Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Process Models Enhancement with Trace Clustering

Wiem HACHICHA[a,b*], Ronan CHAMPAGNAT[b], Leila GHORBEL[a] & Corinne Amel ZAYANI[a]
[a] *MIRACL Laboratory, Sfax University, Tunisia*
[b] *L3i Laboratory, La Rochelle University, France*
*wiem.hachicha1@univ-lr.fr

Abstract: Learning Management Systems collect data (such as event logs) about learners and trainers. There are techniques for analysing this data such as Educational Process Mining. In our previous work, we proposed an approach that extracts knowledge about learning paths based on process mining algorithms by generating process models. The latter are used for learning resource recommendation by taking into account learning features (learning style, interests, learning results, etc.). This approach is available for a limited size of event logs. In fact, process models generated from event logs of large classes are not expressive. Trace clustering is one of the successful methods that lead to overcome this limitation. For this reason, we aim to improve the previous approach by using trace clustering in order to characterise learners before the discovery of the corresponding process models. We applied the proposed approach on a Moodle dataset of 100 undergraduate students. Results show that trace clustering improve the general quality of discovered process models.

Keywords: Moodle, event logs, process mining, process discovery, trace clustering, recommendation

1. Introduction

Process mining aims to discover, monitor, and improve real processes by extracting knowledge from event logs readily available in current information systems (Van Der Aalst et al., 2012). One of the uses of process mining is process discovery. A discovery algorithm takes an event log and produces a model without using any prior information (Van Der Aalst et al., 2012). If process mining techniques are applied to educational data, they are often referred to as Educational Process Mining.
In order to evaluate the results of process discovery algorithms, three main quality dimensions (metrics) are invested (Buijs et al., 2012): 1) Fitness (F) which determines how well the model allows the behaviour present in the event log, 2) Precision (P) that corresponds to the rate of activities in the event logs compared to the total of activities observed in the process model and 3) Generalization (G) measures the ability of the model to generalize the behaviour present in the event log.

In our previous work (Hachicha et al., 2021), we proposed an architecture leading to discover process models and recommend to the learner not only learning resources but also process models, each of which is relative to a specific learning resource. We applied process mining techniques from event logs, learning results, and learner's profiles in order to make useful recommendations to learners that would enable them to build up their own learning scenario. This approach was validated on a Moodle-based course of 100 undergraduate students. Results are encouraging but may be difficult to scale to large classes such as MOOCs. Indeed, process models generated from event logs of large classes are not expressive.

In order to overcome this limitation various works in literature used trace clustering (Song et al., 2009), (Xu & Liu, 2019) and (Faizan et al., 2021). The trace represents a sequence of events of the same process instance (in our case the learners) (Van der Aalst, 2016). An event contains information such as the identifier of each learner, the activity name, the timestamp, etc. Indeed, trace clustering identifies homogeneous sets of traces in a heterogeneous event log and allows the discovery of many, simpler process models. For example, authors in (Faizan et al., 2021) have used the concept of trace profiles in order to cluster traces. They have employed the activity profile where each trace is transformed into a vector of features that represents the frequency of activities.

We aim in this work to improve the previous architecture by applying the trace clustering. The research questions tackled in this study are: is it possible to use trace clustering for e-learning traces? and does it give an advantage compared to applying process mining directly?

2. Proposed Approach

In our previous work, we proposed an architecture that use process mining for learning resource recommendation (Hachicha et al., 2021). This architecture is composed of four layers which are source, client, recommendation, and process mining. The source layer involves distributed databases of event logs and process models. The event logs are files that contain a large amount of raw data about the interaction of learners with Learning Management Systems (LMS). The client layer allows the interaction between the learner and e-learning systems (LMS). The recommendation layer recommends not only learning resources but also process models generated from the process mining layer. The latter allows the discovery of process models based on event logs. Each process model demonstrates the most common usage behaviour by learners in an LMS.

In this work, we propose to extend the process mining layer by the step of trace clustering to enhance the quality of process models. This step aims at finding such homogenous groups of process instances. First, we filter the event log by instances to ensure that only learners' activity logs were kept in the event log. Second, we anonymize the learners. We converted the learners' names into IDs to maintain their anonymity and ensure the principles of ethics. Finally, we used the trace clustering method to split the event log into several homogeneous sub-event logs, which are used to generate the corresponding process models. We created the activity profile which is used to split the event log. The activity profile is an NxM matrix; N represents the number of process instances and M represents the number of activities. Each row in this matrix corresponds to a vector trace which is composed of activity frequencies (the number of occurrences of each activity in the trace).

3. Experimental results

The case study uses an event log extracted from the Moodle platform of the University of La Rochelle (France). The event log includes 42,438 traces of 100 students that learned a course over one semester.

We implemented the different steps of the process mining layer using the PyCharm tool. In the trace clustering method, we created the activity profile. Based on this activity profile, k-means algorithm is employed to split the event log. Actually, we conclude that the optimal number of clusters for k-means clustering in our case study is 5 using the elbow method.

At last, we evaluated process mining algorithm resting on the result of trace clustering. In fact, we used the heuristic miner process discovery algorithm to obtain the process models representing the event log's behaviour. Subsequently, the Fitness (F), Precision (P) and Generalization (G) of discovered models were measured (cf. **Error! Reference source not found.**). These metrics are foregrounded in section **Error! Reference source not found.**.

To compare our new results with the results obtained during our previous work (Hachicha et al., 2021), we have calculated the average of each metric (cf. **Error! Reference source not found.**).

Table 1. *Evaluation metrics obtained on process models mined from the used data set*

	F	P	G
Cluster0	0.9922	0.1951	0.7718
Cluster1	0.9847	0.2218	0.8235
Cluster2	0.9652	0.1624	0.6668
Cluster3	0.9838	0.2884	0.7663
Cluster4	0.9923	0.1486	0.7938

Table 2. *Comparison of results*

	F	P	G
Previous approach	0.9903	0.1597	0.7609
New approach	0.9836	0.2023	0.7644

As plotted in **Error! Reference source not found.**, the heuristic miner used in the new approach proves better performance in the precision value (0.2023) and generalization value (0.7644) than the heuristic miner used in the previous approach.

4. Conclusion

In this work, we improve our proposed architecture by extending the process mining layer by trace clustering step. This step is based on the activity profile. We experimented the proposed architecture on a dataset extracted from the Moodle LMS. We derive 5 clusters corresponding to 5 variant of learning scenarios. Therefore, we compare the quality metrics of extracted process models with our previous approach. The study shows that the results have been enhanced even on a, relatively, small dataset. The results validate the possibility of using trace clustering on data coming from e-Learning traces.

As future work, we would enhance the activity profile. We plan also to conduct experiments on larger e-learning data such as MOOC datasets.

Acknowledgements

This work was financially supported by the PHC Utique program of the French Ministry of Foreign Affairs and Ministry of Higher Education and Research and the Tunisian Ministry of Higher Education and Scientific Research in the CMCU project number 22G1403.

References

Buijs, J. C. A. M., Van Dongen, B. F., & Van Der Aalst, W. M. P. (2012). On the role of fitness, precision, generalization and simplicity in process discovery. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7565 LNCS*(PART 1), 305–322. https://doi.org/10.1007/978-3-642-33606-5_19

Faizan, M., Zuhairi, M. F., & Ismail, S. (2021). Process discovery enhancement with trace clustering and profiling. *Annals of Emerging Technologies in Computing*, *5*(4), 1–13. https://doi.org/10.33166/AETIC.2021.04.001

Hachicha, W., Ghorbel, L., Champagnat, R., Zayani, C. A., & Amous, I. (2021). Using process mining for learning resource recommendation: A Moodle case study. *Procedia Computer Science*, *192*, 853–862. https://doi.org/10.1016/j.procs.2021.08.088

Song, M., Günther, C. W., & Van Der Aalst, W. M. P. (2009). Trace clustering in process mining. *Lecture Notes in Business Information Processing*, *17 LNBIP*, 109–120. https://doi.org/10.1007/978-3-642-00328-8_11

Van der Aalst, W. (2016). Process mining: Data science in action. *Process Mining: Data Science in Action*, 1–467. https://doi.org/10.1007/978-3-662-49851-4

Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., Bose, J. C., Van Den Brand, P., Brandtjen, R., Buijs, J., Burattin, A., Carmona, J., Castellanos, M., Claes, J., Cook, J., Costantini, N., Curbera, F., Damiani, E., De Leoni, M., … Wynn, M. (2012). Process mining manifesto. *Lecture Notes in Business Information Processing*, *99 LNBIP*(PART 1), 169–194. https://doi.org/10.1007/978-3-642-28108-2_19

Xu, J., & Liu, J. (2019). A Profile Clustering Based Event Logs Repairing Approach for Process Mining. *IEEE Access*, *7*, 17872–17881. https://doi.org/10.1109/ACCESS.2019.2894905