CTP Algorithm : An Implementation In the Real World

Priyaswarupa Rani Pati & E. Poovammal

Dept. of Computer Science and Engineering, S.R.M University, Chennai ,TN
E-mail : priya.tulu@gmail.com

***Abstract* - Recently effort has been laid for mining of closed temporal pattern from time interval based data. The importance of time interval based data is manifold since the various interesting pattern a data can take within some interval of time is worth analyzing. The applicability of the algorithm could be analyzed by considering various data item-sets in the real world which comprises of medical, business and networks. Further extension of the study of the closed temporal time based mining with the datasets belonging to different fields for experimentation is carried out that is to find out the varying pattern of different packets in network intrusion detection in the interval based time Again by this new exploit the applicability of the above algorithm is noted. Before only mining of sequential pattern as in was widely used . With the advent of time importance of closed temporal pattern is given importance to. Now mining of closed temporal pattern from interval based will reveal a lot of different pattern over the time which will be of utmost importance in the varying spheres of life.**

Keywords-Closed temporal pattern ,time interval based

I. INTRODUCTION

A. Objective

Mining of closed temporal pattern was what made the researchers to build interest in , many researchers got zealous to work in the field of time point based data where the capability of using compact results to preserve the same expressive power as conventional mining. Mining method to mine time interval based pattern [4],[5],[6] was used earlier. To mine the time interval based data is very difficult since the pair wise relationship between the two interval based events are very intrinsic to obtain. This is the recent endeavor in the field of mining of time interval based data where the change in pattern of different data sets belonging to different field is taken into consideration, their results are studied to find out the different interesting pattern. In the above process temporal pattern mining is done using data sets including that of sequence of DNA, network intrusion detection where the pair wise relationship between the end points are relaxed using the CTP algorithm. To check the different data sets that can be used together for time based analysis and the result is noted which would reveal further different characteristics of the datasets.

B. Existing System

In the existing system that is in [1] the algorithm was used to deal with the datasets of the readers in the library. Here the data sets are limited in number and the closed temporal patterns are developed without candidate key generation based on the proposed end point representation where there is the employment of the three pruning strategies to reduce the search space effectively.

C. Proposed System

In the proposed system the CTP Algorithm will be essentially be dealing with the several datasets which comprises of that of network intrusion detection and tandem repeats of the DNA sequence. The change in the pattern of the dataset is noted from which necessary conclusion can be deduced to recognize new patterns if any from the huge dataset .

II. RELATED WORKS

The researches are mostly concerned with the sequential pattern mining as introduced in [2] [3] which is not very efficient for long frequent sequences .Closed pattern mining was then considered to be a better alternative in dealing with the long frequent sequential pattern which is time point based as illustrated in [7].Now considering the case where mining of the interval based pattern can lead to a different level of mining since with the passage of time the pattern would take up varying forms as it has a definite start and an

end time. Which is considered to be more complicated and difficult in comparison to the time point based data.

III. CTP ALGORITHM

1) Capturing Of Global Information: When the dataset is fed into the system, before being stored in the table first its global information is captured.

2)Comparison With Intervals: ”Before”, ”after”, ”equal” is considered while relating the two end points which generates less number of intermediate candidate sequences.

3) Relationship Among The End Points: End point representation is very much essential to eliminate the ambiguity of any sort. At a particular instant only it will check for the behavior and that will be noted to as the end point representation and beyond that it will not be checked and then the behavior will be stored in the temporal database. This will probably increase the efficiency of the system.

4) Temporal Database : Temporal database is used to keep the record of the entry of the dataset at a particular time and also it will verify the behavior of the datasets. When datasets are stored in the temporal database they are extracted to see for the events which has occurred at different instant of time .

5) *Optimization strategy*: In this phase the search space is reduced and also the redundancy is checked.

6) Closure checking: To check for the type of sequence obtained closure checking is done as because the end point generation is the most important aspect in the interval based temporal mining of the data.

7) Forward extension: To grow the patterns in a forward direction and then checked for the endpoints which can be reached by checking the timeline validity of that particular dataset.

8) Backward extension: Here the backward extension of a dataset is checked and also the closure is defined for this particular phase where the pruning of the search space occurs.

9) Check for semi max period: It is the time between a prefix and a suffix part of a particular dataset which has to be checked to know for the end point generation of a sequence. This can be defined as last in first appearance wrt a prefix.

IV. EXPERIMENTAL RESULT

The various pattern can be identified in a time based interval data mining which otherwise is a very difficult process to find in the very large datasets. This is one of a kind of pattern recognition which is carried out to determine the various pattern which can bring a change in the meaning of the dataset in a time interval based and also will provide ample help in finding out patterns in a huge dataset without impacting the normal process. The experimental data set consists of 325 records which are of the type denial of service one. When it is fed into the system and with minimal support and confidence value it generates the output with the machine id ,the type of attack along with the support value.

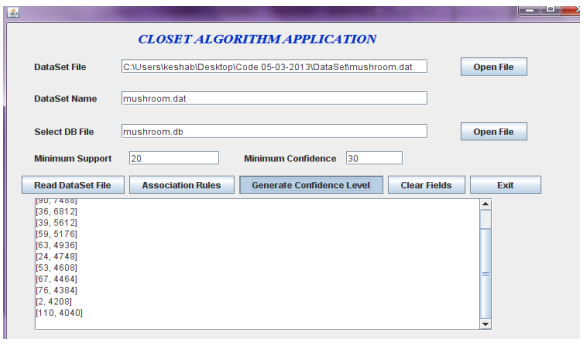

Fig 1 Implementation using dataset.

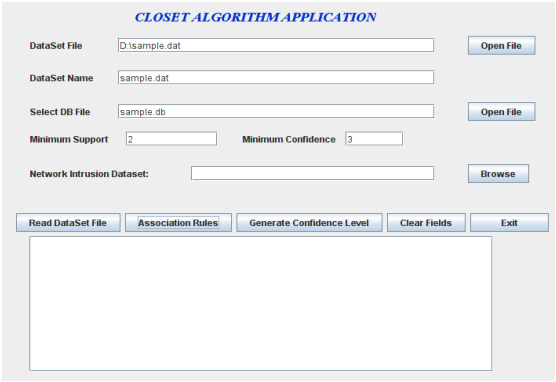

Fig 2 Implementation using network intrusion dataset.

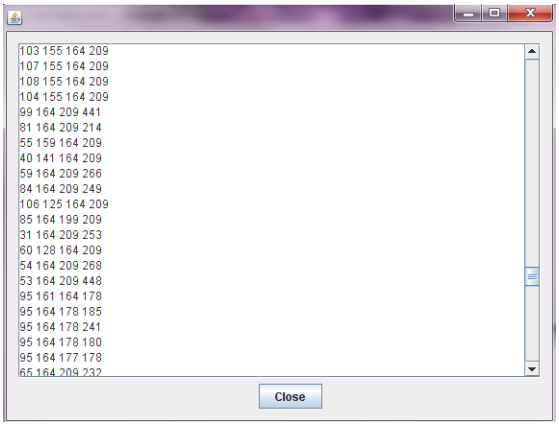

Fig 3 Reading of the dataset file

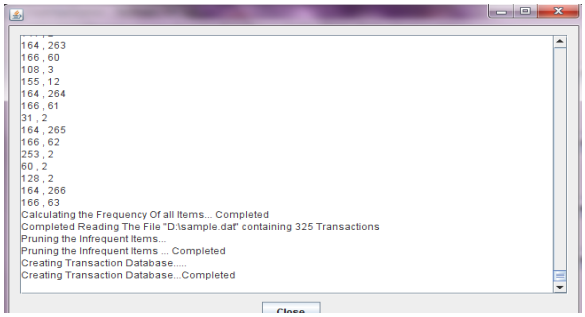

Fig 4 Calculation of maximum data value

V. CONCLUSION

By using the CTP algorithm a real dataset of network intrusion is used. Intrusion detection is a very important process where various cases of intrusion is detected in today's scenario .Any suspicious pattern if generated in a large dataset over the time, it will not be easier to detect. By using CTP algorithm the variance of pattern is recognized in the long run and the packet is not entertained without hampering the normal activities which is being carried out in the organization as such.

VI. REFERENCES

[1] Yi-Cheng Chen,Wen-Chih Peng and Suh-Yin Lee , CE Miner –An Efficient Algorithm for mining closed patterns from time interval-based data of 11th IEEE International Conference on Data Mining(DOI 10.1109/ICDM.2011.31).

[2] R.Agrawal and R. Srikant ,Mining sequential patterns.Proceedings of 11th International Conference on Data engineering (ICDE'95) pp 3-14,1995.

[3] J.Chen An Updown Directed Acyclic Graph Approach for Sequential Pattern Mining,IEEE Transactions on Knowledge and Data Engineering,vol.22,no.7,pp.913-928,2010.

[4] F.Morchen and D.Fradkin Robust mining of time interval with semi interval partial order patterns, proceedings of the SIAM International Conference on data mining (sdm'10),pp.315-326,2010.

[5] P.Kam and W.Fu "discovering temporal patterns for interval based mining".international Conference on Data warehousing and knowledge discovery (DaWaK '00),vol.1874,pp.109-128,2005.

[6] S.Wu and Y.Chen, "Mining Nonambiguos temporal Pattern for interval based events" ,IEEE Transactions on knowledge and data engineering,vol.19,issue 11,pp.1309-1330,2009.

[7] M.Zaki and C.Hsiao ,"CHARM: An Efficient algorithm for Closed Itemset mining ",Proccedings of 2nd SIAM International Conference on Data Mining (SDM'02),pp.457-478,2002.

◈◈◈