Asian Journal of Computer Science and Technology
ISSN: 2249-0701 (P) Vol.9 No.2, 2020, pp.24-27
© The Research Publication, www.trp.org.in
DOI: https://doi.org/10.51983/ajcst-2020.9.2.2171

A New Algorithm for Pattern Based Using Mining Association Rules

S.Selvam
Department of Computer Applications N.M.S.S.Vellaichamy Nadar College, Madurai, Tamil Nadu, India
E-mail: s.selvammscmphil@gmail.com

(Received 3 July 2020; Revised 19 July 2020; Accepted 10 August 2020; Available online 18 August 2020)

***Abstract* - It is indeed an art to match maximum number of preferences by utilizing limited number of resources. During the current academic year 75% of the admissions to Engineering Colleges have gone down, as only 30% to 40% of intake has been filled. Without reaching the breakeven point, the management of the institution becomes a complicated issue. The main aim of this paper is to discover a pattern to identify the choice of preferences of the candidates to seek admissions in any academic institutions. For the purpose of matching optimum number of candidates to suit our existing system, we have designed our algorithmic approach. Here our new system is used to extract frequent item sets from various preferences. By thresholds, it can fix the preferences either decrease or increase the level of frequent. The new algorithm is based on association rule classification which is one of data mining techniques. Here the frequency of itemset2 is combined with frequency to get itemset3 and continues until item set n. the new algorithm is easy to use and implement because its complexity is less. The application is designed to generate association rule until n-antecedent with one consequent. For this study purpose we have identified 15 most frequently used preferences among the students. The samples we have taken to get association rules are 100 students of Pannai College of Engineering and Technology at Sivagangai. The discovered pattern is common to all institutions. The pattern discovery may be accurate because it is computed by using factors like confidence and support. If this intelligent system is followed strictly, definitely the number of outcomes is increased. The applicant would prefer only when the supply is high. The result of this paper is an application that can generalize association rule among various academic institutions.**

***Keywords:* Mining of Association Rules, Newly developed algorithm, Generating Frequent Item sets, Design Admission Diary, Design a Template.**

I. INTRODUCTION

In order to avoid the problem of processing big data into information which useful for user and used data mining techniques. Data mining is the analysis of data with the intent to discover gems of hidden information in the vast quantity of data that has been captured in the normal course of running the business.

[1] Here we develop an application to classify students based on their preferences of academic institutions in one transaction using our new algorithm. We show how to extract data pattern with assumption that data has been in one warehoused database.

The expected outcomes of this paper will be the patterns of student's preferences to study that are bought in one transaction together. These patterns can be used to get admissions for any academic institutions that frequently received in a same time, to design admission diary, to design a template, to plan to get admissions etc.

II. NEW ALGORITHM

```
L1 = {large l-itemsets};
for ( k = 2; Lk-l ≠ 0; k++ ) do begin
ck = mal-gen(Lk-1);
for all transactions t Є D, do begin
Ct = subset,(Ck, t);
For all candidates c Є Ct, do
c.count++;
end
Lk; = {c Є Ck | c.count ≥ MinSup}
end
Answer = UkLk;
```

The newly developed algorithm *is* for mining association rules, takes advantage of structure within the rules themselves to reduce the search problem to a more manageable size.

Suppose that an academic institution is having the following facilities moderate fees, state of the art lab, high volumes of library, experienced faculty, good result, placement, hostel facility, transport facility, sports activity, basic amenities, hi-tech environment, good management, reputation, distance, reference [2].

Denote this set of items as I. One by one, the candidates pull over, pick up their preferences and get admissions through various combinations of these items, subsets of I. Suppose the following table, lists the transactions made during the academic year 2020-21.

Let D be the set of transactions represented in the Table1, where each transaction T in D represents a set of items contained in I. Suppose that we have a particular set of items A (e.g., moderate fees and Placement), and another set of items B (e.g., moderate fees and state of the art lab).

Then an association rule takes the form if A, then B (i.e., A →B), where the antecedent A and the consequent B are proper subsets of I, and A and B are mutually exclusive.
This definition would exclude trivial rules such as if moderate fees and state of the art lab then Placement [3].

$$Support = \frac{No\ of\ transactions\ containing\ both\ A\ \&\ B}{Total\ no\ of\ transactions}$$

$$confidence = \frac{no\ of\ transactions\ containing\ both\ A\ \&\ B}{no\ of\ transactions\ containing\ A}$$

The mining of association rules from large databases is a two-steps process: Find all frequent item sets i.e find all item sets with frequency ≥ Φ. From the frequent item sets, generate association rules satisfying the minimum support and confidence conditions III.

TABLE I TRANSACTIONS MADE DURING THE ACADEMIC YEAR 2020-21

S. No.	**Primary Item (A)**	**Secondary Items (B)**	**No .of Trans. A**	**No.of Trans. A& B**	**Support**	**Confidence**
1	Moderate fees	Placement, Good result	8	5	0.074	0.625
2	State of the art lab	Moderate fees , placement	8	6	0.089	0.75
3	High volume Library	Moderate fees, good result	4	3	0.044	0.75
4	Experienced faculty	Placement, State of the art lab	5	4	0.060	0.8
5	Good result	State of the art lab, Placement	9	5	0.074	0.56
6.	Placement	Moderate fees, Experienced faculty	8	5	0.074	0.625
7	Hostel facility	Moderate fees, Transport facility	3	1	0.015	0.33
8	Sports activity	Experienced faculty, Placement	3	2	0.03	0.67
9	Basic amenities	State of the art lab, Transport facility	2	1	0.015	0.5
10	Transport facility	Placement, Basic amenities	4	2	0.03	0.5
11	Hi-tech environment	Placement, Transport facility	3	2	0.03	0.67
12	Good management	Placement, Library	2	1	0.015	0.5
13	Reputation	Experienced faculty, State of the art lab	3	2	0.03	0.67
14	Distance	Experienced faculty, Transport facility	2	1	0.015	0.5
15	Reference	Moderate fees, State of the art lab	3	1	0.015	0.33
		Total	67	44		

III. GENERATING FREQUENT ITEMSETS

This step is looking for items combination with frequency ≥ Φ. The global candidate set is generated as the union of all local large item sets from all partitions. It also takes n (no of partitions) iterations. The support for a candidate item set in that partition is generated by intersecting all l-subsets of that item set. The cumulative count gives the global support for the item sets. An item set is a set of items contained in I, and a k-item set is an item set containing k items. For example, {moderate fees and placement} is a 2-itemset, and {experienced faculty, good result, State of the art lab} is a 3-itemset, each from the set I. The item set frequency is simply the number of transactions that contain the particular item set. A frequent item set is an item set that occurs at least a certain minimum number of times. We denote the set of frequent k-item sets as F_k. We first find F1, the frequent 1-itemsets, which represent simply the individual items themselves. Since each sum meets or exceeds Φ = 3 (threshold), we conclude that each 1itemset is frequent. In general, to find F_k, our new algorithm first constructs a set C_k of candidate k-item sets by joining F_{k-1} with itself. Then it prunes C_k using the a priori property. The item sets in C_k that survive the pruning step then form F_k. Here, C_2 consists of all the combinations of items. Here no of transactions A&B is called candidates 2-itemset (F_2). Since Φ = 3, we have F_2 = {{Placement, good result}, {Moderate fees, Placement}, {Moderate fees, Good result}, {Placement, State of the art lab}, {Moderate fees, Experienced faculty}}[4].

Next, we use the frequent item sets in F_2 to generate C3, the candidate 3-itemsets. To do so, we join F_2 with itself, where item sets are joined if they have the first k−1 items in common. For example, {Moderate fees, Placement} and {Moderate fees, Good result} have the first k − 1 = 1 item in common, Moderate fees. Thus, they are joined into the new candidate item set {Moderate fees, Placement, Good result}. Similarly, {Placement, good result} and {Placement, state of the art lab} have the second item, Placement in common, generating the candidate 3-itemset {Placement, good result, state of the art lab}. Finally, candidate 3-itemsets {moderate fees, placement}, {Placement, State of the art lab} are generated in like fashion. Thus, C3 = {{Moderate fees, Placement, State of the art lab}, {Moderate fees, Good result, Experienced faculty}}.

C3 is then pruned, using our newly developed algorithm. For each item sets in C3, its size k − 1 subsets are generated and examined. If any of these subsets are not frequent, s cannot be frequent and is therefore pruned. For example, let s = {Moderate fees, Placement, State of the art lab}. The subsets of size k − 1 = 2 are generated, as follows: {Moderate fees, Placement}, {Moderate fees, State of the art lab}, and {Placement, State of the art lab}.

From the above, we see that each of these subsets are frequent and that therefore s = {Experienced faculty, Placement, Well library} is not pruned. Because here the subset {Experienced Faculty, Transport facility} has frequency 1<3 = Φ so that it is not frequent. By our new algorithm, {Experienced Faculty, Placement, Transport facility} cannot be frequent, is therefore pruned, and does not appear in F_3. Also consider s = {Moderate fees, Transport facility, State of the art lab}. The subset {Moderate Fees, Transport facility} has frequency 1<3 = Φ, and hence is not frequent. Again, by the defined property, its superset cannot be frequent and is also pruned, not appearing in F_3. [5].

IV. GENERATING ASSOCIATION RULE

After all of the frequent item set has been found, the next step is generating association rule by using confidence formula. Once the large item sets and their supports are determined the rules can be discovered in a straight forward manner as follows: if I is a large item set, then for every subset a of I, the ratio support (I) / support (a) is computed. If the ratio is at least equal to the user specified minimum confidence, then the rule a =>(L - a) is output. [6]. The local large item sets are generated for the same minimum support as specified by the user. Hence this is equivalent to generate large item sets with that minimum support. For large partition sizes, the number of local large item sets is likely to be comparable to the number of large item sets generated for the entire database. From F3:{Moderate fees, Placement, Good result}, we have 3 candidate rules that have 2 antecedent and one consequent.[7].

TABLE II CANDIDATE RULES OF F3

Sl. No.	Rule	Confidence	
1	If Moderate fees and Placement Then Good result	6/9	67%
2	if Moderate fees and Good result Then Placement	3/8	37.5%
3	if Placement and Good result thenmoderate fees	5/8	63%

If our minimum confidence is 65% then the second and third rule would not be reported. Finally we turn to single antecedent and single consequent. The candidate rule is shown in the below table:

TABLE III CANDIDATE RULES OF F2

Rule	Confidence	
If Moderate fees then Placement	8/8	100 %
If placement then Moderate fees	8/8	100 %
If Placement then Good result	8/9	89%
If Good result then Moderate fees	9/8	112.5%
If Good result then Placement	9/8	112.5%
If Placement then reference	8/3	267%
If Placement, then state of the art lab	8/8	100%
If Placement then Library	8/4	200 %
If State of the art lab then Placement	8/8	100%
If distance then Transport facility	2/4	50 %
If distance then hostel facility	2/3	67 %
If Good result then State of the art lab	9/8	112.5%
If State of the art lab then good result	8/9	89%

From Table II and Table II we can find all rules in this case. They are shown in Table IV below.

TABLE IV FINAL RULES

Rule	Confidence	
If Placement, Moderate fees then good result	6/9	67%
If Moderate fees, Experienced faculty then Placement	5/8	62.5%
If Moderate fees, state of the art lab then good management	1/2	50%
If Experienced faculty, state of the art lab then good result	2/9	22%
if Moderate fees, Placement Then good management	6/2	300%

V. CONCLUSION

We have described an algorithmic approach for discovering a dynamic pattern to improve the number of potentials is fast in various ways. It is exclusively useful for very large databases. An important contribution of our approach is that it drastically reduces the I/O overhead. This feature may prove useful for many real-life database mining scenarios where the data is most often a centralized resource shared by many user groups, and may even have to support on-line transactions. In near future we have a proposal to analyze the disk I/O and CPU overhead. In addition the new algorithm has an excellent scale-up property. The main drawback of this algorithm is estimate the number of partitions for the given available memory. This can be rectified in near future. We can extend this work by parallelizing the algorithm for a shared multiprocessor machine.

REFERENCES

[1] S.ThabasuKannan, “Optimized mining of Very Large database via ClusteredIndexing Method*”, InternationalJournal of Intelligent Optimization Modeling,* , Vol 1,pp 3,June 2009.

[2] K.C.C Chan, A.K.C. Wong and D.K.Y. Chiu, "Learning sequential patterns for probabilistic inductive prediction," *IEEE Trans. Systems, Man and Cybernetics,* Vol. 24, No. 10, pp. 1532, July 1994.

[3] S.ThabasuKannan, “An algorithmic approach for a simple prototype of business system to get customer satisfaction on CRM”, *International Journal of Business Review*, Vol 4, pp-26,December2013.

[4] Jian Pei, Jiawei Han, BehzadMortazavi-asl and Hua Zhu, “Mining access patterns efficiently from web logs”, *In Pacific-Asia Conference on Knowledge Discovery and Data Mining,* pp. 3967,April 2000.

[5] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan and A. S. Tomkins, “The web as a graph: measurements, models and methods”, *Lecture Notes in Computer Science*, Vol. 162, pp: 18, 1999.

[6] Qiang Yang, Hui Wang and Wei Zhang, “Web-log mining for quantitative temporal event prediction”, *IEEE Computational Intelligence Bulletin,* Vol.1, No.1, December 2002.

[7] W. Wang and O. R. Zaïane, “Clustering Web Sessions by Sequence Alignment”, *Proceedings of DEXA Workshops*, pp: 394-398. 2002.

[8] R. Agrawal and R. Srikant. “Fast algorithms for mining association rules in large databases” *In Proceedings of the 20th International Conference on Very Large Data Bases,* Santiago, Chile, September 1994.

[9] K.C.C Chan, A.K.C. Wong and D.K.Y. Chiu, "Learning sequential patterns for probabilistic inductive prediction," *IEEE Trans. Systems, Man and Cybernetics,* Vol. 24, No. 10, pp. 153, July1994.

[10] Savasere, E. Omiecinski, and S. Navathe. “An efficient algorithm for mining association rules in large databases”. *Technical Report GIT-CC-95- 04,* Georgia. Institute of Technology, Atlanta. GA 30332, January 1995.