Name: Océane SOULIER

Name of your Level 1: Morjane SAIDANI

Paper title: Evaluating Predictive Business Process Monitoring Approaches on Small Event Logs

Source: Google scholars

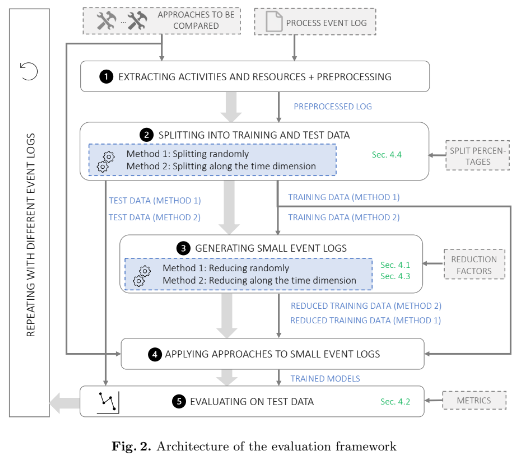
Keywords specific to the paper: business process monitoring, deep learning, predictive business process monitoring, small sample learning, process prediction.

Summary of the main contributions (use text paragraphs, tables and if necessary, figures):

This paper proposes an evaluation framework for comparing existing predictive process monitoring approaches regarding their suitability for small data sets. A lot of situations are unable to provide sufficient qualitative and quantitative data sets (ex. business process management, small and medium companies without sufficient data records) and, according to the authors, this issue is largely neglected. They do not reconsider the techniques used by Machine Learning (ML) but they do state that ML can not be applied to each and every situation. In some situations, small data can also result from a wish of having real-time feedback. Hence, situations in which small data are used are various so there is a need for evaluation of the quality of the process with state-of-the-art approaches. It raises a simple question: does less data also mean less quality in the results ?

Producing small event logs is complicated due to the lack of generally accepted definition of “small event”. Therefore, authors use different reduction factors that enable them to keep the reference log available but they also underline that the way data is removed affects the final results. Either they remove the most recent data (reduction through time dimension), or they remove specific trace variants (if there is 1 variant among 1000 logs it does not have much impact but if there is still this one variant among 100 logs, it is tremendously different). They will be able to compare their results (from a small log) and the results obtained with a reference log and they will know whether or not there is a quality loss.

Data reduction comes with several risks: the executive unit can get lost, the reduction of the data can also mean statistical bias (the issue of trace variant raised above). In case of a random reduction, events can be extremely distorted.



Results show approaches can achieve good performance even with significantly reduced training data. Reduction of performance appears only when it comes to 95% data reduction. Sometimes, a reduced amount of data even achieves better results (not in the case of data reduced randomly).

However, more data is needed to learn rare cases and complex logs, such as prediction of the next roles. They only perform well in predicting frequent activities. Since ML always try to generalize the data in the simplest way, this is not a surprise.

To put it in a nutshell, a reduced amount of data is sufficient when it comes to less complex log events and frequent trace variants. When it comes to learning rare variants, an increased amount of data is necessary. When it comes to complex logs, the current amount of data is not sufficient.

Rare trace variants could be an issue investigated in further works.

AI model used (e.g. Neural network, etc.):

This paper is all about small sample learning (SSL), an area in artificial intelligence that deals with machine learning in order to tackle insufficiency of data sets. It is composed of two main branches: concept learning (attempts to detect new concepts while only using a small number of examples) and experience learning (attempts to solve SSL problems using ML techniques).

Generative Adversarial Networks (GANs) are quoted because of their strong structure and efficiency but authors used conventional (not small) event logs for the tests, so results can not be relied on.

The paper mentions that the most frequently used deep learning architectures for next activity prediction are Long-short-term-memory (LSTM) and Convolutional neural networks (CNN). Gated Recurrent Units (GRUs) are also discussed. These models are used for predictive business process monitoring.

Supported by a software application? (If yes, provide more details):

Python 3.7

Rrun on a system equipped with a Windows 10 operating system, an Intel Core i9-9900K CPU3.60GHz, 64GB RAM, and a NVIDIA Quadro RTX 4000 having 6GB of memory.