RESEARCH ARTICLE   OPEN ACCESS

Analysis of Process Mining Model Using Frequentgroup Based Noise Filtering Algorithm

V. Priyadharshini[*], A. Malathi[**]

[*](PhD Research Scholar, PG & Research Department of Computer Science, Government Arts College Coimbatore – 18)
[**](Assistant Professor, PG & Research Department of Computer Science, Government Arts College Coimbatore – 18)

ABSTRACT
Process mining is a process management system used to analyze business processes based on event logs. The knowledge is extracted from event logs by using knowledge retrieval techniques. The process mining algorithms are capable of automatically discover models to give details of all the events registered in some log traces provided as input. The theory of regions is a valuable tool in process discovery: it aims at learning a formal model (Petri nets) from a set of traces. The main objective of this paper is to propose new concept Frequentgroup based noise filtering algorithm. The experiment is done based on standard bench mark dataset HELIX and RALIC datasets. The performance of the proposed system is better than existing method.
Keywords: Cycle detection algorithm, frequent group based noise filtering algorithm, Helix dataset, Process discovery, Process mining, RALIC dataset

I. INTRODUCTION

Process discovery is one of the most challenging process mining tasks. Based on event log, a process model is constructed by capturing the behavior of the log [4]. Event Logs essentially capture the business activities happened at a certain time period [1]. The basic idea is to extract knowledge from event logs recorded by an information system.

Process mining aims at rising this by providing techniques and tools for locating method, control, data, structure, and social structures from event logs. The research domain that is concerned with knowledge discovery from event logs is called process mining [2]. More traditional data mining techniques can be used in process mining. New techniques are developed to perform process mining i.e. mining of process models. It is the traditional analysis of business processes based on the opinion of process expert [3]. The business process mining attempts to reconstruct a complete process model from data logs that contain real process execution data. Many techniques highlight the likelihood of mixing variety of method mining approaches to mine tougher event logs, like those which contain noise.

The necessary background in Section II describes related work. Section III presents existing alpha algorithm that describes previous work done. Section IV describes the proposed implemented work. The result and discussion is presented in Section V. Conclusion and future work is discussed in Section VI.

II. RELATED WORK

The algorithm in [5] is an approximation algorithm that has proven its efficiency in estimating large number of cycles in polynomial time when applied to real world networks. The algorithm counts the number of cycles in random, sparse graphs as a function of their length. While using it in real world networks, the result is not guaranteed for generic graphs.

The algorithm in [6] presented an algorithm based on cycle vector space methods. This algorithm is slow since it investigates all vectors and only a small portion of them could be cycles.

The algorithm in [9] is DFS-XOR based on the fact that small cycles can be joined together to form bigger cycle. It is more time efficient when it comes to real life problems of counting cycles in a graph because its complexity is not depending on the factor of number of cycles.

III. CYCLE DETECTION ALGORITHM

It is a pointer algorithmic rule that uses solely two pointers that move through the sequence at totally different speeds. The algorithmic rule solely must check for recurrent values of this special type, one double as aloof from the beginning of the sequence because the different, to search out an amount [7]. The function value is used to find

a cycle in a sequence of iterations. Cycles are available in a graph and in much real life application; it is required to know the existence of cycles in a graph [8]. The algorithm is developed in the context of network design problem but useful in any graph application where existence is to be finding out.

IV. FREQUENT GROUP BASED NOISE FILTERING

It finds all Frequentgroup patterns for a given minimum support and minimum confidence and remove the objects that are not a part of any Frequentgroup pattern [10]. The set of Frequentgroup patterns for any data set depends upon the value of minimum support and minimum confidence. Wherever possible, we fix the minimum support to be zero and employ the minimum confidence to control the number of objects that are designated as noise [12].

However, in some data sets setting the support threshold to zero leads to an explosion in the number of hypergroup patterns. For this reason, a low support threshold that is high enough to reduce the number of Frequentgroup patterns to a manageable level is used.

The proposed system automatically decides correct initial weight, noise filtering, feature selection properties [11]. The proposed system is more efficient than the existing system.

4.1 Datasets

Helix and RALIC datasets is a compilation of release histories of a number of non-trivial Java Open Source Software System. It contains class files for each release of the system along with meta-data. A metric history is derived from extraction of releases and this data is directly used in research works.

4.2 Results and discussion

In this work helix and RALIC dataset are used for the experiment. We are comparing the three methods such as existing and proposed works [13]. The existing region based folding and proposed unsupervised noise filtering and Frequentgroup based noise filtering as shown in Fig. 1. The fitness value of existing system is low compared to the proposed approach. By using proposed unsupervised noise filtering and Frequentgroup based noise filtering methods, we obtain the high fitness value. So we can get the better performance comparatively [14].

By doing the experiment based on fitness value calculation and fraction of unconnected transitions (T_u). In Fig.1 experiment based on fraction of unconnected transitions (T_u), we calculate the T_u corresponding to existing and proposed approaches. The T_u is decreased for the proposed methods

unsupervised noise filtering and Frequentgroup based noise filtering methods.

This will also increase the performance of the proposed system compared to existing system. Finally we can conclude as the proposed unsupervised noise filtering and Frequentgroup based noise filtering methods has more effective than the existing system.

V. CONCLUSION AND FUTURE WORK

The datasets were used with three different region based algorithms for unconnected transitions and the result is shown. In future, we will look for improvements of the existing process discovery and visualization techniques that allow for the construction of comprehensible models based on realistic characteristics of an event logs.
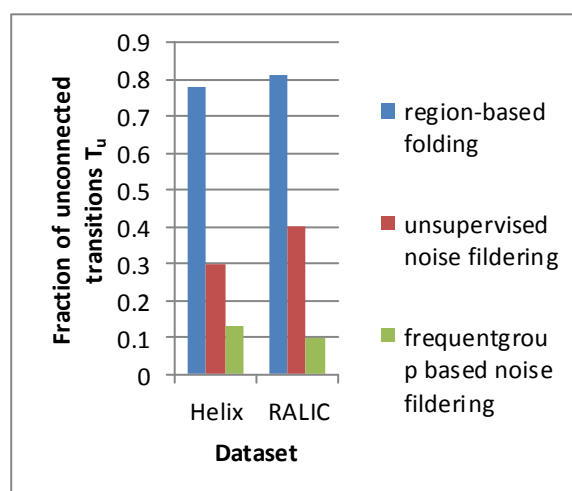

Fig 1. *Fraction of unconnected Transitions using datasets*

REFERENCES
[1] Josep Carmona et al, New region-based algorithms for deriving bounded petri nets, *IEEE Transactions on Computers, Vol. 59, No.3, March* 2010.
[2] Jochen De Weerdt et al, Active trace clustering for improved process discovery, *IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 12, Dec* 2013
[3] Zhiang Wu et al, A novel noise filter based on interesting pattern mining for bag-of-features images, *Expert systems with applications 2013.*
[4] Filip Caron, Jan Vanthienen and Bart Baesens, A comprehensive framework for the application of process mining in risk management and compliance checking, Faculty of Business and Economics.

[5]   K. A. Mahdi, M. Safar, I. Sorkhoh, and A. Kassem, Cycle-based versus degree-based classification of social networks, *Journal of Digital Information Management, vol. 7, no. 6,* 2009.

[6]   Robert Sedge wick, Thomas G. Szymanski, Andrew C. Yao, The complexity of finding cycles in periodic functions, *SIAM journal of computing, Vol. 11, No. 2,* 2006.

[7]   Van der Aalst, W., de Medeiros, A., Weijters, A., Genetic process mining. *Applications and Theory of Petri Nets pp. 985–985,* 2005.

[8]   De Medeiros, A., Weijters, A., van der Aalst, W., Genetic process mining: an experimental evaluation. *Data Mining and Knowledge Discovery vol.14, Issue. 2, 245–304,* 2007.

[9]   Kumar, Anand, Jani, N. N, An Algorithm to Detect Cycle in an Undirected Graph, *International Journal of Computational Intelligence Research, Vol. 6 No. 2,* 2010.

[10]   Maytham Safar, Fahad Mahdi and Khaled Mahdi, An Algorithm for Detecting Cycles in Undirected Graphs using CUDA Technology, *International Journal on New Computer Architectures and Their Applications, Vol. 2 No. 1,* 2011.

[11]   Gianluigi Greco et al, Discovering expressive process models by clustering log traces, *IEEE Transactions on Knowledge and data engineering, Vol. 18, No. 8, Aug* 2006.

[12]   Joonsoo Bae, Ling Liu et al, Development of Distance Measures for Process Mining, Discovery, and Integration, *International Journal of Web Services Research, Vol. 10, No. 10,* 2010.

[13]   Philip Weber, Behzad Bordbar, Peter Tiˇ no, A Framework for the Analysis of Process Mining Algorithms, 2009.

[14]   Wil M.P Van der Aalst, Process discovery: capturing the invisible, *IEEE Computational Intelligence Magazine,* 2010.