Nom L0 : PACCIARELLI Erina
Nom L1 : SAIDANI Morjane
Titre: Predictive Monitoring of Business Processes: A Survey

The most representative approaches to the prediction of business process execution are summarized in this article.

Predictive Business Process Monitoring (BPM) is one of the sub-domains of Process Mining. Its aim is to provide timely data to facilitate proactive and corrective actions aimed at improving process performance and reducing risk.
It is defined as the set of runtime methods for creating predictive models that anticipate a specific value of a current process, based on its current path and past events. Input data includes event logs defining the characteristics required for prediction, as well as process models or contextual elements such as Petri nets. The output is a predicted value for each current process instance, or a series of such values, which can be Boolean (the achievement of a specific goal), categorical or numerical (the remaining time of a process), depending on the purpose of the prediction. By exploiting the exchange of events between information systems, it is possible to predict performance indicators, assess risks or anticipate breaches of service level agreements (SLA).
The prediction is made at a certain point in the run, called the control point. This affects other factors, such as evaluation and the creation of predictive models. Encoding is paramount and usually involves a feature engineering task which is always specific to the concrete process and hampers the initial stages of the predictive monitoring process.
During the learning phase, input data includes event logs, possibly a process model and external information.
They are then encoded into feature vectors that can be interpreted by the predictive algorithm.
The predictive model is then evaluated and applied to running instances during execution to predict output. Event logs, providing information such as instance ID, event and timestamp, are essential and usually provided by systems recording process traces.
Additional information can also be included in the log, such as the name of the resource executing the activity, or the cost of the activity.
Massive amounts of information can be generated by any of these systems and stored in event logs. Consequently, it is necessary to acquire the process characteristics most relevant to data management from four different perspectives. Firstly, the control flow perspective, linked to the order of activities performed in the process. Secondly, the data flow perspective, involving the various attributes attached to events. Thirdly, the temporal perspective, which relates to different types of process duration, such as the duration of an activity or the time remaining in a process. Finally, the resource/organization perspective, which relates to the resource executing a given event.
Predictions from event logs can take a variety of forms, such as the next activity in a process, compliance with time logic constraints, the remaining time of a process instance, or the risk associated with a specific value. These predictions can focus on a single process instance, such as time remaining, or be an aggregation of multiple instances, such as the average cycle time of all processes completed this month.

In terms of the application domain for predictions, the work collected covers four distinct areas: performance indicators (to monitor and improve the process), risk predictions (to anticipate future actions), predictions of SLA violations (service contracts between supplier and customer) and other predicted values (such as abnormal events or prediction of the next event in a current process instance).
Various methods are explored, including regression based on machine learning and annotated transition systems (ATS), and statistical methods. These approaches aim to predict process performance using quality measures such as root mean square error (RMSE). Models based on non-Markovian Petri nets are used to predict the remaining execution time of a process. Machine

learning approaches such as Naive Bayes and SVR are used to predict the remaining time and the future sequence of activities. Statistical techniques are also used to predict events and their correlation with contextual elements. Finally, cost prediction methods, based on queuing theory and regression techniques, are also explored to predict delays in manufacturing processes.
Finally, it is difficult to determine the best methodology because of the variability of the data sets and algorithms used. There are challenges associated with model evaluation due to the lack of comprehensive comparisons and available software.