ICAI 2023
12th International Conference on Applied Informatics
Eger, Hungary, 2023

Application of Neural Network Tools in Process Mining

László Kovács, Erika Baksáné Varga, Péter Mileff, Dávid Polonkai

University of Miskolc, Institute of Informatics
laszlo.kovacs@uni-miskolc.hu

Abstract

Process mining is a technique to discover, analyze, and monitor processes in an objective manner. It uses log data from corporate information systems and turns them into insights and actions. In this sense process mining is a subfield of data science: it requires the availability of data and aims at improving processes [1]. The induction of complex schema or grammar models is a key challenge in knowledge engineering. The input for schema mining is a set of sequences (traces) and the engine determines the general schema graph covering the input set. The input for schema mining is given with an event log, which is a list of event traces. The traces of the event log are generalized, merged into a schema graph. The first software systems for schema induction usually used a pattern matching approach to determine the common section in the sentences. In most cases, the induction was based on a set of transformation rules defined by human experts. Current toolsets for process mining are dominated by the graph or automaton-oriented algorithms [2] usually starting with the construction of the Direct Follower Graph which is a graphical representation of a process. Traditional frequent pattern mining techniques run into their limits when dealing with massive datasets [6]. Therefore, more advanced algorithms apply tree-based representation [3]. These incremental methods derive relevant patterns recursively. One of the most widely used industrial approaches is the inductive mining algorithm [4], which uses a top-down discovery algorithm. The top-down method recursively decomposes the event log into smaller event logs. The method first converts the event log into a corresponding DFG (direct follow graph), then it simplifies the initial graph into a compact schema graph. Beside the graph-based standard approaches, we can find some recent approaches on applica-

tion of neural networks in process mining. We can consider graph schema induction as a special type of classification problem, where the final output category corresponds to a complete schema graph. The schema graph is usually constructed in an incremental way by selecting the next winner event or events and the edges in the graph denote the adjacency relation. Considering simple event sequences, the dominating approach for next event prediction is to use Recurrent Neural Networks (RNN), where the output signal of a processing step is used as input component for the prediction of the next element in the sequence. In [5], the RNN network model was applied for the process discovery task. Current solutions can detect XOR branches in the trace log, but these methods still miss some crucial functionality required by industrial problems. The main goal of the presentation is to show our proposals on the adaptation of neural networks in process mining to cover messing functionalities like AND branches and loop detection. First, we give a neural network architecture for detection of AND (parallel) branches. With the help of this approach, the engine can determine the parallel processes in the traces, and it builds up a process schema graph related to more actors. The proposed engine uses a two-level representation approach where the bottom level corresponds to the single actor level activity chain. These homogeneous segments are merged by using synchronization nodes. The top level contains the synchronization graph of the agent level segments. The second neural network architecture presents a novel loop detection approach which can be used to discover tandem repetition sections. The proposed method first converts the sequence into an image matrix format and this matrix will be sent to a CNN convolutional network. The category labels correspond to the different loop kernel positions in the sequence. The next model integrates these two neural network models to have a more general schema induction engine for process mining. The presented novel network models were implemented in Python Tensorflow/Keras framework. In order to test the induction engines on event logs of different complexity levels, a test set generator application was developed in the research project. The target schema is constructed with a visual editor, and the engine generates the related event log of given size. In the test experiments, we compared the efficiency of our proposed models with some standard graph based engines. In the paper, we analyze the test results showing both the benefits and limitations of the neural network approach.

References

[1] W. M. P. van der Aalst: *Process Mining: Data Science in Action*, Berlin, Heidelberg: Springer, 2016, doi: https://doi.org/10.1007/978-3-662-49851-4.

[2] J. Liu, S. Yan, Y. Wang, J. Ren: *Incremental mining algorithm of sequential patterns based on sequence tree*, in: Advances in Intelligent Systems, Netherlands: Springer, 2016, pp. 61–67.

[3] X. Liu, L. Zheng, W. Zhang, J. Zhou, S. Cao, S. Yu: *An evolutive frequent pattern tree-based incremental knowledge discovery algorithm*, ACM Transactions on Management Information Systems 13.3 (2022), pp. 1–20.

[4] Y. Lu, Q. Chen, S. Poon: *A novel approach to discover switch behaviors in process mining*, in: Proceedings of International Conference on Process Mining, 2021, pp. 57–58.

[5] T. Shunin, N. Zubkova, S. Shershakov: *Neural approach to the discovery problem in process mining*, in: Proceedings of International Conference on Analysis of Images, Social Networks and Texts, 2018, pp. 261–273.

[6] T. Truong-Chi, P. Fournier-Viger: *High-Utility Pattern Mining: Theory, Algorithms and Applications*, in: A Survey of High Utility Sequential Pattern Mining, Netherlands: Springer, 2019, pp. 97–129.