A Novel Approach For Improving Crm With Profitable Itemsets

Lakshmi Surekha T, Rama Devi.P , Srilekha Ch

[1]Assistant Professor, Department of IT, VR Siddhartha Engineering College, Andhra Pradesh, India.

[2]Assistant Professor, Department of IT, VR Siddhartha Engineering College, Andhra Pradesh, India.

[3]Student, Department of IT, VR Siddhartha Engineering College, Andhra Pradesh, India.

Abstract- *In recent years, retaining the customer has becoming a challenging issue for many of the organizations on par with finding profitable item sets. Association rule mining algorithms which are binary in nature like Apriori and FP- Growth were developed but they do not consider quantity and profits (profit per unit) which are considered as important for finding the Profitable Item sets. In this a weighted frame work has been discussed by taking into account the profit (intensity of the item) and the quantity of each item in each transaction of the given dataset. Here we proposed "Profitable Weighted FP Growth" algorithm in this paper, which find outs the profitable Item sets. From this Profitable Item sets we are retrieving the valued customer who plays a key rule for the profit of the organization and maintaining a "Customer Relation Management"(CRM) with those customer by offering some discounts etc.,.*

Keywords (Size 10 & Bold) — Profit, Quantity, Profitable Item sets, Profitable Weighted FP Growth, (CRM)Customer Relationship Management.

I. INTRODUCTION

Mining frequent patterns or Itemsets is an important aspect in the field of data mining .Traditional Algorithms like Apriori and Frequent Pattern Growth algorithms are binary in nature. They only consider whether an Item is sold are not, and they are generating frequent Patterns. These algorithms concentrates on the frequent occurrences of the items and produce frequent items sets, which only consider the occurrence of items but do not reflect any other factors, such as price or profit. In Profitable Itemset Mining, transactions are attached with weighted values according to some criteria, it is important because if we consider support and confidence as the parameters, there are chances that they can ignore some of the profitable patterns. The problem of Profitable Itemset mining is to find the complete set of Itemsets satisfying a minimum profit constraint in the database. . Profitable Itemset mining has been suggested to find Profitable patterns by considering the profits as well as quantity of Items. Here we are considering two important parameters for finding the profitable item sets they are quantity and profit per item.

1.1 BASIC CONCEPTS

Frequent Itemset mining helps us to find the frequent patterns or itemsets . The two most widely used algorithms are Apriori and FP Growth. These two algorithms are binary in nature. They concerned about whether the product is sold or not. The measures considered by these algorithms are support and confidence. But in reality they are not sufficient for decision making in the large organizations. So In this framework we consider two measures named **Quantity** and **Profit**. By using both the parameters we calculate Weight.
Consider the following two transactions:

T1	10Buttermilk packets, 1 Milk shake
T2	1 Buttermilk packets, 10 Milk shake

By using traditional algorithms that are considering support and confidence the above two transactions are considered to be the same, since the quantity of an item is not taken into account. But in reality, it is quite clear that the transaction T2 gives more profit than the transaction T1 if we consider the profit for selling the milk shake is higher than that of butter milk. Thus to make efficient marketing we take in to account the quantity of each item in each transaction. In addition we also consider the intensity of each item, which is represented using profit per item p.
Consider the following two transactions:

T3	10Bread packets, 1 choclate
T4	10 choclates, 5 bread packets

In reality the quantity sold in transaction T4 is greater than transaction T3, but the amount of profit gained by selling a Bread packet is 10 times that of a choclate. So, the profit is also given priority

represented by p. "p" may represent the retail price / profit per unit of an item/ intensity of the disease in case of medical diagnosis. After finding the profitable item sets we need to find out the value customer of our organization because Since retaining the valued customer is a challenging task for the organization. Valued customer is one who purchases profitable items.

II PROPOSED WORK

Our paper consists of two phases . First phase is finding Profitable items by using " Profitable Weighted FP Growth algorithm" and the second phase is identifying valued customers using " Identification valued customers algorithm.

A . PROFITABLE WEIGHTED FP GROWTH

Profitable Weighted FP-Growth algorithm is based on FP-Growth algorithm in data mining. Generally FP-Growth algorithm is binary in nature. It doesn't consider quantity and weight per unit in a transaction. So In our algorithm we consider quantities and profits per items in the transactions.

ALGORITHM

Input:

1. D, a Transactional database that includes customer Id, Items purchased with its quantity.
2. Min_sup, the minimum support count threshold.
3. Profit table that displays profit earned by each Item.

Output:

The complete set of profitable patterns.

Method

Step1:

1. Scan the database and develop a new table by multiplying profit of each item with their corresponding quantity, represented as weight .
2. Calculate the frequency for each item i.e., sum of Weights of each item in all the transactions.
3. Discard infrequent items.
4. Sort frequent items in decreasing order based on their support and a new database table is generated.
5. Use this order when building the FP-Tree, so common prefixes can be shared.

Step 2:

1. Profitable Weighted FP-Growth reads the first transaction at a time and maps it to a path.
2. Fixed order is used, so paths can overlap when transactions share items (when they have the same prefix). In this case, counters are incremented by their corresponding weights
3. Pointers are maintained between nodes containing the same item, creating singly linked lists (dotted lines). The more paths that overlap, the higher the compression. FP-tree may fit in memory.
4. Frequent item sets are extracted from the Profitable Weighted FP Tree.

B. IDENTIFICATION OF VALUED CUSTOMER

1. Profitable Item sets are identified with Profitable Weighted FP-Growth algorithm.
2. Retrieve the Valued customers Information from the database by using these profitable items.
3. Special Discounts are provided to those customers on the next purchases.

C. FLOW DIAGRAM FOR PROFITABLE WEIGHTED FP-GROWTH ALGORITHM.

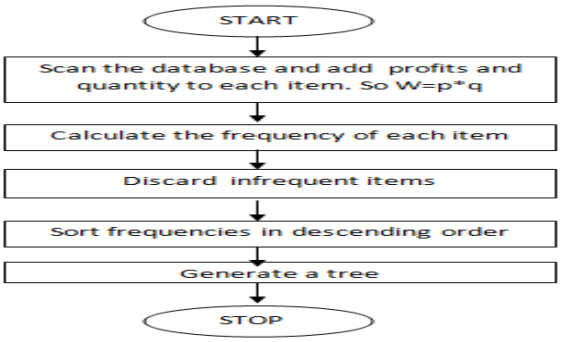

Fig1:Flow chart

III EXAMPLE BASE TABLE

TID	CID	BISCUITS	PASTRIES	ICECREAMS	CHOCOLATE	MILK SHAKE	BUNS	Add New Field
1	101	1	0	1	0	0	0	
2	102	0	0	2	0	1	1	
3	101	1	1	0	2	1	0	
4	105	0	0	0	1	0	2	
5	106	2	2	3	0	2	2	
6	103	0	3	2	2	0	0	
7	102	3	0	0	0	3	0	
8	105	0	1	2	1	0	3	
9	103	0	2	0	0	0	0	
10	109	1	0	1	0	0	2	
11	110	0	3	2	2	1	1	
12	114	2	0	0	1	2	0	
13	107	0	2	3	0	0	2	
14	106	3	0	2	3	3	0	
15	105	0	1	0	0	2	1	
16	103	0	0	0	2	0	2	
17	104	2	2	1	0	0	0	
18	114	0	2	0	0	1	3	
19	101	1	1	2	1	2	0	
20	102	0	0	0	1	0	2	
21	106	0	0	3	0	0	1	
22	107	2	3	0	0	3	0	
23	120	0	0	2	0	2	0	

Fig 2: Bakery dataset

PROFIT TABLE

	ITEM ID	ITEM	PROFIT/UNIT
	BAKERY DATASET(PROFITS)		
	1	BISCUITS	5
	2	PASTRIES	2
	3	ICECREAMS	1
	4	CHOCOLATES	4
	5	MILK SHAKES	3
	6	BUNS	4

Fig 3: Profit dataset

The quantities and profits are multiplied in order to get weights. Thus total profits are obtained.

WEIGHTED DATASET

Problems | Javadoc | Declaration | Console

<terminated> Weights1 [Java Application] C:\Program Files\Java\jre1.8.0_112\bin\j

The Weights table is

I1	I2	I3	I4	I5	I6
5	0	1	0	0	0
0	0	2	0	3	4
5	2	0	8	3	0
0	0	0	4	0	8
10	4	3	0	6	8
0	6	2	8	0	0
15	0	0	0	9	0
0	2	2	4	0	12
0	4	0	0	0	0
5	0	1	0	0	8
0	6	2	8	3	4
10	0	0	4	6	0
0	4	3	0	0	8
15	0	2	12	9	0
0	2	0	0	6	4
0	0	0	8	0	8
10	4	1	0	0	0
0	4	0	0	3	12
5	2	2	4	6	0
0	0	0	4	0	8
0	0	3	0	0	4
10	6	0	0	9	0
0	0	2	0	6	0
15	4	0	8	0	12
0	0	3	0	0	8
10	2	3	12	3	0
0	0	0	0	0	8

Fig 4: Weighted dataset

FORMAT FOR THE INPUT DATA TO THE ALGORITHM

input.txt

```
1 3:6:5 1
2 3 5 6:9:2 3 4
3 1 2 4 5:18:5 2 8 3
4 4 6:12:4 8
5 1 2 3 5 6:31:10 4 3 6 8
6 2 3 4:16:6 2 8
7 1 5:24:15 9
8 2 3 4 6:20:2 2 4 12
9 2:4:4
10 1 3 6:14:5 1 8
11 2 3 4 5 6:23:6 2 8 3 4
12 1 4 5:20:10 4 6
13 2 3 6:15:4 3 8
14 1 3 4 5:38:15 2 12 9
15 2 5 6:12:2 6 4
16 4 6:16:8 8
17 1 2 3:15:10 4 1
18 2 5 6:19:4 3 12
19 1 2 3 4 5:19:5 2 2 4 6
20 4 6:12:4 8
21 3 6:7:3 4
22 1 2 5:25:10 6 9
23 3 5:8:2 6
24 1 2 4 6:39:15 4 8 12
25 3 6:11:3 8
26 1 2 3 4 5:30:10 2 3 12 3
27 6:8:8
28 1 2 3 4 5 6:46:15 4 1 8 6 12
29 3 4 5 6:29:2 12 3 12
30 2 6:14:2 12
```

Fig 5: Input data format

consider the following Example transaction:

3 5 6 :9 :2 3 4

This means that there are three items in the transaction I3, I5, I6 And the corresponding profits obtained by purchasing those items are given after the second colon(i.e., 2 3 4). Finally the total profit obtained by the transaction is 9 i.e. in between the two colon symbols.

***PATTERNS OBTAINED FOR PROFITABLE WEIGHTED FP-GROWTH*:**

Weights1.java | ProfitableFPGrowth.java | patterns.txt

```
1 PATTERN :  1  TOTAL PROFIT : 530
2 PATTERN :  6  TOTAL PROFIT : 508
3 PATTERN :  1 4  TOTAL PROFIT : 477
4 PATTERN :  4 6  TOTAL PROFIT : 524
5 PATTERN :  4 5  TOTAL PROFIT : 463
6 PATTERN :  1 3  TOTAL PROFIT : 427
7 PATTERN :  1 6  TOTAL PROFIT : 530
8 PATTERN :  1 5  TOTAL PROFIT : 601
9 PATTERN :  5 6  TOTAL PROFIT : 476
10 PATTERN :  1 4 5  TOTAL PROFIT : 511
11 PATTERN :  4 5 6  TOTAL PROFIT : 491
12 PATTERN :  1 3 6  TOTAL PROFIT : 484
13 PATTERN :  3 5 6  TOTAL PROFIT : 432
14 PATTERN :  1 3 5  TOTAL PROFIT : 520
15 PATTERN :  1 5 6  TOTAL PROFIT : 530
16 PATTERN :  1 4 5 6  TOTAL PROFIT : 417
17 PATTERN :  1 3 4 5  TOTAL PROFIT : 481
18 PATTERN :  3 4 5 6  TOTAL PROFIT : 445
19 PATTERN :  1 3 5 6  TOTAL PROFIT : 548
20 PATTERN :  1 3 4 5 6  TOTAL PROFIT : 441
21
```

Fig 6: Profitable patterns

The above figure is the output for Profitable Weighted FP-Growth algorithm. The output in the format.

18 **PATTERN : 3 4 5 6 TOTAL PROFIT :445**

This means that the profit of the Items 3,4,5 and 6 in the entire database is 445.

8 **PATTERN : 1 5 TOTAL PROFIT :601**

This means that the profit of the Items 1 and 5 in the entire database is 308.

2 **PATTERN : 6 TOTAL PROFIT :508**

This means that the profit of the Item 6 in the entire database is 508.

PROFITABLE WEIGHTED FP GROWTH STATISTICS

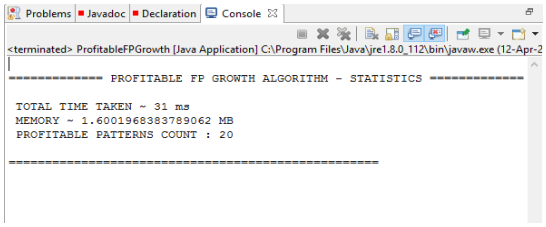

Fig 7: Statistics

The customers who purchased Items 1 ,5 and 6 are considered as most valued customers. And they are provided with some special offers on their next purchase. These are some of the customers.

Customer Id
101
102
103
105
106
107
110
114

Fig 8:Valued Customers

REFERENCES

[1] R. Agrawal, T. Imielinski and A. Swami, “Mining Association Rules Between Sets of Items in Large Databases”, Proc. 1993 ACM SIGMOD,Washington, DC, pp. 207–216, May 1993

[2] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules”, Proc. 20th International Conference on Very Large Databases (VLDB‟94), Santiago, Chile, pp. 487–499, Sept. 1994

[3] J. Han and Micheline Kamber, “Data Mining – Concepts and Techniques”, Morgan Kaufmann Publishers, 2001

[4] T.Lakshmi Surekha and Ch.Srilekha, “Profitable Itemset Mining Using Weights”, International Research Journal of Engineering and Technology (IRJET) Volume 4, Issue 3 ,March 2017.

[5] T.Lakshmi Surekha, P. Rama devi, “ Profitable Association Rule Mining Using Weights”,IJCTT, May 2016.

[6] Robert J. Hilderman, Colin L. Carter, Howard J. Hamilton, And Nick Cercone, “Mining Association Rules From Market Basket Data Using Share Measures and characterized Itemsets”

[7] Feng Tao, Fionn Murtagh, Mohsen Farid, “Weighted Association Rule Mining using Weight ed Support and Significance framework”

[8] Wang, B-Y., Zhang, S-M.: A Mining Algorithm for Fuzzy Weighted Association Rules. In: IEEE Conference on Machine Learning and Cybernetics, 4, pp. 2495--2499 (2003).

[9] Unil Yun, John J. Leggett, ‘WFIM: Weighted Frequent Itemset Mining with a weight range and a minimum weight‟ SDM`05, April 2005