Recent Trends in Intensive Computing
M. Rajesh et al. (Eds.)

doi:10.3233/APC210213

Conformance Checking Techniques of Process Mining: A Survey

Ashok Kumar Saini [a, 1], Ruchi Kamra [b] and Utpal Shrivastava [c]
[a]Dept of CSE, JECRC University, Jaipur, Rajasthan, India
[b]Dept of CSE, Amity University, Manesar, Haryana, India
[c]Dept of CSE, G L Bajaj Group of Inst. Mathura, U.P., India

Abstract. Conformance Checking (CC) techniques enable us to gives the deviation between modelled behavior and actual execution behavior. The majority of organizations have Process-Aware Information Systems for recording the insights of the system. They have the process model to show how the process will be executed. The key intention of Process Mining is to extracting facts from the event log and used them for analysis, ratification, improvement, and redesigning of a process. Researchers have proposed various CC techniques for specific applications and process models. This paper has a detailed study of key concepts and contributions of Process Mining. It also helps in achieving business goals. The current challenges and opportunities in Process Mining are also discussed. The survey is based on CC techniques proposed by researchers with key objectives like quality parameters, perspective, algorithm types, tools, and achievements.

Keywords. Conformance Checking; event log; Petri-net; Process Mining.

1. Introduction

Process Mining (PM) is new research that lies between data science and Business Process Management (BPM)[1]. Generally, BPM processes the model rather than event data. It focuses on the designing, controlling, quantity, and optimization of business processes. Traditional data analytical techniques like machine learning and data mining do not consider the end-to-end process. It focuses mainly on patterns or results. There is a missing link between BPM and data science, namely PM [2][3], to improve the process. Nowadays, most organizations use information systems such as BPM, Enterprise Resource Planning systems, etc. These information systems record each activity and describe a process's underlying behaviour, as shown in Figure 1. Each event is related to a movement that belongs to a particular stage of the process[4][5]. PM uses these events to discover, monitor, and improve the process [4]. The organization of the paper is as follows: Section I is the introduction of PM. Section II is PM techniques and it's applicability in different perspectives. Section III is a discussion of tools used in PM. Section IV is a detailed discussion of various research work carried out in CC techniques. Lastly, section V is the contributions and the scope of future research in this domain.

[1] Ashok Kumar Saini, Department of CSE, India
Email: shksaini@gmail.com.

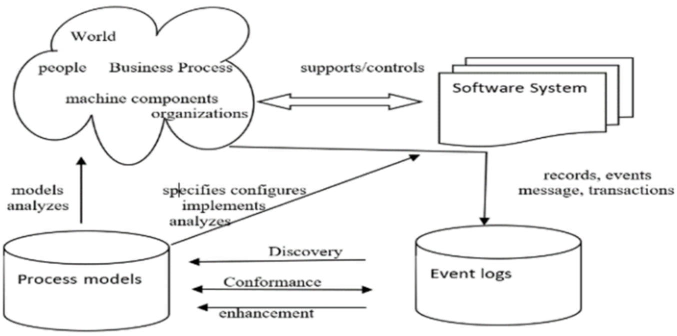

Figure 1. Process mining overview and its techniques

2. Process Mining Techniques Overview

Components of PM in Figure 2 are process discovery that has an event log as input, CC, and enhancement both have log and model as input. Process discovery: First technique of PM is process discovery for discovering the model that replicates log [2]. CC: It checks the conformity of the model with log and assesses whether they described reality. There are four quality parameters [6]: fitness, simplicity, precision, and generalization. A perfect fitness model can replay all traces from beginning to end. For any log (EL) and model (M), then the fitness of the model is:

$$\text{fitness}(\text{EL}, \text{M}) = 1 - \frac{fcost(\text{EL}, M)}{move(\text{EL}) + |\text{EL}| + move(M)}$$

Enhancement: It takes the process model and event log as input and enhances the process model using the observed event log [4][5]. A model is simple if it explains clearly all behaviors [5][6] shown in Figure 3. A precise model does not allow many traces. The flower model is less precise and more generalized. The fitness value varies from 0 to 1. The best-fitted model has one fitness value. The model that is not generalized is also called overfitting.

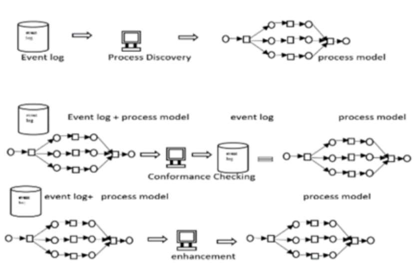

Figure 2. PM techniques in terms of input and output

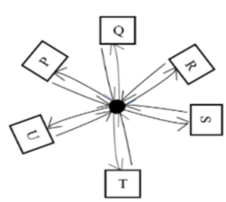

Figure 3. Flower model

- *PM perspective:* There are mainly four types of PM perspective. First is a control-flow, to find the excellent categorization of numerous promising paths

[5]. The second organizational, shows how movements are associated with each other. The third case, emphasizes cases, and the fourth perspective time related to evaluating cases [4][5].

3. Tools and Algorithms Categories

Comparative usability of ProM and Disco tools of PM are shown in table 1. Some other tools are ProM Lite, RapidProM (both are open source), LANA (Lana Labs), SNP (SNP Schneider-Neureither & Partner AG), EDS (StereoLOGIC Ltd), Icris, ProcessGold, ARIS PPM, Fujitsu, Icaro, Minit, myInvenio, QPR, Rialto, etc. At the time of loading event in ProM framework shown in Figure 4.

Table 1. Comparison of Modulation schemes

Tools	ProM	Disco
Class	Open	Commercial
Purpose	General	General
Discovery	supported	supported
CC Checking	supported	Not supported
Societal Mining	supported	Not supported

PM algorithm is categorized into three classes: Deterministic, Heuristic, and Genetic algorithms [6]. Deterministic like α-algorithm provide constant output for the specific input of variables. Heuristic algorithms provide a better solution by trial and error. A genetic algorithm is used when the problem starts with an arbitrary point and tries to find a better solution by introducing random variations [7].

Figure 4. ProM framework when loading event log

4. Proposed Models by Researchers

This section surveys the proposed work on CC. Table 2 has shown the comparative survey. Most CC techniques are created on a control-flow perspective and offline mode, but the conformity of the model also depends on different perspectives like data, time, etc. Online CC framework [9] proposed quantifying the observed behaviour in real-time and controlling the complexity to the constant time of each event. To test the approach, they run a conformance checker for about 70 mins and 256110 events generated by generator PLG about 65 events/sec. Process model [10] discovered from different process discovery algorithms and compared these algorithms on same data streams such as Lossy Counting with Budget, Sliding Window, and Exponential Decay.

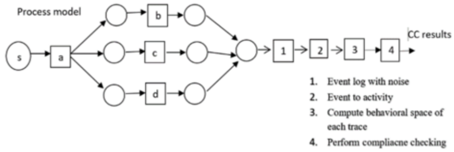

Figure 5. Evolution process

In Figure 5, the first and second step is to build in existing plugins and the probabilistic CC approach is implemented as a plugin in third step [11]. For checking conformity of the model, mapping is required. For noise level 0, their compliance checking technique results in 70.2%, and traditional methods provide 29.8%. They are shown in Figure 6.

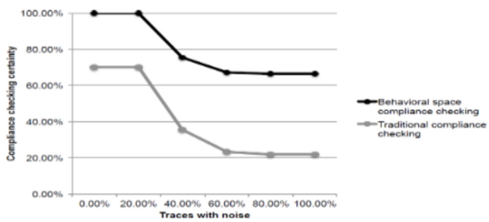

Figure 6. Comparative analysis between traditional and probabilistic Conformance Checking.

The CC technique is based on alignment [13]; for alignment, it is essential to associate the passage of events with the passage of the process model. One more CC technique [14] is based on replay-token. A new approximation CC technique [15] was proposed to compute conformance value in a faster way. [16] found the value of all four parameters fitness f=0.995, precision p=0.996, generalization g=0.958 and simplicity s=0.387. [17] proposed an approach that detects the anomalies in traces stored in PAIS using the ProM tool. [25] proposed a novel framework for PM analysis that uses advances in-memory data processing and graph algorithms that reduce the cost of taking out and converting the event data present in the information system. [28] Increasing the volume of data becomes a challenge as existing PM techniques cannot handle the high volume of data with many activities. Clustering [30] based approach that overcomes the problem of the complex and imprecise model due to large volume of data. PM technique [29] was used to analyze the process of an emergency room in the hospital. But they [27] neither considered real-time concerns in the behavioural domain nor the resources and relationships between actions. [26] expanded the work done in [28], provided approach of reconstructing process model from audit trail logs.

Table 2. Different approaches to Conformance Checking.

Ref	Key Objective	Achievements	Future work	Tool
[10]	Comparing results visually of two different process discovery algorithms.	Users can analyze the internal data structure for handling the event data stream.	Make CC metrics checking the performance algorithms.	ProM
[11]	Creating a mapping between process model and uncertain events.	Applicable on several real-world procedures where traditional CC methods fail	Need to extend the approach of mapping and also help in the selection.	ProM 6
[12]	Detect the deviation between modelled and observed behaviour.	The hierarchical approach is compared with decomposition by manual.	Enhance the projected technique to a more significant class.	ProM
[13]	Maintain alignment between events and model.	Alignment makes it possible to replay the event.	Finding an optimal alignment algorithm.	ProM

[15]	Finding the possible behaviour of subset.	Approximation value is close to actual alignment value.	The best subset method.	ProM
[16]	Find the similarities between PM and event log.	Fitness of the model is 0.87, precision=0.9, simplicity=0.3 8 and generalization=0.98.	Finding learning automata for discovering the process model	ProM
[18]	Design rule sets that show the relationship between tasks. Traditional methods are time-consuming.	The rule set considers noise and imbalance in data and the problem with the alpha algorithm recovered.	Planning of performing more real-world case studies on discovering the model	ProM
[21]	Dealing with the log consisting noise.	Experiments with a synthetic and real-time log.	Future work for dealing with duplicate tasks.	ProM
[22]	An alignment-based replay to enhance the state space.	Handled intertwined state space with the help of alignment-based replay.	Enhance the CC matrix by parameters precision and generalization.	ProM
[23]	Process model generates with minimum information.	A block-structured model that is fit and sound replays all the observed behavior.	Length of two-loop removes the restriction of start and end state.	
[24]	Approach provides the instance graph of an individual log instance.	In this approach, the noise is filtered out from the log.	Developing algorithms that integrated multiple instances.	
[27]	Provided an approach of constructing a PM	The modelling technique is compatible.	Need controlling learning conditions	Flow mark

5. Conclusion and Future Work

The study shows that PM techniques are limited for process discovery and check the conformity of the process model. There are many proposed CC techniques. Most CC techniques are based on control flow and do not provide an actual cause of process deviation. The other parameters like data, resources, time need to be considered. Many CC techniques are token-based, but sometimes they give unpredictable and ambiguous results. The CC techniques based on alignment provide more strong conformity. As the volume of data is increased day by day, the alignment-based approach comes with challenges. The available tools also face difficulties. For massive data, these techniques are inefficient, being a complex process model. To overcome such problems, a decompose alignment technique is used. Such methods are based on the decomposition of model and event log into small components and aligned respectively; the decomposition technique shows better computation time. The detailed study shows the need for future work for the automatic decomposition of the process model and event log with minimum error and execution time.

References

[1] Van der Aalst, W., & Damiani, E. (2015). Processes meet big data: Connecting data science with process science. IEEE Transactions on Services Computing, 8(6), 810-819.

[2] Van Der Aalst, W. (2016). Data science in action. In Process mining (pp. 3-23). Springer, Berlin, Heidelberg.

[3] Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). Fundamentals of business process management (Vol. 1, p. 2). Heidelberg: Springer.

[4] Van, D. (2011). Process mining discovery, conformance and enhancement of business processes. Springer Heidelberg Dordrecht London New York. ISBN.

[5] Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., & Wynn, M. (2011, August). Process mining manifesto. In International Conference on Business Process Management (pp. 169-194). Springer, Berlin, Heidelberg.

[6] Lee, W. L. J., Verbeek, H. M. W., Munoz-Gama, J., van der Aalst, W. M., & Sepúlveda, M. (2018). Recomposing conformance: Closing the circle on decomposed alignment-based conformance checking in process mining. Information Sciences, 466, 55-91.

[7] Mahendrawathi, E. R., Arsad, N., Astuti, H. M., Kusumawardani, R. P., & Utami, R. A. (2018). Analysis of production planning in a global manufacturing company with process mining. Journal of Enterprise Information Management.

[8] Nagy, Z., & Werner-Stark, Á. (2020, March). A Multi-perspective Online Conformance Checking Technique. In 2020 6th Int. Conference on Information Management (ICIM) (pp. 172-176). IEEE.

[9] Burattin, A., & Carmona, J. (2017, September). A framework for online conformance checking. In International Conference on Business Process Management (pp. 165-177). Springer, Cham.

[10] Baskar, K., & Hassani, M. (2019, January). Online Comparison of Streaming Process Discovery Algorithms. In BPM (PhD/Demos) (pp. 164-168).

[11] van der Aa, H., Leopold, H., & Reijers, H. A. (2017, June). Checking process compliance on the basis of uncertain event-to-activity mappings. In International Conference on Advanced Information Systems Engineering (pp. 79-93). Springer, Cham.

[12] Munoz-Gama, J., Carmona, J., & van der Aalst, W. M. (2013, June). Hierarchical conformance checking of process models based on event logs. In International Conference on Applications and Theory of Petri Nets and Concurrency (pp. 291-310). Springer, Berlin, Heidelberg.

[13] Van der Aalst, W., Adriansyah, A., & van Dongen, B. (2012). Replaying history on process models for conformance checking and performance analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(2), 182-192.

[14] Rozinat, A., & Van der Aalst, W. M. (2008). Conformance checking of processes based on monitoring real behavior. Information Systems, 33(1), 64-95.

[15] Sani, M. F., van Zelst, S. J., & van der Aalst, W. M. (2020, June). Conformance checking approximation using subset selection and edit distance. In International Conference on Advanced Information Systems Engineering (pp. 234-251). Springer, Cham.

[16] Naderifar, V., Sahran, S., & Shukur, Z. (2019). A review on conformance checking technique for the evaluation of process mining algorithms. TEM Journal, 8(4), 1232.

[17] Bezerra, F., Wainer, J., & van der Aalst, W. M. (2009). Anomaly detection using process mining. In Enterprise, business-process and information systems modeling (pp. 149-161). Springer, Berlin, Heidelberg.

[18] Măruşter, L., Weijters, A. T., Van Der Aalst, W. M., & Van Den Bosch, A. (2006). A rule-based approach for process discovery: Dealing with noise and imbalance in process logs. Data mining and knowledge discovery, 13(1), 67-87.

[19] Accorsi, R., & Stocker, T. (2012, March). On the exploitation of process mining for security audits: the conformance checking case. In Proceedings of the 27th Annual ACM Symposium on Applied Computing (pp. 1709-1716).

[20] Cook, J. E., & Wolf, A. L. (1996). Process discovery and validation through event-data analysis (Doctoral dissertation, University of Colorado).

[21] de Medeiros, A. K. A., Weijters, A. J., & van der Aalst, W. M. (2007). Genetic process mining: an experimental evaluation. Data Mining and Knowledge Discovery, 14(2), 245-304.

[22] Camargo, M., Dumas, M., & Gonzalez-Rojas, O. (2020). Discovering Generative Models from Event Logs: Data-driven Simulation vs Deep Learning. arXiv preprint arXiv:2009.03567.

[23] Leemans, S. J., Fahland, D., & van der Aalst, W. M. (2013, June). Discovering block-structured process models from event logs-a constructive approach. In International conference on applications and theory of Petri nets and concurrency (pp. 311-329). Springer, Berlin, Heidelberg.

[24] Van Dongen, B. F., & Van der Aalst, W. M. (2004, November). Multi-phase process mining: Building instance graphs. In International Conference on Conceptual Modeling (pp. 362-376). Springer, Berlin, Heidelberg.

[25] Berti, A. (2019). Process Mining on Event Graphs: a Framework to Extensively Support Projects. In BPM (PhD/Demos) (pp. 60-65).

[26] Pinter, S. S., & Golani, M. (2004). Discovering workflow models from activities' lifespans. Computers in Industry, 53(3), 283-296.

[27] Agrawal, R., Gunopulos, D., & Leymann, F. (1998, March). Mining process models from workflow logs. In Int. Conference on Extending Database Tech. (pp. 467-483). Springer, Berlin, Heidelberg.

[28] Van der Aalst, W. M. (2013). Decomposing Petri nets for process mining: A generic approach. Distributed and Parallel Databases, 31(4), 471-507.

[29] Badakhshan, P., & Alibabaei, A. (2020). Using process mining for process analysis improvement in pre-hospital emergency. In ICT for an Inclusive World (pp. 567-580). Springer, Cham.

[30] Sani, M. F., Boltenhagen, M., & van der Aalst, W. (2020, September). Prototype Selection using Clustering and Conformance Metrics for Process Discovery. In Int. Conference on Business Process Management (pp. 281-294). Springer, Cham.