

Weighted Based Frequent and Infrequent Pattern Mining Model for Real-time E-Commerce Databases

Sujatha Kamepalli^{1*}, Srinivasa Rao Bandaru²

¹ Department of Information Technology, VFSTR Deemed to be University, Vadlamudi Village, Guntur District, Andhra Pradesh 522213, India

² Department of Management Studies, VFSTR Deemed to be University, Vadlamudi Village, Guntur District, Andhra Pradesh 522213, India

Corresponding Author Email: sujatha101012@gmail.com

https://doi.org/10.18280/ama_b.622-404

ABSTRACT

Received: 10 October 2018

Accepted: 24 March 2019

Keywords:

data mining, frequent pattern mining, infrequent pattern mining, e-commerce

In modern system e-commerce is developing in fast and it makes the availability of resources and services on the internet colorful. In today's e-commerce world day by day the data is increasing tremendously and this data should be used effectively. Data mining techniques produce useful knowledge for decision makers from high dimensional databases. Association rule mining is a used in e-commerce data analysis to realize cross selling and patterns generated can be used as recommendation system. Numerous models have been studied in both frequent as well as infrequent pattern mining in marketing applications which have some unsolved issues yet. A novel weighted based frequent and infrequent pattern mining model for real time e-commerce databases is proposed to find weighted based frequent and infrequent patterns from large data bases. Here weighted infrequent ranking measure is used to filter the infrequent product from the frequent associations. In this model a real-time e-commerce application is designed for pattern extraction process. This model is implemented in Java on real time e-commerce database (flip cart database). This model generates weighted based frequent and infrequent patterns based on user selected feature product in e-commerce database. This model is also implemented on distributed market database (training database), cloud database and medical database.

1. INTRODUCTION

Data, the collection of raw facts, is considered as basic form of information. This data has to undergo a number of steps in order to create knowledge. The steps are collection, management, mining and interpretation [1]. In today's e-commerce world day by day the data is increasing tremendously and this data should be used effectively. Collected data must be systematically processed and arranged properly by various techniques and tools as databases and data warehouses. Data mining tools were applied on that data for discovering hidden knowledge. As the size of data increases, it creates many problems in detecting the hidden patterns among various attributes. The discovered patterns are so important in e-commerce system in order to make decisions and e-commerce system provides a best work bench for data mining [2-4]. Cross selling analysis is a marketing method which is helpful in improving customer value, for maintaining more relations between the enterprise and the customer, and it demonstrates that the more dependent the customer will be on the enterprise, the higher the loyalty will be [5].

Association pattern mining is one of the data mining processes used in e-commerce. It examines the buying habits of various customers and helps in deciding the layout of the products in the virtual store and how promotions are to be bundled. It can also be used to improve the sales of different products by suggesting additional products that are related to

the purchased products for the customer. In addition, in e-commerce, latest products have to be launched regularly while existing ones are excited to assure the increasing demanding needs of customers. From the discussion it is clear that the items in database are changed often. Unfortunately, existing algorithms assume that the set of unique items is fixed and hence, each time items are added or removed, the algorithms must discard valuable past mined results and re-mine the database. Finally, because the data in e-commerce is far too dynamic and volatile, there is no good way to decide on a suitable support threshold for the mining process. Using too high a threshold may result in too many useless rules while too low a threshold may result in certain important rules being passed over. Therefore, the database must be mined with several different support thresholds before an optimal threshold can be determined [6].

2. LITERATURE REVIEW

2.1 FP-Growth* algorithm

FP-Growth* algorithm basically implemented the data structure of traditional FP-Tree technique and merged it with array-based approaches. This will produce numbers of optimization approaches. Array based approaches are responsible to decrease the overall traversal time of the

conventional FP-Tree model. Additionally, the proposed algorithm also decreases memory consumption rate of FP-Growth Algorithm. The original FP-Growth* algorithm is modified and enhanced to form a new algorithm which is called as Broglet's FP-Growth algorithm [7].

2.2 IFP-min algorithm

Generally, all items presented in the processed database are frequent in nature. The proposed approach detects data access into two separate sets, those are residual database and projected database. Initially, the approach is implemented in case of residual database. If there is only a single item available in database, the item set must contain only that single item. After implementing in residual database, the above algorithm is now implemented in projected database. The 1f-item is appended as prefix to the item sets of projected databases [8]. MIIs are generated from the projected database by implementing the above algorithm recursively. Both the MIIs produced from residual and projected database are analyzed and compared with each other. When an item set is detected in the second set of MIIs, then it is not reported. The LF-item is added with the other terms of item set and produced as an MII of the original database.

2.3 Weighted association mining algorithm

A transaction tree is almost equivalent to FP-tree, except having header table or node link. A transaction tree can be defined as the overall representation of all the transactions in a particular database. Every individual node of the tree is represented with an id and a counter. ID is responsible for representing an item, whereas counter contains records about the number of transactions involving that particular item [9]. The algorithm also provides a way to compress the transactions of every item set into continuous intervals. In order to achieve this, a mapping function is implemented to map transaction ids into a different transaction tree. The overall performance of transaction mapping algorithm is far better than that of all previously developed algorithms.

2.4 Frequent pattern mining and infrequent pattern mining models

Pudi and Haritsa [10] proposed a technique, has many applications that are web usage mining, intrusion detection system and, recommender system. It is also implemented in market basket analysis in order to detect frequent patterns as well as to generate number of association rules. It deals with the market basket analysis to find frequent patterns and generate association rules. In the past decade, a large number of frequent and infrequent mining models have been proposed, that include: [8-10, 11] proposed rule ordering strategies and infrequent mining using weighted measures. These models are applicable to classification-based association mining. The main limitations to these models are static weight measure, which are applicable to limited data size. They also proposed classification based on predictive rule mining model which combines both the association mining and classification model for pattern analysis. The main problems identified in this model are incorrect class rules and class imbalance.

McConnell and Skillicorn [12] proposed an efficient high utility sequential pattern miner for frequent and infrequent rules. UPSpan is the model developed to find frequent and

infrequent rules on the sequential databases. This model fails to improve the performance under large databases with lack of pruning strategies. Albashiri [13] proposed an infrequent weighted item set mining using FP-growth model. Two novel measures for weighted item set measure and IWI (infrequent weighted item sets) are implemented to discover frequent and infrequent patterns. This model requires domain expert knowledge for patterns analysis and lack of candidate pruning strategies. Implemented coherent based frequent mining model with support and confidence measures [14-18]. In this model, a large number of patterns are discovered but missed some interesting infrequent patterns and the pattern's quality.

(1) Infrequent Weighted Item set and Minimal Infrequent Weighted Item set mining is evaluated through maximum IWI-support-min threshold.

(2) Infrequent Weighted Item set and Minimal Infrequent Weighted Item set mining can be calculated through maximum IWI-support-max threshold.

2.5 Existed models in E-commerce

E-commerce sites focus their customers to write feedbacks on the products. It shows the customer's opinions on different features of products. A feature is known and it can be defined as an attribute of a particular product. Product recommended systems were implemented in different e-commerce sites such as flip kart, eBay, and Amazon, in these sites product rankings were calculated based on user and seller reviews. These tremendous volumes of reviews provided by the consumers have important knowledge which is very important both for consumers and organizations. Each individual product can have many features. It may happen in many cases that, a particular feature can have more importance and significance than other features of the same product. It has a remarkable effect on consumers' decision making along with organizations' product development methods. Therefore, detection of vital product features plays a prime role in the process of enhancement of usability reviews. In accordance with consumers' perspective, consumers can be able to make purchase decision by considering the most vital features. At the same time, organizations emphasize on advancement of product quality. Detecting these important product features manually is quite impossible and impractical till date. Hence, auto-determination of these features is required. Many researchers carried out research and proposed many number of product ranking models in recent past.

The whole process of product ranking is divided into sub-phases: feature identification, pre-processing and ranking. First of all, product features are detected out of reviews. Then these reviews are used for the computation of opinions with the help of a pre-processing algorithm. Later, ranking algorithm is implemented in order to rank the products which depend on feature frequencies and consumers' reviews. Product ranking models are the core part of product search engine systems. To maximize the users' interest when using e-commerce search engine, the product ranking model should succeed in ranking the products that users really want to purchase at top positions. In general, e-commerce sites rank top k-products as user recommended products using the product rate or user navigation behavior for product ranking. Also, most of the product search engines today are designed based on the product rating or product reviews using classical similarity measures. However, the process of locating e-commerce products that users really desired to buy is quite

different from that of extracting products using product rate or similarity measures.

In online e-commerce applications, business to business sites such as Flip kart, eBay and Amazon, people can easily configure an online store to purchase or sell items. These e-commerce sites allow users to share their product review or shopping experience via feedback form or reviews on the selected products to improve the e-commerce business. Some of the useful information or attributes on Amazon reviews are labeled as shown in Figure 1 and Figure 2 for Amazon and Flip kart sites respectively [19].



Figure 1. Product Specifications, Review Rate, and Review Count of Moto G Plus in Amazon

Since there is exponential increment in product features and information, it becomes difficult to analyze the product reviews and giving ranks to the products based on its features.



Figure 2. Specifications, review rate, and review count of a product of flip kart

For online retailers, the recommender system helps in expanding online e-commerce and to compete the market research and compare each product with alternative products. There are some research works that describe models to find reviews and phrases from text comments [20].

Ashkan and Clarke [21] presented a product ranking approach for online e-commerce system. This method uses the domain knowledge from previous transactions made by the users. They considered the similarity index of various products for their approach and this similarity index influences the final outcome greatly. Both product-product and user-product relations handled with this ranking algorithm. It outputs vital recommendations for the consumers. Many e-commerce websites implement this ranking scheme in order to generate

real-time recommendations. It also outputs personalized search results for its users. The researchers proved through validation that, this method requires very less amount of time as compared to the other techniques. In future, this work can be optimized on live e-commerce recommendations and provide personalized searching results for customers.

2.6 Product ranking measures for association rule mining

Rong et al. [22] developed a new ranking model named as G_Rank. Tripartite Preference Graph (TPG) topology is used here in order to find the relationships among product preferences and users as shown in Figure 3.

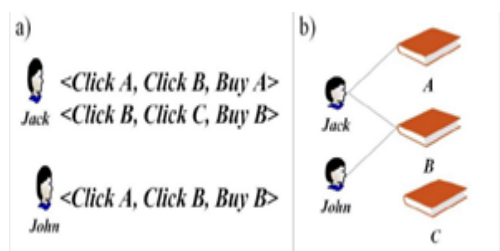


Figure 3. A Topology of Tripartite Preference Graph (TPG)

Paraschaki et al. [23] analyzed and performed a comparative study of top-n recommenders in e-commerce platform. Here, they studied different advanced machine learning approaches to implement in e-commerce platform for the purpose of product ranking. E-commerce platforms implement simple methods for recommendation. They used KNN collaborative filtering and association rule mining models as product recommended system on binary purchase data. A time-oriented splitting mechanism was suggested by them which are responsible for detecting optimal time span on training datasets. Significantly enhanced recommendation accuracy and runtime efficiency can be achieved through this approach. Li et al. [24] implemented a product feature ranking model. They used linear regression and rule-based models on product features as overall product ratings. Future researches are needed to extend this technique to gain more effective feature ranking.

Association patterns in the e-commerce framework tried to enhance product search by best-selling prediction in e-commerce. In order to achieve this goal a new framework was developed for dynamic best-selling in prediction. For the prediction of dynamic best-selling, an algorithm was designed (For example, volume of sales in all items on a particular transaction). The authors suggested a new ranking scheme for best-selling prediction with relevance. The products having highest probability to be purchased and relevant to consumers' need are marked with higher rank. In the future, this work can be extended to real-time product features other than customer's best-selling rates. Arun et.al introduced a new method for computing weighted product features and opinion strength. This approach is very significant and effective in case of online shopping. Further research may be carried out to extend this work and gain optimum performance and accuracy. Flip kart and Amazon, two of the most successful e-commerce sites, pioneered the use of product review systems by integrating user feedback data along with product price and product rate. However, have implemented that the eBay review framework is likely to mislead e-commerce users due to the lack of product features and review information. In contrast to eBay,

the Flip kart and Amazon review system integrates users' rating scores by summing or average of all similar type of product categories. However, this model doesn't consider the product features of the reviews and review rates which are highly skewed towards the product rate and price. In the past if any customer wanted to buy any product, item, services he had to go different stores to compare various prices and offers provides by different stores on different products of various brands. It is very time consuming and costly especially for that product that is not in neighborhood location of customer and is a new arrival in market. Another difficulty with traditional method is also not suitable for old age people or those people who live in remote area from main markets & in extreme weather condition. Thus e-commerce comes into picture. As discussed earlier static recommendation and personalized recommendation are the two types for e-commerce recommendations. Static recommendation shows same product recommendations to all users it is based on previous static transactional database. A personalized recommendation is show different recommendation for different particular user. It uses customer profile, goods commodity relationship, hobby, preference of user, locality, interest, budget, history of customers buying pattern to give personalized recommendation.

3. MATHEMATICAL MODEL

3.1 TF-IDF computation

TF-IDF is a statistical measure in information retrieval.

$$w_i = \text{find}_{tfidf}(D, t_i) = (1 + \log(f(t, D))). \log(N/n(t))$$

where, $f(t, D)$ =probability of the item t in D , N is the total number of items and $n(t)$ is the number of terms.

3.2 Weighted positive rank

Weighted Positive rank measures the relationship between user selected feature and its associated positive items in the $f s_m$ based on highest weighted list.

$$\text{WeightedPRank}(w_i, f s_m, PCS_m) = \frac{\sum \sum w_i \text{Pr ob}(f s_i \cap PCS_m)}{\sum_i w_i \text{Pr ob}(f s_i)}$$

3.3 Weighted infrequent rank

Weighted infrequent rank measures the relationship between user selected feature and its associated negative items in the $i f s_m$ based on highest weighted list.

$$\text{WeightedIRank}(w_i, i f s_m, PCS_m) = \frac{\sum \sum w_i \text{Pr ob}(i f s_i \cap PCS_m)}{\sum_i w_i \text{Pr ob}(i f s_i)} \text{RankedCorr}(\varphi_1 = i f s_m(A \rightarrow B), \varphi_2 = PCS_m(C \rightarrow D)) = \frac{|\sum_{t \in \varphi_1} \text{find}_{tfidf}(\varphi_1)| * (\text{sup}(A \cup C) \cup (B \cup D)) - \text{sup}(\varphi_1) \cdot \text{sup}(\varphi_2)}{\sqrt{\text{sup}(A \rightarrow B) \cdot \text{sup}(C \rightarrow D)}}$$

4. WEIGHTED BASED FREQUENT AND INFREQUENT PATTERN MINING MODEL FOR REAL TIME E-COMMERCE DATABASES

In this proposed model, a novel weighted based frequent and infrequent pattern mining model for real time e-commerce databases is developed to discover the frequent and infrequent

relational patterns on the e-commerce databases. The overall framework is shown in the following Figure 4. In this model a real-time e-commerce application is designed for pattern extraction process [25].

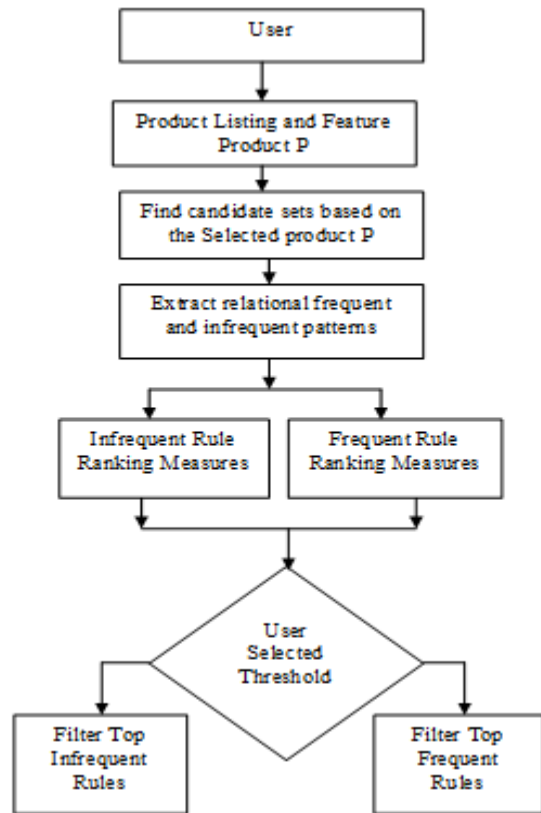


Figure 4. Proposed model flow chart

4.1 Proposed model

Input: E-commerce Database

PL: Product list

FL: Featured product list

P: User selected Featured product.

PP: Positive association patterns

IP: Infrequent association patterns

Output: Infrequent Association patterns and frequent association patterns.

Step 1: Start user login session in ecommerce application.

Step 2: if (user exist in DB)

then

User select's his feature product P.

end if

else

Incorrect user.

Step 3: Generates Dataset using the feature product P and its associated product's list.

Step 4: Load the user generated dataset.

Step 5: Generate candidate sets using the Apriori algorithm.

Step 6: η_{\min} minimum weighted threshold

$PS \leftarrow \emptyset; IS \leftarrow \emptyset$

Generate feature selected 1-item frequent item-set of P as P_1 .

Generate 1-itemset product list associated with selected feature product P_{r1} .

for ($i = 2, P_1 \neq \emptyset, i++$)

for ($j = 2, P_{r1} \neq \emptyset, j++$)

```

do
PCSm ← Join (P1, Pr1);
done
done
Step 7: for each item set  $IS_i \in PCS_m$ 
do
 $w_i = \text{Find\_tfidf}(D, IS_i)$ 
if  $w_i \geq \eta_{\min}$  then
 $fs_m \leftarrow fs_m \cup \{IS_i\}$ 
else
 $if_{s_m} \leftarrow fs_m \cup \{IS_i\}$ 
done
 $P(\varphi_A, \varphi_C) = \text{Positive-rules}(\text{WeightedPRank}(w_i, fs_m, PCS_m))$ 
 $I(\varphi_A, \varphi_C) = \text{Infrequent-rules}(\text{WeightedIRank}(w_i, if_{s_m}, PCS_m))$ 
// Mining all feature relational products using antecedent
and consequent items.
for each item in  $p(\varphi_A)$  (antecedent)
for each item in  $p(\varphi_C)$  (consequent)
do
 $r\sigma_{corr} = \text{FRankedCorr}(p(\varphi_A), p(\varphi_C))$ 
if  $r\sigma_{corr} \geq \text{min } t \text{ / } hres$ 
then
if  $\text{conf}(p(\varphi_A), p(\varphi_C)) \geq \text{conf}_{\min}$  then
PP ← PP ∪ {Join( $p(\varphi_A), p(\varphi_C)$ )}
else if  $\text{conf}(p(\varphi_A), p(\varphi_C)) \geq \text{conf}_{\min}$  and
sup( $\neg p(\varphi_A), \neg p(\varphi_C)$ )  $\geq \rho_{\min}$  then
IP ← IP ∪ {Join( $\neg p(\varphi_A), \neg p(\varphi_C)$ )}
endif
if  $r\sigma_{corr} \leq -\text{min } t \text{ / } hres$  then
if  $\text{conf}(p(\varphi_A), \neg p(\varphi_C)) \geq \text{conf}_{\min}$  then
IP ← IP ∪ {Join( $p(\varphi_A), \neg p(\varphi_C)$ )}
endif
if  $\text{conf}(\neg p(\varphi_A), p(\varphi_C)) \geq \text{conf}_{\min}$  then
IP ← IP ∪ {Join( $\neg p(\varphi_A), p(\varphi_C)$ )}
endif
 $I(\varphi_A, \varphi_C) = \text{Infrequent-rules}(\text{WeightedIRank}(w_i, if_{s_m}, PCS_m))$ 
// Mining top infrequent patterns of all feature relational
products using antecedent and consequent items.
for each item in  $I(\varphi_A)$  (antecedent)
for each item in  $I(\varphi_C)$  (consequent)
do
 $r\sigma_{corr} = \text{IRankedCorr}(I(\varphi_A), I(\varphi_C))$ 
if  $r\sigma_{corr} \geq \text{min } t \text{ / } hres$ 
then
if  $\text{conf}(I(\varphi_A), I(\varphi_C)) \geq \text{conf}_{\min}$  then
PP ← PP ∪ {Join( $I(\varphi_A), I(\varphi_C)$ )}
else if  $\text{conf}(I(\varphi_A), I(\varphi_C)) \geq \text{conf}_{\min}$  and
sup( $\neg I(\varphi_A), \neg I(\varphi_C)$ )  $\geq \rho_{\min}$  then
IP ← IP ∪ {Join( $\neg I(\varphi_A), \neg I(\varphi_C)$ )}
endif
if  $r\sigma_{corr} \leq -\text{min } thres$  then
if  $\text{conf}(I(\varphi_A), \neg I(\varphi_C)) \geq \text{conf}_{\min}$  then
IP ← IP ∪ {Join( $I(\varphi_A), \neg I(\varphi_C)$ )}
endif
if  $\text{conf}(\neg I(\varphi_A), I(\varphi_C)) \geq \text{conf}_{\min}$  then
IP ← IP ∪ {Join( $\neg I(\varphi_A), I(\varphi_C)$ )}
endif
done

```

done
The time complexity will be $O(|k| \log(n))$. where N is the number of complex patterns and k is optimal filtered patterns.

5. EXPERIMENTAL RESULTS

In this model a real-time e-commerce application is designed for pattern extraction process. User selects his feature product to generate frequent and infrequent association patterns. Based on the feature product, all related candidate sets are generated to the user selected feature product P. These candidate sets are used to discover the infrequent and frequent associations with other feature products. Here weighted infrequent ranking measure is used to filter the infrequent product from the frequent associations.

In this model a real-time e-commerce application is designed for pattern extraction process. User selects his feature product to generate frequent and infrequent association patterns. Based on the feature product, all related candidate sets are generated to the user selected feature product P. These candidate sets are used to discover the infrequent and frequent associations with other feature products. Here weighted infrequent ranking measure is used to filter the infrequent product from the frequent associations.

5.1 Web E-commerce database

MOBD8GVKYPVCVAD,'Nokia 200 Pearl White',0.143578
MOBD8U4PFBTA7WDF,'iBall i225',0.722979
MOBD8ZERPNG74FMZ,'Nokia Asha 302 Dark Grey',0.299327
MOBD8ZG28RSCKHWY,'Nokia Asha 302 White',0.707809
MOBD8ZG2BZ8VM9RH,'Nokia Asha 302 Plum Red',0.828714
MOBD8ZG6GNUMU37Q,'Samsung Galaxy Y Duos S6102 Pure White',0.410379
MOBD96CUHCYNMZH8,'Nokia Asha 202 Black',0.230572
MOBD96CUYPYMPGGZ,'Sansui S45 Black & Red',0.829439

5.2 User selected feature products list

| Brand | ItemName | Category | msrp |
|-------------|-------------------------------------|-------------------------------|---|
| Micromax | Micromax A300 Black | Mobiles & Accessories>Mobiles | http://www.flipkart.com/micromax-a300-black/p/... |
| Samsung | Samsung Note 4 Black | Mobiles & Accessories>Mobiles | http://www.flipkart.com/samsung-note-4-black/p/... |
| Micromax | Micromax A290 White | Mobiles & Accessories>Mobiles | http://www.flipkart.com/micromax-a290-white/p/... |
| Micromax | Micromax A290 Black | Mobiles & Accessories>Mobiles | http://www.flipkart.com/micromax-a290-black/p/... |
| Samsung | Samsung Note 4 White | Mobiles & Accessories>Mobiles | http://www.flipkart.com/samsung-note-4-white/p/... |
| iBall | iBall AEC 500 | Mobiles & Accessories>Mobiles | http://www.flipkart.com/iball-aec-500/p/... |
| Advan | Advan CT 6AT A35 Blue White | Mobiles & Accessories>Mobiles | http://www.flipkart.com/advan-ct-6at-a35-blue-white/p/... |
| Advan | Advan CT 6AT A35 Red White | Mobiles & Accessories>Mobiles | http://www.flipkart.com/advan-ct-6at-a35-red-white/p/... |
| Colson | Colson A 136 Black | Mobiles & Accessories>Mobiles | http://www.flipkart.com/colson-a-136-black/p/... |
| Zepp | Zepp Ultra One S55 Sport Black | Mobiles & Accessories>Mobiles | http://www.flipkart.com/zepp-ultra-one-s55-sport-black/p/... |
| Zepp | Zepp S55 Black | Mobiles & Accessories>Mobiles | http://www.flipkart.com/zepp-s55-black/p/... |
| Intel | Intel Atom 12 Pro Blue | Mobiles & Accessories>Mobiles | http://www.flipkart.com/intel-atom-12-pro-blue/p/... |
| Intel | Intel Atom 12 Pro Grey | Mobiles & Accessories>Mobiles | http://www.flipkart.com/intel-atom-12-pro-grey/p/... |
| Intel | Intel Atom Star Power KitKat Grey | Mobiles & Accessories>Mobiles | http://www.flipkart.com/intel-atom-star-power-kitkat-grey/p/... |
| Nokia | Nokia Lumia 830 White | Mobiles & Accessories>Mobiles | http://www.flipkart.com/nokia-lumia-830-white/p/... |
| Nokia | Nokia Lumia 830 Bright Orange | Mobiles & Accessories>Mobiles | http://www.flipkart.com/nokia-lumia-830-bright-orange/p/... |
| Nokia | Nokia Lumia 830 Black | Mobiles & Accessories>Mobiles | http://www.flipkart.com/nokia-lumia-830-black/p/... |
| Nokia | Nokia Lumia 830 Red | Mobiles & Accessories>Mobiles | http://www.flipkart.com/nokia-lumia-830-red/p/... |
| Micromax | Micromax A300 Black | Mobiles & Accessories>Mobiles | http://www.flipkart.com/micromax-a300-black/p/... |
| Micromax | Micromax A300 White | Mobiles & Accessories>Mobiles | http://www.flipkart.com/micromax-a300-white/p/... |
| Wing | Wing T 123 White | Mobiles & Accessories>Mobiles | http://www.flipkart.com/wing-t-123-white/p/... |
| Wing | Wing T 123 Blue | Mobiles & Accessories>Mobiles | http://www.flipkart.com/wing-t-123-blue/p/... |
| Wing | Wing T 123 Black | Mobiles & Accessories>Mobiles | http://www.flipkart.com/wing-t-123-black/p/... |
| 3i Infotech | 3i-Infotech 3i-Infotech CapSonic... | Mobiles & Accessories>Mobiles | http://www.flipkart.com/3i-infotech-3i-infotech-capsonic/p/... |

Figure 5. User selection E-commerce Product List

| Category | msrp | url | Product |
|-------------------------------|---|---|----------------------|
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/micromax-a300-black/p/... | http://www.flipkart.com/micromax-a300-black/p/... | 0.588795048497 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/samsung-note-4-black/p/... | http://www.flipkart.com/samsung-note-4-black/p/... | 0.15170195120881 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/micromax-a290-white/p/... | http://www.flipkart.com/micromax-a290-white/p/... | 0.001486799970590302 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/micromax-a290-black/p/... | http://www.flipkart.com/micromax-a290-black/p/... | 0.00208789963248 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/samsung-note-4-white/p/... | http://www.flipkart.com/samsung-note-4-white/p/... | 0.194525567520478 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/iball-aec-500/p/... | http://www.flipkart.com/iball-aec-500/p/... | 0.1039072204582 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/advan-ct-6at-a35-blue-white/p/... | http://www.flipkart.com/advan-ct-6at-a35-blue-white/p/... | 0.757910310713797 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/advan-ct-6at-a35-red-white/p/... | http://www.flipkart.com/advan-ct-6at-a35-red-white/p/... | 0.000603268800022 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/colson-a-136-black/p/... | http://www.flipkart.com/colson-a-136-black/p/... | 0.9638104448575 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/zepp-ultra-one-s55-sport-black/p/... | http://www.flipkart.com/zepp-ultra-one-s55-sport-black/p/... | 0.627860026014042 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/zepp-s55-black/p/... | http://www.flipkart.com/zepp-s55-black/p/... | 0.04803502520878 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/intel-atom-12-pro-blue/p/... | http://www.flipkart.com/intel-atom-12-pro-blue/p/... | 0.176750167205043 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/intel-atom-12-pro-grey/p/... | http://www.flipkart.com/intel-atom-12-pro-grey/p/... | 0.15170195120881 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/intel-atom-star-power-kitkat-grey/p/... | http://www.flipkart.com/intel-atom-star-power-kitkat-grey/p/... | 0.068630227280784 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/nokia-lumia-830-white/p/... | http://www.flipkart.com/nokia-lumia-830-white/p/... | 0.19747045142948 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/nokia-lumia-830-bright-orange/p/... | http://www.flipkart.com/nokia-lumia-830-bright-orange/p/... | 0.87600410313991 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/nokia-lumia-830-black/p/... | http://www.flipkart.com/nokia-lumia-830-black/p/... | 0.588795048497 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/nokia-lumia-830-red/p/... | http://www.flipkart.com/nokia-lumia-830-red/p/... | 0.184503035003608 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/micromax-a300-black/p/... | http://www.flipkart.com/micromax-a300-black/p/... | 0.8282232451961 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/wing-t-123-white/p/... | http://www.flipkart.com/wing-t-123-white/p/... | 0.72290328514847 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/wing-t-123-blue/p/... | http://www.flipkart.com/wing-t-123-blue/p/... | 0.422400887200003 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/wing-t-123-black/p/... | http://www.flipkart.com/wing-t-123-black/p/... | 0.184503035003608 |
| Mobiles & Accessories>Mobiles | http://www.flipkart.com/3i-infotech-3i-infotech-capsonic/p/... | http://www.flipkart.com/3i-infotech-3i-infotech-capsonic/p/... | 0.94462869381213 |

Figure 6. User selected E-commerce Product List with product rating

5.3 Experimental results evaluation

Precision in mining can be defined as the ratio of obtained relevant rules to the total number of rules detected in a database. Recall is described as the ratio of the number of relevant rules identified to the total number of relevant rules in the database. The term relevant explains the rules which are very firmly associated with each other. On the contrary, irrelevant rules are loosely or weakly associated with each other. The relevant rules are considered as true-positive, whereas the irrelevant are mentioned as false positive. There are some other types of rules which are relevant but can't be detected. Those types of rules are termed as false-negative rules.

Here a real time e-commerce data base is considered for implementing the algorithm. The rules produced out of this database are analyzed and verified thoroughly. If the generated rule is detected within the list of rules, it is considered to be relevant. But if the rules are not found in that specific list of previously-formed rules, it is categorized as irrelevant.

5.4 Generated patterns

The following are the frequent and infrequent patterns identified using the proposed frequent and infrequent pattern mining model.

PredictVal >= 0.002093 AND PredictVal <= 0.997823->
Itemid != MOBDHGYH4ENGRYXE

PredictVal >= 0.002093-> ItemName != Samsung Galaxy Grand Duos I9082 Elegant White, with 2 Flip Covers Color: White and Blue

PredictVal >= 0.002093->ItemName != Karbonn K205 Black and Grey

PredictVal <= 0.997823-> Itemid!= MOBDMHCZTZ8Z6JEC

PredictVal >= 0.002093-> ItemName != Nokia Lumia 620 White

PredictVal <= 0.997823 AND PredictVal >= 0.002093->
Itemid != MOBD3Y8XGKPRH836

5.5 Performance analysis

The experiment is made on a PC with 3.0GHZ CPU and 8GB memory. Proposed model is implemented in Java programming environment with Net beans 8.0 IDE. Experiments were done to evaluate the performance of proposed model with the traditional models such as CP Tree, IWITree in terms of memory, computation cost, computational time and number of patterns generated.

Table 1. Performance of computational search space and time on market database

| Measure | CPTree | IWITree | Weighted Frequent and Infrequent Model (ms) |
|---------|-----------|-----------|---|
| 0.1 | 6797.746 | 6055.974 | 3669.199 |
| 0.2 | 10630.355 | 8457.717 | 4963.015 |
| 0.3 | 13203.586 | 11280.127 | 7435.542 |
| 0.4 | 17153.753 | 13872.852 | 8133.702 |
| 0.5 | 18283.648 | 17456.099 | 9918.954 |

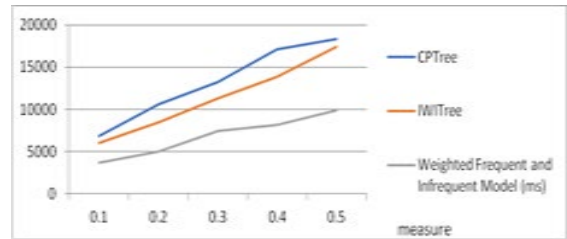


Figure 7. Visualization of performance of computational search space and time on market database

Table 1, illustrates the performance analysis of existing models with the proposed model in terms of candidate sets generation time and pattern mining time. Here, proposed weighted infrequent measure is compared with the traditional infrequent measures in terms of runtime. It is observed that the developed model requires reduces time complexity when compared to the existing pattern mining models on market database.

Table 2. Performance of computational search space and time on cloud database

| Measure | CPTree | IWITree | Weighted Frequent and Infrequent Model (ms) |
|---------|-----------|-----------|---|
| 0.1 | 6493.437 | 5540.862 | 2734.21 |
| 0.2 | 9224.168 | 8122.665 | 3251.177 |
| 0.3 | 12988.888 | 11050.933 | 5750.21 |
| 0.4 | 16942.067 | 13875.944 | 7835.565 |
| 0.5 | 18290.654 | 16394.602 | 7269.856 |

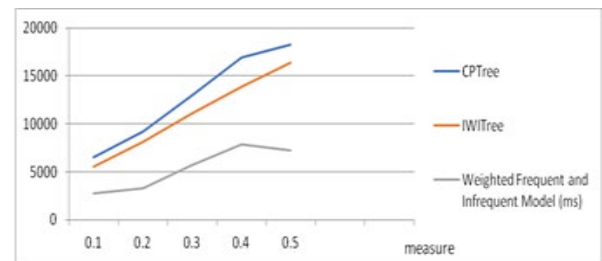


Figure 8. Visualization of performance of computational search space and time on cloud database

Table 2, illustrates the performance analysis of existing models with the proposed model in terms of candidate sets generation time and pattern mining time. Here, proposed weighted infrequent measure is compared with the traditional infrequent measures in terms of runtime. It is observed that the developed model reduces time complexity compared to the existing pattern mining models on cloud database.

Table 3. Performance of computational search space and time on medical database

| Patterns | CPTree | IWITree | Weighted Frequent and Infrequent Model (ms) |
|----------|-----------|-----------|---|
| 500 | 7658.923 | 5922.043 | 3273.708 |
| 1000 | 10773.957 | 8988.299 | 3973.716 |
| 1500 | 14735.01 | 11214.597 | 5002.805 |
| 2000 | 16626.562 | 14090.187 | 7113.348 |
| 2500 | 22198.064 | 16338.435 | 8077.925 |

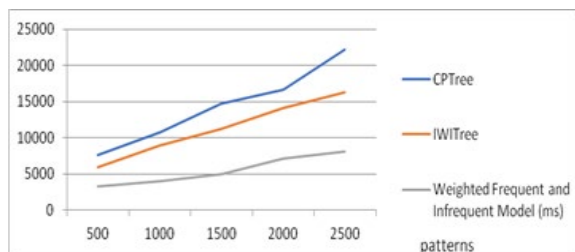


Figure 9. Visualization of performance of computational search space and time on medical database

Table 3, illustrates the performance analysis of existing models with the proposed model in terms of candidate sets generation time and pattern mining time. It is observed that the developed model reduces time complexity compared to existing pattern mining algorithms on medical database.

6. CONCLUSIONS & FUTURE SCOPE

Weighted based frequent and infrequent pattern mining model for real time e-commerce databases is proposed. In this model, efficient weighted positive rank, weighted infrequent rank, TF-IDF and ranked correlation measures are used to find the relational frequent and infrequent patterns from real-time e-commerce database. These patterns are used to filter the e-commerce products and act as a recommended system for online customers.

Similarly, this model is applied on cloud database, medical database and market database for the same purpose of decision making. Experimental results proved that the proposed model minimizes 8-10% search space and memory space when compared to the traditional models.

REFERENCES

- [1] Sadath, L. (2013). Data mining in e-commerce: A CRM platform. *International Journal of Computer Applications*, 68(24): 32-37. <https://doi.org/10.5120/11729-7383>
- [2] Rastegari, H., Md Noor, M. (2008). Data mining and e-commerce: Methods, applications, and challenges. *Journal Teknologi, Maklumat Malaysia*, 20(2): 116-128.
- [3] Ansari, S., Kohavi, R., Mason, L., Zheng, Z. (2001). Integrating e-commerce and data mining: architecture and challenges. *Proceedings 2001 IEEE International Conference on Data Mining*, San Jose, CA, USA, pp. 27-34. <https://doi.org/10.1109/ICDM.2001.989497>
- [4] Kohavi, R. (2001). Mining e-commerce data: The good, the bad, and the ugly (invited industrial track talk). *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 8-13. https://doi.org/10.1007/3-540-45357-1_2
- [5] Venkateswari, S., Suresh, R.M. (2011). Association rule mining in ecommerce: A survey. *International Journal of Engineering Science and Technology*, 3(4): 3086-3089.
- [6] Woon, Y.K., Ng, W.K., Lim, E.P. (2001). Association rule mining for electronic commerce: New challenges. *School of Computer Engineering, Nanyang Technological University Singapore*, 1-5.
- [7] Venkatesan, N., Ramaraj, E. (2011). High performance bit search mining technique. *International Journal of Computer Applications*, 14(2): 15-21. <https://doi.org/10.5120/1817-2371>
- [8] Kaosar, M.D., Xu, Z., Yi, X. (2009). Distributed association rule mining with minimum communication overhead. *Proceedings of the 8th Australasian Data Mining Conference (AusDM'09)*, School of Engineering and Science, Victoria University, Australia, 101: 17-23.
- [9] Coenen, F., Leng, P. (2002). Optimising association rule algorithms using itemset ordering. *Research and Development in Intelligent Systems XVIII*, 53-66. https://doi.org/10.1007/978-1-4471-0119-2_5
- [10] Pudi, V., Haritsa, J.R. (2003). ARMOR: Association Rule Mining based on ORacle. In *FIMI, Florida, USA*.
- [11] Nguyen S.N., Orłowska M.E. (2005) Improvements in the data partitioning approach for frequent itemsets mining. *European Conference on Principles of Data Mining and Knowledge Discovery*, Berlin, Heidelberg, pp. 625-633. https://doi.org/10.1007/11564126_66
- [12] McConnell, S., Skillicorn, D. (2004). Building predictors from vertically distributed data. *Proceedings of the 2004 conference of the centre for advanced studies conference on collaborative research 04-07*. Markham, Ontario, Canada, pp. 150-62.
- [13] Albashiri, K.A. (2014). Agent based data distribution for parallel association rule mining. *International Journal of Computers*, 8: 24-32.
- [14] Paul, S. (2010). An optimized distributed association rule mining algorithm in parallel and distributed data mining with XML data for improved response time. *International Journal of Computer Science and Information Technology*, 2(2): 88-101. <https://doi.org/10.5121/ijcsit.2010.2208>
- [15] Kotsiantis, S., Kanellopoulos, D. (2006). Association rules mining: A recent overview. *GESTS International Transactions on Computer Science and Engineering*, 32(1): 71-82.
- [16] Ng, R.T., Lakshmanan, L., Han, J.W., Pang, A. (1998). Exploratory mining and pruning optimizations of constrained association rules. *ACM SIGMOD Record*, 27(2): 13-24. <https://doi.org/10.1145/276305.276307>
- [17] Shaw, G., Xu, Y., Geva, S. (2008). Deriving non-redundant approximate association rules from hierarchical datasets. In *Proceedings of the 17th ACM conference on Information and knowledge management*, Napa Valley, USA, pp. 1451-1452. <https://doi.org/10.1145/1458082.1458328>
- [18] Bayardo, R.J., Agrawal, R., Gunopulos, D. (2000). Constraint based rule mining in large, dense databases. *Data Mining and Knowledge Discovery*, 4(23): 217-240. <https://doi.org/10.1023/A:1009895914772>
- [19] Fong, A.C.M., Zhou, B., Hui, S.C., Hong, G.Y., Do, T.A. Web content recommender system based on consumer behavior modeling. *IEEE Transaction on Consumer Electronics*, 57(2): 962-969. <https://doi.org/10.1109/TCE.2011.5955246>
- [20] Li, L., Wang, D., Li, T., Knox, D., Padmanabhan, B. (2011, July). SCENE: a scalable two-stage personalized news recommendation system. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 125-134. <https://doi.org/10.1145/2009916.2009937>
- [21] Ashkan, A., Clarke, C. (2016). Modeling browsing behavior for click analysis in sponsored search. *Proceedings of the 21st ACM International Conference*

- on Information and Knowledge Management, 2015-2019. <https://doi.org/10.1145/2396761.2398563>
- [22] Rong, W., Peng, B.L., Ouyang, Y.X., Xiong, Z. (2016). Collaborative personal profiling for web service ranking and recommendation. *Information Systems Frontiers*, 17(6): 1265-1282. <https://doi.org/10.1007/s10796-014-9495-4>
- [23] Paraschakis, D., Nilsson, B.J., Holländer, J. (2015). Comparative evaluation of top-n recommenders in e-commerce: An industrial perspective. In 2015 IEEE 14th International Conference on Machine Learning and Applications, Miami, FL, pp. 1024-1031. <https://doi.org/10.1109/ICMLA.2015.183>
- [24] Li, S.K., Guan, Z., Tang, L.Y., Chen, Z. (2011). Exploiting consumer reviews for product feature ranking. *Journal of Computer Science and Technology*, 27(3): 635-649. <https://doi.org/10.1007/s11390-012-1250-z>
- [25] Ramesh Kumar, G., Arulanandam, K., Kavitha, K. (2018). Frequent itemset mining with TM algorithm and Tree creation. *Advances in Modelling and Analysis B*, 61(4): 171-175. https://doi.org/10.18280/ama_b.610401