Discovering a Pattern in Effective Manner

Avinash S Shegokar[1], Rohit S Jachak[1], Ashish Yerawar[1].

[1] *Student, Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune - 411 048 Maharashtra, India*

avinashsshegokar@gmail.com
jachakrohit07@gmail.com
yerawarashish@gmail.com

Abstract— **There are lots of Text mining techniques proposed for identifying desired Pattern present in the various documents. But Developing a Technique which is Efficient as well as innovative is still a Research Issue.**

To find pattern in the given dataset we can use the frequent pattern relation. Finding a frequent pattern in the given set of data is like relating two or more items with each other with a specific kind of associationship between them. Basically the frequent pattern is efficient technique but as the size of fed data more specifically the size of pattern increases the performance of technique decreases.

Various types of techniques have already been proposed using frequent pattern technique. But this Paper introduces an efficient and leading edge Pattern Based Technique which uses the frequent item set matching method. This results in an innovative and effective approach for finding a desired term in Data mining field.

Keywords— **Text Mining, Pattern Evolving, Pattern Matching, Information Filtering.**

I. INTRODUCTION

Due to the super-colossal Growth in information technology sector we are surrounded by tons and tons of Digital Information. If we wish to find out a Particular Information from a Huge Pile Data, Manual Observation will become inefficient as well as time consuming. This is where the concept of Pattern Matching (Information Retrieving) is introduced. To find an Effective way to search a word or a pattern from chunk of data; various

Data Mining or Information Retrieving Techniques are proposed and still lots of Research are going on to Find an Efficient and Innovative Process for Pattern Discovery. So as to find pattern in the given dataset we can use the frequent pattern relation. Finding a frequent pattern in the given set of data is like relating two or more items with each other with a specific kind of associability between them. Basically the frequent pattern is efficient technique but as the size of fed data more specifically the size of pattern increases the performance of technique decreases. Various types of techniques have already been proposed using frequent pattern technique. Though many of them fails to fully utilise the use of association rule of mining. It's not possible to extract the exact amount of information from a junk of data without properly understanding and analysing

The maximal frequency sets. In this paper we are proposing an innovative and efficient algorithm called as Discovering Pattern Effectively by Using Pattern Decomposition. The algorithm found to be effective as the data transferred from one phase to another phase is reduced. The set of data was sorted and deflated by breaking the transactions and linking similar transactions resulting in decrease in time complexity and increase in performance. Moreover if a transaction contains isolated patterns they can be broken further if altogether they does not constitute minimum requirement to be in the interested set of data. After breaking all the transactions and collecting the indistinguishable patterns results in the decreased space complexity thus resulting in space saving. The algorithm used here proved not only to be an innovative one but also effective one.

II. ALGORITHM

The Discovering Pattern Effectively using Pattern Decomposition algorithm constricts data set each time when Stray patterns are discovered. Particularly it searches for usual items by manipulating bottom-up search approach. Now for given proceeding data set D1, there are two stages at the first pass as:

1) In this phase the algorithm creates two set of items. Item set F1 which is a set of frequent item and item set I1 which is a set of infrequent or stray items. This is done by counting the occurrence of a particular pattern in the given data set.

2) Now we are further breaking the D1 set to D2 set in such a manner that D2 doesn't contains the items in I1. By doing this for further more passes let's say N passes, frequent item set F1 and I1 are created by reckoning for all N item sets in DN. Then DN+1 is made by disintegrating DN using IN such that DN+1 doesn't comprises item set in IN.

Now we try to clarify all the process for discovering effective patterns. In the figure 1.0, it is shown how the algorithm is used to find the patterns which are continual in given data set. Let's assume that we are having autochthonous set of data, which is D1 and basal support as 2. Lets first start with counting the support of all the items in data set D1 to determine F1 and I1. Consider here in this situation we got two item sets F1 and I1 where frequent item set F1={a,b,c,d,e} and infrequent item set I1={f,g,h,k}. After that we will putrefy each pattern in D1 using I1 to get D2. In the next phase precisely in the second pass, we

Originate and sum all 2-item sets containing in D2 to persuade F2 and I2 which is illustrated in the displayed figure. After that we further crumble each pattern in D2 to get next set D3. This keeps on until and unless we ordain D5 from D4, which is an empty set and we come to an end. The end result is the united result of all frequent sets F1 through F4.Here is the example shown which indicates three ways to reduce the dataset. As denoted by δ1, δ2 and δ3 in the figure 1.

In δ1, when patterns after festering yield the same item set, we unite them by totalling their occurrence. Here abcg and abc scale down to abc. As both of their existences are 1, the final pattern is abc: 2 in D2.

In δ2, we will dismiss the patterns which are having lesser size than that of the next dataset. Here consider the patterns are abc and abd. Both are having size of 3. But as our dataset D4 can't have pattern with size 3. Hence they are eliminated.

In δ3, when part of a given pattern is having an identical item set as that of another pattern after putrefaction, we unite them by totalling their occurrence. Here bcde is the item set of pattern 4 and is a part of pattern 1's item set after degradation, so end result obtained is bcde: 2 in D4.

There is also a lucid way to putrefy the item set S by an scant N-item set t, as elaborated in [6], is to reimburse S by N item sets, each is attained by eliminating a single item in t from S. For example S=abcd and t=ad, we get reduced s by removing a, d respectively from S as {bcd, abc}. This is generally more popular as the simple split. Now when we encounter a large set this method is proved to be not efficient one. This is the sole reason we need a quicker splitting technique to dissolve pattern.

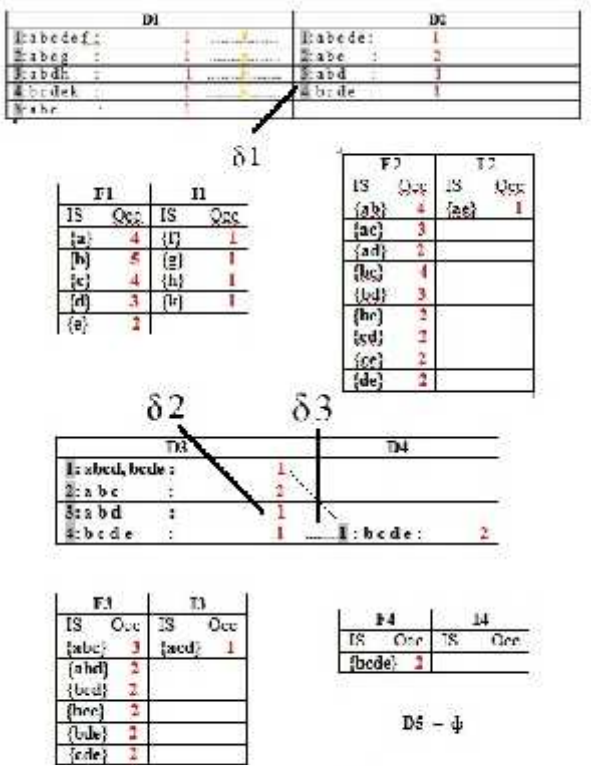

Fig. 1

III. Conclusion

Thus we are nominating a Pattern Discovery algorithm to detect frequent patterns. The algorithm not only reduces the time but it also saves the costly candidate set generation procedure. The algorithm efficiently reduces the space requirement as the dataset used in the algorithm is reduced during each single pass. Our experiments show that the algorithm is not only innovative and efficient one but it also creates a lot new opportunities in the field of Data Mining.

References

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In VLDB'94, pp. 487-499.

[2] R. J. Bayardo. Efficiently mining long patterns from databases. In SIGMOD'98, pp. 85-93

[3] Lin, D.-I and Kedem, Z. M. 1998. Pincer-Search: A New Algorithm for Discovering the Maximum Frequent Set. In Proc. of the Sixth European Conf. on Extending Database Technology.

[4] Park, J. S.; Chen, M.-S.; and Yu, P. S. 1996. An Effective Hash Based Algorithm for Mining Association Rules. In Proc. of the 1995 ACM-SIGMOD Conf. on Management of Data, pp.175-186.

[5] Zaki, M. J.; Parthasarathy, S.; Ogihara, M.; and Li, W. 1997. New Algorithms for Fast Discovery of Association Rules. In Proc. of the Third Int'l Conf. on Knowledge Discovery in Databases and

Data Mining, pp. 283-286.

[6] Brin, S.; Motwani, R.; Ullman, J.; and Tsur, S. 1997. Dynamic Itemset Counting and Implication Rules for Market Basket Data. In Proc. of the 1997 ACM-SIGMOD Conf. On Management of Data, 255-264.

[7] J. Han, J. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. Proc. 2000 ACM-SIGMOD Int. Conf. on Management of Data, Dallas, TX, May 2000.