A new weighted based frequent and infrequent pattern mining method on real-time E-commerce

Thulasi Bikku

Vignan's Nirula Institute of Technology & Science for Women, Peda Palakaluru, Guntur 522009, Andhra Pradesh, India

thulasi.jntuk@gmail.com

ABSTRACT. *The purpose of this research is to perform infrequent pattern mining on E-Commerece Data. Association mining is an interesting model of data mining which is responsible for retrieving correlations, frequent patterns and infrequent associations from la rge datasets. The main objective of infrequent pattern mining is to discover the top infrequent items from the positive and negative association patterns with minimum support and confidence measures. Generally, association rule mining process is processed in 2 phases. Initially itemsets having high threshold values are identified and then secondly generates association patterns from these frequent candidate sets. Association rules can be represented in two forms, one is positive association rules and the other is negative association rules. In this proposed approach, user recommended frequent and infrequent mining model was developed to discover the top frequent and infrequent relational patterns on the e-commerce dataset. 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 frequent and infrequent associations with other feature products. Here weighted infrequent ranking measure was used to filter the infrequent product from the frequent associations. Experimental results show that proposed model has high computational prediction compared to traditional infrequent mining models.*

RÉSUMÉ. *Le but de cette recherche est de réaliser une extraction de motif peu fréquente sur les données E-Commerece. La fouille par associations est un modèle intéressant de fouille de données, chargé de récupérer des corrélations, des modèles fréquents et des associations peu fréquentes à partir de grands ensembles de données. L'objectif principal de l'extraction de modèles peu fréquents est de découvrir les éléments les moins fréquents parmi les modèles d'association positifs et négatifs, avec des mesures de soutien et de confiance minimales. Généralement, le processus d'extraction de règles d'association est traité en 2 phases. Initialement, les ensembles d'éléments ayant des valeurs de seuil élevées sont identifiés et ensuite génèrent des modèles d'association à partir de ces ensembles candidats fréquents. Les règles d'association peuvent être représentées dans deux formes: l'une est une règle d'association positive et l'autre une règle d'association négative. Dans cette approche proposée, un modèle d'exploration de données fréquent et peu fréquent a été développé pour* *permettre de découvrir les principaux modèles relationnels fréquents et peu fréquents sur l'ensemble de données de commerce électronique. L'utilisateur sélectionne son produit caractéristique pour générer des modèles d'association fréquents et peu fréquents. L'utilisateur sélectionne son produit caractéristique pour générer des modèles d'association fréquents et rares. Sur la base du produit caractéristique, tous les ensembles de candidats associés sont générés pour l'utilisateur sélectionné le produit caractéristique P. Ces ensembles de candidats permettent de découvrir les associations fréquentes et peu fréquentes avec d'autres produits. Une mesure de classement peu fréquente a été ici utilisée pour filtrer le produit peu fréquent des associations fréquentes. Les résultats expérimentaux montrent que le modèle proposé a une prédiction de calcul élevée par rapport aux modèles d'extraction peu fréquents traditionnels.*

KEYWORDS: *market data, infrequent association rules, support.*

MOTS-CLÉS: *données de marché, règles d'association peu fréquentes, support.*

DOI:10.3166/ISI.23.5.121-138

1. Introduction

Many algorithms are proposed since years to retrieve these data with the help of association rules and minimum threshold value. Threshold value depends on domain knowledge and statistical methods of retrieval. High threshold signifies that some items are missing, whereas low threshold results inconsistency. Association rule mining is one of the significant approaches of data mining which is responsible for retrieving correlations, frequent patterns and associations from the pool of itemsets in databases. It has its applications wide spread in different fields like telecommunication, market basket, risk management, and so on. Various research efforts have been done for the development of these association mining schemes. The main objective is to generate association rules satisfy minimum support and confidence.

Let us consider one frequent itemsets I = {I1, I2, …, Ik}. The first association rule is {I1, I2, …, Ik-1}⇒ {Ik}. The interestingness of the rule is decided by the confidence value. Other rules are generated by the same way and interestingness of each rule is checked by confidence and these steps are repeated multiple times till antecedent ends. The first phase is again split into two, those are: - candidate frequent itemsets generation and frequent itemsets generation.

The itemsets having support more than threshold is known as frequent itemsets and the expected itemsets to be frequent are known as candidate itemsets. Sometimes algorithms generate extensive amount of large associations which adds difficulties for end users.

Numbers of techniques have been introduced to decrease the numbers of association rules, those are: - generation of rules based on interestingness, non-redundancy, coverage, leverage, and so on.

Market basket analysis is a significant application of data mining. Let us consider a market with huge consumers' transactions. Suppose an association rule

X⇒Y is generated, where X stands for antecedent and Y stands for consequent. The probability that consumers purchasing item X will also purchase item Y is %c, where c is termed as confidence. It is required to check the dependency between items.

By analyzing the transactional data and studying the consumers purchasing behaviours, the business professionals gain idea for market planning and advertising.

Steps of association rule mining is given in Figure-1 below: -

Step-1: The set of candidate k-itemsets is generated by 1 extensions of the large (k -1)-itemsets generated in the previous iteration.

Step-2: Supports for the candidate k-itemsets are generated by a pass over the database.

Step-3: Itemsets that do not have the minimum support are discarded and the remaining itemsets are called large k-itemsets.

Step4: This process is repeated until no more large itemsets are found.

Figure 1. General association mining algorithm

The efficiency of association mining algorithms can be increased by reducing computational cost and the following four factors are responsible for this.

- Decrease numbers of passes over database
- Sampling
- Additional constraints
- Parallelization

Proper studies of many works by various researchers in the area of association rule mining, the following issues are identified and appropriate solutions to resolve these issues are described below.

- Most of the association rule mining approaches generate large numbers of rules including both relevant and non-relevant ones. Support and confidence metrics, exceed the threshold are used to find interestingness. It is also responsible for association pruning. Some other metrics used to evaluate interestingness are: - Ch i-square, Laplace, Correlation coefficient etc. More appropriate use of these metrics is in the area of ranking. Subjective metrics which are controlled by users have become more famous and widely accepted. Examples of subjective metrics are: - Unexpectedness: If the users cannot expect the rules based on his previous knowledge, its considered as interesting. Acitionability: If users can use the rules in their favour, those are marked as interesting. By the use of this kind of subjective metrics, the numbers of rules can be decreased upto some extent.

Besides the numbers of rules, quality of the rules also plays vital role in association rules mining and adds an important drawback to these algorithms. Because subjective nature of metrics, the quality of the rules always gains least attention. The accuracy of rules is calculated by users independently.

- These algorithms are configured before execution. Thus, the users need to specify the parameters much before. Users are needed to take two thresholds (minima l support and minimal confidence) in association rules mining and all rules exceeding these thresholds are selected. So the users should have proper knowledge and experience of domain. Algorithms with least numbers of parameters or no parameters can overcome this problem.

2. Related works

Lin *et al.* (2011) tried to increase the performance of FP growth algorithm and introduced an Improved FP growth approach. They emphasize on three important features of this algorithm, those are: - 1) Construction of FP-tree involves an address table to reduce complexity. 2) In order to minimize recursive FP-trees, FP-tree + is formed. 3) It makes the algorithm more efficient by reducing memory need and increasing performance. The authors carried out experiments and found that, because of less memory requirement IFP growth algorithm approximately uses half memory than that of conventional FP growth algorithm. With minimum supports the execution time of the said method is 300 times less than FP growth. The major drawback of this approach is “potential problem in hybrid FP-tree” due to lack of appropriate tree- level value. Further work is needed to find this tree-level value for better performance.

Lee *et al.* (2013) worked to include both statistical correlation and semantic significance in their new utility-based technique. The proposed approach evaluates association rules by considering the business benefits. Utility function based on users’ preference are defined by three major elements (opportunity, effectiveness and probability). The application of their approach on huge transactional databases involves functional algorithms and pruning methods. Other previously existing methods emphasizes on frequency of itemset and ignores its significances which may e xc lude the low frequency items from the association rules. They experimented and found that the presented technique has better advantages in the field of business as compared to other mining approaches. The above said algorithm performs better than the other methods by 133% (best), 120% (normal) and 113% (worst) case scenarios.

Rodríguez-González *et al.* (2013) proposed two algorithms for frequent pattern mining with similarity functions fulfilling f- downward property and also association rules are generated from these frequent patterns. They presented various features and validated the propositions for pruning search space. To overcome the problem of lost and false rules, GenRules algorithm with similarity function was used for construction of association rules from frequent patterns. A data structure was introduced for storing object sub- descriptions and the similarities among them. The researchers validated their work by numbers of experiments and found this more effective than that of others. Further work can be done to decrease set of frequent patterns and generate non-lossy association rules. The said algorithm can be extended in future by generating association rules for non-Boolean similarity functions.

Agrawal *et al.* (2015) identified two major problems of association rule mining for market basket dataset, those are:- 1) Association mining only considers positive association rules and ignores the negative association rules. 2) They also generate large numbers of candidate itemsets and scan the database repeatedly. The authors developed a new algorithm and termed it as SARIC algorithm. By taking negative association attributes into account, it applies SET-PSO (SET- Particle Swarm Optimization) for construction of association rules. Instead of using minimum support and confidence, the said algorithm uses itemsets range and correlation coefficient. The researchers simulated their theory on two variable-sized databases and showed that it outperforms Apriori, Eclat, HMINE and GA in terms of performance, computational time and efficiency. Further work needed to be carried out for implementation of the proposed method on spatial and temporal databases.

In market basket dataset, Arvind and Badhe (2015) indentified the existence of uncertanty in database and included some uncertainty in association rules mining. The association rules consisting of uncertainty is known as Vague Association Rules. This technique converts uncertain rules to certain in database. They applied Vague Set Theory to characterize association rule mining and the rules generated are termed as Vague Association Rules. The authors stated that their said approach can be efficiently applicable in the field of Profit Pattern Mining and Knowledge-Based Recommender System. In Profit Pattren Mining, the vague values are responsible for making decisions in different market scenarios. Instead of using collaborative and content filtering methods, the said method provides a better scenario for market basket by generating uncertainty-based association rules.

Cavique (2007) proposed a new algorithm to identify huge itemsets in market basket approach. He used condensed data in this method which are generated from the transformation of market basket to maximum-weighted clique. His approach is carried out in two important steps, those are:-1) The entered input dataset is converted to graphical representation. 2) A meta-heuristic technique is used to resolve maximum- weighted clique problem. The three variables affecting computation time are: - numbers of transactions, basket size and number of items. Both the conventional Apriori and FP- growth considers numbers of items and basket size as significant metrics, FP-growth scales more efficiently as compared to Apriori. But in case of Similis algorithm huge datasets with better scalability is encountered. He decided to extend their work in future by implementing Primal-Tabu meta-heuristic for optimized performance.

Narmadha *et al.* (2011) analyzed and surveyed on various association rule mining approaches. They also highlighted advantages and limitations of each approach. The authors identified the problem of choosing appropriate association rule mining among vast number of techniques. They also stated that selection of proper approach may give optimized performance as compared to others.

Ouyang (2014) presented a sliding window approach for mining both positive and negative association rule mining. He introduced a one-pass algorithm and termed it as MPNAR- SW. The algorithm is split into four important sub-steps, those are- window in itialization, window slid ing, frequent- infrequent itemsets generation and positive-negative association rule mining. He validated his scheme on indirect temporal sequential patterns and showed that the said algorithm is both scalable and effective.

Said *et al.* (2012) introduced a new technique in association rule mining on transactional datasets. Their approach has six major phases, those are:- business understanding, data assembling, nanodrop technique, Apriori mining algorithm, model building, post-processing associations and result interpretations. They applied their scheme on market basket analysis and it was helpful to monitor customers purchase patterns for further market research and planning. Support, correlation and confidence are three vital metrics used in association rule mining methods. According to users' want any of these metrics are chosen appropriately. They simulated their approach and noticed that their approach requires less computational time and less memory. In future it can be implemented in some of other applications like intrusion detection, customer relationship management, etc.

Trnka (2010) implemented market basket approach in six sigma. He argued that, by integrating market basket analysis with one phase of six sigma method the resulting performance is increased remarkably. The author introduced Genera l rule Induction algorithm to generate association rules. The inter dependency of the products are represented by a web plot. Another rule-based algorithm C5.0 was also used. By the integration of these two approaches, the performance of sig sigma is increased significantly. Further works need to be done to extend this proposed approach by combining other mining techniques with sig sigma, so that better performance can be found.

Waghmare and Mukhopadhyay (2014) developed a location-based shopping application for mobile platform. They implemented their proposed approach on bakery shops. Cloud platform was chosen for deployment. It contains three tiers, those are- front tier, middle ware and back tier. A location-based shopping application for android devices constitutes the front tier and the associations are also generated in this tier. Middle ware is responsible for production of JSON output from relational databases. It also has another important responsibility, that is transformation of information between application and cloud server. The back tier contains Apache Tomcat server with MySQL database. This approach also utlizes Google Cloud Messaging to update the status of purchase to the seller. Further work can be done to integrate GPS to provide direction and distance between the sellers and customers.

Yang *et al.* (2016) tried to improve Naïve Bayes classifier by using association mining rules. They stated that their approach can be applicable in all business datasets. First they applied Naïve bayes algorithm on business datasets, but it did not give satisfactory results. Later by merging attributes, they applied association rule mining to reduce features. The aim of involving Apriori algorithm is to merge the related attributes by the frequent patterns. This time the result was satisfactory. The authors merged related attributes in order to decrease the numbers of attributes to satisfy independent assumptions of Naïve Bayes algorithm. In future further work can be done in the process of implementation of other algorithms to merge related attributes.

Duggirala and Narayana (2013) developed a new association rule mining framework based on coherent rules approach. There is no need of pre-set support threshold. No expert ise or prior knowledge is required, because coherent rules use features of propositional logic. Association rules are generated out of these coherent rules without considering minimum support threshold. The proposed approach is a threshold free mining technique which maps to logical equivalences to extract all interesting associations without loss. Itemsets involve in this framework can be either frequent or infrequent in nature. They validated their approach with experiments and concluded that their method can identify association rules in between required execution time. Thus, it makes framework feasible with respect to time. Further researches can be carried out to integrate various classification techniques with the proposed framework.

Setiabudi *et al.* (2011) presented hybrid-dimension association for application of association rules mining in market basket datasets. They performed their experiments in minimarket X. Apriori algorithm was applied on frequent datasets and responsible for pairing of items in individual transactions. Frequent itemsets having values more than minumum support generate association rules. Confidence is also calculated from frequent datasets and it uses hybrid- dimension association. The experimental outcome shows that, which products are purchased combinely often by the customers.

3. Top ‘k’ infrequent mining algorithm using ranking measurers

In this proposed approach, user recommended frequent and infrequent mining model was developed to discover the top frequent and infrequent relational patterns on the e-commerce dataset. The overall framework was shown in the figure 2. In this model a real-time e-commerce application was 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 was used to filter the infrequent product from the frequent associations.

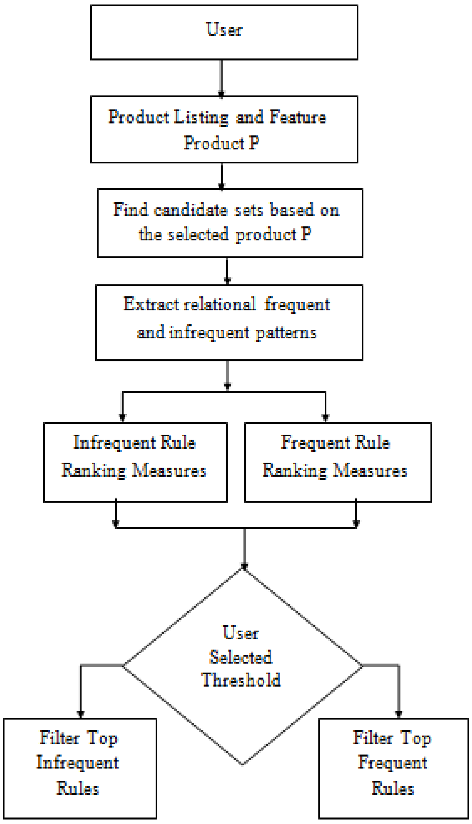


Figure 2. Proposed model

Algorithm:

Input: E-commerce Dataset **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 $\leftarrow \varnothing$; IS $\leftarrow \varnothing$
Generate feature selected 1-ite m 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 != $\varnothing$, i++)
for (j = 2, P_{r1} != $\varnothing$, j++) do
$PCS_m \leftarrow Join(P_1, P_{r1})$;
Done
Done
Step 5: For each itemset $IS_i \in PCS_m$
do
w_i=Find_tfidf(D, IS_i)
// Mining top infrequent patterns of all
if $w_i \geq \eta_{min}$ then
$fs_m \leftarrow fs_m \cup \{IS_i\}$
else
$ifs_m \leftarrow fs_m \cup \{IS_i\}$

done

P(φ_A, φ_C) =Positive-rules(WeightedPRank(w_i, fs_m, PCS_m))
// Mining all feature relational products using antecedent and consequent items.
For each item m in p(φ_A) (antecedent)

For each item m in $p(\varphi_C)$
(consequent) do

$r\sigma_{corr}$ = $FRankedCorr$(p(φ_A), $p(\varphi_C)$)
if $r\sigma_{corr} \geq minthres$
then
if conf(p(φ_A), $p(\varphi_C)$) $\geq$ then
$conf_{min}$
PP $\leftarrow$ PP $\cup$ {Join(p(φ_A), $p(\varphi_C$
else if conf(p(φ_A), $p(\varphi_C)$) $\geq conf_{min}$

For each item m in I (φ_A) (antecedent) For each item m in I (φ_C) (consequent) do

$r\sigma_{corr}$ = IRankedCorr (I (φ_A), I (φ_C))

If $r\sigma_{corr} \geq minthres$

Then

if conf(I(φ_A), I(φ_C)) ≥ $conf_{min}$ the

n PP ← PP ∪ {Jo in(I(φ_A), I(φ_C))}

else if conf(I(φ_A), I(φ_C)) ≥ $conf_{min}$

and sup(¬I(φ_A), ¬I(φ_C)) ≥ ρ_{min}

then

IP ← IP ∪ {Join ($\neg I(\varphi_A)$, $\neg I(\varphi_C)$)}

endif

if $r\sigma_{corr} \leq -minthres$ then

if conf(I(φ_A), $\neg I$ (φ_C)) ≥ $conf_{min}$ then IP ← IP ∪ {Join(I(φ_A), $\neg I$(φ_C))} endif

If conf ($\neg I$ (φ_A), I (φ_C)) ≥ $conf_{min}$ then

IP ← IP ∪ {Join ($\neg I$ (φ_A), I (φ_C))} endif

done

done

4. Experimental results

Web ecommerce Dataset:

MOBD3FAHZRRAQRZH,'Samsung Primo S5610 Metallic Silver',0.636282

MOBD3HPBWH7GPTKX,'Nokia 100 Legion Blue',0.558762

MOBD3NG5RBHHRVRH,'Sa msung Ga la xy Note N-7000 Blue',0.94691

MOBD3UW 2BYGZDEUQ,'No kia 100 Ocean Blue',0.685656 MOBD3V9VNM NT HWXJ,'No kia Asha 300

Graphite',0.077503

MOBD3Y8XGKPRH836,'iBall Shaan S207 Black & Gun Metal',0.298332

MOBD3Y8XXYVNFXMA,'iBa ll Shaan S315 Black & Da rk

Silver',0.109448

MOBD43T8HNNRUZFQ,'Samsung Ga la xy S 2 I9100

Ceramic White',0.598589

MOBD444TC2FZFA4H,'BlackBerry Bold 9790 Black',0.04611

MOBD47HSHDPGEHNQ,'Samsung S5253 Metallic

Silver',0.293666

MOBD47HUXYPUPXMT,'Hua wei Ideos X3 U8510 Black',0.100846

MOBD48G6UJSDAZMB,'Samsung S5610 Metallic Go ld',0.955599

MOBD4AV67HHHMNZX,'Sony Ericsson Live with

Walkman - WT19i Black',0.853872

MOBD4BTPC7GW HBMW,'Karbonn K 280 Black',0.083173

MOBD4BTPPN2YVQFF,'Karbonn K 409 Blac k & Red',0.782305

MOBD4GCNABDKQG9V,'HTC Sensation XL White Silver',0.392984

MOBD4GCXGCQFWSUU,'LG Optimus Sol E730 Black',0.862086

MOBD4GCXPZHP7TUE,'LG Optimus Hub E510

Black',0.903598

MOBD4GCXWJDBUBAQ,'LG Opt imus Net Dua l Sim P698 Titanium',0.701116

MOBD4M YJTCZMSF7A,'Nokia C5-05 Black A lu minu m Grey',0.39885

MOBD4P7SZCCSNBUA,'Micromax X256 Black

Silver',0.596181

MOBD4P8EGEXS3BEX,'Micromax X490 Metallic Grey',0.039943

MOBD4PA5XKGT8ZBW,'Micro max X207 Black & Yellow',0.071803

MOBD4Q2W GPTWQQJP,'Alcatel OT 318D White',0.995826

MOBD4Q2WZFE8NFU7,'Alcate l OT 810D Victorian Blush',0.619561

MOBD4Z3NDUBH6VXW,'Nokia Lumia 710

White',0.560559

MOBD54PGFZBWWKUH,'Samsung Gala xy Y S5360 Pure White',0.227649

MOBD54T8DFKAPEW8,'Samsung Ga la xy Ace S5830 Pure White, with 2 GB Micro SD Card',0.98603 MOBD5EHHVC8WA95V,'Samsung Ga la xy Note N-7000 Ceramic White',0.791362

MOBD5FNYSGCSNFDQ,'Ka rbonn K 4 Black & Red',0.167776

MOBD5G5WUJC942HZ,'Nokia 710 Black',0.997823 MOBD5G7DDCRHQCHS,'Nokia 200 Light Pink',0.94812 MOBD5G7REBGZJPFZ,'No kia Asha 200 Aqua',0.85797

MOBD5H9ZZPEHN7GP,'Samsung Wave Y S5380 Sand Silver',0.53997

MOBD5HGZ822GM3ZB,'Micromax X78 White',0.743679

MOBD5J6TG7QM GJVS,'BlackBerry 9330 Tata Indicom Black',0.899927

MOBD5UBXTYQUSXEP,'Samsung I9100G Ceramic White',0.809618

MOBD65FANVFS7AEF,'BlackBerry Curve 9360

White',0.592698

MOBD6F7AVGQNXHTH,'Nokia 100 P. Black',0.722331 MOBD6FUNW DJHQUFK,'Nokia X2-02 Bright Red',0.170038

MOBD6GVZ GMXKYQAD,'Videocon V1575 Black & Yellow',0.071049

MOBD6KRMXUTFJGMN,'Samsung Gala xy Pop Plus S5570i Steel Grey',0.544553

MOBD6KSRZTYUVG9D,'Samsung Ga la xy Ace Plus S7500

Dark Blue',0.036457

MOBD6UZXMPBJWZGS,'Intex Avatar 3D White',0.55781

MOBD6VVSP8JZRYDN,'Samsung Ga la xy Y Duos S6102 Strong Black',0.389209

MOBD6VVSUH6M YFXS,'Sa msung Ga la xy Y Pro Duos B5512 Metallic Black',0.537658

MOBD6YXNR8ESA VBH,'Nokia X2-02 Dark

Silver',0.676336

MOBD72HXQ2KRX4H8,'Samsung Gala xy Ace Duos I589 Metallic Grey',0.303711

MOBD75SW6HJC8EZ9,'Karbonn KT 60 Black & Grey',0.231084

MOBD75SW8SA 9FAR6,'Karbonn Symphony Black and Red',0.833202

MOBD75SW GUZU5FWP,'Karbonn KT 51 Black',0.063458

MOBD786FJ6NHUAFZ,'Samsung Star 3 Duos S5222 Modern Black',0.740503

MOBD7NNA VDZEZYJH,'Sa msung Ga la xy Y S5360

Absolute Black',0.611299

MOBD7SDF8QURBZMH,'BlackBerry 9900 White',0.279194

MOBD835Y7E4TFMAE,'Lava KKT 34i Black and Silver',0.064599

MOBD835YGEGFBGGN,'BlackBerry Curve 9220 Black',0.750649

MOBD83JYR2YF67GD,'Nokia Asha 500 White',0.53173 MOBD85N2DHCYDAXV,'HTC ONE X S720E Brown Gray, with 32 GB',0.830631

MOBD85N2VA ETHGYR,'HTC One V T320 Jupitor Rock,

with Bundled Sandisk 8GB Memory Card',0.661096 MOBD85ZAAWNFHTVV,'Huawei G7010 Black',0.31649 MOBD87MH3ZXGQBX9,'Huawei Ideos X5 U8800 Pro Black',0.63438

MOBD87XUANGZGMW B,'Karbonn Ga mester K3000 Yellow',0.391502

MOBD8B6MPZHF3ZZE,'Sony Xperia S Black',0.330109

MOBD8EGGRDW YA5ZC,'Sony Xperia S White',0.517094

MOBD8FEA YHNVGYUZ,'iBall Gla m 3 Exclusive White',0.243925

MOBD8GVKYPVC VCA D,'No kia 200 Pearl White',0.143578

MOBD8U4PFBTA7WDF,'iBa ll i225',0.722979 MOBD8ZERPNG74FMZ,'Nokia Asha 302 Dark Grey',0.299327

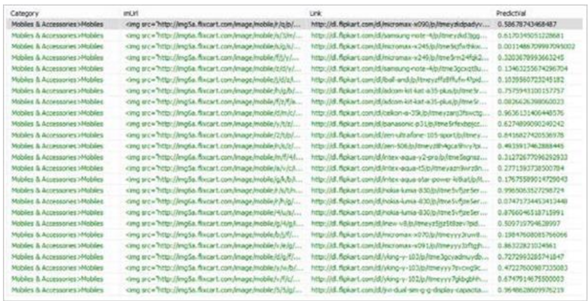
MOBD8ZG28RSCKHW Y,'Nokia Asha 302 White',0.707809 MOBD8ZG2BZ8VM 9RH,'Nokia Asha 302 Plum Red',0.828714

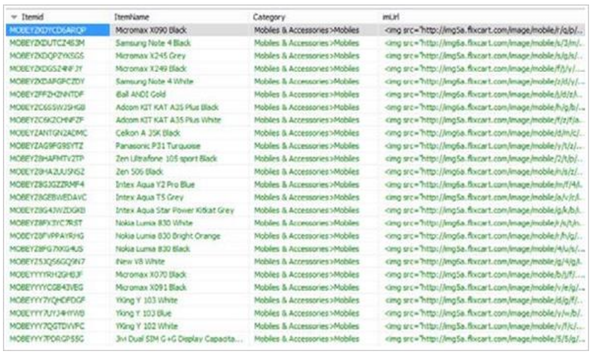
MOBD8ZG6GNUUU37Q,'Samsung Ga la xy Y Duos S6102 Pure White',0.410379

MOBD96CUHCYNMZH8,'Nokia Asha 202 Black',0.230572

MOBD96CUYPYM PGGZ,'Sansui S45 Black & Red',0.829439

User selected Feature Products List





Generated Patterns:

Predict Va l >= 0.002093 AND Predict Va l <= 0.997823 -> ItemName != Celkon A 118 HD Signature Smart Phone White Predict Val < 0.997823 -> ItemName != Celkon A 118 HD Signature Smart Phone White

Itemid != MOBDNBCMZ5VRW QXD -> Ite mName != Celkon A 118 HD Signature Smart Phone White

Ite mName != Karbonn Sma rt A5* Black -> Predict Val >= 0.002093

Itemid != MOBDDCBM5NT UWZCH -> Predict Va l >=

0.002093

Itemid != MOBDM VGQZCEM PBYJ -> Pred ict Va l >= 0.002093

Itemid != MOBDH9W5VSTGY5C G -> Predict Va l >= 0.002093

ItemName != Intex Aqua i5 Black -> Predict Va l >= 0.002093

ItemName != Karbonn Flair K102+ Go lden||White AND Itemid != MOBDMKDAKQGCYZ6D -> Predict Va l >= 0.002093

ItemName != Karbonn Flair K102+ Go lden||White AND Itemid != MOBDMKDAKQGCYZ6D -> Predict Va l >= 0.002093

ItemName != Icon G9 Dual SIM Qwerty Mobile Black AND Itemid != MOBDMKDAKQGCYZ6D -> Predict Va l >= 0.002093

Itemid != MOBCYQHSTVD9HJXH AND Itemid != MOBDM KDAKQGCYZ6D -> Predict Va l >= 0.002093

Predict Val <= 0.997823 -> Ite mName != Icon G9 Dual SIM

Qwerty Mobile Black

Predict Val >= 0.002093 AND Predict Val <= 0.997823 -> ItemName != Karbonn A9+ Pearl White

Predict Val <= 0.997823 AND Predict Val >= 0.002093 -> ItemName != Celkon A9+ Smart Phone Black

Predict Val >= 0.002093 -> Ite mid !=

MOBDHGHFW NGZHF34

Predict Val <= 0.997823 -> ItemName != Lenovo S920 Blue Predict Val <= 0.997823 AND Predict Val >= 0.002093 -> ItemName != Karbonn Flair K102+ Go lden||White

Predict Val <= 0.997823 AND Predict Val >= 0.002093 -> Itemid != MOBDHPF8AE3HQKAB

Predict Val >= 0.002093 AND Predict Val <= 0.997823 -> ItemName != Videocon V1531+ Black & Red

Predict Val <= 0.997823 -> Ite mName != Samsung Ga la xy Y

Duos S6102 Strong Black

Predict Val <= 0.997823 -> ItemName != Nokia Lumia 620 White

Predict Val <= 0.997823 AND Predict Val >= 0.002093 -> ItemName != Celkon A9+ Smart Phone Black

Predict Val >= 0.002093 AND Predict Val <= 0.997823 ->

Itemid != MOBDHPF8AE3HQKAB

Predict Val <= 0.997823 AND Predict Val >= 0.002093 -> Itemid != MOBDHPF8AE3HQKAB

Predict Val <= 0.997823 AND Predict Val >= 0.002093 -> Itemid != MOBDHPF8AE3HQKAB

Predict Val >= 0.002093 -> Ite mid !=

MOBDKZJA6SQGE79D

Predict Val <= 0.997823 AND Predict Val >= 0.002093 -> ItemName != Karbonn Flair K102+ Go lden||White

Predict Val <= 0.997823 -> ItemName != Huawei Ascend G300 Dark Tarnish

Predict Val <= 0.997823 -> Ite mName != LG Optimus L9

P765 Black

Predict Val >= 0.002093 -> ItemName != Samsung Ga la xy Grand Duos I9082 Elegant White, with 2 Flip Covers Color: White and Blue

Predict Val >= 0.002093 -> Ite mid != MOBDHGYH4ENGRYXE

Predict Val <= 0.997823 AND Predict Val >= 0.002093 -> ItemName != HTC A310E Explorer Metallic Black

Predict Val >= 0.002093 -> Ite mid != MOBDKFGYVSE2ZTPH

Predict Val <= 0.997823 -> Ite mName != LG Optimus L9 P765 Black

Predict Val <= 0.997823 -> Ite mid != MOBD9P2MHVBSFFFW

Predict Val <= 0.997823 -> Ite mName != LG Optimus L9

P765 Blac k

Predict Val >= 0.002093 AND Predict Val <= 0.997823 -> Itemid != MOBDESCY27FEKXKB

Predict Val >= 0.002093 AND Predict Val <= 0.997823 -> Itemid !=

MOBDHGYH4ENGRYXE

Predict Val >= 0.002093 AND Predict Val <= 0.997823 ->

Itemid != MOBDGPRJWHJYT YZJ

Predict Val <= 0.997823 AND Predict Val >= 0.002093 -> ItemName != Videocon V1531+ Black & Red

Predict Val >= 0.002093 -> ItemName != Samsung Ga la xy Grand Duos I9082 Elegant White, with 2 Flip Covers Color: White and Blue

Predict Val >= 0.002093 -> ItemName != Karbonn K205 Black and Grey

Predict Val <= 0.997823 -> Ite mid !=

MOBDMHCZTZ8Z6JEC

Predict Val >= 0.002093 -> ItemName != Nokia Lumia 620 White

Predict Val <= 0.997823 AND Predict Val >= 0.002093 -> Itemid!= MOBD3Y8XGKPRH836

Predict Val <= 0.997823 AND Predict Val >= 0.002093 ->

Itemid != MOBD3Y8XGKPRH836

Elapsed time: 23.243s

Number of Iterations :42 F-Measure: 0.97413

Recall: 0.93570

TP rate: 0.98995

FP rate: 0.010050000000000003

Classification Accuracy 0.96729

5. Conclusion

In this paper, we have implemented a novel method for efficiently mining frequent and infrequent patterns in real-time ecommerce database. Our model generates infrequent patterns based on user selected feature product, which is novel and optimal from traditional research models on infrequent association models for ecommerce database. In this proposed approach, user recommended frequent and infrequent mining model was developed to discover the top frequent and infrequent relational patterns on the e-commerce dataset. Experimental results proved that proposed model has high prediction of infrequent patterns in real-time compared to traditional infrequent mining models.

References

Agrawal J., Agrawal S., Singhai A., Sharma S. (2015). SET-PSO- based approach for mining positive and negative association rules. *Knowledge and Information Systems*, pp. 453-471. https://doi.org/10.2141/kis.32141

Amiri E. O. (2018). Application of computational experiments based on the response surface methodology for studying of the recirculation zone in the Y-shaped channel. *Mathematical Modelling of Engineering Problems*, Vol. 5, No. 3, pp. 243-248. https://doi.org/10.18280/mmep.050317

Arvind T. S., Badhe V. (2015). Discovery of certain association rules from an uncertain database. *International Conference on Computational Intelligence and Communication Networks*, pp. 827-831. https://doi.org/10.1109/CICN.2015.168

Cavique L. (2007). A scalable algorithm for the market basket analysis. *Journal of Retailing and Consumer Services*, Vol. 14, No. 6, pp. 400-407. https://doi.org/10.1016/j.jretconser.2007.02.003

Duggirala R., Narayana P. (2013). Mining positive and negative association rules using coherent approach. *International Journal of Computer Trends and Technology*, Vol. 4, No. 1, pp. 1-8.

Lee D., Park S., Moon S. (2013). Utility-based association rule mining: A marketing solution for cross-selling. *Expert Systems with Applications*, pp. 2715-2725. https://doi.org/10.2741/esa.22147

Lin K., Liao I., Chen Z. (2011). An improved frequent pattern growth method for mining association rules. *Expert Systems with Applications*, pp. 5154-5161. https://doi.org/10.2741/esa.11974

Narmadha D., NaveenSundar G., Geetha S. (2011). A novel approach to prune mined association rules in large databases. *2011 3rd International Conference on Electronics Computer Technology*, Vol. 5. https://doi.org/10.1109/ICECTECH.2011.5942031

Ouyang W. (2014). Mining positive and negative association rules in data streams with a sliding window. *4th Global Congress on Intelligent Systems*, pp. 205-209. https://doi.org/10.1109/GCIS.2013.39

Rodríguez-González A. Y., Martínez-Trinidad J. F., Carrasco- Ochoa J. A., Ruiz-Shulcloper J. (2013). Mining frequent patterns and association rules using similarities. *Expert Systems with Applications*, pp. 6823-6836.

Said A. M., Dominic P. D. D. (2012). A new scheme for extracting association rules: Market basket analysis case study. *International Journal of Business Innovation and Research*, Vol. 6, No. 1, pp. 28-46. https://doi.org/10.2214/ijbir.21421.

Setiabudi D. H., Budhi G. S., Purnama I. W., Noertjahyana A. (2011). Data mining market basket analysis' using hybrid-dimension association rules, case study in minimarket X. *International Conference on Uncertainty Reasoning and Knowledge Engineering*, pp. 196-199. https://doi.org/10.1109/URKE.2011.6007796

Trnka A. (2010). Market basket analysis with data mining methods. *International Conference on Networking and Information Technology*, pp. 446-450. https://doi.org/10.1109/ICNIT.2010.5508476

Waghmare V., Mukhopadhyay D. (2014). Mobile Agent based market basket analysis on cloud. *International Conference on Information Technology*, pp. 305-310. https://doi.org/10.1109/ICIT.2014.21

Yang T., Qian K., Lo D. C., Xie Y., Shi Y., Tao L. (2016). Improve the prediction accuracy of naive bays classifier with association rule mining. I*EEE 2nd International Conference on Big Data Security on Cloud, IEEE International Conference on High Performance and Smart Computing, IEEE International Conference on Intelligent Data and Security*, pp. 129-133. https://doi.org/10.1109/BigDataSecurity-HPSC-IDS.2016.38