International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2019): 7.583

A VBC Approach for Remaining Time Prediction in Business

Vishnu P[1], Ajeesh S[2]

[1] Assistant Professor, Mount Zion College of Engineering.
vishnupankajakshan[at]hotmail.com

2Assistant Professor, Mount Zion College of Engineering.
ajeesh.s3[at]gmail.com

Abstract- ***Accurate predictions of the remaining time, defined as the required time for an instance process to finish, are critical in many systems for organizations being able to establish a priori requirements, for optimal management of resources or for improving the quality of the services organizations provide. Our approach consists of i) extracting and assessing a number of features on the business logs, that provide a structural characterization of the traces; ii) extending the well-known annotated transition system (ATS) model to include these features; iii) proposing a partitioning strategy for the lists of features associated to each state in the extended ATS; and iv) applying a linear regression technique to each partition for predicting the remaining time of new traces.***

Keywords: Business processes enhancement, predictive business process monitoring, business processes management, business intelligence

1. Introduction

Massive growth of business processes automation as well as increasing information technology adoption in business process management is producing a vast amount of process execution data that are stored in the form of event logs. By applying process mining techniques to these logs, real hidden processes can be discovered and/or existing processes can be monitored and improved. There are three main types of process mining techniques process discovery, conformance checking, and process enhancement. Process discovery takes an event log and produces a model without using a priori information Conformance checking makes a comparison between a designed process model and the process discovered from the event log, to show where the real process deviates from the designed one Process enhancement aims to extend or improve an existing process, using information related to the process which is usually extracted from the recorded events logs .The problem of predicting the remaining time is a part of a more general problem known as predictive monitoring. In the last years, several proposals focused on predictive monitoring and, more specifically, on the prediction of remaining time have been presented Initially, these proposals have focused on the representation of the process executions or traces under the hypothesis that traces with different characteristics have different remaining times. Several of these approaches are based on Annotated Transition Systems (ATS), where each (partial) trace is associated to a state Other approaches use a partial trace-based or index-based representation More recently, approaches have been proposed for applying machine learning methods for predicting the remaining time In all these approaches, the problem encoding includes information about the context of the process execution state, such as the duration of the activities, or about domain variables. The main problem with all these approaches is that their trace representation (or encoding) does not include all the relevant information related to the traces execution, such as repetition of activities, the distance between activities or co-occurrences. Without this information about the structural features of the traces, it is difficult to make accurate predictions about the remaining time.

2. A New Prediction Model for Remaining Time Estimation

In the previous section, we presented the most relevant elements of an ATS. Based on it, we propose and describe in detail in this section our extension of the ATS model that includes structural features which are extracted from the traces.

A. Trace Features

Firstly, we extend the ATS model by considering a number of features (or attributes) extracted from the analysis of the event log traces. Each of these features is related to a measurement with which the ATS model will be extended, that is, each state of the ATS model will be annotated with both a set of attributes and the remaining time. A key difference between our approach and others in the literature is that the attributes we consider provide specific structural information about traces, such as the occurrence of the activities, its elapsed time, or the existence of loops, among others. This structural information will act as predictor variables that will be taken into account by a regression model, aiming to improve the accuracy in the calculation of the remaining time prediction in a running process.

B. Extended Annotated Transition System

Previously, we introduced the definition of an Annotated Transition System (ATS) model, as well as the definitions of the attributes we consider to be extracted from each trace to build our model. In this section, we show how we integrate the previously introduced attributes into an ATS in order to define our Extended Annotated Transition System (EATS).

C. Estimation of the Remaining Time: Dataset Partitioning

Before describing the details of the regression model we use for estimating the remaining time, we should take into account the following considerations. Let us recall, in the first place, that according to Definition 15, each state S in the EATS is annotated with a list of elements (vectors) List(S) which include the values of the attributes for all the partial traces represented by S. Each of these lists contains all the data (predictors or independent variables) needed for performing the remaining time estimation. Therefore, we have a single dataset associated to each state. Applying a linear regression technique to each of these datasets has proved to produce poor estimations since, usually, traces in the dataset have great variability in terms of size, number of activities and execution times. This is the usual general scenario for real business process data, such as administrative procedures or applications, industrial incidents management or processes in a hospital or other big organisations/institutions, In many cases, the remaining time values range is vast (e.g. from a few seconds to hundred thousand seconds) even for traces that are very similar or even identical.

D. Linear Regression Model

Once the need for partitioning the dataset was established and our partitioning strategy was described we are now in conditions to describe our remaining time estimation model. Basically, it is made up of linear regression functions which are obtained for each of the partitions of the dataset described before. The independent variables of the regression functions are the values of attributes (elements) in each partition, being the dependent variable the remaining time

Validation and Experiments

In this section, we will describe the experiments we have conducted for validating our proposal. Ten Real-life event Algorithm 3 Time Prediction Model (TPM) of a new trace Input: PTNEW: new Partial Trace Output PRT : Predicted Remaining Time for PTNEW . 1: S ← S(PTNEW) F State which represents PTNEW 2: PL = {P1, . . . , Pn}:= Partitions List associated to S, as returned by Algorithm 2 3: RL = {R1, . . . , Rn}:= List of Linear Regression functions associated to PL, as indicated in Section IV-D 4: ENEW := Element associated to PTNEW F 5: distanceMin ← +∞ 6: for k = 1, . . . , n do F For all partitions in PL 7: for each PT ∈ Pk do 8: dist = X M m=1 |valueOfAttm(PT) – value of Attm (PTNEW)| 9: if dist < distanceMin then F New min 10: partitionIndex = {k}; 11: distance Min ← dist 12: else if dist == distanceMin then 13: partitionIndex = partitionIndex ∪ {k}; 14: end if 15: end for 16: end for 17: PRT ← Average of the estimations obtained with the regression models {RLk , k ∈ partitionIndex} applied to ENEW 18: return PRT logs were used for validation: BPIC12w , BPIC13 ,BPIC15 (5 logs), BPIC17 Hospital Billing and Road Traffic Fine Management Process Eight of the datasets are taken from the Business Process Intelligence Challenges (BPIC), a de-facto standard for testing and validation of business processes approaches, since these datasets are proposed for the yearly Business Process Intelligence Contest. We have performed three types of experiments with the following aims: i) to compare our method, for different threshold values, with the baseline ATS-based proposal in order to validate it in terms of accuracy and mean absolute error and also to have experimental evidence about the influence of the threshold values in the results; ii) to determine a single threshold value which could be labelled as the most appropriate choice to use for estimating the remaining time for new event logs, in terms of precision of the results and simplicity of the model; and iii) to perform a comparison between our approach and to the sixteen state of the art approaches described in in order to prove the validity of our approach and assess its quality, confronting it to ATS-based, non-ATS based and machine learning approaches.

Results Considering Different Threshold Values

In this section, we present the validation results of our approach for different threshold values of the partitioning strategy (TBP) described in Section IV-C. Our method is compared to the ATS baseline approach described in [6] for the ten datasets in Table 5. This validation aims to provide a general overview of the dependence of our remaining time estimation results with the TBP threshold values. We used two metrics to compare our results to the Mean Absolute Error (MAE) to measure the error between real remaining and the predicted remaining time and the Accuracy to assess the regression quality in objective terms, using RMSE as a metric should be avoided in this context, since it is very sensitive to outliers.

Threshold Choice

In this section, we discuss how to choose an appropriate threshold value which could be used, in general, for any new dataset. We will support the discussion with the experimental analysis and results we describe in what follows. In a first analysis, it seems straightforward that the best threshold choice should be the most precise one, i.e., that produces the highest accuracy or lowest MAE. In order to experimentally determine which is, in general, the most precise threshold, we need to consider the results in Tables 7 and 8, rank them, and calculate the average rank through all the datasets and thresholds, for both MAE and Accuracy. These results are summarised in Table 9. According to this, the most precise threshold (MPT) is the one with the lowest ranks, which is 0.975 for both Accuracy and MAE.

3. Conclusion

In this paper, we proposed a new approach for predicting the remaining time of the running process in business process management. Our approach consists of two perspectives: firstly, we define a number of attributes that are evaluated from the process traces and capture quantitative and structural information about them. Secondly, a linear regression model is used for remaining time prediction, using these attributes. The attributes are added to the well-known annotated transition system (ATS, [6]), thus producing a new Extended ATS which takes into account structural information of the traces. Furthermore, to deal with the trace variability in terms of size, the number of activities and execution times, a threshold based partitioning method of the dataset logs is proposed. For each of the partitions, a different linear regression model is obtained. The evaluation of our approach was made using ten real-life event logs, showing that our model outperforms the results in the state of the art [13], particularly the LSTM Deep Learning approach [18], in terms of Mean Absolute Error and Accuracy metrics. The scalability of our approach has been addressed by considering and validating an attribute selection method and a model selection method for obtaining a single threshold value that can be recommended as an appropriate choice for new estimation problems.

References

[1] A. Rogge-Solti and M. Weske, ‘‘Prediction of business process durations using non-Markovian stochastic Petri nets,’’ Inf. Syst., vol. 54, pp. 1–14, Dec. 2015. doi: 10.1016/j.is.2015.04.004.

[2] B. F. van Dongen, R. A. Crooy, and W. M. P. van der Aalst, ‘‘Cycle time prediction: When will this case finally be finished?’’ in Proc. Int. Conf. Move Meaningful Internet Syst. (OTM), vol. 5331, R. Meersman and Z. Tari, Eds. Berlin, Germany: Springer, 2008, pp. 319–336. doi: 10.1007/978-3-540-88871-0_22

[3] W. M. P. van der Aalst, Process Mining: Data Science in Action, 2nd ed. Berlin, Germany: Springer, 2016. doi: 10.1007/978-3-662-49851-4.

[4] S. Pandey, S. Nepal, and S. Chen, ‘‘A test-bed for the evaluation of business process prediction techniques,’’ in Proc. 7th Int. Conf. Collaborative Comput. (COLLABORATECOM), D. Georgakopoulos and J. B. D. Joshi, Eds., 2011, pp. 382–391. doi: 10.4108/icst.collaboratecom.2011.247129.