Attention Based LSTM with Multi Tasks Learning for Predictive Process Monitoring

Thuzar Hnin [1+] and Khine Khine Oo [2+]

[1,2] University of Computer Studies, Yangon, Myanmar

Abstract. Today, in the artificial intelligence research field, Deep Learning (DL) is one of the fastest-growing techniques because of the power of learning features, that gives a higher level of abstraction of the raw attributes; and related research in Recurrent Neural Networks (RNN) and Long Short-Term Memory Networks (LSTM) have shown exemplary results in neural machine translation, neural image caption generation, NLP and so on. For our research, in order to detect potential problems and to facilitate proactive management, we focus on predictive process monitoring (PPM) as domain area, by predicting business behaviour from historical event logs. Recent research works, LSTM networks have gained attention in PPM and have been proved that they can highly improve prediction accuracy in PPM. According to the literature, we have learned that PPM resembles to an early sequence classification problem in NLP. And, recent trends in DL based NLP, attention mechanism is mostly embedded in neural networks. Inspired by these results, this paper proposes to firstly use Attention Based LSTM with Multi Tasks Learning for PPM.

Keywords: LSTM, Attention, Predictive Process Monitoring

1. Introduction

Predictive (business) process monitoring is a valuable techniques that apply to early detect potential problems during process execution before they occur so that these problems can be handled proactively. In certain, predictive process monitoring exploits event logs, which are more and more widespread in modern information systems, to allow users to predict how current (uncompleted) cases will unfold up to their completion. Certainly, the event logs generated during process execution can serve as a valuable source to predict of business process with various purposes such as estimation of technical production parameters for more robust production plan optimization, analyzing the behavioral patterns of customers, risk management by predicting compliance violations, resource allocation and etc.

Therefore, currently, there is a number of approaches data mining, machine learning and statistical techniques used for a variety of business process prediction tasks: next activity prediction ,control flow prediction, remaining cycle time prediction, deadline violations prediction, cost prediction and also including risks prediction, ranging from the domain of health, logistic, . Besides, most prior work is based on an explicit process model, e.g. mined from event logs, and augmented with probability tables and execution time information. After introduced J. Evermann [5] of DL to apply in this field, an implicit process model (neural networks) is based on prediction. The paper [5] pointed out that predictive process monitoring is conforming to the NLP because of the sequential nature of process traces. As a result, the getting motivation from the successful achievement of DL to NLP is a natural fit to the problem of process prediction. In the paper [9], the authors have done systematic literature review (SLR) for PPM in details. The authors performed a systematic literature review of outcome-oriented predictive process monitoring methods and have provided a taxonomy of existing methods. And they also performed a comparative experimental evaluation of eleven identified techniques using a unified experimental set-up and 24 predictive monitoring tasks constructed from nine real-life event logs. Likewise, the author from [9] has highlighted that LSTM neural networks had been applied in PPM for predicting remaining time a next activity of a running case of a business process. And the author have planned to study how LSTMs can be used for outcome prediction.

[+] Corresponding author.
 E-mail address: thuzarhninn99@gmail.com; khinekhineoo@ucsy.edu.mm

After that, the author has published another paper [8] to define a notion of temporal stability for predictive process monitoring and evaluates existing methods with respect to both temporal stability and accuracy. The experiments have been done on 12 prediction tasks formulated on 6 real-life publicly available datasets. And the authors have found that the highest temporal stability is achieved by a single classier approach with XGBoost, followed by LSTM.

Inspired by these, we have also done briefly literature review only focusing on DL applied in PPM. To our knowledge, this is first work on using recur-rent neural networks (RNNs) for outcome-oriented predictive process monitoring though, RNNs with long short term memory units (LSTMs) have been used for predicting remaining time and next activity prediction [13,14]. Thus, this paper aims to explore another potentials of LSTM using Attention mechanism by adopting multi-tasks learning for outcome oriented predictive process monitoring.

In this paper, background of predictive process monitoring problem is discussed in Section 2. Section 3 is discussed about related work and motivation for the proposed method. In the section 4, we briefly explain our proposed method and conclusion is made in the final section.

2. Background on Predictive Process Monitoring

2.1. Predictive Process Monitoring

Typically, predictive process monitoring is concerned with anticipating the future behaviors from execution (business) process instances. The author, M. Dumas defined in [11] that "a business process is a collection of inter-related events, activates, and decision points that involve a number of actors and objects, which collectively lead to an outcome that is of value to a customer". Mostly, It is represented as a workflow, e.g. with the standard business process model and notation (BPMN). Here, the figure 1 is the example of business process described by BPMN.
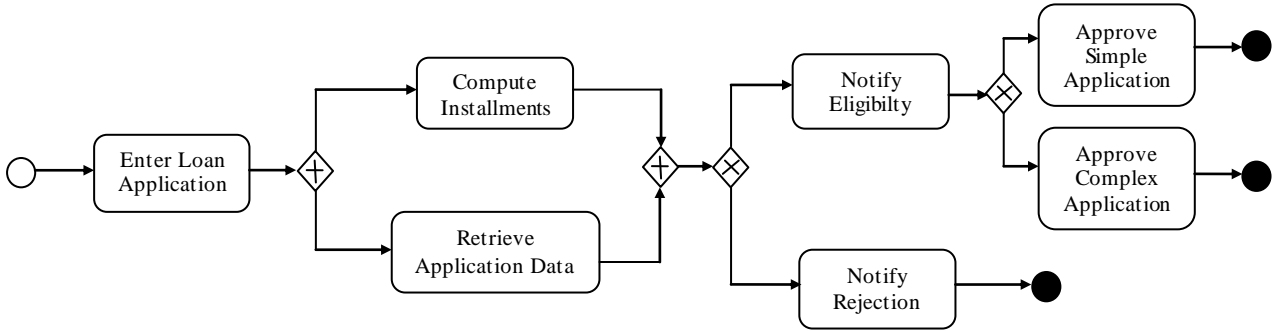

Fig. 1: Example of Business Process

Table. 1: Example of Event Log

CID	Activity (Task)	Timestamp
13219	Enter Loan Application	2007-11-09 T 11:20:10
13219	Retrieve Applicant Data	2007-11-09 T 11:22:15
13220	Enter Loan Application	2007-11-09 T 11:22:40
13220	Compute Instalments	2007-11-09 T 11:24:35
…..	…..	……

The identical business process can generate a different sequence of activities every time it is executed. A specific executive of an activity is referred to as event; a specific execution of the entire process, which corresponds to a path in the workflow model, is referred to as process instance. In formal, the event records of process instance are initial point of predictive process monitoring. Formally, an event can be defined as a tuple: $e=(a,c,t,(d_1,v_1),…(d_m,v_m))$ where a is the activity name, c is the case id, t is the timestamp and $(d_1,v_1)…,(d_m,v_m)$ (where $m \geq 0$) are the event or case attributes and their values. The Table 1 and shows the example of event record. In paper [7], a general PPM workflow is derived from studying all the methods. It includes two main phases: *offline (training or learning phase) and online (testing) phase*, which are shown in figure 2 and 3 respectively.
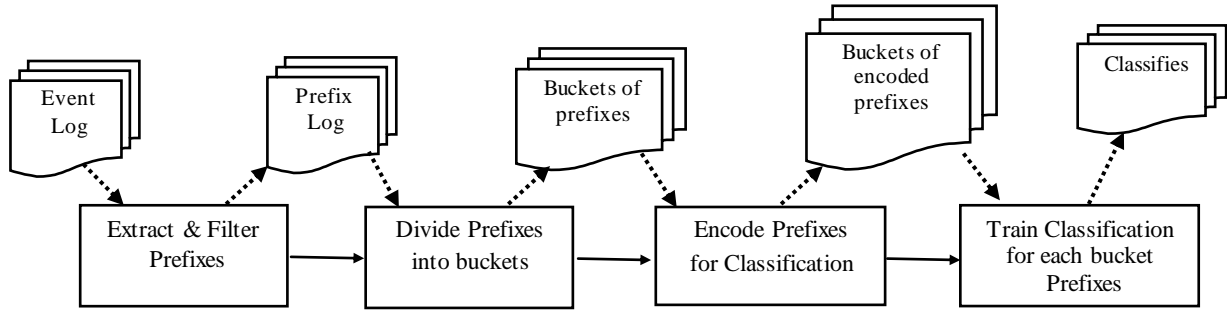
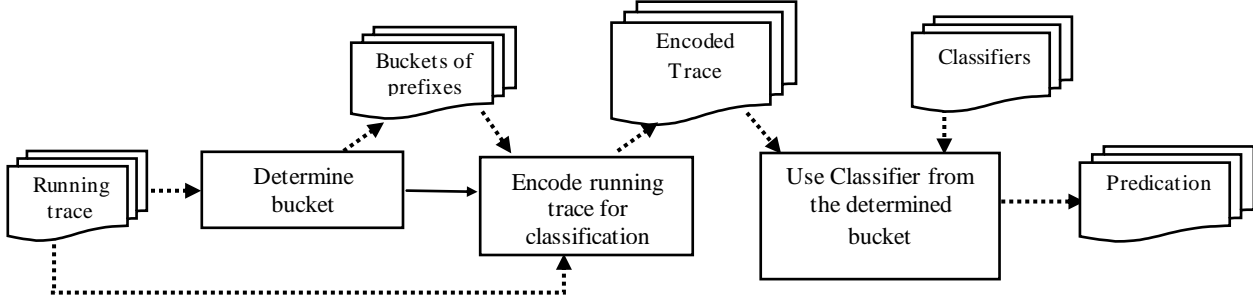Fig. 2: Offline Phase in Predictive Process Monitoring Workflow

Fig. 3: Online Phase in Predictive Process Monitoring Workflow

Based on the discussion from the paper [7,9], we have learned that the researchers should not only pay careful attention and their contributions in the two main steps: trace bucketing and sequence encoding of PPM framework. Additionally, we have also known that the careful usage of specific prefix filtering and classification algorithm is very important to obtain good predictions, with their performance being influenced by the particular settings used. Actually, an event or a case has its own attributes such as : Numeric (e.g. Patient's age), Categorical (e.g. patient's gender) or Textual data type (e.g. patient's medical history). While some existing methods have tried to handle categorical and numeric attributes, some researchers have used text mining techniques in order to handle text attributes. Likewise, according to the paper [4], the researchers need to focus on the prediction target (results) from different perspectives. As the type of prediction, Numeric Prediction (Time Predictions, Cost Predictions), Categorical Prediction (Categorical Outcome, Risk Prediction), Activity Sequence Prediction are categorized.

Among them, though time is importantly needed to predict while business process is executing, we plan to focus on the outcome(normal or deviate) oriented prediction for business process behaviors by taking into account process metrics, especially cost prediction.

3. Related Work

In recent years, deep learning has only successfully started to apply in predicting process behaviour problem because of historical success in other areas. Further improvements are achieved by using more advanced models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). The main problem of these networks concerns the learning algorithm. The most adopted one is back-propagation through time, and shows the well-known vanishing gradient and exploding gradient problems. These problems make the learning of long-time dependencies very difficult for (vanilla) RNNs. While some values of the parameters able to deal with such sequences exist (i.e. a human could carefully pick the model parameters in order to solve problems with long-time dependencies), in practice learning algorithms seem not to be able to find them.

3.1. LSTM Networks

To deal with the problem of difficulty of RNNs to represent long-term information, LSTM network is introduced by having some internal contextual state cells that act as long-term or short-term memory cells. The output of the LSTM network is modulated by the state of these cells. This is a very important property

when we need the prediction of the neural network to depend on the historical context of inputs, rather than only on the very last input. The LSTM transition functions are as follows:

$$i_t = \sigma(x_t \, U^i + h_{t-1} \, W^i), \ f_t = \sigma(x_t \, U^f + h_{t-1} \, W^f)$$

$$o_t = \sigma(x_t \, U^o + h_{t-1} W^o), \ \tilde{C}_t = \tanh(x_t \, U^g + h_{t-1} W^g)$$

$$C_t = \sigma(f_t * C_{t-1} + i_t * \tilde{C}_t), \ h_t = \tanh(C_t) * o_t$$

are the input gate, forget gate, output gate and the memory cell respectively, and refers to element-wise multiplication. As is shown in the equations, the input gate decides how much information from the new input will be added to the memory cell. Similarly, the forget gate **f** controls how much information to forget from the previous states, and the output gate limits the amount of information to expose. By balancing the incoming and outgoing information amount, LSTM is able to prevent the gradient vanishment and explosion problems. There are a lot of variants of LSTM networks on structure, namely BLSTM and Multilayer LSTM, LSTM with peephole connections, Gated Recurrent Unit (GRU), Phased LSTM and so on. They have developed as an effective and scalable model for several learning problems related to sequential data.

In order to predict business process 's behaviors, long short-term memory (LSTM) neural networks have been employed in the literature for deriving accurate and robust representations of prediction[13, 14]. Tax, Verenich [13], Evermann [6] and Navarin,Polato [12,13] are three state-of-the-art models utilizing LSTMs for predicting of remaining time and next activity. In this paper, we aims to study LSTM networks by taking inspiration from these papers.

3.2. Multitask Learning (MTL)

R. Caruana [16] has defined that multitask learning is an approach to inductive transfer that improves generalization by using the domain information contained in the training signals of related tasks as an inductive bias. It does this by learning tasks in parallel while using a shared representation; what is learned for each task can help other tasks be learned better. Usually, it is incorporated with Deep Learning to improve robustness, where the network learns to classify the encoded features and performs feature enhancement at the same time. It can be considered as a form of inductive transfer that is typically done with either hard or soft parameter sharing of hidden layers [17]. In [14], we have studied that predicting using Multitask learning (MTL) for next activity can yields higher accuracy than others.

4. Proposed Method: Attention Based LSTM with Multi-tasks Learning

Though the ordinary LSTM can remember the past information for future prediction, it is limited to only a few latest steps, with more impact from later ones, and may not be able to discover major influences from earlier timestamps. It encodes the input sequence to a fixed length internal representation. This imposes limits on the length of input sequences that can be reasonably learned and results in worse performance for very long input sequences. To address these problems, attention is need recent trend in DL.

Likewise, for PPM, the authors [14] have also identified a limitation of LSTM models when dealing with traces with multiple occurrences of the same activity, in which case the model predicts overly long sequences of the same event. Therefore, to address that problem, we propose the first attention-model based architecture for predictive process monitoring as our main contribution.

4.1. Attention

Certainly, attention is simply a vector, often the outputs of dense layer using softmax function. By assigning attention weights on neural networks, AB-LSTM has achieved great success in various areas such as machine learning tasks, electronic health records recommendation systems, QA systems, image caption generation , stock price prediction and so on. To our knowledge, there is no existing work of AB-LSTM for outcome oriented prediction of business process. Because of these motivations, here we propose to put attention mechanism in LSTM Neural Networks model for predictive process model, as shown in the following figure 4.
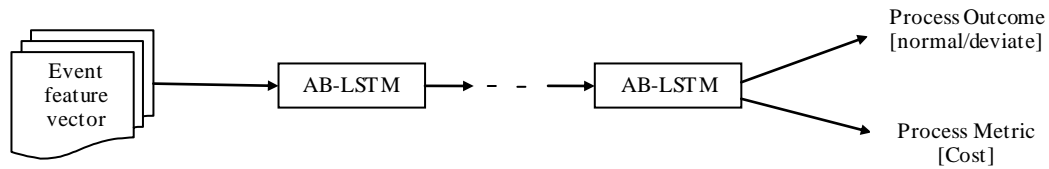
Fig. 4: Proposed Attention Based LSTM with Multitasks Learning for PPM

Due to the mentioned success of MTL and reason in section 3.2, multitask learning is also proposed to use to build an attention based LSTM model for predicting not only the outcome of running case of the business process but also cost as an extra care attribute.

4.2. Experimental Analysis

For the next step of our research, we intend to do experimental analysis of our proposed AB-LSMT model by using nine real-life benchmark datasets, out to which eight are public datasets, assessable from the 4TU Center for Research Data and one is a private dataset, event logs of Claim Handling Process at Australian Insurance Company. To make comparative study, "accuracy, earliness of prediction and time performance are used in predictive process monitoring process.

5. Conclusion

To sum up, in this paper, we propose to explore another potentials of LSTM using Attention mechanism by adopting multi-tasks learning for outcome oriented predictive process monitoring. From the good support of systematic literature review of papers [9], we believe that ABLSTM with MTL can improve predictive accuracy by doing experimental analysis on the benchmark datasets. Thus, our future research work is to do a comparative experimental evaluation of our proposed method with other prior methods.

6. Reference

[1]   A. Metzger, P. Leitner , D. Ivanovic , E. Schmieders , R. Franklin , M. Carro ,S. Dustdar , K. Pohl, " Comparing and combining predictive business process monitoring techniques"

[2]   B. Schwegmann, M. Matzner, and C.Janiesch, "A Method and Tool for Predictive Event-Driven Process Analytics", Wirtschaftsinformatik Proceedings 2013, Association for Information Systems R. Dewri, and N. Chakraborti. Simulating recrystallization through cellular automata and genetic algorithms. *Modelling Simul. Mater. Sci. Eng.* 2005, **13** (3): 173-183.

[3]   C. Koshy,"A literature Review on Predictive Monitoring of Business Processes, Master Thesis

[4]   C. Di Francescomarino, C. Ghidini , F. M. Maggi and F. Milani, "Predictive Process Monitoring Method: Which One Suits Me Best? "

[5]   J. Evermann, J.-R. Rehse, and P. Fettke.: "A deep learning approach for predicting process behaviour at runtime". In: Proceedings of the 1st International Workshop on Runtime Analysis of Process-Aware Information Systems. Springer (2016)

[6]   J. Evermann, J.-R. Rehse, and P. Fettke. "A Deep Learning Approach for Predicting Process Behaviour at Runtime", pages 327–338. Springer International Publishing, Cham, 2017

[7]   I Verenich, M Dumas, M La Rosa, FM Maggi, C Di Francescomarino ," A general framework for predictive business process monitoring", International Conference on Advanced Information Systems Engineering, 186-202, 2016. 5, 2016.

[8]   I. Teinemaa, M. Dumas, A. Leontjeva , F. M.Maggi , "Temporal Stability in Predictive Process Monitoring"

[9]   I. Teinemaa, M. Dumas, M. L. Rosa, and F. M.Maggi, "Outcome-Oriented Predictive Process Monitoring: Review and Benchmark"

[10] M.-Chamorro AE, Resinas M, Ruiz-Cortes A (2017) Predictive monitoring of business processes: a survey. IEEE Transactions on Services Computing

[11] M. Dumas, MarcelloLaRosa, JanMendling, andHajoA. Reijers. 2018. Fundamentals of Business ProcessManagement (2nd ed.). Springer.

[12] M. Polato, A. Sperduti, A. Burattin, and M. de Leoni. "Data-aware remaining time prediction of business process instances." In 2014International Joint Conference on Neural Networks (IJCNN), pages 816–823. IEEE, jul 2014

[13] N. Navarin, B, Vincenzi, M. Polato and A. Sperduti, "LSTM Networks for Data-Aware Remaining Time Prediction of Business Process Instances ".

[14] N. Tax, I. Verenich, M. L. Rosa, and M. Dumas, "Predictive business process monitoring with LSTM neural networks," in International Conference on Advanced Information Systems Engineering (CAiSE). Springer, 2017, pp. 477–492.

[15] Q. Suo , F. Ma , , G. Canino, , J. Gao, A. Zhang , A. Gnasso, , "An Attention-based Recurrent Neural Networks Framework for Health Data Analysis "

[16] R. Caruana , Multitask learning: "A knowledge-based source of inductive bias. In Proceedings of the Tenth International Conference on Machine Learning" , 1993

[17] S. Ruder , "An Overview of Multi-Task Learning in Deep Neural Networks"