

Analysis of Process Mining Model Using Improved HMM

V.Priyadharshini¹, Dr. A. Malathi²

^{1,2}Ph. D. Research Scholar, Assistant Professor²

¹²Government Arts College (Autonomous) Coimbatore – 18, India

Abstract: Process mining is a process management system that analyzes business processes built on event logs. The knowledge is extracted from event logs by using knowledge retrieval techniques. The process mining algorithms are capable of automatically discovering models to provide 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 improved hidden markov model 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, Sequence Clustering, Hidden Markov Model, Helix 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 [5]. 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 raising 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 [6]. The business process mining attempts to reconstruct a complete process model from data logs that contain real process execution data [1]. 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 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.

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

Ferreira et al [4] demonstrated the usefulness of applying the Sequence Clustering technique to Process Mining, which can withstand noise and provide added flexibility. The algorithm is a Model Based Clustering technique, which makes use of first-order Markov chains. But in order to capture additional knowledge about business processes, such as their main stages, or their distinctive profiles, there are more elaborated models such as HMMs. Sequence Clustering can also be used as a pre-processing technique. After separating the traces into separate clusters, existing Process Mining techniques can be used to retrieve the process model from each cluster.

III. EXISTING SYSTEM

The use of HMMs as a framework for Sequence Clustering is a relatively unexplored area, and there are relatively few references in the literature. Initial work was presented by Smyth [8], where a distance measure using HMMs was used to cluster the sequences assuming the HMM structure was known *a-priori*, as well as the number of clusters. The optimal number of clusters is then determined maximizing the Partition Mutual Information (PMI), which is a measure of the inter-cluster distances. Discover natural groups in the data by performing sequence clustering [9]. It is well known that data clustering is inherently a more difficult task than supervised organization, and this difficulty exacerbates if sequential data are considered: the structure of the underlying process is often difficult to infer, and typically different length sequences have to be dealt with.

IV. PROPOSED SYSTEM

Process mining deals with the discovery of process models (i.e., structures that model behavior) from event-based data. The goal is to construct a process model that react the behavior it has been observed in some kind of event log. An event log is a set of finite event sequences, whereas each event sequence corresponds to one particular materialization of the process. Process modeling languages, such as Petri nets [7], can then be used to capture the causal relationships of the steps, or events, in the process. While many different process mining approaches have been proposed over the last decade, no standard measure is available to evaluate the quality of such a learned model [10].

IV. 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

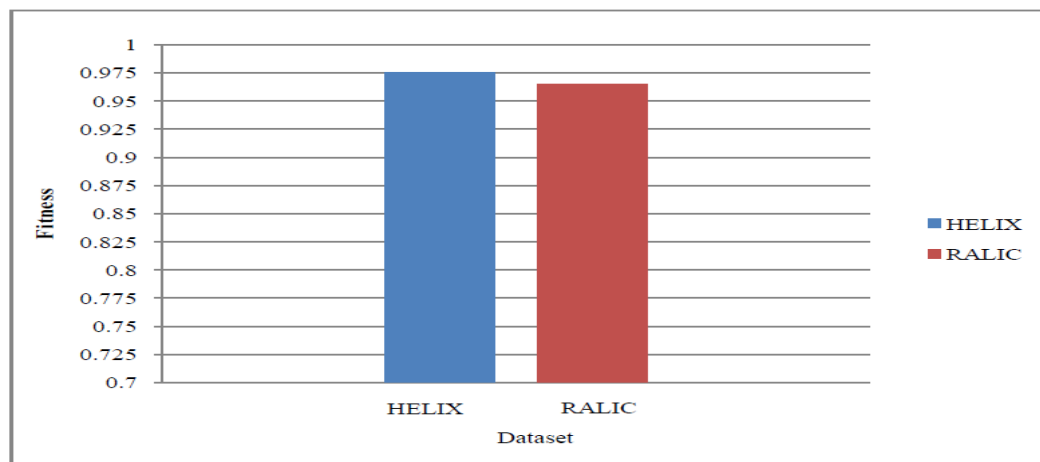


Fig 1. Fitness value of improved HMM

VI. 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].Arya Adriansyah, Peter van den brand et al, “Process mining manifesto”, Business process management, Vol. 99, 2012, 169-194.
- [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, Vol.10, No.10, 2010.
- [4].Process Mining Manifesto. IEEE Task Force on Process Mining.
- [5].Aalst, W.M.P.v.d., et al.,” Workflow Mining: Discovering Process Models from Event Logs”. IEEE Transactions on Knowledge And Data Engineering Vol. 16, No.9, 2004 1128–1142.
- [6].A. Rozinat, M. Veloso, W.M.P. van der Aalst, “Using Hidden Markov Models to Evaluate the Quality of Discovered Process Models”.
- [7].Ferreira, D., Zacarias, M., Malheiros, M., Ferreira, P. “Approaching Process Mining with Sequence Clustering: Experiments and Findings”, Vol. 4714, 360–374, Springer, 2007.
- [8].Smyth, P., “Clustering Sequences with HMM”, Advances in Neural Information Processing, MIT Press 9, 1997.
- [9].Bicego, M., Pekalska, E. Tax, D. Duin,” Component-based discriminative classification for hidden markov models”, Pattern Recognition in press, 2009.
- [10].Gil Aires Da Silva, Diogo R. Ferreira, “Applying Hidden Markov Models to Process Mining”.