International Journal of Engineering & Technology, 7 (2.33) (2018) 796-798

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Impact of modified frequent-pattern generation on cross-selling

Ms. Sarmistha Saha [1] *, Mr. G. P. Saradhi Varma [2]

[1] Ph.D. Scholar, Computer Science, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India
[2] Principal, Computer Science, S.R.K.R. Engineering College, Bhimavaram, Andhra Pradesh, India
**Corresponding author E-mail: sarmist@rediffmail.com*

Abstract

The banking sector is an integral part of the economy. Hence this sector plays a key role in the well being of the economy. In the modern era, all the banks are being computerized to handle smoothly ample of data. In this regard, they all need data mining techniques to discover patterns for unknown relationship of customer data on daily basis. Frequent – Pattern tree construction is one of the popular method. This paper represents new approach for frequent-pattern tree construction where more than one item has same support count and it is mainly useful for cross-selling of products by the banks.

Keywords: *Confidence; Cross-Selling; FP-Tree; Market Basket Analysis; Support.*

1. Introduction

Data mining is an emerging technology for the automatic extraction of patterns, associations for large data sets. It is the one of the key technologies which enable business to select, filter and correlate data automatically. Cross-selling for banks uses the wonderful technique of Data Mining i.e. Market Basket Analysis, where bank sells related products to their existing customers.

Market Basket Analysis is a modeling technique based upon the theory that if we buy a certain group of items, we are more (or less) likely to buy another group of items. The set of items a customer buys is referred to as an itemset, and market basket analysis seeks to find relationships between purchases.

Typically the relationship will be in the form of a rule:

IF {milk, sugar} THEN {bread}.

The probability that a customer will buy bread without milk or sugar (i.e. that the antecedent is true) is referred to as the support for the rule.

i. e support{(milk, sugar)→bread}

#_tuples _containing milk, sugar, bread
= --
Total_# _ of tuples

The conditional probability that a customer will purchase bread crisps is referred to as the confidence.

i.e confidence{(milk, sugar)→ bread}

#_ tuples _ containing milk, sugar, bread
= ------------------------------------
#_tuples _containing milk, sugar

The above concept is applicable for finding patterns to the related products offered by banks to the customers. FP-Growth tree is one of the key method for market basket analysis and it is also used immensely by the financial company. Hence it is now become popular all over the world . Now this modified Frequent-Pattern generation has a very large impact on Cross-Selling.

2. Research methodology

FP-Tree is constructed using [2] passes over the data-set (item set).

Data -Set (item set)

E→Product1
B→Product2
J→Product3
Bt→Product4
M→Product5
S→Product6

Table 1: Transactional Database

Transactions	Itemsets
T1	E,B
T2	E,B,J
T3	M,B,S
T4	M,B,Bt,S
T5	M,B,J,S
T6	B,Bt
T7	M,B,S
T8	M,B,Bt,J,S
T9	B,J
T10	M,S

Table 2: A Binary 0/1 Representation of Market Basket Data

TID	Prod.1	Prod.2	Prod.3	Prod.4	Prod.5	Prod.6
1	1	0	0	0	1	0
2	1	0	1	0	1	0
3	1	0	0	1	0	1
4	1	1	0	1	0	1
5	1	0	1	1	0	1
6	1	1	0	0	0	0
7	1	0	0	1	0	1
8	1	1	1	1	0	1
9	1	0	1	0	0	0
10	0	0	0	1	0	1

2.1. FP-Tree is constructed using two passes over the data set (item set)

Pass 1:

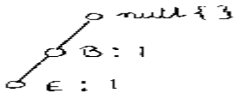
a) We scan data and find support for each item.
b) We discard infrequent items.
c) We Sort frequent items in decreasing order

Based on their support.
For our example: B,M,S,J,Bt,E
We use this order when building the FP-Tree,
So common prefixes can be shared.
Pass 2:
We construct the FP-Tree similar to method
Given by Data Mining author Kamber except for
Items having same support count.
In our present example, the support count for
Both Product5 and Product6 is six(6).

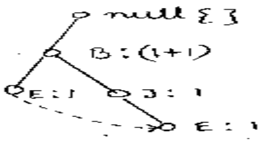
2.2. New methodology for creating nodes in FP-tree

According to data mining FP- tree construction method,
We have learnt that whenever there is different support count of item sets, we use to put those items in different level. But there is no such method discussed till now, when we have items with same support count.
In this section, we are going to discuss about our new methodology, where we are introducing a new technique to draw FP- tree with items with same support.
Transactions T3, T4, T5, T7, T8, and T10 consist of items product5 and product6 along with other items.
For these above transactions we have kept both items product5 and product6 in the same level.
Taking root node as a null node designated as null{},we have proceeded to draw a complete FP-tree by following steps (step-1 to step-10) as shown.
Taking into account, Table-1 as transactional database, the following complete FP-tree has been constructed where numbers are assigned for corresponding support count in the appropriate level and addition has been done whenever it is necessary.`

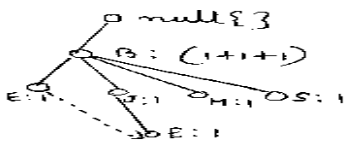
step-1 :



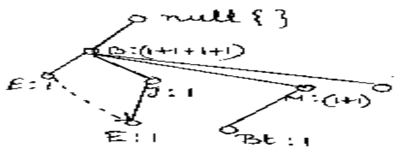
step-2 :



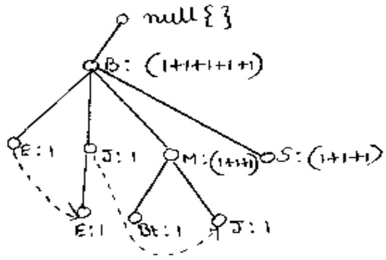
step- 3 :



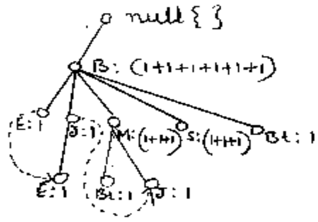
step- 4 :



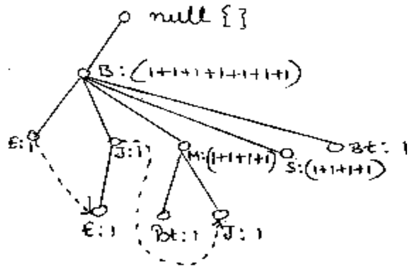
Step-5:



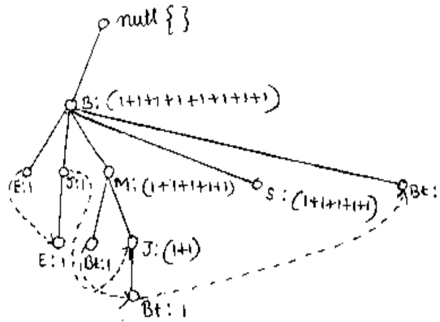
Step-6:



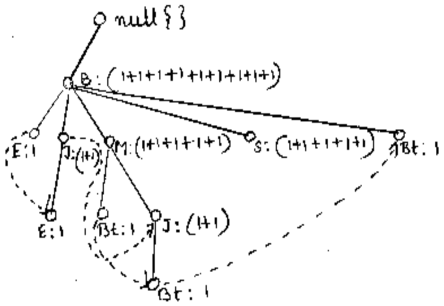
step-7 :



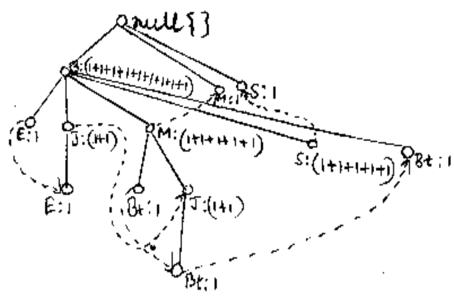
step-8



Step-9



Step-10



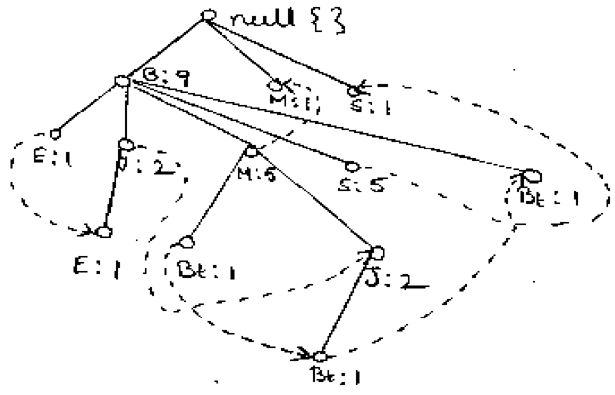


FIGURE I

Fig. 1: Represents The Complete FP-Tree According New Methodology.

Time taken for searching an element in a tree =O (b^d) where

b=branching factor,

d=depth of a tree where the

Element lies.

For searching S (Product6) according to our method:

Time complexity = O ($b_1^{d_1}$) where $b_1 = 5$, $d_1=1$

= O (5^1) (putting value of b_1, d_1) = O (5)

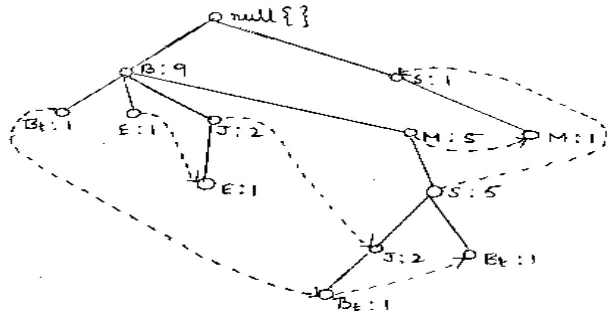


FIGURE II

Fig. 2: Represents the Complete FP-Tree According to Data Mining Author Jiawei Han and Micheline Kamber.

For searching S (product6) according to Han and Kamber,

Time complexity = O ($b_2^{d_2}$) where $b_2 = 4$, $d_2=2$ = O (4^2) = O (16)

Table 3: Mining Frequent Pattern without Candidate Generation

It Em	Conditonal Pattern Base	Conditional Fp-Growth Tree	Frequent Patterns Generated
M	B:5	B:5	M:B:5
S	B:5	B:5	S: B:5
J	{B:2}, {B:M:2}	B:2	J:B:2
Bt	{B:M:1},{B:M:J:1},{B:1}	B:1	Bt:B:1
E	{B:1},{B:J:1}	B:1	E:B:1

2.3. Advantage: time complexity = O (b1d1)

Where

b_1 = branching factor

d_1=depth of a tree

Time complexity according to earlier algorithm proposed by Han Kamber is O ($b_2^{d_2}$),

Where $b_2 < b_1$ and $d_2 > d_1$

I.E O ($b_1^{d_1}$) < O ($b_2^{d_2}$) (i)

3. Conclusion

This is an efficient and scalable method for constructing Frequent-Pattern tree.It shows some variations in earlier method (proposed by Han and Kamber). It is proved that this method has better performance than earlier method for mining frequent patterns (shown in (i)).

Acknowledgement

It is my pleasure to express our sincere gratitude to Acharya Nagarjuna University for providing me an opportunity to do our work. I sincerely thank to my guide,Prof. Partha Saradhi Varma for guiding and encouraging in carrying out this paper work.

References

[1] Brieman. L., J. Friedman, and R. Olshen, C.Stone(1994), Classification and Regression Trees,Pascific Grove: Wadsworth.

[2] Steinberg, D. and Colla, P.L., (1995), CART: Tree- Structured Nonparmetric Data Analysis, San Diego, CA: Salford Systems.

[3] Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann , 2008

[4] PangNing Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson Education, 2008.