DOI: http://dx.doi.org/10.26483/ijarcs.v9i1.5402

ISSN No. 0976-5697

Volume 9, No. 1, January-February 2018

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

IMPROVED EVENT DATA SCHEDULING FRAMEWORK THROUGH OPTIMIZED FP-GROWTH ALGORITHM

Dr. J. Chockalingam
M.Sc., M.B.A.,M.Phil., Ph.D.,
Associate Professor
Research Supervisor,
Computer Science department
Khadir Mohideen College, Adirampattinam-India

C. Seetharaman
M.Sc., M.Phil.,B.Ed.,(Ph.D)
Associate Professor Govt. Arts College(A), Kumbakonam
Research Scholar,
Computer Science department
Khadir Mohideen College, Adirampattinam-India

***Abstract*:** Education is the important tool and plays a vital role in the nation development. In recent days, modern technologies are used for the educational tool. Nevertheless the students and trainers should be trained thoroughly about their subjects. In this work, data mining techniques help to perform decision making such as who are the trainee for the program and which are the skill programs need to be offer and schedule the training sessions based on the trainee time schedule. Officially the events can be conducted for the students as well as for the faculties. In this proposed work, a novel algorithm will be produced for the event data scheduler using Association Rule Mining. The Optimized Frequent Pattern Growth algorithm can be proposed to obtain the market basket analysis. The main contribution of the proposed work is to obtain the optimized decision making for the training sessions to make the event scheduling with improved performance. Thus the decision making can be obtained through the knowledge obtained from the behavior patterns identified by the heuristics detector. The behavior pattern helps to provide efficient event data scheduling framework those who are in the educational sector. The experiment results can be generated for the analyzing the performance and effectiveness of the proposed work in the tools defined for the data mining using education dataset as evaluation values for teaching assistant from UCI repository.

Key terms: Association Rule Mining, Market Basket Analysis, Frequent Patter Growth Algorithm, heuristics detector and Behavior Pattern

I. INTRODUCTION

Data mining techniques [1] [2] [3] are used to obtain the analytical results for very huge amount of data stored in the dataset as data warehouse. The results obtained after the processing the data mining is so called as interestingness knowledge and information. Many proposed existing techniques [8] for market basket analysis have been proposed such as decision tress, neural networks and association rules. Association Rule Mining is the important measure for obtaining the market basket analysis through the algorithms.

Association Rules [3] were first introduced in the form of “if-then” statements. That the rules can be generated as based on the dataset which derives from the support and confidence value given as input from the user. The most important and familiar ARM algorithms such as SETM Algorithm, Apriori Algorithm [10] were used for the generation of market basket analysis to obtain the support and confidence measures. This might be used for most of the vital applications if it needs to generate the candidate generations or it need to generate the candidate using frequent itemsets. The problem of finding the results over the association rules is usually decomposed into two sub problems: (a) Find all frequent data sets as feedback details about the teaching aids using Minimum support. (b) Find Association rules from frequent data sets for training sessions using Minimum Confidence.

In existing, they had overcome the association rule mining through the traditional methods [1] [3] [4] like SETM algorithms, Apriori algorithms and Frequent Pattern Growth Algorithm. Thus the finding of minimum support and minimum confidence can be done through the proposed Frequent Pattern Growth Algorithm [2] [4] [12]. However it needs optimization for the event scheduling based on market basket analysis.

In FP Growth algorithm, it generated the frequent it emsets based on the transactional databases itself. Here the event scheduling can be done as like as frequent itemset mining but it should schedule based on their performance appraisal.

In this paper, an improved event scheduling framework has been designed using an optimized Frequent Pattern Growth algorithm. This can optimize the process of event scheduling based on the frequency of training sessions attended on the basis of training topic and feedback evaluated.

II. RELATED WORK

Han and Kamber [1] were proposed most important association rule mining techniques and evaluation strategies for the frequent itemset mining. Shamila Nasreen et.al. [2] were proposed pattern recognition for pronouncement the patterns with the rationale of discover how these algorithms can be used to obtain the frequent patterns over large transactional databases here we deal with the education training schedule databases.

M.Suman et.al. [2] proposed an new framework for analyzing the frequent itemsets using two top most algorithms such as Apriori algorithm and Frequent Pattern growth algorithm named as Apriori Growth Algorithm. In reference paper [11], they overcomes the disadvantages met in the Apriori algorithm and efficiently mine association rules without candidate generations from itemsets and with this it also overcomes the disadvantages in the FP Growth algorithm.

Paresh Tanna et.al. [12] determines the frequent patterns from a database. The aim of this study paper is to analyze the existing techniques for mining frequent patterns and evaluate the result performance of them by comparing the DHP algorithm and Apriori algorithm in terms of candidate generation.

Meera Navarkar et. al [6] were proposed optimized frequent pattern growth algorithm. It overcomes the existing frequent pattern growth algorithm [7] such as without generating huge number of conditional FP tree [9]. Moreover they only concentrates on the transactional databases and deals with the market itemsets. Here we concentrate on the educational training databases.

III. BACKGROUND KNOWLEDGE

A. FP Growth Algorithm

In the area of data mining, the most familiar algorithm for pattern discovery [4] [5] is obtained through the Frequent Pattern Growth algorithm. It deals with the most important two drawbacks of traditional Apriori algorithm. The constructed FP tree is used for the development of traditional frequent pattern growth algorithm [7]. The FP Tree [9] is constructed based on the dataset using the two passes are as follows:

Pass 1:

1. Scan the data content and find the support value for each item
2. Discard all the infrequent items.
3. Sort the frequent items in maximum to minimum order based on the support value.

Pass 2:

1. Here the nodes in the tree correspond to the items and it has incremental value as counter.
2. FP growth reads the transaction at a once in a time and then maps it to a path.
3. Use the fixed order, so that the paths can overlap when the transactions share the items.

In this algorithm, the counters are incremented frequently. Thus the algorithm for obtaining the pattern is as follows:

Input: Transaction Database T
Output: FP tree F

Algorithm: FP-growth Algorithm

[1] Function FP-growth(Tree F, A)
[2] {
[3] If FP tree F contains a single path P
[4] Then for each combination of nodes in the path P do
[5] Hence support=minimum support of nodes in B
[6] Generate pattern B U A
[7] Construct B’s conditional pattern base and B’s conditional FP tree that is FP Tree P;
[8] Else for each H in the header of the FP Tree P do
[9] {
[10] Hence H support;
[11] Generate pattern B =H U A
[12] Construct B’s conditional base and B’s conditional tree i.e., FP tree P;
[13] If FP Tree P ≠ □
[14] Then invoke FP-growth(FP Tree P, B);
[15] }

By this above algorithm, the frequent pattern tree can be constructed through the minimum support value. The growth can be constructed based on the transactional database with the itemsets to obtain the frequent itemsets. This is can be illustrate by this figure Fig 3.1.

Consider a transactional database T in the table 1 consists of set of transactions based on their transaction ID and list of items are displayed in the transaction. Then scan the entire data items in the T. Count the items in the database T. Then sort the items in the decreasing order based on the occurrences in the transaction.

Table 3.1: Transactional Database T

Trans. ID	Items
T1	A,B,D,E
T2	A,C,D
T3	E,F,H,I
T4	A,B
T5	C,E,F

The frequent itemlist based in the number of occurrences for the transactional database T from Table 3.1.

Table 3.2: Counts of Frequent Itemset in Table 3.1

Items	Count
A	3
B	2
C	2
D	2
E	3
F	2
H	1
I	1

The items that does not meet the minimum threshold support value has been discarded. Here the minimum threshold support value is Table 3.1 in the transactions. The minimum support solutions as follows in Table 3.3:

Table 3.3: Sorted Frequent itemsets

Items	Count
A	3
E	3
B	2
C	2
D	2
F	2

The transaction can be reordered based on the occurrences of the data items. Thus they are generated as based on the data items.

Table 3.4: Reordered itemsets with the transactions

Trans. ID	Items
T1	A,E,B,D
T2	A,C,D
T3	E,F
T4	A,B
T5	E,C,F

Now, we can scan again the transactional database T in the Table 4.3. The frequent pattern can be generated based the above table. The details can be increased in the number of transactions means it makes more complex in the FP Tree Construction. The FP tree can be constructed as follows.

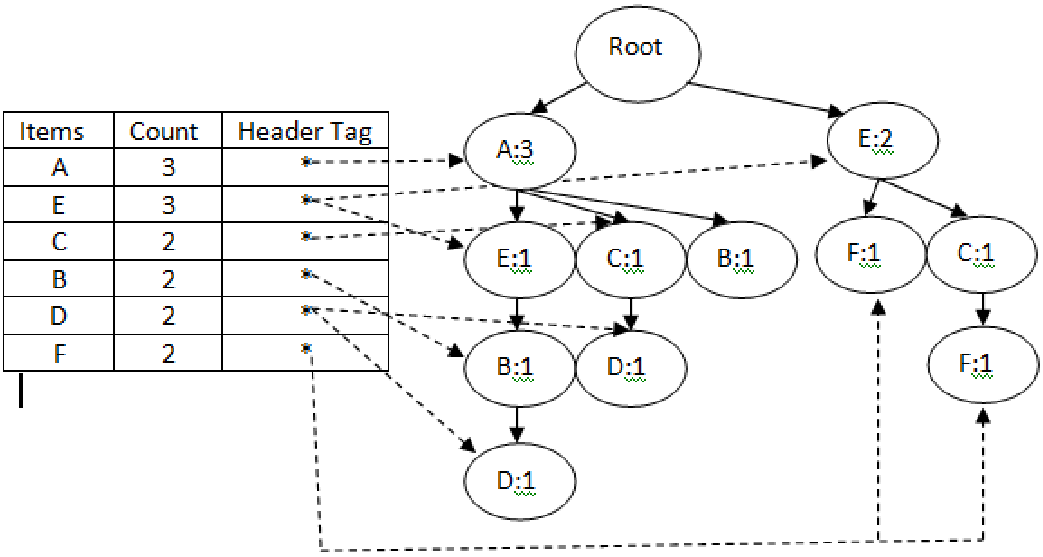

Figure 4.1 Frequent Pattern Tree for the Transactional Database T Table 3.4

IV. IMPLEMENTATION

Construct a Frequent Pattern Growth for the educational database. In this educational database, the frequent itemsets are considered based on the event scheduled for the training sessions. This can be generated with their overall average feedback obtained from the students those who are attended the training sessions.

Here we propose this work as modified optimized Frequent Pattern Growth algorithm which considers the feedback value to construct an FP Tree from the event database. Here we consider the event database for the training sessions

In this optimized FP Growth Algorithm, the contents needed are considered for the construction of the FP Tree based their performance appraisal. The algorithm can be created as follows to optimize the event scheduling tasks of the training sessions.

Input: Event Database E

Output: FP tree F

Algorithm: Optimized FP-growth

[1] Function Opt_FP-growth(Tree F, A, FV)
[2] {
[3] If FP tree F contains a single path P
[4] Then for each combination of nodes in the path P do
[5] Hence support=minimum support of nodes in B
[6] Check whether the support value satisfy based on FV
[7] Generate pattern B U A
[8] Construct B's conditional pattern base and B's conditional FP tree that is FP Tree P;
[9] Else for each H in the header of the FP Tree P do
[10] {
[11] Hence H support;
[12] Check whether the support value satisfy based on FV
[13] Generate pattern B =H U A
[14] Construct B's conditional base and B's conditional tree i.e., FP tree P;
[15] If FP Tree P ≠ □
[16] Then invoke FP-growth(FP Tree P, B);}

By this above proposed modified FP Growth algorithm we can construct the FP tree for the database E as follows in Table 4.1

Table 4.1: Event table with Avg_feedback

Event ID	Teaching Assistant
E1	A:5,E:3,B:3,D:4
E2	A:4,C:3,D:4
E3	E:3,F:4,H:2
E4	A:4,B:4,I:4
E5	E:3,C:4,F:5

Now we can scan the table and the number of occurrences can be calculated as follows:

Table 4.2: Training assistant with count

Teaching Assistant	Count
A	3
B	2
C	2
D	2
E	3
F	2
H	1
I	1

After the elimination of Teaching Assistant based on the feedback obtained the prune table should be like as follows:

Event ID	Teaching Assistant
E1	A:5,D:4
E2	A:4,D:4
E3	F:4,H:4
E4	A:4,B:4,I:4
E5	C:4,F:5

VI. CONCLUSION:

In this proposed work, a novel modified and optimized FP Growth algorithm may used to obtain the pattern discovery. Here we construct an event scheduling framework using this algorithm to obtain the schedule based on the training assistant those who having their feedback values. Most of the traditional methods are failed to do their analysis without generation the frequent itemsets based on occurrence and the performance feedback. In future, it should be implemented with the real time values of the top university details.

REFERENCES:

[1] J Han and M.Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publisher, San Fransisco,CA,USA, 2001.

[2] S Nasreen, M A Azam, K Shehzaf, U Naeem and M Ali, "Frequent Pattern Mining Algorithms for finding Associated Frequent Patterns for Data Streams: A Survey", Elsevier Proceedings of Computer Science, Vol.37, 2014, pp.109-116.

[3] Trupti A.Kumbhare and Santosh V.Chobe, "An Overview of Association Rule Mining Algorithms", Intl. Journal of Computer Science and Information Technologies, Vol.5, No.1, 2014, pp.927-930.

[4] M Suman, T Anuradha, K Gowtham and A Ramakrishna, "A Frequent Pattern Mining Algorithm based on FP-Tree Structure and Aprori Algorithm", Intl. Journal of Engineering Research and Applications, Vol.2, Issue 1, January 2012, pp.114-116.

[5] Jingbo Yuan and Shunl Ding, "Research and Improvement on Association Rule algorithm on FP-Growth", Intl. Conf. on Web Information systems and mining, 2012, pp.306-313.

[6] Meera Narvekar and S Fatma Syed, "An Optimized Algorithm for Association Rule Mining using FP tree", Elsevier proceedings on Computer Science, Vol.45, 2015, pp.101-110.

[7] Yi Seng, Shiqun Yin, Jiangyue Liu and M Zhang, "Research of Improved FP-growth algorithm in Association Rules Mining", Scientific Programming - Article ID.910281, vol.2015, 2015.

[8] Princy Shrivastava, Ankita Hundet, Babita Pathik and Shiv Kumar, "A Survey on FP (Growth) Tree using Association rule mining", Intl. Research Journal of Engineering and Technology, Vol.04, No.07, July 2017.

[9] Ruchi Bhargava and S Lade, "Effective Positive Negative Association Rule Mining using Improved Frequent Pattern Tree", Intl. Journal of Advanced Research in Computer Science and Software Engineering, Vol.3, No.4, April 2013.

[10] C.Borgelt, "Efficient Implementations of Apriori and Eclat", In Proc. 1st IEEE ICDM Workshop on Frequent Item Set Mining Implementations, CEUR Workshop Proceedings 90, Aachen, Germany 2003.

[11] Han J., Pei H., and Yin. Y., "Mining Frequent Patterns without Candidate Generation", Intl. Conference on Management Data, ACM SIGMOD, vol.29, No.2, June 2000,pp-1-12.

[12] Paresh Tanna and Yogesh Ghodasara, "Foundation for frequent Pattern Mining Algorithms Implementation", Intl. Journal of Computer Trends and Technology, Vol. 4, No.7, July 2013.

