ISSN: 0975-766X
CODEN: IJPTFI
Available Online through
Research Article
www.ijptonline.com

ANALYTICAL STUDY ON CONFORMANCE CHECKING IN TERMS OF FITNESS AND BEHAVIROAL APPROPRIATENESS IN CONTROL FLOW SEQUENCE PATTERN USING PROCESS MINING ALGORITHMS

R. Ramya[1*], Dr. M. Suriakala[2],
[1*]Research Scholar, Bharathiar University, Coimbatore, Tamil Nadu, India; rv_ramya@yahoo.com.
[2]Assistant Professor, Department of Computer Science, Dr. Ambedkar Government Arts College, Chennai, India.
Email: suryasubash@gmail.com

Received on 06-08-2016 *Accepted on 10-09-2016*

Abstract

Process mining comprises the research area which is concerned with knowledge discovery from event logs. Process mining technique focuses on the conformance checking by comparing the discovered process models with the original real-life event logs in order to evaluate the goodness of the process model. This paper makes a comparative study between the discovered process models with respect to the real-life behavior as captured in event logs using the heuristic miner algorithm. To achieve this, the metrics of fitness and appropriateness is considered and the metrics are evaluated and measured by comparing the process flow of planned process model and its observed model using the heuristic miner algorithm.

In Business processes, Flow of events in particular planned sequence is said to be discovered or designed process model (Work Flow Model) and observing the sequence of real time event logs is observed model. Process mining is the young research discipline; it is used to discover knowledge from event logs. Event logs are the processes that are extracted from the information systems like transaction log, Ms-Excel spread sheet or a normal database tables. A study is conducted on the observed event logs from the Business process outsourcing organization which deals with computer technical faults. The output predicts the ratio of deviated fitness and behavioral appropriateness, of process flow with respect to the work flow model.

Key Words:

Business Process Mining, Conformance Checking, Control Flow Bench Mark, ProM, Workflow Model.

1. Introduction

Process Mining is a process management technique which allows analysis of Business Process based on event logs. It is a method to analyze event logs for discovery, conformance and enhancement of the process. Process oriented information systems store process related data in ordinary databases or even normal Ms-Excel work sheets. But the relation between process and data is non-trivial to the rescue.

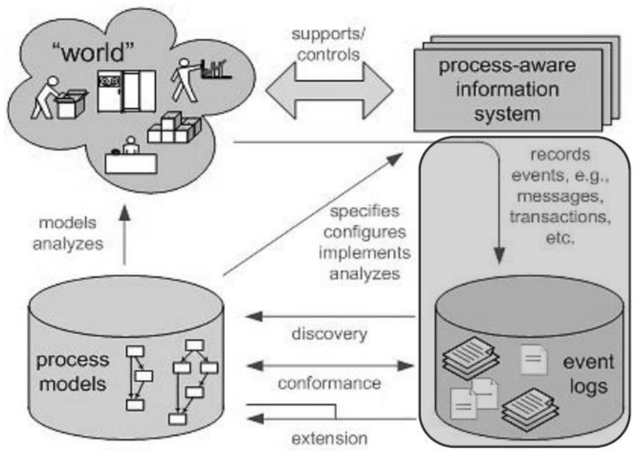


Figure 1. Three types of process mining: discovery, conformance and enhancement.

Process Mining Perspective:

i) Process Perspective: Process Perspective focuses on the control flow i.e., ordering of the activities. The goal is to find the good characterization of all possible paths.

ii) Organizational Perspective: Organizational Perspective focuses on the originator field i.e., which performers are involved and how they are related. The goal is find the social network in the organization.

iii) Case Perspective: Case Perspective focuses on the properties of the cases. Cases can be categorized by their paths in the process or by the originators working on a case.

[25]Over the last decade, event data have become readily available and process mining techniques have matured.

Moreover, managements trends related to process improvement, e.g., Six Sigma, TQM (Total Quality Management), CPI (Continuous Process Improvement), and CPM (Corporate Performance Management) can benefit from process mining.

The idea of process mining is to discover the process model, monitor and enhancement of real processes by getting knowledge from event logs readily available in the systems. In the three powerful areas of process mining application, many organizations are interested in information about conformance of its processes to rules that should be observed.

[27]However, process mining research shows that process executions in reality often deviate from documented process models, potentially violating security and compliance policies. As models enable various analysis techniques ranging from verification to simulation, it is essential to provide diagnostic information about the conformance of process models with respects to event logs recording the real behavior. So, among the three areas of process mining this paper conducts an observation and a case study on conformance checking that is checking how far the observed behavior matches with the modeled behavior of the workflow model in an organization.

2. Materials and Methods

In this paper a case study based on a log of computer related service request processes are presented. The event log was generated by Workflow Management System supporting company operation which concentrates on providing service to the requests based on implementation of computer, hardware and its software queries. This support is provided by a BPO company and it has four management systems namely Request Management, Incident Management, Problem Management and Change Management. Workflow and process management techniques are used in this company to add new facilities faster, as well as additional capacity on demand for customers of any size.

The event log used in the study described in this paper includes 3000 events recorded from January 1st, 2014 to February 20th, 2014. The log contains information about 5281 process instances. The analysis of the event log aimed at answering the questions raised by Incident Management managers concerning organization operation and its difference between the work flow model and the observed instances. The main aim of the paper is to concentrates on the quality analysis of the Incident Management workflow process by using the process mining algorithms.

Table No.1: Example of Event log.

No. of	Log Traces
256	abdga
179	acdga
58	acdfga
63	achefa
29	acdhefa

The process used throughout the paper concerns the Incident Management Process within a business process outsourcing domain (Fig.2.).Incidents or breakdowns are inevitable features of an organization. Incidents affect the normal service processes and in turn impact the productivity and business of the organization. The Incident Management plays an important role in resolving incidents and restoring the normal service operations as soon as possible. The workflow process of Incident Management involves the following types of activities that are performed: Incident gets logged through User. Service Desk/Resolver Group identifies the job type whether it is an Incident Request. The Verification is made for the user details and for any pending incident to the same user is unresolved .If the incident unresolved is available, the service desk will close the new incident request and prioritize the incident. Categorized and prioritized information will be updated to the user. Here, the job is categorized and prioritized based on the issue description and it determine the scope of the job, and it will be updated to the incident table, and then it is informed to user and the job is closed.

Incident Request – it refers to request submission activity by the user to the technical team.

Log Incident – it refers to the incident gets logged for the recovery.

Categorize Incident – it refers to categorize the service problem

Prioritize Incident – it refers to prioritize the incident

Initial Diagnosis – it refers to initial problem identification

Resolve and Recover – it refers to the problem rectification.

Close – it refers to closing incident.

The above events have taken place in the Incident Management system of business process domain. The real time event logs are recorded for this work flow model and the comparison with the above work flow model is presented in detail in this paper.

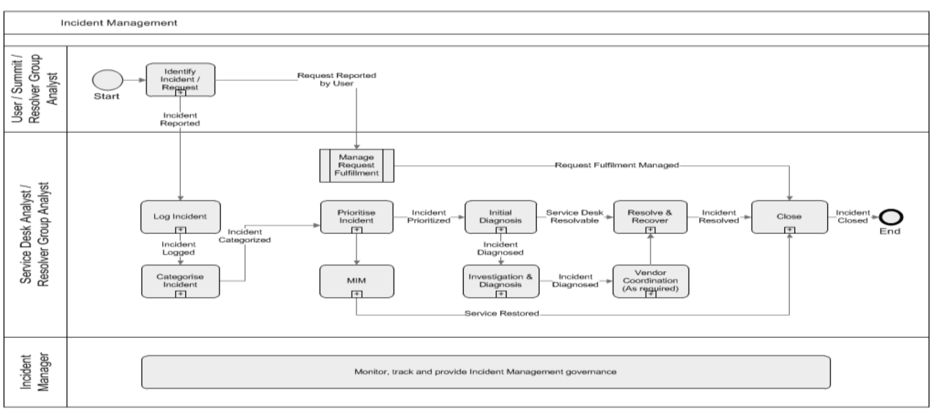


Figure 2. Incident Management Workflow diagram.

Table 2. Event Log.

case_id	event id	timestamp	activity	Resource
152	39464	14/2/2013 05:01:22 PM	register request	Karthikrishnan
152	39467	14/2/2013 02:12:20 PM	log incident	Mike

152	39468	14/2/2013 04:13:10 PM	categorise incident	John Stephen
152	39500	15/2/2013 09:06:23 AM	initial disgniosis	Praveen
152	39501	15/2/2013 03:04:24 PM	resolve and recover	Ravi
152	39504	16/2/2013 07:06:26 AM	fulfill request	Petchi Ganesan
323	39059	16/1/2013 02:02:30 PM	register request	Mark
323	39060	16/1/2013 03:11:20 PM	log incident	Karthikrishnan
323	39063	17/1/2013 04:10:10 PM	categorise incident	Mike
323	39065	18/1/2013 05:02:13 PM	prioritise incident	Mark
323	39066	19/1/2013 09:13:21 AM	initial disgniosis	Mike
323	39067	19/1/2013 02:01:14 PM	resolve and recover	Karthikrishnan
323	39070	20/1/2013 03:07:16 PM	fulfill request	Karthikrishnan
977	39190	16/1/2013 05:01:22 PM	register request	John Stephen
977	39193	17/1/2013 02:12:20 PM	log incident	Praveen
977	39191	17/1/2013 04:13:10 PM	categorise incident	Mike
977	39195	18/1/2013 05:20:11 PM	prioritise incident	John Stephen
977	39196	19/1/2013 09:06:23 AM	initial disgniosis	Praveen
977	39199	19/1/2013 03:04:24 PM	resolve and recover	Mike
977	39201	20/1/2013 07:06:26 AM	fulfill request	Mark

2.1. Alpha Miner Algorithm

[11]Alpha-algorithm used in process mining, aimed at reconstructing causality from a set of sequences of events and it constructs a workflow nets from event logs. It orders events sequentially, such that each event refers to a case and activity. It has problem with noise, infrequent behavior and complex routing constructs. [26]Existing commercial tools, such as Perceptive Process Mining and Fluxicon Disco, and academic tools, such as Inductive Visual Miner, mainly focus on the control-flow perspective, and provide for data-aware process exploration. So the event logs are traced with alpha miner algorithm in PROM tool and the Petri net model of the observed event logs with complexity and deviations in the control flow are identified.

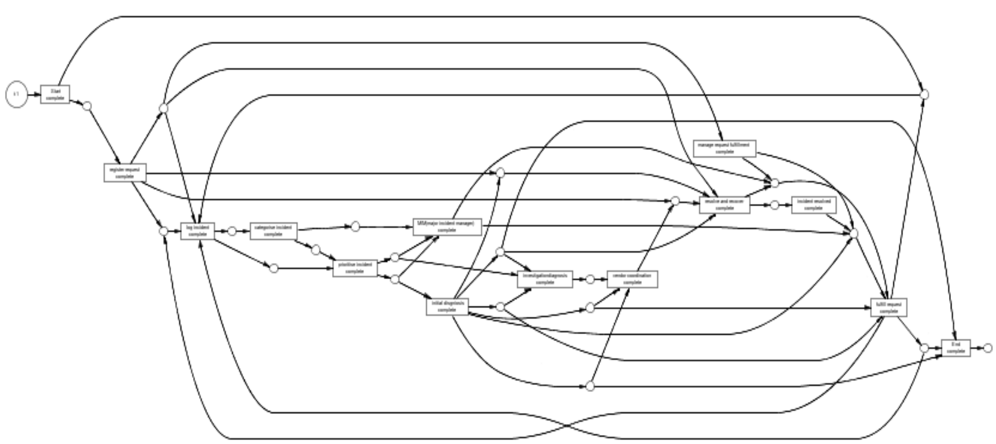


Figure 3. Original Process Model.

2.2. Heuristic Miner Algorithm

Heuristic Mining algorithm is a practically applicable mining algorithm that can deal with noise, and it gives the principal behavior of the system, registered in an event log. Heuristic Miner is the extension of alpha algorithm and it considers the frequency of traces in the log. To get the process model it considers the sequence of the events within a case. Control flow perspective of The Heuristic Miner plug-in is used to find the deviation in the observed event sequences to form the observed work flow model using the PROM tool. Figure 5. presents deviated control flow from the original control flow constructs. This is achieved through Heuristic Miner algorithm and its conversion plug-ins. This control flow provides a long distance dependency of events and their fitness measures.

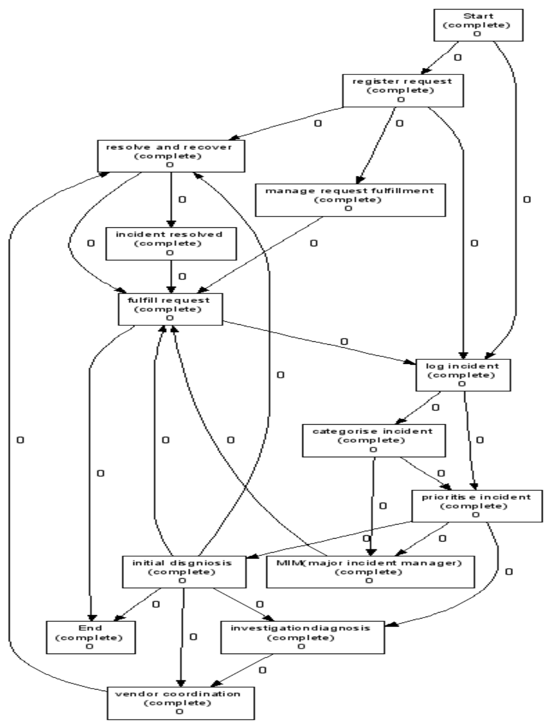


Figure 4. Mined Process Model using Heuristics Miner.

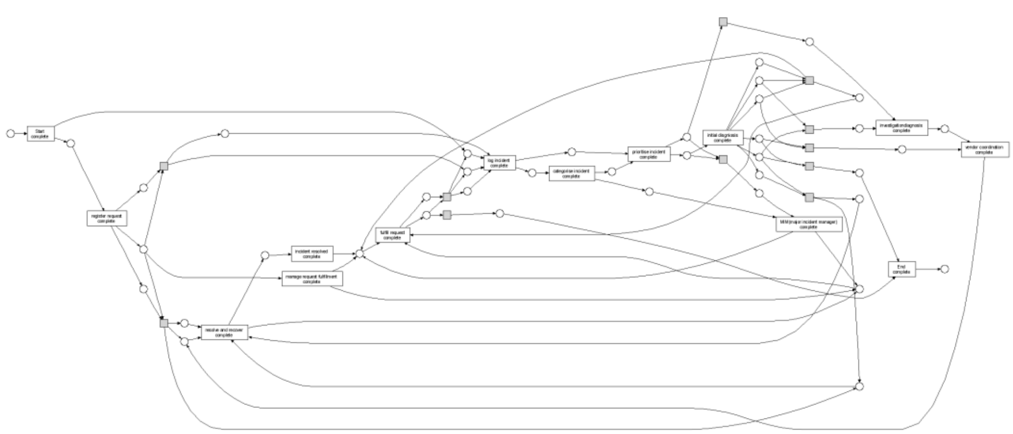
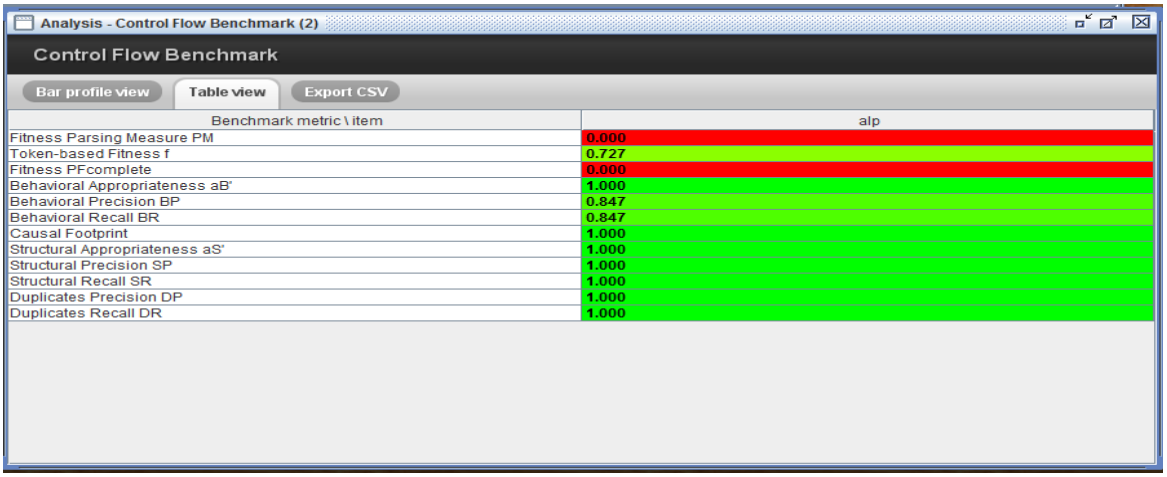


Figure 5. Heuristics Net to Petri Net with deviations.

Table 3. Control Flow Benchmark.



Analysis - Control Flow Benchmark (2)

Control Flow Benchmark

Bar profile view | Table view | Export CSV

Benchmark metric \ item	alp
Fitness Parsing Measure PM	0.000
Token-based Fitness f	0.727
Fitness PFcomplete	0.000
Behavioral Appropriateness aB'	1.000
Behavioral Precision BP	0.847
Behavioral Recall BR	0.847
Causal Footprint	1.000
Structural Appropriateness aS'	1.000
Structural Precision SP	1.000
Structural Recall SR	1.000
Duplicates Precision DP	1.000
Duplicates Recall DR	1.000

The following Table No.3, presents the value of deviation from the work flow model and the value should exist between 0to1. Lowest value shows the highest deviation in all the metrics. The Metrics are:

Table 4. Bench Mark Metric with Control flow deviation.

Bench Mark Metric / Item		**Control**
Fitness Parsing Measure (PM) – This is the basic fitness measurement technique. It calculates the fitness of models generated using the Heuristic Miner algorithm.	0 and 1	0
Token-based Fitness (f) - It measures the fitness of the model by replaying every type of trace.	0 and 1	0.727
Fitness PF Complete – It checks for the number of events that could be parsed without problems during replay.	0 and 1	0
Behavioral Appropriateness (aB') – It measures how much behavior is allowed by the model which is not present in the log.	0 and 1	1
Behavioral Precision (BP) – It checks whether enabled activities in the model actually correspond to observed executions in the log.	0 and 1	0.847
Behavioral Recall (BR) – It is used in pattern recognition and information retrieval, through the construction of a confusion matrix.	0 and 1	0.847
Causal Foot Print - A footprint is a matrix showing causal dependencies between activities.	0 and 1	1
Structural Appropriateness (aS') – To express the presence of same behavior in the process model which results in complex model due to duplicate task (Transition with the same label, invisible task, transition without a label or label T)	0 and 1	1
Structural Precision (SP) – It assesses how many causality relations the mined model has that are not in the original model.	0 and 1	1

Structural Recall (SR) – It checks how many causality relations from the original model are not included in the mined model.	0 and 1	1
Duplicates Precision (DP) – It is similar to precision. It checks how many duplicate tasks are mined.	0 and 1	1
Duplicates Recall (DR) - It is similar to Recall. It checks how many duplicate tasks are in the referenced model.	0 and 1	1

3. Results and Discussion

The following metrics to measure the control flow of the real time event logs to match with the work flow model produces control flow deviations as 1 which implies the flow is not deviated in terms of following measures:

1. Behavioral Precision BP
2. Behavioral Recall BR
3. Casual Foot Print
4. Structural Appropriateness aS'
5. Structural Precision SP
6. Structural Recall SR
7. Duplicates Precision DP
8. Duplicates Recall DR

The above metrics are explained in table no 2. It is observed that the metric Fitness Parsing Measure shows the value 0.000 which implies that there is a complete deviation for the observed model from the work flow model planned. It means the real time event traces are recorded not as planned in the work flow model. Fitness PF Complete shows the value of 0.000 which implies that there is a deviation in fitness of complete event log due to missing of sequential events Token-based Fitness (f) shows the value of 0.727 which implies that there is a deviation in the fitness of single event. It means the complete event log is separated as tokens and each token is tested for its sequential occurrence.

From the results it is learnt that observed event log (process model) is deviated from the work flow model. The events are not taken place in the planned work flow in the business environment. Most of the times the events are not taken place in the planned sequential order. The results are useful for future improvement to avoid deviation in the work flow in the business environment.

4. Conclusion

In this paper, the metrics of fitness, behavioral appropriateness, token based fitness, quality, relevant event traces, structural appropriateness and structural quality are measured for a set of event logs. It is observed the fitness of this observed model is completely 0 and the result shows that the sequence of events are not occurring as per plan in the work flow model. In addition the result produced in token based study also proves that the traces are not occurred in order with its low value results. It is suggested that the control flow constructs of the observed event logs have to follow the workflow model more closely to improve the fitness and the appropriateness. Future research may be carried out in complexity metrics of the same event logs.

5. Reference

1. Ana Karla Alves de Medeiros, "Process Mining-control flow mining algorithm", Eindhoven University of Technology.
2. W.M.P. van der Aalst, A.K. Alves de Medeiros, and A.J.M.M. Weijters, 2005, "Genetic Process Mining", Department of Technology Management, Eindhoven University of Technology, 26th International Conference, ICATPN 2005.
3. Saravanan M.S. Rama Sree R.J "A Role of Heuristics Miner algorithm in the Business Process System", Comp. Tech. Appl., Vol2 (2), 340-344.
4. W.M.P. van der Aalst*, A.J.M.M. Weijters, 2004,"Process mining: a research agenda", Department of Technology Management, Eindhoven University of Technology.
5. Geetika T. Lakshmanan and Rania Khalaf, September/October 2013, "Leveraging Process mining techniques", IBM T.J.Watson Research Center, IT Pro.
6. ZahraFarzanyar, MohammadrezaKangavari, 2012, "Efficient Mining of Fuzzy association Rules from Preprocessed Data set", Computing and Informatics, Vol. 31, , 331-347.
7. Esmita.P.Gupta, November 2014, "Process Mining A Comparative Study", International Journal of Advanced Research in Computer and Communication Engineering, Vol.3, Issue 11.
8. StB Prof. Dr. Nick Gehrke, Michael Wener, Dipl.-Wirt-Inf. "Process Mining", WISU – die Zeitschriftfur den wirtschaftsstundenten 7/13.

9. Van der Aalst, W., Van Dongen, B., Gunther, C., Mans, R., de Medeiros, A.A., Rozinat, A., Rubin, V., Song, M., Verbeek, H., and Weijters, 2007, A.ProM 4.0: Comprehensive support for Real Process Analysis. In kleijn, J and Yakovlev, A editors, Application and Theory of Petri Nets and other Models of Concurrency (ICATPN 2007), Lecture Note in Computer Science, Published by Springer-Verlag, Berlin, Vol 4546, pp.484-494.
10. W.M.P. van der Aalst and A.J.M.M. Weijters, 2004, Process Mining, Special Issue of Computers in Industry. Elsevier Science Publishers, Amsterdam, Vol.53.
11. Gianluigi Greco et al, Aug 2006, "Discovering expressive process models by clustering log traces, IEEE Transactions on knowledge and data engineering, Vol.18, No.8.
12. A. Rozinat and W.M.P. Van der Aalst., March 2008, Conformance checking of processes based on monitoring real behavior. Information Systems, 33(1):64{95.
13. M. Becker and R. Laue.A, February 2012, comparative survey of business process similarity measures. Computers in Industry, 63(2):148{167.
14. AJMM Weijters, Wil MP van der Aalst, and AK Alves De Medeiros., 2006, Process mining with the heuristics miner-algorithm. Technische Universiteit Eindhoven, Tech. Rep. WP, 166.1{34.
15. S. Goedertier, D. Martens, J. Vanthienen, and B. Baesens., June 2009, Robust Process Discovery with Artificial Negative Events J.Mach. Learn. Res, 10:1305{1340.
16. J. De Weerdt, M.DeBacker, J.Vanthienen, and B.Baesens, April 2011, A robust F-measure for evaluating discovered process models. In 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), pages 148{155.
17. J. Munoz Gama and J. Carmona., 2010, A Fresh Look at Precision in Process Conformance. In Business Process Management, number 6336 in Lecture Notes in Computer Science, pages 211{226. Springer Berlin Heidelberg.
18. J. Wang, R.K. Wong, J. Ding, Q. Geo, and L. Wen., October 2013, Efficient Selection of Process Mining Algorithms IEEE Transaction on Services Computing, 6(4):484{496.
19. J. Bae, L. Liu, J. Caverlee, L. Zhang, and H.Bae., 2007, Development of Distance Measures for Process Mining, Discovery and Integration: International Journal of Web Services Research, 4(4):1{17.
20. B. Vazquez-Barreiros, M. Mucientes, and M. Lama, February 2015, ProDiGen: Mining complete, precise and minimal structure process models with a genetic algorithm. Information Sciences, 294:315{333.

21. J. Wang, T. He, L. Wen, N. Wu, A.H.M. terHofstede, and J. Su., 2010, A Behavioral Similarity Measure between Labeled Petri Nets Based on Principal Transition Sequences. In on the Move to Meaningful Internet Systems: OTM 2010, number 6426 in Lecture Notes in Computer Science, pages 394{401. Springer Berlin Heidelberg.

22. T. Murata., 1989, Petri nets: Properties, analysis and applications. Proceedings of the IEEE, 77(4):541{580.

23. W.M.P. Van der Aalst, A. Adriansyah, and B.F. Van Dongen, March 2012, Replaying History on Process Models for Conformance Checking and Performance Analysis. Wiley Int. Rev. Data Min. and Knowl. Disc., 2(2):182{192.

24. L. Sanchez-Gonzalez, F. Garcia, J. Mendling, F. Ruiz, and Mario Piattini, 2010. Prediction of Business Process Model Quality Based on Structural Metrics. In Conceptual Modeling ER 2010, number 6412 in Lecture Notes in Computer Science, pages 458{463. Springer Berlin Heidelberg.

25. Wil Van der Aalst, Arya Adriansyah, and Boudewijn van Dongen, 2012. Replaying History on Process Models for Conformance Checking and Performance Analysis, Department of Mathematics and Computer Science, Technische Universiteit.

26. Leemans, S.J.J., Fahland, D., van der Aalst, W.M.P., 2014.: Process and deviation exploration with inductive visual miner. In: Proceedings of the BPM Demo Sessions 2014. Volume 1295 of CEUR Workshop Proceedings., CEUR-WS.org (2014) 46.

27. Conformance Checking Based on Partially Ordered Event Data, 2015. X. Lu, D. Fahland, W.M.P. van der Aalst Department of Mathematics and Computer Science Eindhoven University of Technology P.O. Box 513, 5600 MB Eindhoven, The Netherlands (x.lu,d.fahland,w.m.p.v.d.aalst)@tue.nl.

Corresponding Author:

R. Ramya*,

Email: rv_ramya@yahoo.com