Chiang Li et al proposed a paper Isolated Items Discarding Strategy (IIDS), which can be applied to any existing level-wise utility mining method to reduce candidates and to improve performance[7]. Hsin-Yun Huang et al proposed two efficient one-pass algorithms, MHUI-BIT and MHUI-TID, for mining high utility itemsets from data streams within a transaction-sensitive sliding window[8][12].

Long Jin et al have studied multiple sliding windows for mining frequent patterns on data stream in this paper and propose an efficient discounting method with different lengths of time-sensitive sliding-window[9]. Dr. Govardhan proposed a paper to find all the frequent itemsets from given data sets using genetic algorithm with lack of accuracy[10]. Thus by analyzing these papers the provider needs a better solution for viewing frequent high profit itemsets. So the itemsets using utility pattern gowth is an easy technique for identifying frequent itemsets. Based on the genetic algorithm provider can easily find the frequent itemsets from the large data stream.

FREQUENT ITEMSETS MINING

In existing algorithms they acquire the difficulty of producing huge numbers of candidate itemsets reduces the extracting performance in conditions of time and memory requirement. The drawback is it cannot satisfy the requirements of users who are interested in discovering the itemsets with high sales profits, since the profits are composed of unit profits, i.e., weights, and purchased quantities[11].

We propose two novel algorithms as well as a compact data structure for efficiently discovering high utility item sets from transactional databases. The genetic algorithm used for discovering high utility itemsets and maintaining important information related to utility patterns and gives out compact frequent item sets that is comparatively a good result when compared to already by pruning the candidate item sets and therefore time and memory requirement has been reduced.

The overall scheme of the project is to find and view high profit frequent itemsets from the large database. To view a frequent itemsets first user and provider have to login into the account. Login is mainly for the security purposes to maintain the customer purchase details secretly. If user wants to purchase the product they have to login with their unique id and password. Unique password will be given to the individual person to personalize their data's. Before login the user would have to register their details in customer registration. So only the registered users can able to login the page.

Message to admin option used to each customer send their opinion to admin. View original item option is used show original item to customer. User can purchase the product using their login. From the available product list customer can purchase according to the available count of the product. Before login the provider would have to register their details in provider registration. Multiple providers can login into their page using unique id and password. can add their item details with their item profit in database and they can update and delete their item details after add the item. Here there is a function to help provider to find concurrently available items and frequently buying item. Fig 2 shows the login page. Admin login option is used to show the available total item, current registered user, Customer registration details are maintained by the admin. Customer can give their reviews about the products in the message box and it can be seen by the admin.

Fig.2 Login page

Here admin can see registration details of customer and available product list. Then entire data set can also seen by the admin. Multiple providers can find their frequent itemsets using their customer purchased details.

The frequent item is finding by UP-Growth and UP-Growth+ algorithm for each Provider's items. Providers can find frequent itemsets in either monthly or daily basis. But there is a problem for finding frequent itemsets using this algorithm. UP-Growth generates only minimum utility itemsets and UP-Growth+ generates maximum utility itemsets. But there is difficulty for the provider to find frequent itemsets based on profit from the entire dataset. Our main aim is to view the frequent itemsets from the combination of minimum and maximum utility itemsets using GENETIC algorithm. Pre-processing is a process of filtering. Cleaning the data's in pre-processing helps to remove the unwanted data. By applying the ETL the unwanted candidate itemsets are get removed.

A pre-processing technique extraction transformation loading ETL is used for secure processing and maintaining the datasets efficiently. Extract, transform, and load (ETL) refers to a process in database. Remove data from outside sources and convert it to fit operational needs, and loads it into the end target i.e database. This process is very efficient and scalable[13].

Fig.3 Datasets with working of genetic algorithm

Fig 3 shows the datasets with working of genetic algorithm method gives out compact frequent item sets that is comparatively a good result when compared to already existing method. Hence thereby, increasing the efficiency by pruning the candidate item sets and therefore time and memory requirement has been reduced. It does not only reduce the number of candidates effectively but also outperform other algorithms substantially in terms of runtime, especially when databases contain lots of long transactions. The transaction database shows the entire dataset purchased by the customer with their name, product id, product name and product rate. The dataset also includes the purchase cost and count bought by the customer with the time of purchase. The customer can purchase their product through three different mode via debit, credit or using net banking through their card name and password.

RESULTS

The experimental task shows the result of individual provider high profit frequent itemsets from the large database. Here I am using genetic algorithm to find out the frequent items in provider login from the large transaction data stream according to the customer purchase details. Here we eradicate some problems. Namely it generates huge numbers of candidate itemsets reduces the extracting performance in conditions of time and memory requirement. By using UP-Growth and UP-Growth+ algorithm we can overcome this problem. Another drawback is UP-Growth generates minimum utility itemsets and UP-Growth+ generates maximum utility itemsets. Genetic algorithm is used to view the frequent itemsets from the combination of minimum and maximum utility itemsets. And time and memory is the main drawback in existing papers. But in this itemsets using utility pattern growth methodology the UP-Growth and Genetic algorithm minimize the time and memory problem. Fig 4 shows the final result of frequent itemsets.

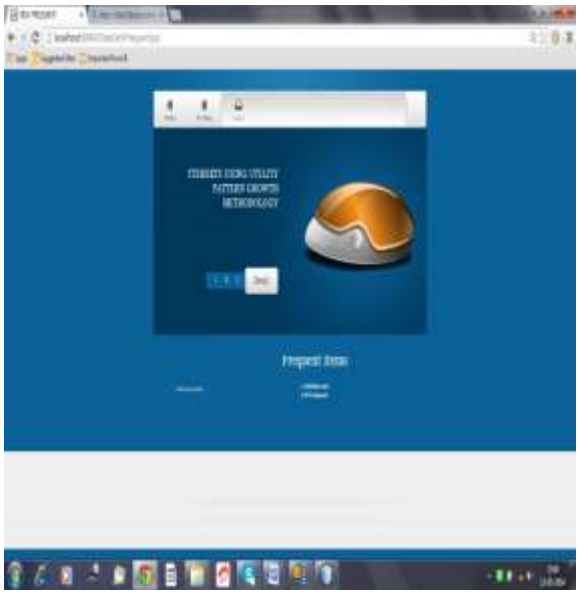

Fig. 4 Final result

Fig 5 shows the performance graph with transactions and time taken to find frequent items using genetic algorithm. The measure includes the Apriori computation time, and the time to identify frequent itemsets, as described. In apriori algorithm it generates more number of candidate itemsets and it takes more time for finding frequent itemsets. Thereby adding to the advantage of the proposed work, genetic algorithm reduces the number of candidate sets and time taken to find frequent itemsets from large transaction database get decreased.

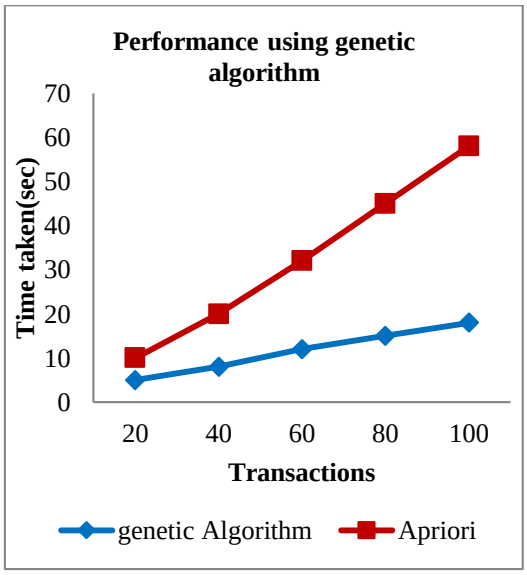

Fig. 5 Performance graph

CONCLUSION AND FUTURE WORK

In this paper, we have discussed various techniques used for finding frequent itemsets. Therefore utility pattern growth (UP-Growth) and UP-Growth+ method gives out compact frequent item sets that is comparatively a good result when compared to already existing method. . Even though it gives out compact frequent itemsets, UP-Growth generates minimum utility itemsets and UP-Growth+ generates maximum utility itemsets. But there is a difficulty for the provider to find frequent itemsets based on profit from the entire dataset. Hence thereby, increasing the efficiency by pruning tce he candidate item sets and therefore time and memory requirement has been reduced. Genetic algorithm is used to view the frequent itemsets from the combination of minimum and maximum utility itemsets. And the results show the clear output of high profit frequent itemsets from the large transaction database.

REFERENCE

[1] Vincent S. Tseng, Bai-En Shie, Cheng-Wei Wu, Philip S. Yu, Efficient Algorithm for Mining High Utility Itemsets from Transactional Databases,IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 8, pp. 1772-1786, Aug. 2013, doi:10.1109/TKDE.2012.59

[2] R. Agrawal and R. Srikant, Mining Sequential Patterns, Proc.11th Int'l Conf. Data Eng., pp. 3-14, Mar. 1995.

[3] J. Han, J. Pei, and Y. Yin, Mining Frequent Patterns without Candidate Generation, Proc. ACM-SIGMOD Int'l Conf. Management of Data, pp. 1-12, 2000.

[4] Mahesh S. Raisinghani, Laura Meade, and Lawrence L. Schkade(2007), Strategic e-Business Decision Analysis Using the Analytic Network Process, IEEE Transaction, on engineering management,vol.54,no.4.

[5] Tamir Tassa,(2013) Secure Mining of Association Rules in Horizontally Distributed Databases, IEEE Transaction on knowledge and data engineering.

[6] Beomjin Choi, T. S. Raghu, Ajay Vinze, and Kevin J. Dooley(2010), Process Model for e-Business Standards Development: A Case of ebXML Standards, IEEE Transaction, on geoscience and remote sensing, vol. 50, no. 5.

[7] Yu-Chiang Li, Jieh-Shan Yeh, Chin-Chen Chang(2008), Isolated items discarding strategy for discovering high utility itemsets, ROC Data & Knowledge Engineering 01/2008; DOI:10.1016/j.datak.2007.06.009.

[8] Hua-Fu Li, Hsin-Yun Huang, Yi-Cheng Chen, Yu-Jiun Liu, Suh-Yin Lee(2008), Fast and Memory Efficient Mining of High Utility Itemsets in Data Streams, Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008), December 15-19, 2008.

[9] Long Jin, Duck Jin Chai(2007), Mining Frequent Itemsets over Data Streams with Multiple Time-Sensitive Sliding Windows, ALPIT '07 Proceedings of the Sixth International Conference on Advanced Language Processing and Web Information Technology (ALPIT 2007).

[10] R. Vijaya Prakash, Dr. Govardhan(2011), Mining Frequent Itemsets from Large Data Sets using Genetic Algorithms, IJCA Special Issue on "Artificial Intelligence Techniques - Novel Approaches & Practical Applications"AIT, 2011.
[11] M.-S. Chen, J.-S. Park, and P.S. Yu, Efficient Data Mining for Path Traversal Patterns, IEEE Transaction on Knowledge and Data Eng., vol. 10, no. 2, pp. 209-221, Mar. 1998.
[12] www.ijcem.org/papers42011/42011_14.pdf.
[13] http://en.wikipedia.org/wiki/Extract,_transform,_load.