International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org Volume 3 Issue 4 ǁ April 2014 ǁ PP.47-49

Analysis of process mining model for label splitting problem

V.Priyadharshini[1] A.Malathi[2]
[1](Ph.D. Research Scholar, PG & Research Department of CS, Govt. Arts College, Coimbatore – 18,
[2](Assistant Professor, PG & Research Department of CS, Govt. Arts College, Coimbatore – 18,

ABSTRACT: *Process mining is a process management system that analyzes business processes builtusing event logs. The information is extracted from event logs by using knowledge retrieval techniques. The process mining algorithms are capable of inevitably 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 foremost objective of this paper is to propose new concept for label splitting. The experiment is done based on standard bench mark dataset HELIX and RALIC datasets. The performance of the proposed system is better than other existing methods.*

KEYWORDS: *HELIX Dataset, Label splitting, Process Discovery, Process Mining, and 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 [1]. Event logs essentially capture the business activities happened at a certain time period [2]. 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 [3]. 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 [4]. The business process mining attempts to reconstruct a complete process model from data logs that contain real process execution data [5]. Many techniques highlight the possibility of combining a number of process mining approaches to mine more stimulating event logs, such as those that contain noise.

Petri nets are a widespread formalism to model concurrent systems. By labeling transitions with symbols from a given alphabet, transitions can be interpreted as the occurrence of events or the execution of tasks in a system [9]. Labeled Petri Nets have been used in numerous applications: design and specifications of asynchronous circuits, resource allocation in operating systems and distributed computation, analysis of concurrent programs, performance analysis and timing verification and high-level design. Petri Nets are popular due to their inherent ability to express concurrency, choice and causality between events in a system, without explicit enumeration of global states. Although checking properties of Petri Nets could be difficult in general, for some subclasses of Petri Nets there are efficient verification algorithms [10]. In this paper, we will deal with safe Petri Nets. Safe nets have high expressive power; in particular every finite state system can be expressed as a safe labeled PN. On the other hand, safe nets are also well suited for verification.

The necessary background in Section II describes related work. Section III presents existing ECTS 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

In [6] J. Carmona et al describes the label splitting problemto find an optimal sequence of splitting's that induces the minimal number of transitions in the derived Petrinets. It isdefined on the sets of states computed when searching for regions in state-basedsynthesis methods. In [7] J. Cortadella et al describehow label splitting for excitation closure is done by computing the optimal coloring of a graph, i.e., the *chromatic number*. In[8] the author J. Carmona definesthe conditions under which an optimal label splitting can be derived to accomplish excitation closure.

III. ECTS ALGORITHM

Non – minimal regions can be removed from a state in TS to derive a state in the reachability graph of the synthesized net. The ECTS ensures that this preserves the transition in both transition systems. Based on this, a dissimulation is defined between the states of TS and states of reachability graph of Petrinets. The algorithm is complete if excitation closure holds the initial transition system and a set of minimal region to derive a Petrinets.

IV. PROPOSED SYSTEM

This work is applying label splitting is a technique that is used to improve the visualization of the PN, it is also used to avoiding over generalization. In some cases the derived Petri net equivalent to the initial transition system is required. This cannot be always accomplished. We solve this problem by defining a property excitation closure that TS must satisfy in order to derive a PN with bi-similar behavior. A label splitting technique is presented here is used to repair excitation closure violations in a Transition system [11]. This allows to present a method for obtain bi-similar k-bounded PNs; excitation closure is used to reduce the number of necessary regions while preserving the properties in the derived net.

Two open problems are formulated in the previous work,
1) If the existence of an optimal net can be characterized in terms of the TSs and
2) If there exist at most one optimal net which could be considered canonical.
Our work gives a positive answer to the first problem and a negative answer to the second problem.

- They did not address the problem of *merging* and *splitting* "equivalent" labels, which model the same event, but must be split in order to yield a valid or/and efficient Petri Net. We, for the first time, provide a method for label splitting which seems to satisfactorily solve the problem [13].
- They were inadequate to *elementary* TSs, which are moderately restricted, while we can handle the full class of TSs by means of transition splitting.
- They produce a PN with an RG *isomorphic* to the TS, which appears to be too strong a degree of correspondence. Our method extends this result to *excitation closed* TSs (ECTS), which include not only Elementary Transition Systems (ETS)s, but also all those TSs that have *dissimilar*ETS.

4.1 Steps for synthesis of PNs from TSs

- Initially, the labels of TS are split to obtain a split-morphic ECTS. Next, all minimal pre-regions that are predecessors of some event are generated to derive a minimal saturated PN. This restriction to event predecessors only is due to our region generation mechanism and has been shown to be sufficient in practice to obtain good results using a reasonable amount of computation time [14].
- Then, an irredundant subset of pre-regions is calculated and a place-irredundant PN obtained.
- Finally, by merging minimal pre-regions, further minimization of regions can be obtained. Exploring all place-irredundant nets can be computationally very expensive. Hence, we use only a greedy place merging starting from a place-irredundant net, thus yielding a quasi-place-minimal PN [12].

The function generate_min_preregions generates all minimal preregions for one event. The function find_irredundant_cover produces an irredundant set of regions. From this set, a place-irredundant net is generated. The function split_labels performs the splitting of labels if the initial TS are not an ECTS. The function map_to_PN is the final step for constructing a PN from the set of regions.

4.2 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.

V. RESULT AND DISCUSSION

In the work helix and RALIC dataset are used for the experiment. By using proposed label splitting in this system, we obtain the high fitness value. So the effective performance is obtained comparatively. The experiment is based on fitness value calculation and fraction of unconnected transitions (T_u). The second experiment is 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 label splitting technique. This will also increase the performance of the proposed system compared to existing system. Finally we can conclude as the proposed label splitting approach has more effective than the existing system.

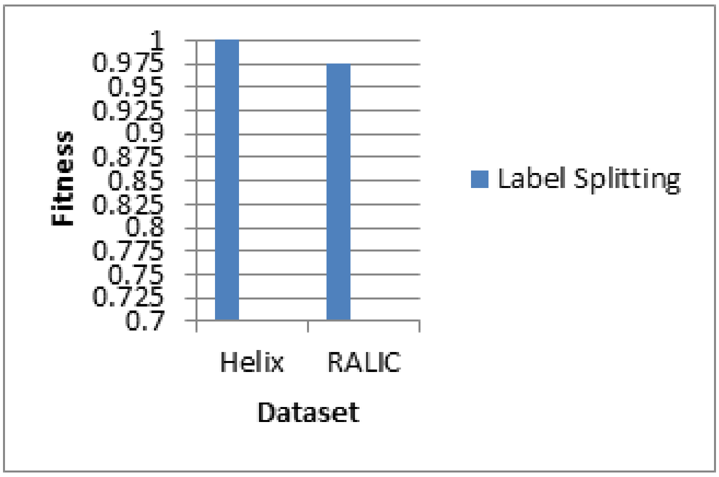

Fig.1 Positive and Negative Fitness value

VI. CONCLUSION AND FUTURE WORK

In this work both positive and negative value are obtained using the proposed algorithm, which shows performance of proposed system. The datasets were used with real time events 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] 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.
[2] S.J.J.Leemans et al, "Discovering block-structured process models from event logs- A constructive approach".
[3] Joonsoo Bae et al, "Development of distance measures for process mining, discovery and integration" International journal of web services research, 10(10), 2010.
[4] Process Mining Manifesto. IEEE Task Force on Process Mining.
[5] Arya Adriansyah, Peter van den brand et al, "Process mining manifesto", Business process management, 99, 2012, 169-194.
[6] J. Carmona, J. Cortadella, M. Kishinevsky,"New region-based algorithms forderiving bounded Petri nets", IEEE *Transactions on Computers*, 59(3), 2009,371–384.
[7] J. Cortadella, M. Kishinevsky, L. Lavagno, A. Yakovlev,"Deriving Petri netsfrom finite transition systems", *IEEE Transactions on Computers*, 47(8), 1998,859–882.
[8] The Label Splitting Problem, J. Carmona
[9] A. A. Kalenkova and I. A. Lomazova, "Discovery of Cancellation Regions withinProcess Mining Techniques".
[10] J. Carmona, J. Cortadella, M. Kishinevsky, "A Region-based Algorithm for DiscoveringPetri Nets from Event Logs", 2009.
[11] Jianmin Wang, Tengfei He, Lijie Wen, Nianhua Wu et al, "A Behavioral Similarity Measure betweenLabeled Petri Nets Based on PrincipalTransition Sequences", 2009.
[12] Wil M.P. van der Aalst, "Decomposing Petri Nets for Process Mining - A Generic Approach", Technical Report, BPM Center, 2012.
[13] Wil M.P. van der Aalst, V. Rubin et al, "Process Mining: A Two-Step Approach usingTransition Systems and Regions", 2006.
[14] W.M.P. van der Aalst, H.A. Reijers, M. Song,"Discovering Social Networksfrom Event Logs", Computer Supported Cooperative work, 14(6), 2005, 549–593.