Next Activity Prediction and Elapsed Time Prediction on Process Dataset

Vincenzo Dentamaro [1], Donato Impedovo [1] , Giuseppe Pirlo [1] and Gianfranco Semeraro [1-2]

[1] *Department of Computer Science, University of Studies of Bari, Via Edoardo Orabona, Bari, 4, Italy*
[2] *University School for Advanced Studies IUSS Pavia, Piazza della Vittoria, Pavia, 15, Italy*

Abstract

Process mining is a field of research that has gained much attention in recent years because of its ability to analyze and improve processes. Indeed, one of the key aspects of process mining is its ability to predict the activities in the future and the time spent on these activities. In this work is proposed the use of Bidirectional LSTM and Multi-Speed Transformer on a recent dataset called BPIC-2020 related to reimbursement process of the University of Technology of Eidenhoven. Results shows that Multi-Speed Transformer is more capable to performs next activity prediction than the Bi-LSTM. Meanwhile, for the elapsed time prediction, the viceversa is true.

Keywords

Bi-LSTM, Multi-Speed Transformer, Next-Activity Prediction, Elapsed Time Prediction, University Processes, Public Administration

1. Introduction

Process Mining is a research domain that gains a lot of interest thanks to the application of Data Mining and Machine Learning methods to the processes [1]–[3]. The application is performed on processes recorded as timeseries data called “event log”, accordingly to [4]. In this way several analysis can be performed, starting from an initial discovery study [4] (building new models of the processes recorded) to a conformance and modeling check to analyze the evolving situation of processes and, thus, align models, and related software, to business processes [5].

The analysis of timeseries is a problem broadly manage using several types of models. Starting from the using of classical LSTM [2], [3] to the use of the Convolutional Neural Networks [1], [6].

The importance of the Process Mining techniques is also highlighted by the Public Administration (PA) that starts to use such technology in order to build models that helps to enhance the quality of the PA’s processes [7]. The application of Process Mining starts from the process modelling, using several techniques as design processes based on collected data [8], or analysing processes in order to improve them [9].

The PA involved in the use of Process Mining techniques is not only related to governance[7]–[9], but also related to health [10], [11] and educational system [12], [13].

In this work, it is proposed the use of an neural network architecture depicted in Figure 1. The architecture is based on the use of “block” meant to be the same neural network architecture used firstly to analysis the timeseries and then using a block to predict the next activity and the other one to predict the elapsed time to the next activity. This model was presented in [2] using a Bi-LSTM. The work is structure presented related works in Section 2. Then the dataset and the used method are explained in section 3. The experimental set-ups to perform experiments are explained in Section 4. Results and their

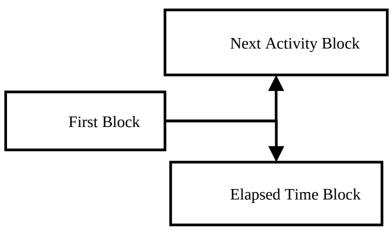

Figure 1. Architecture of the model. Despite the different name of each block, they are intended to be equal

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29-31, 2023, Pisa, Italy
EMAIL: vincenzo.dentamaro@uniba.it (A. 1); donato.impedovo@uniba.it (A. 2); giuseppe.pirlo@uniba.it (A. 3); gianfranco.semeraro@iusspavia.it (A. 4)

discussion are presented in Section 5. Finally in Section 6 there are the conclusions.

2. Related Works

The analysis of the event logs can be performed for several reason. In this work, it was proposed the use of the architecture presented by Gunnarsson B et al in [2]. The main architecture is depicted in Figure 1. It is noticeable that such architecture is based on the use of three blocks that represents the same underlying neural networks. In this way, a first application performs a sort of features extraction from the timeseries, meanwhile the other two application are specialized to predict the next activity and the elapsed time to the next activity. As in the majority of works using timeseries (called also event log), in [2] it was proposed the use of a Bidirectional LSTM (Bi-LSTM) in order to capture temporal patterns looking forward and backward.

Dentamaro V. et al, propose a new architecture called Multi-Speed Transformer [6]. Such architecture is based on a firstly application of a multi branch analysis in order to analysis with different level of details the data. In this way, analogously to the use of microscopy, it possible to identify fine patterns and more gross patterns.

The Multi-Speed Transformer using information improved the State of the Art in terms of prediction metrics.

Hence, in this work is proposed the use of both Bi-LSTM and Multi-Speed Transformer to compare both models on a recent dataset.

3. Materials and Methods

In this section the used methods to perform experiments about next activity and elapsed time prediction are explained along with the used dataset.

3.1 Materials

In this work, it is used a dataset related to the activity performed by University of Technology of Eindhoven (TU/e). In particular, the dataset that comes from the Business Process Intelligence Challenge 2020 (BPIC-2020) [14].

The dataset BPIC-2020 provides real-life event log of reimbursement requests from the staff of the University of Eindhoven (TU/e). The data were collected from the 2017 to the 2018.

Furthermore, the collected data are organized in several event log. Specifically, the event logs are:

- "Domestic Declarations" that is related to domestic travel (within the same country). The event log related to "Domestic Declaration" contains 56437 events recorded and the recorded events are about 10500 cases;
- "Request for Payement" that contains cases that could be not related to travels. The event log related to "Request for Payment" contains 36796 events for 6886 cases;
- "International Declarations" is related to travel outside the country. The event log contains 72151 events and the recorded events are about 6449 cases;
- "Travel Permit Data" is related to the permission to travel. This event log is composed by 86581 events for 7065 cases;
- "Prepaid Travel Costs" contains data related to the travel costs prepaided. This events log is composed 18246 events and the recorded events are related to 2099 cases.

Hence, each event log is composed with several events. Within each event log is possible to identify and separate cases in order to obtain, for each case, a separate event log, called "trace".
The "trace" is used to preprocess data in order to apply methods explained in the next section.
A "trace" T is composed by a sequence of events e_i with $1 < i < N_T$ where N_T is the number of events for the "trace" T. The event e is composed as a record of information as: "case id" c, "activity label" a, "timestamp" t, "attributes j-th" d_j where $1 < j < M$ and M is the number of attributes. The attributes can contains information related to the case (this information are shared from all the activity and they don't change long the trace) and to the activity (this information are specific to the recorded activity).

3.2 Methods

In this subsection, the used methods to perform prediction about the next activity and the elapsed time are explained.

The used methods are based on the use of a new State of Art neural network called "Multi-Speed Transformer" [6] and a well-known

Bidirection LSTM (referred also Bi-LSTM). This two neural networks were used as block in a more complex architecture in order to perform both next activity and elapsed time prediction.

The main architecture is presented in ***Figure 1***.

4. Experimental Set-Up

The architecture takes as input a preprocessed trace as in Table 1. Hence, the sequence of prefixes is feed to the first block. In this way a first analysis is performed. The output of the first block is then passed to both "next activity block" and "elapsed time block". Such blocks are equals to the first block, because they use the same structure. But the "next activity block" gives as ouput a distribution of probability to predict the next activity, meanwhile the "elapsed time block" gives the predicted elapsed time to the next activity.

4.1 Bi-LSTM

Bidirectional LSTM is a neural network architecture based on the processing of timeseries. In particular, the core part is the "LSTM" that analyse data in order to find temporal patterns.

Commonly, such analysis is performed in forward setup, i.e. from the past to the present.

In Bidirectional LSTM, two layer of LSTM are used in order to analyse data in forward and backward way. In this way, both forward and backward temporal patterns are learnt by the model.

4.2 Multi-Speed Transformer

The Multi-Speed Transformer was presented by Dentamaro V. et al [6]. This model is based on the concept of fine and gross analysis of data. Such analysis are done at different levels of detail as analysis performed by microscopies at different resolutions. In this way, different kind of information are extracted and then concatenated to obtain the prediction.

5. Experimental Set-Ups

In this section the setups used to perform experiments are explained.

5.1 Data Preprocessing

In order to use the dataset BPIC2020, each trace, in each event log, is firstly sorted accordingly with the timestamps of each event, then it is preprocessed extracting information as "prefix", "suffix" and "elapsed time". In the following, this information are defined:

- "Prefix": Prefix is defined as a function that given a trace T, a position k and a windows size w, it return a sequence of events from $k - w$ to k:
  $$Prefix(T,k,w) = < e_{\{k-w\}}, \dots, e_{\{k\}} >$$
- "Suffix": Suffix is also defined as a function that given a trace T, a position k and a windows size w, it return a sequence of events from $k + 1$ to $k + w$:
  $$Suffix(T,k,w) = < e_{\{k+1\}}, \dots, e_{\{k+w\}} >$$
- "Elapsed Time": Elapsed time is the time remaining to the next activity. Hence it is defined as a function that given a trace T, a position k it returns the difference between the timestamp of e_{k+1} and e_k:
  $$Elapsed(T,k) = e_{k+1}.t - e_k.t$$
  The "." is intended as the operator that recall the timestamp t values of the event. The elapsed time is computed in seconds.

For both "prefix" and "suffix", if the windows size excides the number of events to be selected zero-padding is applied, e.g. for the prefix, if the given position is lower than the windows size it means that "k-w" is negative, hence to overcome this problem zeros are added.

Finally, in the Table 1 it is shown an example of the results obtained from the preprocessing. It is supposed that the given trace is composed by 5 events: $T = < e_1, e_2, e_3, e_4, e_5 >$, and the windows size is $w = 3$:

Table 1 Example of preprocessing of a trace. For the Suffix the windows size is 1 to obtain only the next activity to predict

t	w	Prefix(T,p,w)	Suffix(T,p,1)	Elapse Time(T,p)
1	3	$0, 0, e_1$	e_2	$e_2 - e_1$
2	3	$0, e_1, e_2$	e_3	$e_3 - e_2$
3	3	e_1, e_2, e_3	e_4	$e_4 - e_3$
4	3	e_2, e_3, e_4	e_5	$e_5 - e_4$

Each event log is also divided in train and test split following the set-up used in [2]. Specifically, the cases are ordered by the timestamp of the related recorded event. Then a 75/25 split is applied. After such division, from the train set are eliminated the cases which are not ended when the

first case of the test set started. Successively, the train set is further divided in train and validation set in 75/25 setup.
For each event log, the activity label are one hot encoded.

6. Results and Discussion

Once the experiments were performed, predictions were evaluated accordingly to the following metrics:

- Categorical Accuracy: The proportion of the correctly predicted “next activity label” on the total number of predicted “next activity label”.
- Root Mean Squared Error: The root applied at the sum of the squared differences between predicted “elapsed time” and real “elapsed time”.

The results of the performed experiments are reported in Table 2:

Table 2. Experimental Results Performed with both models on the BPIC-2020

	Domestic Declaration	International Declarations	Travel Permit	Prepaid Travel Cost	Request For Payment
			Bi-LSTM		
Categorical Accuracy	0,4241	0,2466	0,1918	0,1218	0,6107
Root Mean Squared Error	**480136,44**	**1184055,00**	**1621978,13**	**802258,56**	**490469,88**
			Multi-Speed Transformer		
Categorical Accuracy	**0,8969**	**0,7106**	**0,7515**	**0,7288**	**0,8394**
Root Mean Squared Error	480288,97	1184189,25	1622560,63	802295,56	490572,91

From the results showed in ***Table 2*** it is possible to notice the in terms of accuracy to predict the next activity, the proposed use of the Multi-Speed Transformer is better than the proposed use of the Bidirectional LSTM (Bi-LSTM). Indeed, the best performance of the Bi-LSTM is 10% lower than the lower performance using the Multi-Speed Transformer.

Related to the elapsed time, it is possible to notice that the Root Mean Square Error (RMSE) is lower for the models using the Bi-LSTM. In this sense, the model using the Bi-LSTM seems to be more suitable to predict the elapsed time to the next activity.
A more deeply analysis of the results highlights that the model using Bi-LSTM is truly better than the model using Multi-Speed Transformer to predict elapsed time to the next-activity, but the RMSE in all the models and for all the used event log are high.

7. Conclusion

In this work, the use of Bi-LSTM is compared with the use of Multi-Speed Transformer as block of a more complex architecture to build systems capable of both prediction of the next activity and elapsed time to the next activity.

The results shows that multi-speed Transformer is suitable to predict next activity given a sequence of activity. Meanwhile, the prediction of the elapsed time seems to be a task more suitable for the model using the Bi-LSTM.

Future works could include more information about the case or the activity in order to improve the quality of predictions.

8. Acknowledgements

This Word template was created by Aleksandr Ometov, TAU, Finland. The template is made available under a Creative Commons License Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).

9. References

[1] E. Obodoekwe, X. Fang, and K. Lu, “Convolutional Neural Networks in Process Mining and Data Analytics for Prediction Accuracy,” *Electronics (Switzerland)*, vol. 11, no. 14, Jul. 2022, doi: 10.3390/ELECTRONICS11142128.

[2] B. R. Gunnarsson, S. vanden Broucke, and J. De Weerdt, “A Direct Data Aware LSTM Neural Network Architecture for Complete Remaining Trace and Runtime Prediction,” *IEEE Trans Serv Comput*, 2023, doi: 10.1109/TSC.2023.3245726.

[3] N. Tax, I. Verenich, M. La Rosa, and M. Dumas, “Predictive Business Process Monitoring with LSTM Neural Networks,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture*

Notes in Bioinformatics), vol. 10253 LNCS, pp. 477–492, Dec. 2016, doi: 10.1007/978-3-319-59536-8_30.

[4] W. Van der Aalst, "Process mining: Data science in action," *Process Mining: Data Science in Action*, pp. 1–467, Jan. 2016, doi: 10.1007/978-3-662-49851-4/COVER.

[5] L. Aversano, T. Bodhuin, and M. Tortorella, "Assessment and impact analysis for aligning business processes and software systems," *Proceedings of the ACM Symposium on Applied Computing*, vol. 2, pp. 1338–1343, 2005, doi: 10.1145/1066677.1066978.

[6] M. Cheriet, V. Dentamaro, M. Hamdan, D. Impedovo, and G. Pirlo, "Multi-speed transformer network for neurodegenerative disease assessment and activity recognition," *Comput Methods Programs Biomed*, vol. 230, p. 107344, Mar. 2023, doi: 10.1016/J.CMPB.2023.107344.

[7] C. Forliano, P. De Bernardi, A. Bertello, and V. Temperini, "Innovating business processes in public administrations: towards a systemic approach," *Business Process Management Journal*, vol. 26, no. 5, pp. 1203–1224, Oct. 2020, doi: 10.1108/BPMJ-12-2019-0498.

[8] P. J. Kiss and G. Klimkó, "A Reverse Data-Centric Process Design Methodology for Public Administration Processes," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11709 LNCS, pp. 85–99, 2019, doi: 10.1007/978-3-030-27523-5_7.

[9] G. Polancica, B. Šumaka, and M. Pušnik, "A Case-based Analysis of Process Modeling for Public Administration System Design," *Frontiers in Artificial Intelligence and Applications*, vol. 321, pp. 92–104, Dec. 2019, doi: 10.3233/FAIA200009.

[10] J. Munoz-Gama *et al.*, "Process mining for healthcare: Characteristics and challenges," *J Biomed Inform*, vol. 127, p. 103994, Mar. 2022, doi: 10.1016/J.JBI.2022.103994.

[11] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," *J Biomed Inform*, vol. 61, pp. 224–236, Jun. 2016, doi: 10.1016/J.JBI.2016.04.007.

[12] M. A. Ghazal, O. Ibrahim, and M. A. Salama, "Educational process mining: A systematic literature review," *Proceedings - 2017 European Conference on Electrical Engineering and Computer Science, EECS 2017*, pp. 198–203, Jul. 2018, doi: 10.1109/EECS.2017.45.

[13] A. Bogarín, R. Cerezo, and C. Romero, "A survey on educational process mining," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 8, no. 1, p. e1230, Jan. 2018, doi: 10.1002/WIDM.1230.

[14] "BPI Challenge – Process Mining Conference 2020." https://icpmconference.org/2020/bpi-challenge/ (accessed Mar. 18, 2023).