

Pattern Based Sequence Classification

Ms. Reshma Hasan Atar¹ and Dr. D. S. Bhosale²

Student, ME (Computer Science Engg.)¹

Associate Professor, Computer Science Engg.²

Ashokrao Mane Group of Institutions, Vathar Tarf Vadgaon, Kolhapur, India

Abstract: *The text highlights the challenges of analyzing massive amounts of data, particularly in the context of business processes using process mining. Outliers or irregular behavior in the data can negatively impact processing and clutter process models, leading to less useful paths. The objective is to automatically extract process models from the data, and an automated method for removing irregular behavior from event logs is introduced. This method significantly improves the quality of identified process models and scales well to large datasets. Since the effectiveness of filtering strategies depends on the event log, users can interactively filter activities and directly view the filtered process model from the event log using a slider-based approach. Ultimately, the choice of included activities is left to the user. The method is tested using actual occurrence log collections from enterprise process oversight and hospital environments. The results demonstrate that the newly proposed activity filtering approaches yield process models that are more behaviorally specific compared to conventional frequency-based filtering methods.*

Keywords: Bigdata, Process mining, Outliers, Process model, Event logs, Filtering strategies, Frequency-based filtering

REFERENCES

- [1]. B. Liu, W. Hsu, and Y. Ma, "Integrating classification and association rule mining," in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 1998, pp. 80–86.
- [2]. Z. Xing, J. Pei, and E. Keogh, "A brief survey on sequence classification," ACM SIGKDD Explorations Newslett., vol. 12, no. 1, pp. 40–48, 2010.
- [3]. C. Zhou, B. Cule, and B. Goethals, "Itemset based sequence classification," in Machine Learning and Knowledge Discovery in Databases. New York, NY, USA: Springer, 2013, pp. 353–368.
- [4]. N. Lesh, M. J. Zaki, and M. Ogihara, "Scalable feature mining for sequential data," IEEE Intell. Syst., vol. 15, no. 2, pp. 48–56, Mar./Apr. 2000
- [5]. B. Cule, B. Goethals, and C. Robardet, "A new constraint for mining sets in sequences," in Proc. SIAM Int. Conf. Data Mining, 2009, pp. 317
- [6]. J. Wang and G. Karypis, "Harmony: Efficiently mining the best rules for classification," in Proc. SIAM Int. Conf. Data Mining, 2005, pp. 205–216.
- [7]. V. S. Tseng and C.-H. Lee, "Effective temporal data classification by integrating sequential pattern mining and probabilistic induction," Expert Syst. Appl., vol. 36, no. 5, pp. 9524–9532, 2009.
- [8]. W.M.P. van der Aalst. Process Mining - Data Science in Action, Volume 2nd Edition. Springer, 2016
- [9]. J. Wang, S. Song, X. Lin, X. Zhu, and J. Pei. Cleaning structured event logs: A graph repair approach. In Proc. of ICDE, pages 3041, 2015.
- [10]. W.M.P. van der Aalst, T. Weijters, and L. Maruster. Workflow mining: Discovering process models from event logs. IEEE TKDE, 16(9):11281142, 2004.
- [11]. 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 5564, 2011.
- [12]. R. Conforti, M. Dumas, L. Garca-Banuelos, and M. La Rosa. Beyond tasks and gateways: Discovering BPMN models with subprocesses, boundary events, and activity markers. In Proc. of BPM, pages 101117, 2014.

- [13]. S.J.J. Leemans, D. Fahland, and W.M.P. van der Aalst. Discovering block-structured process models from event logs containing infrequent behavior. In Proc. of BPM Work- shops, pages 6678, 2014.
- [14]. C.C. Aggarwal. Outlier Analysis. Springer, 2013.
- [15]. M. Gupta, J. Gao, C.C. Aggarwal, and J. Han. Outlier detection for temporal data: A survey. IEEE TKDE, 26(9):22502267, 2014.
- [16]. A. Adriansyah, J. Munoz-Gama, J. Carmona, B.F. van Dongen, and W.M.P. van der Aalst. Alignment based precision checking. In Proc. of BPM Workshops, pages 137149, 2012.
- [17]. R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proc. of IJCAI, pages 11371145, 1995.
- [18]. 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. Rev.: Data Mining and Knowledge Discovery, 2(2):182192, 2012.
- [19]. M. Gupta, A. Mallya, S. Roy, J.H.D. Cho, and J. Han. Local Learning for Mining Outlier Subgraphs from Network Datasets, pages 7381. 2014.
- [20]. Raaele Conforti, Marcello La Rosa, and Arthur H.M. terHofstede Filtering out Infre- quent Behavior from Business Process Event Logs. Transactions on Knowledge and Data Engineering IEEE Feb 2017.
- [21]. C.C. Aggarwal. Outlier Analysis. Springer, 2013.
- [22]. M. Gupta, J. Gao, C.C. Aggarwal, and J. Han. Outlier detection for temporal data: A survey. IEEE TKDE, 26(9):22502267, 2014.
- [23]. M. Gupta, A. Mallya, S. Roy, J.H.D. Cho, and J. Han. Local Learning for Mining Outlier Subgraphs from Network Datasets pages 7381. 2014.
- [24]. W.M.P. van der Aalst. Process Mining - Data Science in Action, volume 2nd Edition. Springer, 2016.
- [25]. Tax, Niek, Natalia Sidorova, Wil MP van der Aalst, and ReinderHaakma. "Heuristic ap- proaches for generating local process models through log projections." In Compu- tational Intelligence (SSCI), 2016 IEEE Symposium Series on, pp. 1-8. IEEE, 2016.
- [26]. Werner, Michael. "Financial process mining-Accounting data structure dependent control flow inference." International Journal of Accounting Information Systems 25 (2017): 57- 80.
- [27]. AlShathry, Omar. "Process Mining as a Business Process Discovery Technique." Com- puter Engineering Information Technology 2016 (2016).
- [28]. Xia, Xiaoxu, Wei Song, Fangfei Chen, Xuansong Li, and Pengcheng Zhang. "Ea: a proM plugin for recovering event logs." In Proceedings of the 8th Asia-Pacific Sym- posium on Internetware, pp. 108-111. ACM, 2016