Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 10, No. 1, October 2021

AN ASSOCIATION RULE ALGORITHM FOR ONLINE E-COMMERCE RECOMMENDATION SERVICE

#1RAVULA HARITHA, ***Research Scholar,***
#2Dr. G.THIPPANNA, ***Supervisor,***
#3Dr. NALLA SRINIVAS, Co-***Supervisor,***
Department of Computer Science and Engineering,
NIILM UNIVERSITY, KAITHAL, HARYANA, INDIA.

ABSTRACT: In this study, researchers introduce an innovative association rule method that online store owners can implement to increase revenue and provide consumers with the most suitable product options. Our primary objective is to assist these managers in developing effective e-commerce advice instruments. By increasing profits, this approach will successfully accomplish both objectives and provide consumers with enhanced knowledge. This page contains a wealth of information regarding the formula of principles that underpins the association of profit. To ascertain association principles, a minimal support comprising of multiple minimum supports and a distinct profit is assigned to each item. Multiple studies have demonstrated that these improvement strategies significantly accelerate the entire process.

Index Terms*—Association rules, data mining, information extraction, recommendation system.*

1. INTRODUCTION

Association rule mining is being utilized to improve compatibility in a growing variety of business solutions. Their primary objectives have been to increase income and provide excellent customer service. As a result, they've worked really hard to achieve both goals. Decision makers employ these mining approaches to increase the accuracy of their investment cost and return forecasts. Despite these constraints, conventional association rule algorithms continue to have trouble working. As a result, classic association rule algorithms exclusively use sales data. This implies that attempting to obtain possible goods in small numbers serves no purpose. Users of recommendation systems, on the other hand, have access to data that remains consistent throughout multiple transactions. These programmes may make it more difficult to obtain money.

Managers continue to believe that using information processing techniques will improve organizational performance by lowering costs, increasing revenues, and enhancing customer service, despite advances in information systems designed to enable market transactions. Association rule mining-based commerce recommendation systems can significantly reduce market managers' workloads, saving them time and effort.

The literature on data mining provides detailed explanations of a number of algorithms designed specifically for association rule mining. Nonetheless, they lack the necessary expertise to provide commercial advice to others. Using client purchase records, association rule mining efficiently tackles the challenge of analyzing shopping baskets in commercial contexts.

However, there is an issue with the association rule algorithms now in use for market suggestions. Apriori prioritizes sales volume over profit, as do the vast majority of rule-mining algorithms. Our data mining methods revealed a hitherto unknown association between sales volume and profit. The algorithm being discussed is called the profit-support association rule.

2. RELATED WORK

In order to find any rules with a reliability higher than the user-specified minimal confidence (minconf) and minimum support

(minsupp) values, association rule mining examines a transaction database T. Link mining techniques usually involve two distinct phases.
1) Achieving a significant production volume is required to meet the demand for reasonably priced items.
2) When creating association rules, use large sets of things, but proceed with caution.
A great deal of research has been done on association rule mining. The two studies mostly agreed on the process for developing connection rules that included user limitations, despite using different models (minsupp and minconf).
In order to extract data, association norms must be put into practice. Market basket analysis has been used for a longer period due to its credibility and reliability. Customers' purchases are categorized by the technology into "item sets" in an attempt to identify meaningful patterns like item correlations. The associations formed by association principles between two unconnected entities, A and B. The purpose of support and assurance is to strengthen a rule. To compute the confidence ratio, divide the total of transactions that include A by the total of transactions that include B. On the other hand, the support ratio indicates the portion of all transactions that contain both A and B. An analysis can be made on the buying behaviors of individuals within the same demographic as "Beer" on the purchase of "DVD Player" and "DVD Disk." An explicit demonstration of the association between "DVD Player" and "DVD Disk" may be observed in beer.

3. PROBLEM ANALYSIS

M is an important factor in the application of association rule mining. The two major aims are to reduce the number of rules that must be created and to improve the usefulness of the search area. In contrast, outdated algorithms only used a single minimum support number. The research is based on the assumption that all records in the database are identical.
1) The profit per item remains unchanged.
2) The database contains identical entries, implying a possible relationship between them.
This type of behavior is improper in the real world. Even if they do not initially sell as many as required, there may be times when particular commodities yield more money than expected. This could be owing to the much higher pricing. A CD player, for example, costs significantly more than CDs. Two problems exist when employing common association rule mining approaches, such as the Apriori algorithm: Initially, our goal was to develop a limited number of concepts for the least amount of money. Furthermore, components with a high profit potential but low value are eliminated from the first iteration of the Apriori algorithm, which provides a list of unique things.
For example, setting the minimum support (minsupp) at 2% allows for the use of two more components.

1. The DVD player fails 0.5% of the time, whereas the DVD disc fails 3% of the time.
2. According to Rule B, there is a 4% chance of someone stealing beer and 3% chance of someone taking DVDs.

There's no denying that Rule A will increase earnings. In contrast, Rule B can only be found using the processes described above. To make a one-itemset, numerous elements from the first version must be deleted. Prioritizing finances above the quantity of objects is critical while performing association rule mining. This is related to the fact that data mining aims to increase earnings. Furthermore, the dataset contains a few elements that are common in certain apps but not in others. When the price of an item varies in two different directions, two separate issues arise. When the minimum support values are very high, it becomes more difficult to find patterns in the data that are associated with objects that are only encountered sometimes or never. Furthermore, the minimum supply values must be extremely low in order to establish the rules that regulate both common and unique entities. Many connections between everyday things will be possible, the most of which will be pointless. This could result in the emergence of multiple complex and

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 10, No. 1, October 2021

different regulations. The "rare item problem" might also be stated in these words. In these circumstances, application specialists categorize the data based on how frequently each item appears. The association rules are then derived from each segment using various levels of basic assistance. To bring attention to an abstract subject, integrate a range of unusual components into a single work. The first strategy was deemed inadequate since it was difficult to discern which restrictions applied to products from different geographical locations. Similarly, the second technique does not provide clear limits for both regular and unusual items. Both approaches are undeniably tedious and ineffectual in reaching their objectives. One technique to addressing this issue is to utilize an algorithm to generate connection rules from various minimum supports. This project seeks to create a mechanism for mining association rules with N-itemsets.

4. ALGORITHM DESCRIPTION

Priority is the basic strategy behind this method. This study differs from earlier studies in that it uses profit as an evaluative criteria for each object. Minimal support was found by using specialized procedures to compute item counts or percentages. As a result, we use a unique benefit metric to detect unsupported areas of our algorithm before completing the procedure. Then large collections of products with several extra supports are purchased.

A. Generation of 1-Itemset

Data mining systems used frequency of occurrence to estimate the significance of transaction database items. As previously stated, the data collection procedure is founded on two essential assumptions. The first is that no variables change. The profitability and frequency of each object in the database are identical. However, a different approach is taken with relation to material goods. Any manager in charge of a retail or internet-based organization must first and foremost ensure profitability.

Financial gain was our primary goal for studying association limits. Assembling an itemset is intended to reduce overall profit. The final stage is to compare item profits to the smallest total profit. Organizations with a low total profit despite a large sales volume are less competitive. Items are disposed of when their revenue falls short of the bare minimum requirement. For example, Item 4 is no longer listed in Table 1.

Table I: Example of Setting Minimum Support For Item

	Item 1	Item 2	Item 3	Item 4
Amount of items	4	40	150	150
Unit profit	100$	2.5$	0.5$	0.5$
Total profit	400$	100$	75$	75$
Minimum total	100$			

B. Minimum Support for Each Item Using Profit

During this phase, the N-itemset is created by computing the minimum support for each item, known as the Profit-support Amount, using Unique Profit. Previous level-based mining systems required extracting a single minimum support value after generating a collection of one item. It is assumed that the quantity of each object is equal. This clearly does not reflect the current condition of circumstances. The most obvious and practical factor in judging the relevance of a specific item is its profitability. Hence, profit should be viewed as the benchmark or criterion. The quantity of items, on the other hand, is directly proportional to the variety of income sources. Despite its tiny quantity, beer is a popular item in the market basket research for supper, with a high profit margin. Unlike alcohol, the DVD player is a great product with a huge profit margin. One specific criterion we employ is to determine the quantity of things that generate a profit from this standpoint. Simply said, the profit margin is calculated in relation to the number of items. When reassessed based on profitability, the item with a high amount but a low profit margin faces a proportional decrease in quantity. Furthermore, when a small product generates a large profit, its quantity increases. Another advantage of reevaluating the quantity is that it enables the continued use of the level-wise Apriori approach for extracting association rules, which uses the number of items as a threshold. This algorithm shares similarities with prior techniques. However, this method is motivated by a desire for financial benefit. Prior to delving into the precise specifics, the following explanation is provided:

Definition 1 (Appointed Profit) The profit-support quantity for each item is determined based on the user's set profit value.

Table I is a computational representation of an item's minimum support. It is often set to an integer number that equals or exceeds the maximum profit per unit item. The bulk of the ultimate profit-support amount can be expressed as an integer in this arrangement, making computation preparation easier.

Definition 2 (Minimum Profit Support (MPS)) The definition follows the previous one (MPS: Minimum Profit Support).

We want to see how altering the frequency affects profitability versus requiring the same minimum support for all items. The MPS is calculated with

$$MPS = \frac{Appointed\ profit}{Unit\ profit} \qquad (1)$$

For example, in Table 1, minimum profit support of Item2 is calculated as follow, suppose Appointed Profit is 100

$$MPS(Item2) = \frac{Appointed\ profit}{Unit\ profit} = \frac{100}{2.5} = 40 \qquad (2)$$

To make \$100 from item 2, 40 units must be sold. The same method is used to determine the minimum profit assistance for Item L. The outcome is one. As a result, selling a single item can yield a profit of \$100. To make the same amount of profit, different quantities of Items 1 and 2 must be sold. This simple example demonstrates the enhanced logic of the profit-support algorithm.

C. Mine of Large Itemsets

As stated in reference the process of discovering vast collections of objects with MPS involves adjusting MPS for each individual item. The proposed method, like the Apriori algorithm, employs level-wise searching. Each full item collection is created through multiple cycles of data processing. The system calculates the amount of supporters for each item and determines if it will be a primary emphasis during the initial phase. The seed set for each iteration serves as the initial reference for item sets recognized as relevant in the previous cycle. Using this initial set of data, candidate itemsets are created, which can include vast collections of items. The current supports for these future item sets are calculated as the data traverses. At the end of the pass, it determines whether the candidate item sets are genuinely large. Below, we will look over the second phase in further detail and find one important oddity. The proposed solution relies heavily on sorting the list entries in ascending order of MPS. Every subsequent action of the algorithm strictly adheres to this ordering. The objects in each set are organized in the same way. Table I has four entries. Items 1 through 4 have MPS values of 10%, 20%, 4%, and 4%, respectively. Finally, the pieces might be put in the following sequence: three, four, one, two.

5. EXPERIMENTAL RESULTS

In this section, we conducted several experiments to assess the effectiveness and practicality of the association rule mining algorithm (PARMA) that was proposed for usage in an e-commerce environment. Our findings show that the proposed approach does not generate an excessive number of meaningless rules relating to popular goods. Instead, it enables us to identify rules that require little support, even for uncommon products.

Our experiment used the IBM synthetic data generator to generate a dataset that met the given conditions. The dataset consists of 10,000 transactions, each with an average of ten items. The dataset contains a total of 1,000 unique objects. Furthermore, each frequent item set has four items. Furthermore, the profit margins for each individual item have been computed as

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 50, Issue 10, No. 1, October 2021

follows: 10% of the products have a profit margin of $5 to $10, with the remaining 10% having a profit margin of $0.1 to $1. The remaining 80% of products have a profit margin ranging between $1 and $5. This illustration of the normal distribution is more straightforward. The precise profit for each item is derived by randomly selecting from that item's individual profits. The calculation for each profit is as follows: Ten percent of the items make a significant profit ranging from $5 to $10, another ten percent generate a small profit ranging from $0.1 to $1, and the remaining eighty percent generate a moderate profit ranging between $1 and $5. This representation of the normal distribution is simpler. The precise profit for each article is estimated using a random selection from the chosen profit range. We only examined sales of individual units for each item. We investigated the efficacy of the proposed method in comparison to MMSapriori, a strategy for mining association rules based on multiple minimum support thresholds. Both systems frequently rely on a range of minimum supports. Our proposed method prioritizes the profit of each item as the optimization target, as opposed to MMSapriori, which just considers the quantity of items. MMSapriori is undoubtedly an important component of the suggested strategy. This indicates that N-itemsets (where N > 2) with numerous minimal supports can be calculated in a single step. Simply put, the only difference is the mechanism for creating a certain group of pieces. This includes a comparison of the two algorithms, as seen in Tables II and III.

TABLE II: Number of large itemsets found. Ls stands for the lowest support

	MMSapriori			PARMA		
LS	a=2	a=10	a=20	a=2	a=10	a=20
0.1%	5140	25030	29570	5070	24000	28470
0.2%	4920	12850	13100	4750	11250	12020
0.3%	2400	5040	5910	2300	4980	5900

TABLE III: Number of Candidate Itemsets

	MMSapriori			PARMA		
LS	a=2	a=10	a=20	a=2	a=10	a=20
0.1%	326520	356840	402560	345450	369510	422100
0.2%	269540	275100	290110	285640	285000	310050
0.3%	225460	235420	241050	235010	250550	259840

Based on the data presented in Table II, we discovered that the results produced by our proposed PARMA algorithm are consistently inferior to those produced by MMSapriori. This is due to PARMA's capacity to delete the majority of 1-items that generate lesser revenues on the first database traversal. Furthermore, MMSapriori creates larger item sets than PARMA, as seen in Table III. On the other side, PARMA generates a greater number of possible item sets. This shows that PARMA is more effective than MMS. To eliminate most things of little or no value, PARMA generates a one-itemset at the start of the process that includes the item with the lowest total profit value. Furthermore, sets of N objects are made for profit. This phase essentially eliminates the lowest-income item sets.

6. CONCLUSION

The profit-support association rule improves user efficiency and allows online shops to raise their revenue and sales. The first portion of this study introduces a revolutionary profit measure that assures each item has a minimum support value. Subsequently, we combine this approach with the traditional way of obtaining association rules from a large set of fundamental supports. This boosts the rules' usability and utility. Finally, this approach for identifying successful link strategies is used in the field of e-commerce.

We discovered many difficulties with our Profit-support Association Rule method that need to be addressed. Tasks such as finding the lowest total profit, increasing efficiency, and shortening the duration of the operation provide substantial hurdles. In order to effectively meet the upcoming problems, we must adjust our current plan.

REFERENCES

1. S. C. Young, C. M. Song, and H. R. Keun, “Mining association Rules using RFM scoring method for personalized u-commerce recommendation system in emerging data,” in Proc. the 2nd Communications in Computer and Information Science, pp. 190-198, 2012.
2. J. Dongre, G. L. Prajapati, and S. V. Tokekar, “The role of Apriori algorithm for finding the association rules in data mining,” in Proc. the 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India: IEEE, pp. 657-660, 2011.
3. Y. L. Chen and C. H. Weng, “Mining fuzzy association rules from questionnaire data,” Knowledge-Based Systems, vol. 22, pp. 46-56, Jan. 2009.
4. R. Belohlavek, J. Dvorak, and J. Outrata, “Fast factorization by similarity in formal concept analysis of data with fuzzy attributes,” Journal of Computer and System Sciences, vol. 73, pp. 1012-1022, May. 2007.
5. Y. S. Chen and C. H. Cheng, “A soft-computing based rough sets classifier for classifying Ipo returns in the financial markets,” Application of Soft Computation, vol. 12, pp. 462-475, Apr. 2012.
6. X. H. Li and S. Jin, “An improvement of the new Apriori algorithm,” Computer Science, vol. 34, pp. 196-198, Apr. 2007.
7. G. F. Kuang and Y. C. Li, “Using Fuzzy Association Rules to Design E-commerce Personalized Recommendation System,” Indonesian Journal of Electrical Engineering, vol. 12, pp. 1519 -1527, Feb. 2014.
8. A. A. Kardan and M. Ebrahimi, “A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups,” Information Sciences, vol. 219, pp. 93-110, Oct. 2013.
9. S. X. Bai and S. L. Duan, “Association rule mining algorithm DB-growth based on relational database,” International Journal of Granular Computing, Rough Sets and Intelligent Systems, vol. 4, pp. 105-121, Jan. 2015.
10. H. Fan and K. Ramamohanarao, “Fast discovery and the generalization of strong jumping emerging patterns for building compact and accurate classifiers,” IEEE Transaction of Knowledge of Data Engineering, vol. 18, pp. 721-737, Aug. 2006.
11. M. Garcia-Borroto, J. Martinez-Trinidad, and J. A. Carrasco-Ochoa, “A survey of emerging patterns for supervised classification,” Artificial Intelligence Review, vol. 42, pp. 705-721, Aug. 2014.
12. E. R. Omiecinski, “Alternative interest measures for mining associations in databases,” IEEE Transaction of Knowledge Data Engineering, vol. 15, pp. 57-69, Aug. 2003.