International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 9 (2015) pp. 22853-22861

http://www.ripublication.com

Handling Unstructured Event Logs Using Process Discovery Algorithms

K.Lalitha devi, M.Sc., M.Phil.,

Assit.prof in Computer Science,
M.R.Govt.Arts College, Mannargudi.

M.Suriakala, M.Sc., M.Phil., Ph.D.,

Assit.Prof., in Computer Science,
Dr.Ambethkar govt college, Chennai

Abstract

The aim of process mining is the construction of process models based on available log of data. A process model is a graphical representation of a business process that describes the dependencies between activities that need to be executed collectively for realizing specific business objectives. Based on an event log, a process model is constructed and capturing the behavior seen in the log. Today's information system logs consist of enormous amounts of events. Most information system stores such information in unstructured form. This paper discusses about process discovery and supported algorithm for handling unstructured event logs.

Keywords: Event Logs, α-algorithm, Heuristic mining, Fuzzy mining.

Introduction

The goal of process mining is to use event data to extract process related information e.g., to automatically discover a process model by observed events recorded by some enterprise system. Most of the data stored in the digital universe is unstructured and organizations have problem dealing with such data. One of the main challenges of today's organization is to extract information and value from real data stored in their information systems which may have noise and incomplete data. In process mining techniques uses many process models to automate the business processes such as existing approaches are BPM (Business Process Management),WFM (Work Flow Management). Process aware information systems (PAISs) include the traditional

WFM systems and also include that provide more flexibility or support specific tasks. Many notations are available to model operational business processes (e.g., Petrinets, UML and EPCs). The model is automated by using process mining techniques with supporting algorithms for process discovery. The goal of process model is to decide which activities need to be executed and in what order activities can be executed sequentially, activities can be optional or concurrent and the repeated execution of the same activity may be possible. Process discovery is one of the most challenging process mining tasks. There are three types of process mining: Discovery, Conformance and Enhancement. The combination of discovery task and the control flow perspective is often referred to as process discovery. This paper discusses about which algorithm is suitable for unstructured event logs to discover the hidden knowledge.

Event Log

An event log is basically a table. It contains all recorded events that relate to executed business activities. Each event is mapped to a case. A process model is an abstraction of the real world execution of a business process. A single execution of a business process is called process instance. They are reflected in the event log as a set of events that are mapped to the same case. The sequence of recorded events in a case is called trace. The model that describes the execution of a single process instance is called process instance model. A process model abstracts from the single behavior of process instances and provides a model that reflects the behavior of all instances that belong to the same process. Cases and events are characterized by classifiers and attributes. Classifiers ensure the distinctness of cases and events by mapping unique names to each case and event. Attributes store additional information that can be used for analysis purposes.

Table 1: Example Event Log

Case ID	Event ID	Timestamp	Activity	Resource
1	1000	01.01.2013	(A)Order Goods	Peter
	1001	10.01.2013	(B) Receive Goods	Michael
	1002	13.01.2013	(c) Receive Invoice	Frank
	1003	20.03.2013	(D) Pay Invoice	Tanja
2	1004	02.01.2013	(A)Order Goods	Peter
	1005	03.02.2013	(B) Receive Goods	Michael
	1006	05.02.2013	(c) Receive Invoice	Frank
	1007	06.01.2013	(D) Pay Invoice	Tanja
3	1008	01.01.2013	(A)Order Goods	Louise
	1009	04.01.2013	(c) Receive Invoice	Frank
	1010	05.01.2013	(B) Receive Goods	Michael
	1011	10.01.2013	(D) Pay Invoice	Tanja
4	1016	15.01.2013	(A)Order Goods	Peter
	1017	20.01.2013	(c) Receive Invoice	Claire

	1018	25.01.2013	(D) Pay Invoice	Frank
5	1023	01.01.2013	(A)Order Goods	Michael
	1023	10.01.2013	(B) Receive Goods	Michael
	1023	13.01.2013	(c) Receive Invoice	Michael
	1023	20.01.2013	(D) Pay Invoice	Michael

The figure1 illustrates process model of simple purchasing process and figure2 shows mining procedure for above given an event logs. Before applying process mining algorithm an event logs may be involved to log filtering to get mining results. It was created manually so we cannot hope the model will reflect reality.

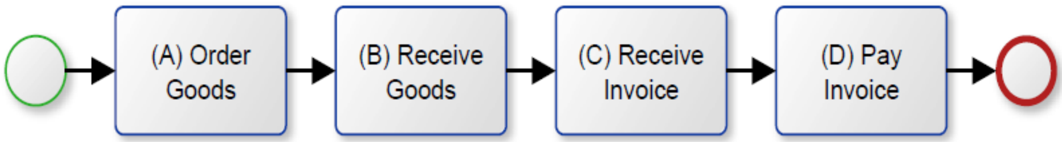

Figure 1: Process Model of Simple Purchasing Process

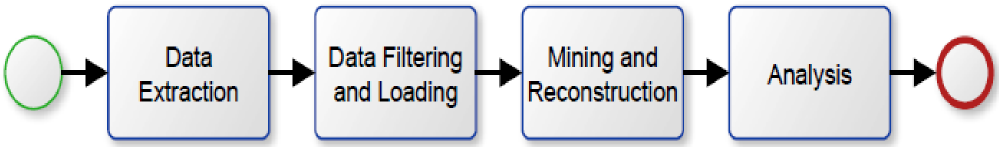

Figure 2: Mining Procedure

α -Algorithm

The α-algorithm was one of the first process discovery algorithms that could effectively deal with concurrency. However the α-algorithm has problems with dealing noise, infrequent /incomplete behavior and complex routing constructs. The basic 8 line algorithm has some limitation when it comes to a particular process patterns e.g., short loops and non-local dependencies. Some of these problems can be solved using various refinements. The α-algorithm guarantees to produce a correct process model for the underlying processes by a WF net that does not contain duplicate activities (two transitions with the same activity label) ,silent transitions(activities that are not recorded in the event log) and does not use the two constructs for the precise requirements. Figure 3 illustrates the alpha algorithm produces a process model with concurrent activities B and C. No execution sequence in the model is able to reproduce the trace case 4.So it has poor fitness. This notation in BPM (Business Process Management) for better comparability. The petrinet cannot express OR-split and joins directly. Figure 4 illustrates under-fitting process model .The traces ABCD, ACBD, ACD are possible from an event log. The sequence

ABCBCD is also possible although it is not included as a trace in event log. If a process model is too general it is called under-fitting. It has poor precision. In process mining when finding results the major challenges are met and have to find solution between fitness, precision, simplicity and generalizability

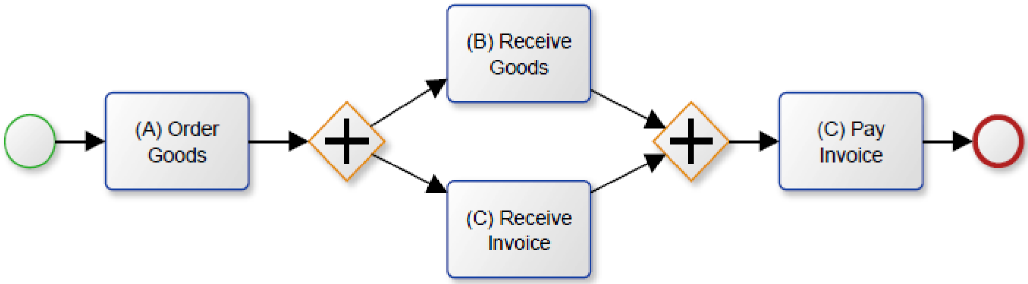

Figure 3: Mined model reconstructed by α-algorithm

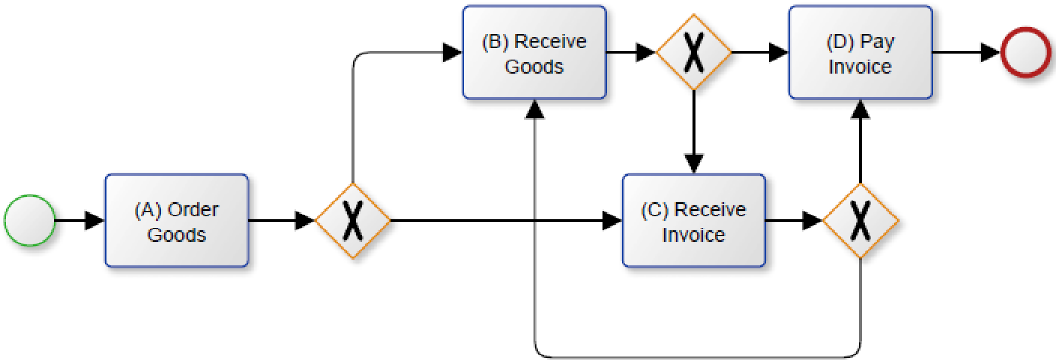

Figure 4: Illustrates Under-Fitting Process Model

Heuristic Mining

Heuristic mining algorithms use a representation similar to causal nets. The notation of causal nets referred to as C-nets. Cnets are a more suitable representation for process discovery. It takes frequencies into account for heuristic mining .We can calculate the value of the dependency relation between any pair of activities is referred to as dependency graph. The dependency graph does not show routing log. The different model can be generated by adjusting the thresholds in an event log. It is used to user can decide to focus on the mainstream behavior or to also include low frequent (i.e., noisy) behavior. The two thresholds cannot be used to remove low frequent activities. This should be done by preprocessing the event log.

Fuzzy Miner:

The Fuzzy miner is one of the younger process discovery algorithms, and was developed by Fluxicon co-founder Christian W. Günther in 2007. It is the first algorithm to directly address the problems of large numbers of activities and highly unstructured behavior. Fuzzy mine approach provides an extensible set of parameters to determine which activities and arcs need to be included.This approach can construct hierarchical model .i.e., less frequent activities may be moved to subprocesses.The fuzzy miner has many parameters that allow the user to influence the resulting model using different settings of these parameters.It is also be possible to abstract from activities rather than aggregated them. Activities can also be removed by filtering an event log before applying discovery algorithm. Figure 5 shows dependency between activities, A is followed three times by B and two times by C. The thickness paths of the dependency graph illustrate the highways in the process model.

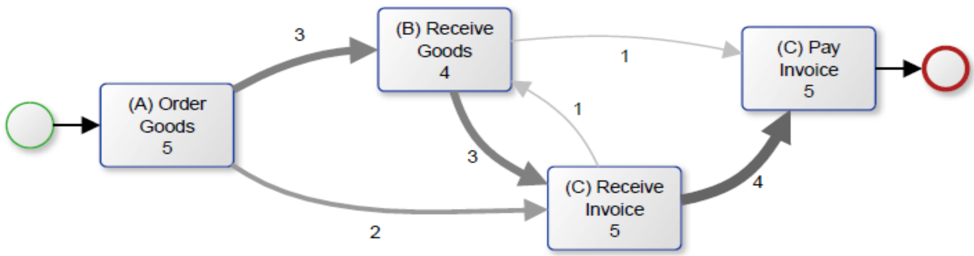

Figure 5: Mined Process Model Using Fuzzy Miner Algorithm.

Case Study

There is popular a process mining tool is ProM is available, that is an open source toolkit Technology. Prom is a good choice to explore process mining, because it has consistently been at the forefront of that technology. In case study, when we applied an event log in ProM tool, we find out some results from real event log which has got from Business processes of whole sale medical distributor. An event logs have 22 cases, 16 originators, 646 events. It is less structured event logs.(which has infrequent and incomplete).When applied mining algorithm and alpha algorithm is applied on an event logs, the result was flower model. This type of model allows for any sequence starting with start and ending with end and also containing any order of activities in between. But this model does not contain any knowledge. Although, the result model is simple and has a perfect fitness. But not sufficient. The following figures show the results.

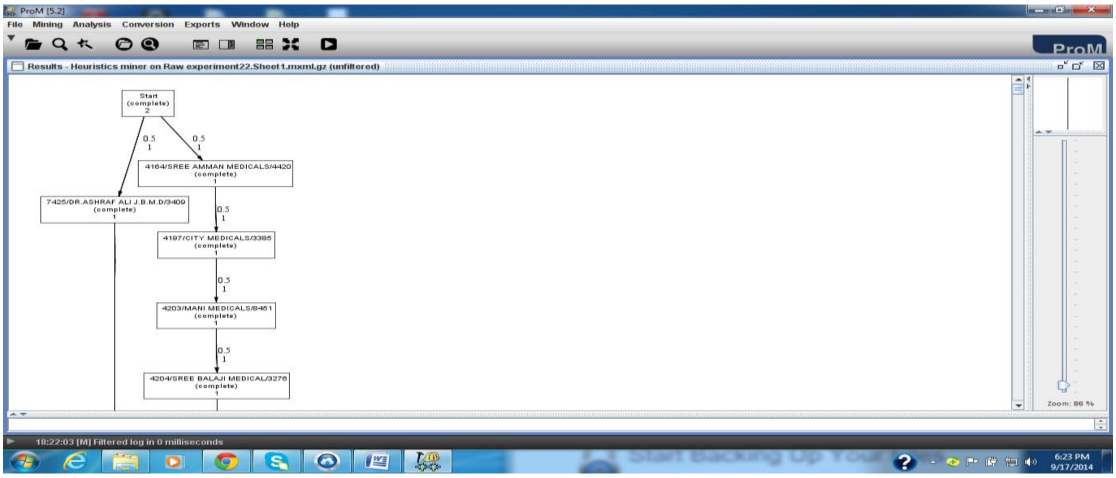

Figure 6: Result of Heuristic Algorithm Applied In Prom Tool

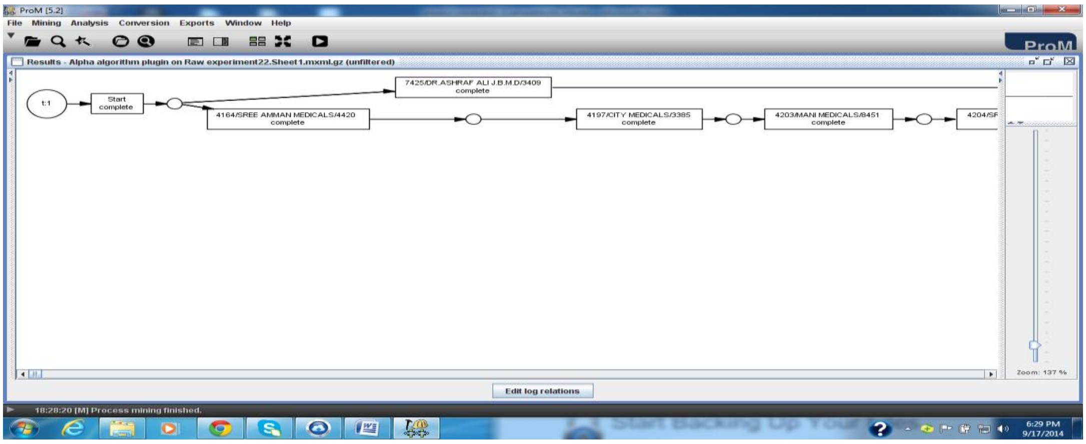

Figure 7: Result of Alpha algorithm applied in ProM tool

In this situation, we can try to apply other techniques to find optimal solution from available existing algorithm for handling real event logs. So we have applied fuzzy miner algorithm on an event logs. The Mining results are as follow for adjusting the parameters Node filter, edge filter which values denote the dependency relations between the activities in the form of hierarchical.when applying an event log using heuristic algorithm,we cannot adjusting the parameters value. We can get optimal result only by filter the real logs when preprocessing steps. The Figure 8 illustrates the node filter, the cut off setting is 0.381.when we set parameter value 0.381 in node filter, the result will be show like all the activities of value 0.381 which denotes the dependency relations between activities and above values will be visible. Otherwise activities of below value 0.381 will be aggregated and put into the corresponding clusters depends upon the dependency relation values called cluster nodes.

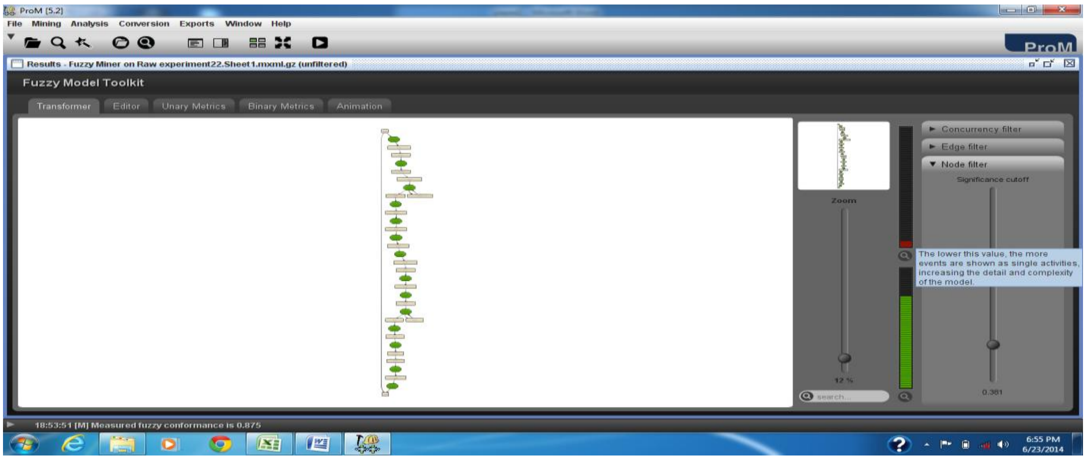

Figure 8: Node Filter

A figure 9 and 10 illustrates the edge filters. It denotes the model after the parameter adjustment of cutoff is 0.606, utility is 0.231, and fuzzy edges value is 0.415. This adjustment of edge filter is used to show the model without low frequent activities or remove from the model for best analysis of the event logs.

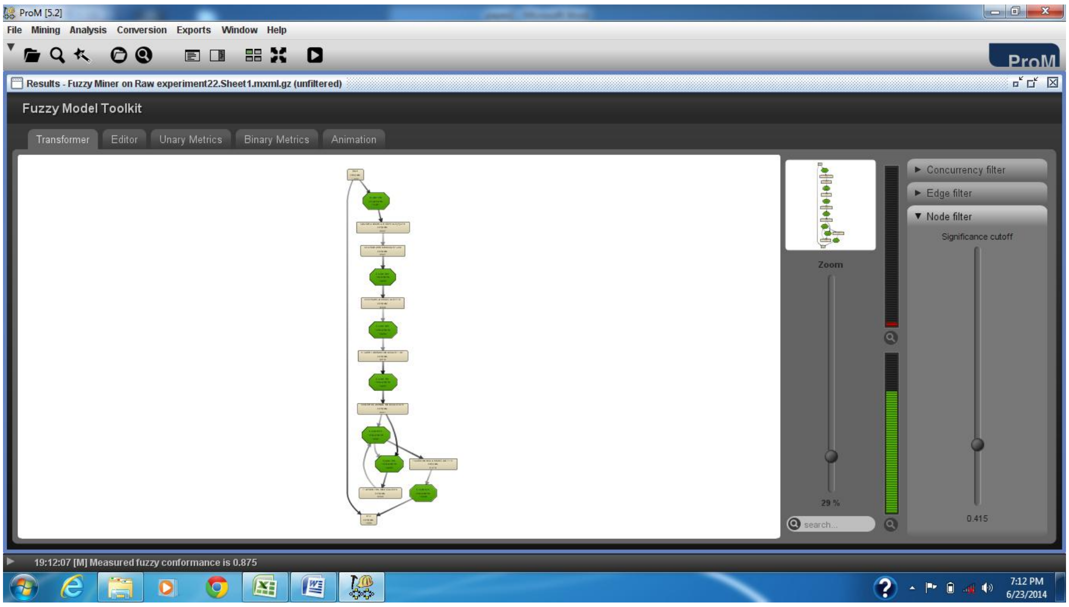

Figure 9: Best Edge Filter

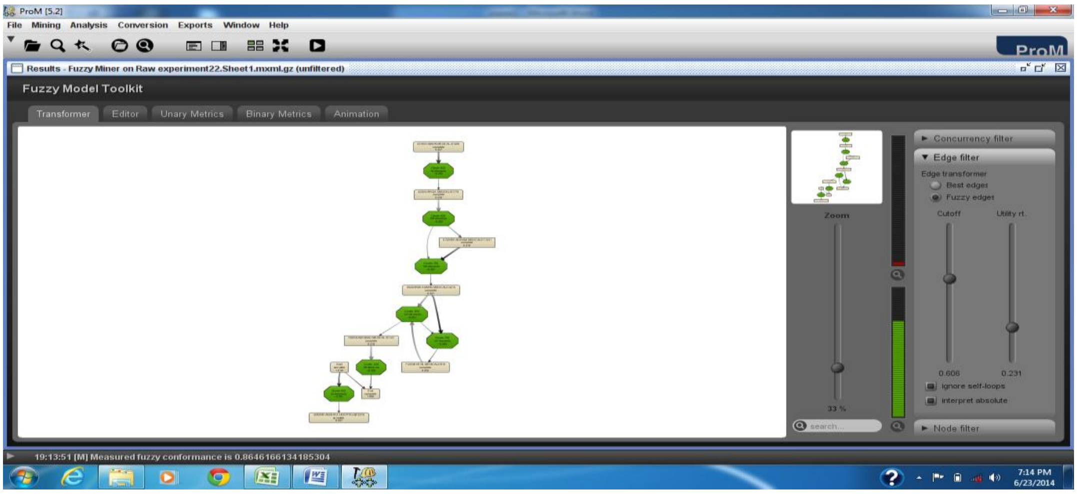

Figure 10: Fuzzy Edges

Conclusion

For real-life logs, the alpha-algorithm is almost never the right choice. Although, one of the advantages is that a Heuristic net can be converted to other types of process models, such as a Petri net for further analysis in PROM and derives XOR and AND connectors from dependency relations. It can abstract from exceptional behavior and noise (by leaving out edges) before find dependency graph. It is also suitable for many real-life logs.Although, the fuzzy model cannot be converted to other types of process modeling languages, but we can use it to animate the event log on top of the created model to get a feeling for the dynamic process behavior.

Reference

[1] Van der Aalst, W. M. P. 2011. Process Mining: Discovery, Conformance and Enhancement of Business Processes, (1st Edition) Berlin; Heidelberg:Springer.

[2] Van der Aalst, W. M. P., Andriansyah, A., Alves de Medeiros, A. K., Arcieri, F., Baier, T., Blickle, T., Bose, J. C., Van den Brand, P., Brandtjen, R., and Buijs, J. 2012."Process mining manifesto," In BPM 2011 Workshops Proceedings, pp. 169–194.

[3] Weske, M. 2012. Business process management concepts, languages, architectures, Berlin; New York: Springer.

[4] PROCESS MINING Article StB Prof. Dr. Nick Gehrke, Nordakademie, Chair for Information Systems Köllner Chaussee 11 D-25337 Elmshorn , nick.gehrke@nordakademie.de Michael Werner, Dipl.-Wirt.-Inf. University of Hamburg Chair for Information Systems Max-Brauer-Allee 60 D-22765 Hamburg michael.werner@wiso.uni-hamburg.de.

[5] Fuzzy Mining – Adaptive Process Simplification Based on Multi-Perspective Metrics Christian W. G¨unther and Wil M.P. van der Aalst Eindhoven University of Technology P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands fc.w.gunther, w.m.p.v.d.aalstg@tue.nl.

[6] Minaei-Bidgoli B., Punch W.F., Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System, GECCO 2003 Conference, Springer-Verlag, Vol2, Chicago, USA; July 2003. pp.2252-2263.

[7] Mostow J., Beck J., Cen H., Cuneo A., Gouvea E., and Heiner C., An educational data mining tool to browse tutor-student interactions: Time will tell! Proceedings of the Workshop on Educational Data Mining, Pittsburgh, USA; 2005. pp.15-22.

[8] Rusu, L., Breşfelean, V.P., Management prototype for universities. Annals of the Tiberiu Popoviciu Seminar, Supplement: International Workshop in Collaborative systems,Volume 4, 2006, Mediamira Science Publisher, Cluj-Napoca, Romania, pp. 287-295.

[9] Vanda me J.P., Meskens N., Superby J.F., Predicting Academic Performance by Data Mining Methods, Education Economics, Volume 15, Issue 4 December 2007 , pp.405 – 419.

[10] Van der Aalst, W.M.P., Weijters, A.J.M.M., Maruster, L. Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data. Engineering* 16(9), p.1128–1142, 2004.