Analysis of Process Mining Model Using Unsupervised Noise Filtering Algorithm

V. Priyadharshini[1] A. Malathi[2]

Ph.D Research Scholar[1] Assistant Professor[2]

PG & Research Department of CS

Govt. 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 Unsupervised 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:
Process Mining, Process Discovery, alpha algorithm, Unsupervised Noise Filtering, Helix and RALIC Dataset.

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 [10]. Event logs essentially capture the business activities happened at a certain time period [7]. The basic plan is to extract data from event logs recorded by a data system. The method 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 [8]. 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 [9]. The business process mining attempts to reconstruct a complete process model from data logs that contain real process execution data [6]. Many techniques highlight the possibility of combining a number of process mining approaches to mine more challenging event logs, such as those that 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.

Related Work

Fabio in [1] compared business process models with metric that is based on a representation of local relations between pairs of activities in a model. Saravanan in [2] developed a control flow process mining algorithm that can discover all the

common control flow structures. Boleslaw in [4] described workflow mining algorithms that are used to improve and/or refine design of existing workflows. The time and space complexity is dependent on workflow's internal structure and on the number of workflow tasks.

Alpha algorithm

The **α-algorithm** is an algorithm used in process mining, aimed at reconstructing causality from a set of sequences of events. It constructs P/T nets with special properties (workflow nets) from event logs [3]. Each transition in the net corresponds to an observed task. It orders events sequentially, such that each event refers to a case and activity. It has problems with noise, infrequent behavior and complex routing constructs [13]. It scans event logs for particular patterns. It makes single pass through workflow log to identify which tasks directly follow each other. It is sensitive to noise and incompleteness of event logs [12]. The algorithm takes a workflow log as input and results in a workflow net being constructed. It will thus by examining causative relationships determined between tasks [14]. For instance, one specific task would possibly continually precede another specific task in each execution trace, which might be helpful data.

An unsupervised Noise filtering Algorithm

The proposed system automatically decides correct initial weight, noise filtering, and feature selection properties [11]. This proposed algorithm not only does noise filtering, it also measures the relationship of members in all calculated clusters to find the clarity of relationship. If any record has uncertain membership, it will be considered as the noise. It allows factoring out the cycles and therefore reduces the size of the transition system [5]. A lot of considerably, in distinction with approaches that will yield transition system with cycles, this reduction in size comes with the guarantee that necessary subsets of regions square measure preserved. Here we use a low support threshold that is high enough to reduce the number of frequent group patterns to a manageable level. The proposed system is more efficient than the existing alpha algorithm.

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.

Result and discussion

In the work helix and RALIC dataset are used for the experiment. We are comparing the two methods such as existing and proposed works. The existing alpha algorithm and proposed unsupervised noise filtering algorithm is shown in Fig. 1. The fitness value of existing system is low compared to the proposed approach. By using proposed unsupervised noise filtering and method, we obtain the high fitness value. So we can get the better performance comparatively. By doing the experiment based on fitness value calculation and fraction of unconnected transitions (T_u).

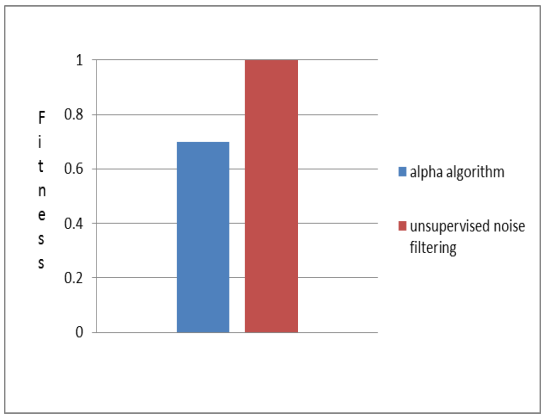

Fig. 1 *Comparison of two algorithms using fitness value*

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.

References

1. Fabio Aiolli, Andrea Burattin, Alessandro Sperduti, "A Business Process Metric Based on the Alpha Algorithm Relations",
2. Saravanan .M.S, Rama Sree .R.J, "A Role of Heuristics Miner Algorithm in the Business Process System", International Journal of Computer Technology and Applications, Vol. 2 No. 2, Page. 340-344.
3. C.J. Turner, A. Tiwari, R. A. Olaiya and Y, Xu, "Business Process Mining: From Theory to Practice", Business Process Management Journal, 2012, Vol. 18 Issue: 3, pp.493 – 512.
4. Boleslaw Mikolajczak, Jian-Lun Chen, "Workflow Mining Alpha Algorithm — A Complexity Study", Intelligent Information Processing and Web Mining Advances in Soft Computing Volume 31, 2005, pp. 451-455.
5. W. M. P. Van der Aalst, "The application of petri nets to workflow management", Journal of circuits, systems and computers, Volume 08, Issue 01, February 1998.
6. Arya Adriansyah, Peter van den brand et al, "Process mining manifesto", Business process management, Volume 99, 2012, pp. 169-194.
7. S.J.J.Leemans et al, "Discovering block-structured process models from event logs- A constructive approach".
8. Joonsoo Bae et al, "Development of distance measures for process mining, discovery and integration" International journal of web

services research, Vol.10, No.10, 2010.

9. Process Mining Manifesto. IEEE Task Force on Process Mining.
10. Aalst, W.M.P.v.d., et al.," Workflow Mining: Discovering Process Models from Event Logs". IEEE Transactions on Knowledge and Data Engineering 16(9) (2004) 1128–1142.
11. Zhiang Wu et al, "A novel noise filter based on interesting pattern mining for bag-of-features images", Expert systems with applications 2013.
12. 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.
13. Josep Carmona et al, "New region-based algorithms for deriving bounded petri nets", IEEE Transactions on Computers, Vol. 59, No.3, March 2010.
14. Bergenthum, R., Desel, J., Mauser, S., Lorenz, R.," Synthesis of Petri nets from term based representations of infinite partial languages". Fundamentals in Information System. 95(1), 187–217 (2009).