BEST: International Journal of Management Information
Technology and Engineering (BEST: IJMITE)
ISSN (P): 2348-0513, ISSN (E): 2454-471X
Vol. 7, Issue 7, Jul 2019, 1-6
© BEST Journals

INFREQUENT BEHAVIOUR FILTERING FROM BUSINESS PROCESS EVENT LOGS

Dr. K.Priyadharshini[1] & Dr. G.Gayathri[2]

[1]Department of Computer Science, Prist University, Thanjavur, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Science, Prist University, Thanjavur, Tamil Nadu, India

ABSTRACT

Among all recent technologies "Big Data" has the capability to deal or analyze with the bulk amount of both structured and unstructured data. Analysis of a particular type of data is called "Process mining" (i.e, the data resulted after the execution of specified business processes).Examine the resulted output from process mining had created a negative impact, due to the presence of outliers. Because of the presence of outliers in the resulted output, "noise" or "infrequent behavior" produced. The main objective of the process discovery is automatically extracting the process models from discovered data; which may lead a result of rarely traveled pathways that fill process models. The proposed idea to present an online-based automated technique which intern will remove infrequent behavior from business process event logs. The application which is proposed on recent process discovery algorithms will significantly reform the discovered process models to a better extend and it scales well to huge data-sets.

KEYWORDS: Process Mining, Infrequent Behaviour, Discovery, and Management Process

INTRODUCTION

The goal of process mining extracting action-able knowledge's from event logs of software application which is most commonly available in software organizations. Different type algorithms have been proposed addressing problem deriving process models from the business process event log. Algorithms provide various level tradeoffs between degree they accurately capture recorded behavior logs & deriving complexity process model.

Process discovery algorithm operates the assumption that business process log faithfully represents the behavior of the business process of the organization during a particular time period. Unfortunately, real life logs very often have more possible outliers. These kinds of outlier represent infrequent behavior, which very often referred to as "noise" and their presence due to data entry mistakes which may lead to data quality issues. The noise presented data lead to derived model exhibiting execution able paths infrequent in nature, which clutters the model or models that simply not a true representation of the original behavior. To limit these kinds of produced (–ve) effects, logs are typically subjected to data mining technique called pre-processing phase where data manually cleaned noise. However performing this operation is quite challenging and time taking tasks, which not even guarantee the effectiveness of produced result, especially in the context of large logs exhibiting complex behavior.

In the current paper, we are trying our level best to resolve the challenges which we are processing various process discovery model along very high noises. Online automation used for filtering infrequent behaviors systematically. The proposed automation filtering technique builds abstract process behavior which is recorded in the business process event log as an automaton. Automaton always captures direct follow dependencies between event labels from the business process event log. This automaton, infrequent transitions are subsequently removed which called as noise. Again original logs replayed reduced automaton identifying business process events no longer fit. Then business process event removed logs. The proposed technique aims to remove max number of infrequent transitions in the automaton, while minimizing a number of events which removed business process event log. Filtered business process event logs fit automaton perfectly.

We had implemented our proposed technique after "**ProM Framework**". We had evaluated the results extensively in combination with various baseline discovery algorithms using a 3-pronged approach. 1st, we injected artificial logs at various levels of noise and measured the accuracy of the proposed technique identifying infrequent behavior. 2nd, in varying levels of noise we had evaluated the process discovery accuracy and reduction of process model complexity. For an 'n' number of baseline process discovery algorithms, we had compared results obtained from two different baselines by automated filtering techniques. 3rd experiment is repeated using a variety of real-life logs by exhibiting different characteristics such as overall size and number of distinct events. The accuracies for process discovery measured in terms such as fitness, precision, etc. Results show that the proposed technique will lead to a statistically significant improvement of fitness, precision, and complexity.

BACKGROUND & RELATED WORK

Pre-Processing Technique

Pre-processing a data mining technique used for analyzing logs before starting the process. Typically, pre-processing includes log filtering technique. "ProM Framework3" offers a various plug-in for log filtering. Particularly two different plug-in deals with the removal of infrequent behavior. They are Simple Heuristics (SLF) plugin and Prefix-Closed Language (PCL) plugin. The Simple Heuristics plug-in removes trace which does not start or end with a specific event on the basis of frequency, e.g. removing traces which start with the infrequent event. The Prefix-Closed Language plug-in removes possible events from traces to obtain a log as output. If the obtained trace is a prefix of another trace in business process event logs which is based on the basis of user-defined frequency threshold. Various other log filtering plugins are available in ProM, which doesn't deal with the removal of infrequent behavior. In literature, Wang et al have addressed the filtering of noise from process event logs. Wang et al, the proposed approach uses a reference process model to repair business process event log whose events are affected by inconsistent labels. This approach requires the availability of the reference model.

Discovery Algorithms

Alpha algorithm automates process discovery model. Its defined P > Q, where 'P' & 'Q' 2 different tasks. There exists event 'P' directly follow the event 'Q'. Several noise-tolerant discovery algorithms proposed to overcome limitations of the current algorithm. They are Heuristics, Inductive, Fuzzy, ILP Miner, etc. **Heuristic** discovers a particular model using relationships. **Inductive** conquers approach which always results in sound process models. **Fuzzy** applies noise filtering posteriori directly on the discovered model. **ILP** follows a different approach to filter noise.

Outlier Detection

Advanced Outlier Detection Algorithm" has built-in data model such as statistics, probability model by normal behavior. These approaches classified into 3 major groups. 1st group deals problem of finding out entire sequences of the event in the outlier. 2nd group identifies single data points or sequences on the basis of data models of normal behavior from the log, example statistical model. Finally, the third group used to identify anomalous patterns within sequences. Eg: if subgraph made of 2 different events “A” and “B” considering outliers, these 2 events will be removed from each trace log they occur, regardless of their relative order or frequency. While this filtering mechanism may work with captured logs, removal of infrequent behavior would again coarse-grained.

Model Dimensions

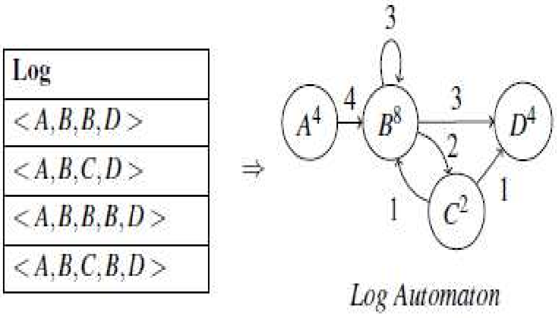
Dimensions that used to measure the quality of discovered models are fitness, precisions, generalization, and complexity. Fitness measure can reproduce business process behavior contained in logs. 0’s indicate in-ability to reproduce any behavior logs while 1’s indicate the ability to reproduce all behavior. The precision measure is the degree to which behavior made by a model is found logs. 0’s indicate model produces lots of behavior not observed in logs while 1’s indicate model allows behavior observed in logs. Generalization measure capability of model behaviour not observed in the log. We measure generalization using 10 fold cross-validation technique in data mining, which established approach. Finally, complexity quantifies structural process model.

DETECTION APPROACH

Detection of infrequent behavior depends on the identification of anomalies by Log automaton technique. Anomalies represent relationship which occurs infrequently. An automaton nothing but directed graph states represents task which can occur in business process event log under consideration of each arc connecting 2 states indicates direct follow dependency between respective tasks.

Infrequent Behavior Detection

Figure 1; generates log automaton from a single business process event log. Each log converted into states, along ‘A’ initial & ‘D’ final states. Moreover, 2 states connect with arc if these 2 events follow each other. Finally, annotation showing frequency added to each state and arc.



Log
$<A,B,B,D>$
$<A,B,C,D>$
$<A,B,B,B,D>$
$<A,B,C,B,D>$

Figure 1: Example of Infrequent Behavior Detection

Infrequent Behavior Removal

The infrequent behavior in a business process event log cause events recorded in the wrong order at an incorrect point of time. Such behavior cause derivations of direct follow dependencies which do not hold or cause direct follow dependencies. Hence, removal of infrequent behavior is to focus on incorrect recorded events.

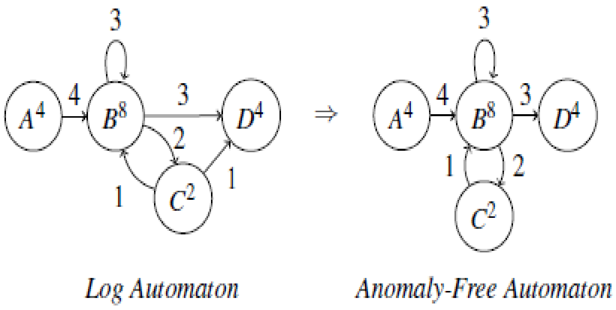


Figure 2: Example of Anomaly-Free Automaton

RESULTS AND DISCUSSION

The study purpose is that business process event logs with high levels of infrequent behavior pose contradictions high level of infrequent behavior corresponds to frequent behavior. Our proposed technique identifies infrequent behavior with accuracy 90 percentages from a total number of logs which is shown in Figure 3.

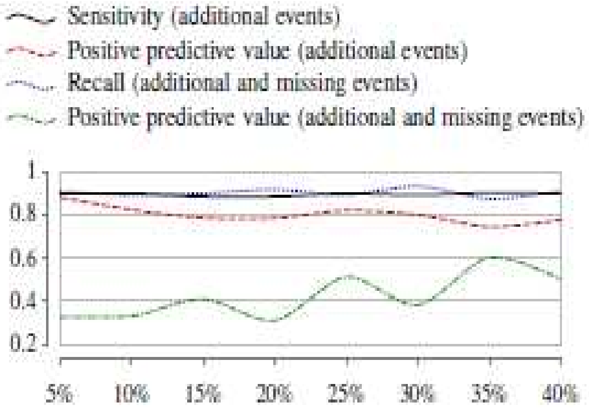
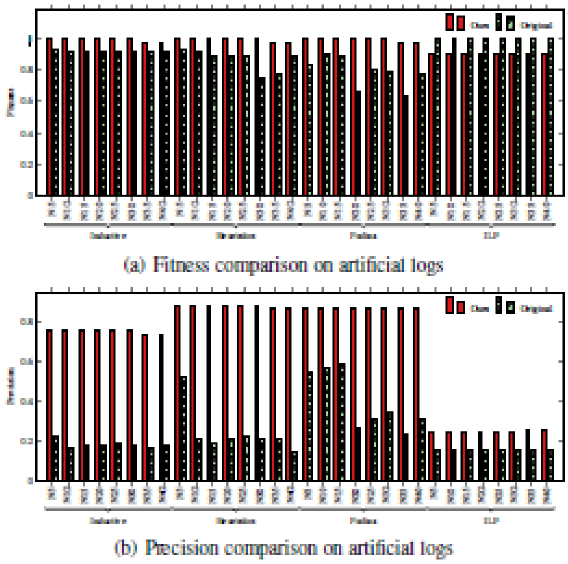


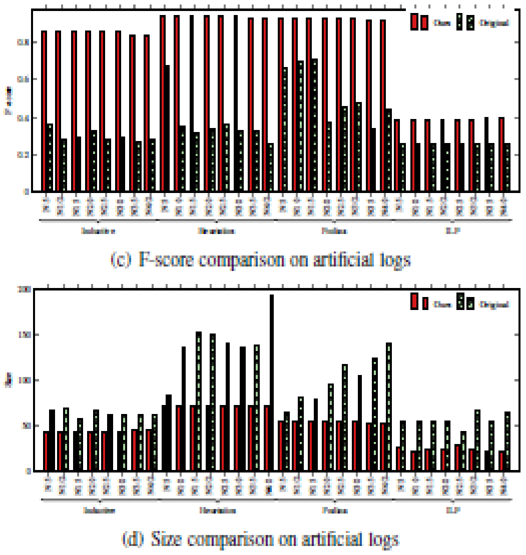
Figure 3: Comparision of Sensitivity, Positive Predictive Value, Recall, Etc.



(a) Fitness comparison on artificial logs

(b) Precision comparison on artificial logs

Figure 4: Shows Result Obtained From Baseline Discovery Algorithm



(c) F-score comparison on artificial logs

(d) Size comparison on artificial logs

Figure 5: Comparison of Different Artificial Logs

CONCLUSIONS

We had implemented advanced systematic filtering technique segregating in-frequent behavior. In-frequent dependency use between event labels as proxies for in-frequent behaviors. The detected dependencies from above process remove automaton which built on top of the event log. The obtained result shows improvement over fitness, precision, and complexity without negative effect on generalization.

REFERENCES

1. J. Wang, S. Song, X. Lin, X. Zhu, and J. Pei. Cleaning structured event logs: A graph repair approach. In Proc. of ICDE, pages 30–41, 2015.
2. J.M.E.M. van der Werf, B.F. van Dongen, C.A.J. Hurkens, and A. Serebrenik. Process discovery using integer linear programming. Fundam. Inform, 94(3-4):387–412, 2009.
3. S. Suriadi, R. Andrews, A.H.M. ter Hofstede, and M. Wynn. Event log imperfection patterns for process mining - towards a systematic approach to cleaning event logs. Information Systems, July 2016.
4. W.M.P. van der Aalst, A. Adriansyah, and B.F. van Dongen. Replaying history on process models for conformance checking and performance analysis. Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery, 2(2):182–192, 2012.
5. T. Lane and C.E. Brodley. Sequence matching and learning in anomaly detection for computer security. In Proc of AI, 1997.
6. M. Gupta, A. Mallya, S. Roy, J.H.D. Cho, and J. Han. Local Learning for Mining Outlier Subgraphs from Network Datasets, pages 73–81. 2014.
7. V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection for discrete sequences: A survey. IEEE TKDE, 24(5):823–839, May 2012.

8. A. Adriansyah, B.F. van Dongen, and W.M.P. van der Aalst. Conformance checking using cost-based fitness analysis. In Proc. of EDOC, pages 55–64, 2011.

9. A. Adriansyah. Aligning Observed and Modeled Behaviour. PhD thesis, Technische Universiteit Eindhoven, 2014.

10. A. Adriansyah, B.F. van Dongen, and W.M.P. van der Aalst. Conformance checking using cost-based fitness analysis. In Proc. of EDOC, pages 55–64, 2011.