ISSN 2395-1621

Intelligent Customer Analysis System Using Sentiment Analysis by Apriori with Multithreading and SVM

#1Avani Joshi, #2Supriya Gawali, #3Priyanka Khandage, #4Akshata Jagnade, #5Prof. Nisha Kimmatkar

1joshiavani23@gmail.com,
2supriya.gawali@gmail.com,
3priyakhandge10@gmail.com,
4akshatajagnade@gmail.com,
5nishakimmatkar@gmail.com

#1234Student, Computer Department,
#5Professor, Computer Department

JSPM's Rajarshi Shahu College of Engineering
Tathawade, Pune, Maharashtra, India.

ABSTRACT

In the past few years, the term shopping has been completely redefined. Masses have adopted shopping through malls, shopping districts and also online shopping websites. This creates a need for analysis of the buying behavior of people so that the shopping experience can be enriched and also the distributors/mall owners gain profit. Lucrative discounts and combo offers have always attracted shoppers. In this approach, we provide a new way of analysis of product sale and then offer calculated discounts. By using frequent item-set mining on customer transactions, we identify product buying patterns. Then products alongside the higher profit margin are given discounts, so that we increase the likelihood of such products being bought. Customer reviews regarding the shopping experiences are collected, so that sentimental analysis reports can be generated to get the customer satisfaction level.

ARTICLE INFO

Article History

Received: 17th April 2017

Received in revised form : 17th April 2017

Accepted: 19th April 2017

Published online : 22nd April 2017

Keywords: Apriori, Multithreading, Discount Generation.

I. INTRODUCTION

The e-commerce or shopping mall owners need to keep a large inventory of items for availability to the masses. There is a need to identify buying behavior of customers, so that inventory is managed according to products that are more likely to be bought. Also, the behavior can be used to recommend discounts on certain products. As, customer relationship is important, so sentimental analysis on the customer reviews is also necessary.

Online or retail stores have become a rising affair in the recent past. Everyday hundreds of people buy products, avail interesting discounts. The discounts are largely set based on stocks, inventory or popularity of a product.

In this paper, we have proposed a novel idea to recommend smart discounts based on analysing the buying behavior of the people of a particular store. It may happen that a supply chain has declared discount on say 5 product. But not all products are popular in demand in all stores. It may happen that some stores sell more of a particular product. Our system will recognize these trends in buying of local crowd and then recommend a smart discount scheme so that store owners maximize their chances of selling more products.

We are using a multithreaded approach for calculating frequent item sets and generating association rules. This provides a rich and fast experience for the long list of transactions in the

database. The web interface will be user friendly and will have charts/graphs and a list of products indicating the discount recommendation.

Sentimental analysis is a new field in data/information retrieval which helps one to find out whether a particular text has a positive or negative sentiment. There are various ways to calculate this but the most effective way is to use SVM method. We will sentimentally analyse the comments or reviews given by customers and hopefully maintain a healthy customer-seller relationship.

II. BACKGROUND WORK ON ITEMSET MINING

In paper [1], a new closed frequent item sets mining algorithm based on GPU, a faster and efficient way of calculating frequent item sets is discussed. From this paper, we have picked up the idea of calculating frequencies of items in transactions in a parallel fashion. Instead of GPU we will go on with multithreaded CPU environment because it will reduce dependency of hardware.

In paper [2], an improved version of the frequent item set mining algorithm, a new innovative way of generating less number of candidate item set and faster way using multithreading/multi node system is discussed. In this project, we are using the algorithm from this paper to calculate frequent item sets in least amount of time. This improves response time of system and helps in better user acceptance.

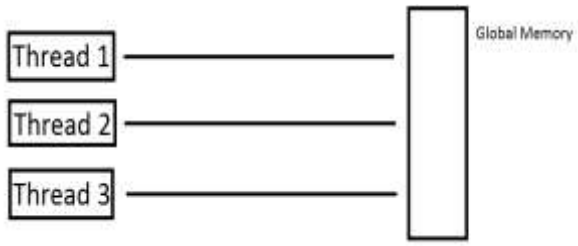
In paper [3], Affective-feature-based Sentiment Analysis using SVM Classifier, we see that the textual content is parsed and classified in positive and negative features. We are using this part of SVM sentiment classification to achieve the classification of user comments/reviews in either good/bad sentiment category.

III. FAST CANDIDATE ITEMSET GENERATION

From the paper [2], we have decided to use a new way of generating candidate itemsets. This new way, uses multiple threads (nodes) to generate distinct lists of candidate itemsets. Then every thread calculates a final list which surpass the threshold value and then forward the list to the global memory. This method reduces the number of candidate itemsets and also splits the task of checking against threshold between threads, so it is much faster than traditional approaches.

Along with its generalization, this algorithm for association rule discovery was designed to be used in parallel and distributed environments. The improvements made to the core formulas have a substantial impact on the overall performance of the algorithm, by reducing to a bare minimum the candidate generation across the entire chain of processing nodes, without missing any potential valid candidates.

These modifications make an exclusive use threads to compute candidate itemsets of different cardinality at once. This greatly reduces the effort needed to calculate candidate itemsets and increases the performance.



In the above diagram, Thread 1 will calculate candidate items of cardinality 2, Thread 2 will calculate candidate items of cardinality 3 and Thread 3 will calculate candidate items of cardinality 4 and so on.

IV. DISCOUNT RECOMMENDATION

Considering the interest of the customer to lie in discounts, our approach is to suggest intelligent discounts on product combos by analysing buying patterns.

For the benefit of both retailers and buyers, we have come up with a technique to recommend discounts in a combination of popular and non-popular products.

Firstly, we read all the transactions and generate statistics which will give the list of popular products and non-popular products which are often bought together.

We then provide the customer with the discounts on popular product if a non-popular product is also bought such that all the products are frequently bought together.

For example, if {X,Y,Z} are a set of products which are frequently bought together but then X and Y are very popular but Z is not so popular. So we offer some discounts on X and Y if Z is also bought together.

But, the recommended scheme may or may not be friendly to the customers which will affect the sales. So we have developed a questionnaire, by which we can judge the sentiment of the customers regarding the discount scheme applied. We have used WEKA Java API to build and train an SVM model so that we can perform sentimental analysis and then the store owners/admins can change the discounted products as per the reviews received.

V. CONCLUSION

As a result of multithreading itemset mining can be done in very quick time. We have implemented multithreaded frequency count, candidate itemset generation and association rule generation. The discount recommendation takes place by using a combination of frequency counts and association rules. Further by using sentimental analysis by SVM, we can iterate over the discount schemes and keep making it better.

REFRENCES

[1] Yun Li, Jie Xu Ling Chen, 2015. “A new closed frequent itemsets mining algorithm based on GPU”, in IEEE Conference 2015. IEEE 2015 978-1-4673-8537-4/15.

[2] Cristian Nicolae Butincu, Mitica Craus, 2015. “An improved version of the frequent itemset mining algorithm”. *In IEEE Conference 2015*. IEEE 2015 978-1-4673-8180-2/15.

[3] Fang Luo, Cheng Li, Zehui Cao, 2016. “Affective-feature-based Sentiment Analysis using SVM Classifier”. *IEEE 2015* 978-1-5090-1/16.

