An Efficient Approach Using FM-Weight for Revenue Prediction on Rare Itemsets

M. Jeyakarthic, S. Selvarani

Abstract--- ***A pattern is a social event of exercises that happen together in a database. Past examinations in the field are normally committed to the issue of regular pattern investigation where just precedents that appear as frequently as conceivable in the data are mined. In this manner, structures including exercises that appear in couple of instructive accumulations are not retrieved. We propose a structure to address different orders of intriguing precedents and after that instantiate it to the specific occasion of extraordinary models. Our focus is strived towards rare itemsets mining. The goal of this work is to show that applying additional measures support, confidence and lift framework to disclosure of rare association rules. Moreover, we proposed new algorithm for finding rare itemsets with high revenues from FM-weighted transactions with client analysis. The experimental results outperformed those found utilizing the traditional approach in the prediction of revenue from clients in next-period transactions.***

Keywords--- ***Association rule mining, Rare Association Rule mining, RFM Analysis, Weighted Association Rule.***

I. INTRODUCTION

In the knowledge discovery in data domain, association rule mining (ARM) is an essential data mining approach that can empower the revelation of consumer purchasing behaviors from transaction databases. The issue of ARM describing it as perceiving all rules from the transaction data that fulfill the minimum support and confidence constraints[1]. The exposure of interesting associations or correlations is valuable in many business decision-making processes and various fields[2].

Regardless, general ARM does not take into consideration the relative benefit or hugeness of transactions belonging to different clients, and rather expect that the noteworthiness of each client is identical. In other words, every client is of equal weight amid the mining process. However, numerous studies in customer relationship management (CRM) have revealed that the contributions of clients to businesses and profit maximization fluctuate. Therefore, the evaluation of client esteem is necessary before sketching out effective marketing strategies.

In the information disclosure in information space, affiliation rule mining (ARM) is an essential information mining approach that can empower the disclosure of shopper acquiring practices from exchange databases.

Businesses have started applying data mining innovations to marketing planning. They will probably pick up client reliability and discover the contribution of client value. Recency, frequency, monetary (RFM) analysis relies upon recency (R), frequency (F), and monetary (M) measures and is one of the most noticeable database marketing metrics for assessing client transaction histories. RFM scoring is a methodology for determining the score of current clients on in light of their R, F, and M esteems, and has been wound up being incredibly convincing in promoting database applications[3]. Moreover, RFM analysis is an exceptional, behavior based data mining technique, which removes client profiles by using specific criteria. Starting late, the RFM model has been used for CRM applications such as client segmentation[4][5].

Since RFM analysis and market basket analysis (i.e., association rule mining) are the two most vital errands in database marketing, this study broadened the traditional association rule mining issue by associating a client esteem (i.e., frequency monetory related (FM) weight, which is determined by applying the FM scoring strategy) with a transaction to reflect the premium or intensity of client esteems. This energizes the association of a FM weight parameter with every transaction, engaging the disclosure of valuable patterns. Furthermore, we propose new rare itemsets frequency money (RIFM) weighted algorithm for identifying rare itemsets from FM weighted transactions for the forecast of client revenue.

We tended to the accompanying two inquiries identified with discovering infrequent itemsets from FM-weighted transactions:

1. Do the top k infrequent itemsets found utilizing the proposed RIFM algorithm outflank those found utilizing the conventional AprioriRare algorithm as far as predicting clients purchasing itemsets?
2. Do the top k rare itemsets found utilizing the proposed RIFM algorithm outflank those found utilizing the conventional AprioriRare algorithm in foreseeing client revenue?

II. RELATED WORK

The primary motivation behind this study was to discover rare itemsets from transaction data with client esteems (FM weights). In this section, we for the most part investigate some techniques frequent association rules, rare association rules, weighted association rules and client value (RFM value).

Association rule mining

Association rule mining (ARM) is a vital data mining approach that empowers the disclosure of client purchasing behavior from transaction databases. Association rules are fascinating and unforeseen association relationships among attributes in a database that fulfill minimum support and confidence constraints[2].

Manuscript received February 01, 2012. (Fill up the Details)

M. Jeyakarthic, Department of Computer and Information Sciences, Annamalai University, Annamalai Nagar, Chidambaram, Tamil Nadu, India.

S. Selvarani, PG Department of Computer Science, Government Arts College, C-Mutlur, Chidambaram, Tamil Nadu, India

The problem, defining it as distinguishing all rules from the transaction data that fulfill the minimum support and confidence constraints[1]. In a nutshell, an ARM algorithm includes two steps: (1) Generation of all frequent itemsets that fulfill the minimum support constraint; and (2) Generation of all association rules that fulfill the minimum confidence constraints from the already found frequent itemsets.

Rare Association Rule Mining (RARM)

In the current years several novel algorithms have been developed to extricate strong association rules, fulfilling the minimum support and minimum confidence requirements. These algorithms are basically centered around the frequent itemsets generation phase and prepared for finding just the frequent itemsets from the dataset, and it drops the infrequent itemsets. In any case, the principle objective here is to generate the rare rules which may give valuable information. Some algorithms were developed for separating rare itemsets and/or infrequent itemsets from a dataset. [a]Apriori-Rare[6]: The principle focus of this algorithm is to generate the frequent as well as rare itemsets. It is the modification of Apriori algorithm to produce negligible rare itemsets. It utilizes a sub-routine called Supportcount to find this algorithm is that it reestablishes all the negligible rare itemsets. Notwithstanding, it neglects to find all the rare itemsets. [b]Apriori-Inverse[7]: This algorithm determines only the sporadic rules using one Minsup value and one Maxsup value. The sporadic rules have the property that they underneath user define Maxsup but above the Minconf value.The primary preferred standpoint of Apriori-Inverse is that it can find the sporadic itemsets extensively more rapidly than apriori. Nonetheless, a noteworthy constraint is that it is incapable of finding all the rare itemsets.

Another essential issue relating to the current methodologies rare association rule mining is that they are based only on a support, confidence structure to evaluate those rules mined, however none of them utilize other measures, such as the lift, to filter uninteresting rules[8].

Weighted association rules

Each item is managed reliably by most ARM algorithms. In any case, in real applications, a client may be more keen on the rules that describe the frequently happening "fashionable" items. Moreover, the client may wish to mine the association rules however put more accentuation on some items. As such, frequent itemsets are typically mined from binary databases, and each item in a transaction may have an alternate noteworthiness. The mixed-weighted association rules algorithm to find the issue of mining mixed-weighted association rules[9]. From that point forward, researchers have proposed weighted frequent itemset mining (WFIM) algorithms that mirror the significance of items. The conventional association rule problem by associating a weight with every item in a transaction to reflect the intrigue or force of each item within the transaction[10]. WFIM algorithm to generate more brief and vital weighted frequent itemsets in large databases[11]. Several algorithms for the quick mining of frequent weighted itemsets (FWI) from weighted item transaction databases[12].

Recency frequency monetary (RFM)

In view of the CRM theory[13][14], distinct techniques have been created for enhancing shopping rates, expanding offers of high-profit or price products, and holding clients as long-term clients. RFM stands for: R is defined as the "last purchasing time;" F is defined as the "purchasing frequency in a particular period"; and M is defined as the "average amount of purchase in a particular period[15]." The RFM model can be utilized to effectively play out way toward clustering based on client values. Business plans can be formulated to extend the clients life cycle by executing marketing projects[16]. RFM scoring is a strategy for deciding the score of current clients from their R, F, and M values, and has been ended up being particularly intense in marketing database applications[3]. Various studies have talked about the use of RFM values in proposal frameworks. Consolidating client lifetime value (CLV) and RFM to analyze clients' utilization properties and to give a suggestion based on these properties. In their investigations, clustering techniques were utilized to segment clients according to the weighted RFM (WRFM) value[17]. The timely RFM (TRFM) method rather than the WRFM for thinking about product property and purchase periodicity[18]. RFM ideas have been connected in various domains. The use of an RFM engine for anomaly detection to minimize false alerts in network attacks and decrease the time required to respond to hacking events[19]. Combined RFM with a CLV model to evaluate the segmented clients and subsequently utilized a genetic algorithm to select more appropriate clients for each campaign methodology[20]. Hsieh(2004) used a self-organizing map (SOM) neural network to distinguish segments of bank clients based on of reimbursement behavior and RFM behavioral scoring predictors[21]. Lin and Tang combined the RFM model inspect client's values and group similar values together. They consolidated users recent behavior with incremental mining as per weight to mine relations based on the weight rather than analyzing all the data, therefore diminishing the computation cost and time[22]. This strategy can likewise be applied to ARM by using the Apriori algorithm. In addition, Chiang (2011) proposed new methodology and an enhanced Recency–Frequency–Monetary–Discount–Return Cost (RFMDR) model to mine the association rules of client values[23]. In any case, clients purchasing behavior may not be obvious from the association rules generated in this study.

Since numerous retailers record transactions without gathering client information, RFM client patterns cannot be found using the current methodologies. Hu and Yeh defined the RFM pattern and built up a novel algorithm to find a complete set of RFM patterns that can inexact the set of RFM client patterns without client identification information. RFM pattern is characterized as a pattern that happens frequently but also involves a current purchase and a higher income rate.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

In spite of the fact that this proposed approach is proficient and can be successfully connected to discover RFM patterns, not all RFM client patterns could be found utilizing this approach[24]. Cheng-Hsiung Weng proposed a new algorithm, for discovering frequent itemsets with high revenues from FM-weighted transactions[25].

From the preceding analysis, we got the accompanying in-development: (1)In conventional rare itemsets mining, the relative advantage or importance of transactions of different clients is not considered. (2) The set of RFM patterns that can approximate the set of RFM client patterns without client identification information was found[24]. However, the relative benefit or essentialness of transactions of different clients was not reflected totally. (3) The integration of client value into the transactions to reflect intrigue or force of each client and accordingly find valuable patterns is imperative.

To address the previously mentioned confinements, in this study, we expanded the traditional rare association rule problem by permitting a client value (FM weight) to be related with a transaction to reflect the premium or power of client values. In this manner, we mined infrequent itemsets with high revenues from FM-weighted transactions. At last, we investigated the distinctions in the revenue of patterns found utilizing the proposed approach.

The rest of this paper is organized as follows. A review of related work is exhibited in Section 2. The problem definitions are provided in Section 3. The proposed algorithm and an example are outlined in Section 4. Section 5 utilizes survey data exhibit the handiness of the proposed algorithm. Conclusions and future work are examined in Section 6.

III. PROBLEM DEFINITIONS

In this section, we characterize the issue of the technique for discovering infrequent itemsets from FM-weighted transactions. Let I={it_1, it_2, …, it_m} be a set of itemsets. Let D be a set of database transactions in which every transaction T is a set of items with the end goal T⊆I. A transaction T is considered to contain X if and just if X⊆T. It is critical to determine weight values of the transactions of clients before discovering rare itemsets from FM-weighted transactions. For averting attributes (frequency and monetary) with at first ranges from exceeding, we standardized frequency and monetary of a client by utilizing min-max normalization strategy.

Table 1: Transaction data

Transaction ID	Client ID	Itemsets	Revenue A	Revenue B	Revenue C	Revenue D	Revenue	Date of purchase
1	1001	A,B,C	30	10	20		60	2017/1/1
2	1002	A,B,D	30	10		30	70	2017/1/2
3	1001	A,B,C	30	10	20		60	2017/1/3
4	1002	A,B,D	30	10		30	70	2017/1/2
5	1003	A,C,D	30		20	30	80	2017/1/1

Definition 1. The transaction weight of the transaction Ti of client c is defined as follows:

$$W(Ti) = \frac{F_{count}{}^{w} + M_{count}{}^{w}}{2} \quad (1)$$

$$F_{count}{}^{w} = \frac{F_{count} - F_{mini}}{F_{maxi} - F_{mini}} \quad (2)$$

$$M_{count}{}^{w} = \frac{M_{count} - M_{mini}}{M_{maxi} - M_{mini}} \quad (3)$$

F_{maxi}: the most extreme tally of transactions for all clients;
F_{mini}: the least tally of transactions for all clients;
F_{count}: the count of transactions of client c;
M_{maxi}: the most extreme monetary sum for all clients;
M_{mini}: the least monetary sum for all clients;
M_{count}: the monetary sum for client c.

Example 1. Five transactions are exhibited in Table 1, and the client FM value after estimation is presented in Table 2. Table 2 shows that the FM weight of transactions {1001, 1003} is 0.83, that of transactions {1002, 1004} is 1.00, and that of transaction {1005} is 0.00. At long last, the transaction data with the FM weights after estimation are exhibited in Table 3.

Definition 2. The FM-weighted support of an itemset X is defined as follows:

$$FM_{support}\ (X) = \frac{\sum T_i\ (X \subseteq T) \wedge (T \in D)\ W\ (Ti)}{\sum_{Ti\ (T \in D)} W\ (T_i)} \quad (4)$$

Example 2. From Table 1, $FM_{support}$ (A) =0.83 + 1.00 + 0.83 + 1.00) / 3.66 = 1.00, $FM_{support}$ (AB) = (0.83 + 1.00 + 0.83 + 1.00) /3.66 = 1.00, and $FM_{support}$ (C)=0.45, $FM_{support}$(AC) = (0.83 + 0.83 + 0.00) / 3.66 = 0.45. $FM_{support}$ (AD)=0.5 and $FM_{support}$(BD)=0.55

Definition 3. Given a user-specified $FM_{support}$ threshold $\sigma_{FMsupport}$, an itemset X is a low $FM_{support}$ itemset if $FM_{support}$ (X) ≤ $\sigma_{FMsupport}$.

Example 3. Let the $FM_{support}$ threshold be set as $\sigma_{FMsupport}$ ≤ 0.60. In Example 2, itemset A with $FM_{support}$ (A) = 1.00 is a high $FM_{support}$ itemset; however, itemset AC with $FM_{support}$(AC) = 0.45, itemset AD with $FM_{support}$(AD)=0.55 is a low $FM_{support}$ itemset.

Table 2: Clients FM values

Client ID	Frequency	Monetory	FM
1001	(2-1)/(2-1) = 1/1	(120-80)/(140-80) = 40/60	((1/1) + (40/60))/2 = 0.83
1002	(2-1)/(2-1) = 1/1	(140-80)/(140-80) = 60/60	((1/1) + (60/60))/2 = 1.00
1003	(1-1)/(2-1) = 0/1	(80-80)/(140-80) = 0/60	((0/1)+ (0/60))/2 = 0.00

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 3: Transaction data with FM weight

Transaction ID	Client ID	Itemsets	FM-Weight
1	1001	A,B,C	0.83
2	1002	A,B,D	1.00
3	1001	A,B,C	0.83
4	1002	A,B,D	1.00
5	1003	A,C,D	0.00
		Sum	3.66

Definition 4. The confidence of an association rule is defined in following Equation as the proportion of the number of transactions that include X and Y among all the transactions that comprise A. Here, X stands for the antecedent, Y stands for the consequent and X and Y are disjoint itemsets, that is, they have no components in like manner: $X \cap Y = \emptyset$. Rule $X \Rightarrow Y$ has a confidence $FM_{confidence}(X \Rightarrow Y)$ in D, where $FM_{confidence}(X \Rightarrow Y)$ is the percentage of transactions in D containing X that likewise contain Y.

$$\text{FMconfidence}(X \Rightarrow Y) = \frac{\text{FMsupport}(X \Rightarrow Y)}{\text{FMsupport}(X)} \quad (5)$$

Example 4. Let the maximum $FM_{support}$ and minimum $FM_{confidence}$ to 60% and 80%, respectively. In Example 2, $FM_{support}(C)=0.45$ and $FM_{support}(CA)=0.45$. In addition, $FM_{confidence}(C \Rightarrow A) = 0.45/0.45 = 1.0$

$C \Rightarrow A$(FMsupport $\leq$ 60%, FMconfidence = 100%)

Definition 5. Despite the way that most proposition in ARM depend on a support, confidence framework. The measure 'lift' was defined how many times the antecedent and the consequent occur together more frequently than would be normal on the off chance that they were factually autonomous. An association rule is interesting if its confidence is higher than the support of its consequent. On the other hand, if the confidence of the rule is equal to the support of its consequent, then both antecedent and consequent are independent. Let the $FM_{confidence}$ threshold be set as $\sigma_{FMconfidence} = 1.00$

$$\text{FMlift}(X \Rightarrow Y) = \frac{\text{FMconfidence}(X \Rightarrow Y)}{\text{FMsupport}(Y)} \quad (6)$$

Example 5. Let the maximum $FM_{support}$, minimum $FM_{confidence}$ and minimum FM_{lift} to 60% ,100%, 100%, respectively. In Example 2, $FM_{support}(A)=1.00$ and $FM_{confidence}(CA) = 1.00$ In addition, $FM_{lift}(C \Rightarrow A)= 1.00/1.00 = 1.00$. Let the FM_{lift} threshold be set as $\sigma_{FMlift} = 1.00$. In this manner, the following association rule is found.

C $\Rightarrow$A (FMsupport $\leq$ 60%, FMconfidence = 100%, FMlift=100%)

RIFM algorithm

Step1. Call the FM_Cal Subroutine

1. *For each client, calculate the FM value*
2. *For each transaction, set the FM value according to client ID*

Step2. Call the Rare_ itemsets_gen Subroutine

1. *For each item iT_i, calculate its FMsupport.*
2. *Check whether the FMsupport of each item iT_i is less than the maximum FMsupport(σFMsupport). If it is, put into the set of rare-itemsets (L_1).*
3. *Generate candidate set $C_{k+1}$from L_k.*
4. *Compute the FMsupport values of all itemsets in C_k and determine L_k.*
5. *If $L_{k+1}$is null, go to Step3; otherwise, set k=k+1 and repeat steps (3)-(5).*

Step3. Call the AR_gen Subroutine.

1. *If the FMconfidence of rule $X \Rightarrow Y$ is no less than the minimum FMconfidence(σFMconfidence), then*
2. *If the FM_{lift} of rule $X \Rightarrow Y$ is no less than the minimum $FM_{lift}(\sigma FM_{lift})$, then generate association ($X \Rightarrow Y$).*

IV. ALGORITHM FOR MINING RARE ITEMSETS FROM FM-WEIGHT TRANSACTIONS

We now clarify the proposed approach (RIFM) and give a case to delineate the strategy for finding the rare (low - $FM_{support}$) itemsets from FM-weighted transactions.

Proposed algorithm

The proposed RIFM algorithm is sketched out in Fig.1. The method can be partitioned into three steps: (1)Count of the FM value for every client and ensuing attaching of the FM value to each transaction; (2)Computation of the $FM_{support}$ of each itemset and generation of low-$FM_{support}$ itemsets; and (3)Generation of association rules from low-$FM_{support}$ itemsets. The well ordered figuring process is as per the following:

Step 1: Preprocess data

For discovering rare (low-$FM_{support}$) itemsets from FM-weighted transactions, each client's FM value must be calculated in advance. Subsequently, each client's FM value must be appended to the transactions in which the client's purchasing behaviors are recorded.

Step 2: Discover low-$FM_{support}$ itemsets from FM-weighted transactions

In this study, a level-wise approach is utilized in the first phase to iteratively create candidate itemsets of k items (C_k), and to then find low-$FM_{support}$ itemsets of k items (L_k). To continue to the following level, we produce a candidate set C_{k+1} from L_k and repeat. Dissimilar to conventional RARM approaches, the itemsets $FM_{support}$ is in the vicinity of 1 and 0, instead of 1 or 0. Moreover, we should whole the aggregate $FM_{support}$ for all transactions.

Step 3: Generate association rules from all rare (low- $FM_{support}$) itemsets.

In this step, we calculate the $FM_{confidence}$ values no less than the minimum $FM_{confidence}(\sigma_{FMconfidence})$, then the we distinguish the association rules with $FM_{lift}(\sigma FM_{lift})$ values is no less than 1.

Example 1. An example is given to show the proposed data mining algorithm. The data set utilized as a part of this case is exhibited in Table 3

Step 1: Preprocess data

We first figure every client's frequency value (F). Table 3 demonstrates that the most extreme tally and least tally of transactions for all clients. In Likewise, the tally of exchanges for clients 1001, 1002, and 1003 are 2, 2, and 1, individually. Definition 1 shows that every client's F can be computed as 1/1, 1/1, and 0/1.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Second, we figure every client's monetary value (M). Table 1 indicates that the most extreme monetary sum and least monetary sum of all clients are 140(70 + 70) and 80, individually. Furthermore the monetary sums for clients 1001, 1002, and 1003 are 120, 140, and 80, separately. Definition 1 demonstrates that every client's M can be processed as 40/60, 60/60, and 0/60.

At long last, each client's FM value can be registered as 0.83, 1.00, and 0.00. Besides, we append each client's FM value to the transactions in which the client's purchasing behaviors are recorded.

Step 2: Discover rare (low-FM) itemsets from FM-weighted transactions

Expect that we set the maximum $FM_{support}$ ($\sigma_{FMsupport}$) to 0.60. For each itemset put away in the transactions, we compute the itemset's $FM_{support}$ and look at whether the $FM_{support}$ of each itemset is lesser than or equivalent to the maximum $FM_{support}$ ($\sigma_{FMsupport}$). The rare itemsets found in this investigation are exhibited in Tables 5.

Step 3. Generate association rules from all rare (low-$FM_{support}$) itemsets

We can generate rules from all rare (low-$FM_{support}$) itemsets. We set the maximum $FM_{support}$($\sigma FM_{support}$) to 60%, minimum $FM_{confidence}$($\sigma_{FMconfidence}$) to 100% and minimum FM_{lift}(σFM_{lift}) to 100%. For brevity, we present only some generated association rules in Table 6.

By using the proposed RIFM approach, the revenues of top k (k=7) rare itemsets found. In addition both low-frequency and high revenue itemsets can be found, rather than only low frequency itemsets.

Table 4: Top k (k=7) Rare itemsets found using RIFM

Itemsets	FMsupport
C	0.45
D	0.55
A,C	0.45
A,D	0.55
B,C	0.45
B,D	0.55
A,B,C	0.45

Table 5: Revenue of Top k (k = 7) rare itemsets found using RIFM

No	Itemsets	FMsupport	Revenue
1	D	0.55	90
2	D,A	0.55	180
3	D,B	0.55	80
4	A,B,D	0.55	70
5	C	0.45	60
6	C,A	0.45	150
7	C,B	0.45	60
		Sum	690

Table 6: Some association rules identified using RIFM

No	*Association rule*	*FMsupport*	*FMconfidence*	*FM lift*
1	*D⇒A*	*60%*	*100%*	*100%*
2	*C⇒A*	*60%*	*100%*	*100%*

V. EXPERIMENTAL RESULTS

We applied online retailer data sets to assess the execution of the proposed RIFM algorithm. This datasets have unique sale properties (such as products, clients, and purchasing transactions).We theorized whether the top k infrequent itemsets with lower support found utilizing the proposed RIFM algorithm are more appropriate for foreseeing clients purchasing itemsets and existing revenue in the next-period transactions. In light of the training dataset (D_A) and test dataset (D_B), we investigate the performance in prediction of clients purchasing itemsets existing in the next-period transactions. In this manner, we characterize two metrics (Precision and Recall) for predicting clients purchasing itemsets in the next-period transactions. In addition, we characterize the metric (Revenue) to examine the performance in anticipating income of clients purchasing itemsets in the next-period transactions.

Consequently, there are three metrics (Precision, Recall, and Revenue) characterized to assess the performance of the infrequent itemsets that existed in the transaction dataset in the following periods found utilizing the proposed approach. Let A and B be the sets of patterns generated from training dataset (D_A) and test dataset (D_B), respectively. Revenue(A∩B) represents the revenue generated from the sets of patterns (A∩B). The three metrics can be characterized as takes after:

$$Precision = \frac{|A \cap B|}{|D_A|} \quad (7)$$

$$Recall = \frac{|A \cap B|}{|D_B|} \quad (8)$$

$$Revenue = Revenue\ (A \cap B),$$

Where

|A∩B| indicates the transaction numbers of itemset (A∩B) in transaction set (D_B);

$|D_A|$ indicates the transaction numbers in transaction set (D_A); $|D_B|$ means the transaction numbers in transaction set (D_B); Revenue(A∩B) signifies the revenue value of itemset (A∩B) in transaction set (D_B).

Dataset description

The client transaction dataset held by the trader has 11 variables and it contains all transactions happening in years 2010 and 2011. Every transaction contains information about the retail records in the online-retailer store. After we played out the vital information preprocessing errands, the dataset included 990 records of transactions identified to distinct product types. It is intriguing to see that the average number of distinct products (items) contained in each transaction occurring in 2011 was 18.3 (=406830/22190). The product cost per unit in sterling 45.23. The transaction sizes in the first, second, third, and fourth weeks were 173, 231, 275, and 311 separately. We utilized the transactions in the first week as the training dataset. Accordingly, we chose the top k infrequent itemsets with lower support to investigate the numbers of patterns in L_1, L_2, and L_3 for top k infrequent itemsets with lower support.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

In this way, exploring the performance difference between the two methodologies (AprioriRare and RIFM) in predicting clients purchasing itemsets and the revenue from client ahead of time is enlightening.

Comparison of revenue predictions

We set the maximum support to 0.001 and varied the top k from 10 to 100 to examine the revenue differences between the two methodologies. Figs.1-3 exhibit the differences between the two methodologies in predicting revenues produced from clients purchasing itemsets. The experimental results of the revenue value predictions uncover that the proposed approach outperformed AprioriRare in predicting revenues of itemsets from clients in next-period transactions. What's more, the proposed approach marginally outflanked AprioriRare in predicting revenues in the top 200 patterns.

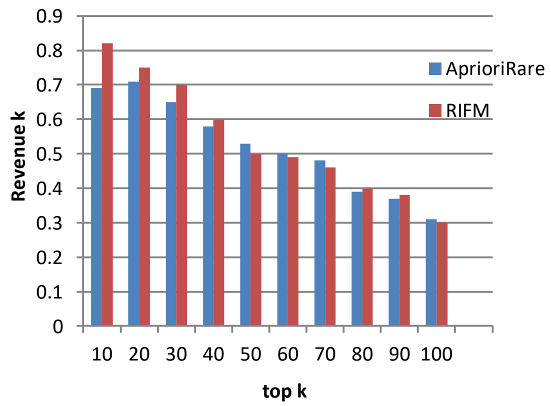

Fig. 1: Second quarter (revenue)

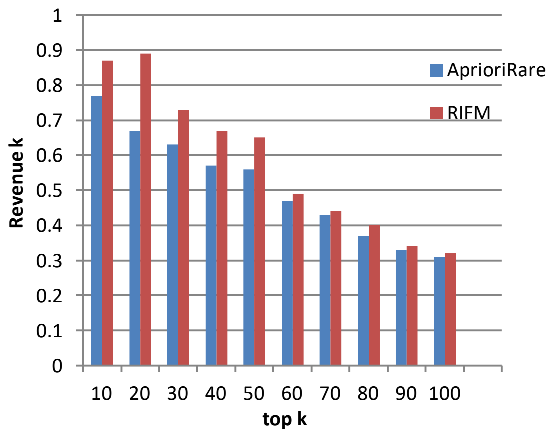

Fig. 2: Third quarter (revenue)

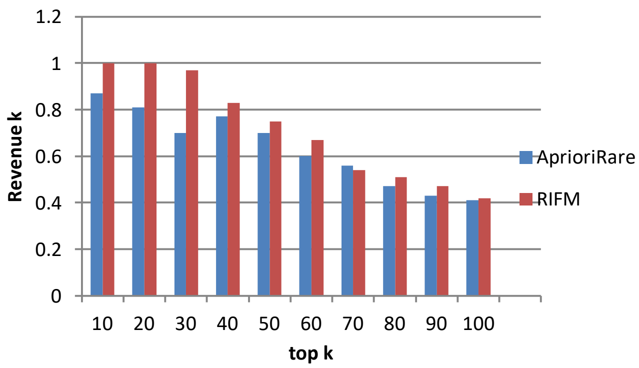

Fig. 3: Fourth quarter (revenue)

VI. CONCLUSION

In light of their practicality, ARM algorithms have been as a part of different applications and data sets. This study is the first to introduce the most capable technique for finding rare itemsets from transactions by considering the client's FM value. Moreover, we proposed a new algorithm RIFM, to find rare itemsets from FM-weighted transactions. Experimental results from the survey data uncover that the proposed approach can empower the disclosure of intriguing and significant patterns that have not been found using conventional approaches. In addition, the top k rare itemsets found utilizing the proposed RIFM approach outperformed those determined utilizing the traditional approach in the prediction of revenue from clients in next period transactions.

A few issues remain to be tended to. To begin with, we concentrated on finding rare itemsets from FM-weighted transactions instead of utilizing all the transactions with equal weights. In some applications, business heads may be enthused about things with various weights. What's more, applying new extra measures to extend the support-confidence and lift framework to discover interesting association rules could be profitable. To address this issue, a more effective algorithm ought to be planned. At last, we recommend that the proposed approach be refined to discover rare itemsets from FM-weighted transactions.

REFERENCES

1. Agrawal, R., Imieliński, T., Swami, A. 1993. Mining association rules between sets of items in large databases. In: Proceedings of ACM SIGMOD. Washington, DC, USA. pp. 207–216.
2. Han, J.W., Kamber, M., 2006. Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco.
3. R.C. Blattberg B.D. Kim S.A. Neslin. 2008. Database Marketing: Analyzing and Managing Clients (Chapter 12, Series eds. J. Eliashberg). Springer. New York, USA.
4. Shim, B., Choi, K., Suh, Y., 2012. CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns. Expert Syst. Appl. 39 (9), 7736–7742.
5. Dursun, A., Caber, M., 2016. Using data mining techniques for profiling profitable hotel clients: an application of RFM analysis. Tour. Manag. Perspect. 18, 153–160.
6. Szathmary L, Valtchev P, Napoli A(2010) Generating rare association rules using the minimal rare itemsets family. Int J Softw Inf 4(3):219-238.
7. Koh YS, Rountree N(2005) Finding Sporadic rules using apriori inverse. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 3518:97-106
8. J.M.Luna, J.R.Romero, S. Ventura 2014, On the adaptablility of G3PARM to the extraction of rare association rules. Knowl. Inf Syst 38: 391-418.
9. Lu, S., Hu, H., Li, F., 2001. Mining weighted association rules. Intell. Data Anal. 5 (3), 211–225.
10. Wang, W.,Yang, J.,Yu, P., 2004. WAR:weighted association rules for item in-tensities. Knowl. Inf. Syst.6(2),203–229.
11. Yun, U., Leggett, J.J., 2005. WFIM: weighted frequent

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

itemset mining with a weight range and a minimum weight. In: Proceeding of the 2005 SIAM International Conference on Data Mining (SDM'05). Newport Beach, CA. pp. 636–640.

12. Vo, B., Coenen, F., Le, B., 2013. A new method for mining frequent weighted itemsets based on WIT-trees. Expert Syst. Appl. 40 (4), 1256–1264.
13. Kalakota, R., Robinson, M., 1999. e-Business Roadmap For Success. Addison Wesley Longman Inc, New York, USA.
14. Peppard, J., 2000. Customer relationship management (CRM) in financial services. Eur. Manag. J. 18 (3), 312–327.
15. Hughes, A.M., 2006. Strategic Database Marketing. McGraw-Hill.
16. Linoff, G.S., Berry, M.J.A., 2002. Mining the Web: Transforming Client Data into Client Value. John Wiley and Sons, New York, NY.
17. Liu, D.R., Shih, Y.Y., 2005. Integrating AHP and data mining for product re-commendation based on client lifetime value. Inf. Manag. 42 (3), 387–400.
18. Li, L.H., Lee, F.M., Liu, W.J., 2006. The timely product recommendation based on RFM method. In: Proceedings of International Conference on Business and In-formation. Singapore.
19. Kim, H.K., Im, K.H., Park, S.C., 2010. DSS for computer security incident response applying CBR and collaborative response. Expert Syst. Appl. 37 (1), 852–870.
20. Chan, C.C.H., 2008. Intelligent value-based client segmentation method for campaign management: a case study of automobile retailer. Expert Syst. Appl. 34 (4), 2754–2762.
21. Hsieh, N., 2004. An integrated data mining and behavioral scoring model for analyzing bank clients. Expert Syst. Appl. 27 (4), 623–633.
22. Lin, C.S., Tang, Y.Q., 2006. Application of incremental mining and client's value analysis to collaborative music recommendations. J. Inf. Technol. Soc. 6 (1), 1–26.
23. Chiang, W.Y., 2011. To mine association rules of client values via a data mining procedure with improved model: an empirical case study. Expert Syst. Appl. 38 (3), 1716–1722.
24. Hu, Y.H., Yeh, T.W., 2014. Discovering valuable frequent patterns based on RFM analysis without client identification information. Knowl. Based Syst. 61, 76–88.
25. Cheng-Hsiung Weng 2017 Revenue prediction by mining frequent itemsets with customer analysis, Engineering applications of artificial intelligence 63 (2017) 85-97.