Summary, Application Example and Critique of the Paper: Predictive Process Monitoring Methods - Which one Suits Me Best?

Fabian Braun

Ludwig-Maximilians-Universität München

Masterseminar “Recent Developments in Data Science - Process Mining”
Wintersemester 2018/19

Abstract. The abstract should summarize the contents of the paper using at least 70 and at most 150 words. It will be set in 9-point font size and be inset 1.0 cm from the right and left margins. There will be two blank lines before and after the Abstract. . . .

Keywords: Keyword1, Keyword2, ...

1 Introduction

Compliance monitoring in businesses, the assessment whether the execution of intern processes is on par with policies, regulations and laws, is generally bound to a reactive approach. Violations are only identified after their occurrence, without the possibility of preventing them in the first place [1]. In Contrast to these reactive approaches, predictive process monitoring methods allow the user to continuously monitor currently ongoing processes. These Predictive approaches provide predictions about an expected outcome, based on the current state and the preliminary information of a process. With a large amount of literature describing these techniques becoming available, the paper by Di Francescomarino, Ghidini, Maggi and Milani aims to order and classify the different approaches currently developing in the field, by developing a value-driven framework to help researchers identifying upcoming challenges and companies finding suitable approaches for their requirements.

In this review, first a brief summary of the paper *”Predictive Monitoring Methods - Which one Suits Me Best?”* is given. In Section 3 the specified concepts are then further explained by an application example. At last Section 4 comprises a critique of the work, depicting three strong and weak points of the paper.

2 Summary

The papers main goal is the implementation of a framework that allows researches to get a quick overview of and identify upcoming challenges in the field of predictive process monitoring. To develop such a framework, a systematic literature review was carried out to identify the current state of the art in the field. This process was divided into two separate phases. In the first step the review protocol was defined, by formulating the underlying research question, identifying the collections of papers to be analysed, defining the inclusion and exclusion criteria and formulating the data extraction strategy [1].

In the second phase the final list of papers was produced by applying the protocol setup in the first phase. By applying a keyword based search on electronic libraries, including terms such as *"predictive", "business process"* or *"process mining"*, a primary collection of 780 papers was identified. These papers where then reviewed considering the inclusion and exclusion criteria, resulting 55 works chosen to be analysed and categorised, according to the underlying research question: *"How can the body of relevant academic publications within the field of predictive process monitoring be classified as a framework?"*. By further refining this research question, four main categories for the distinction of the approaches were defined [1]:

- "What aspects of the process are predicted?"
- "What input is required?"
- "What families of algorithms are used?"
- "What existing tools support the approach?"

The final 55 papers where then further analysed, by firstly extracting standard meta-data, like the author or year of publication. After that, according to the four main categories deducted from the research question, for each work the type of prediction, the required data, the family of algorithms used and tools that support the approach were extracted. Furthermore domains for which the approaches are applicable, such as *"automotive"* or *"financial"* were determined and used as an additional attribute. The prediction type was emphasised as the main dimension to classify predictive process monitoring techniques and further subdivided into three categories of prediction types [1].

Numeric Predictions are grouped into time and cost predictions. Time predictions are a versatile group, consisting of approaches that report information about elapsed, sojourn and remaining time for each state within the transition system. The collected information is applied to make predictions about the completion time of a process. Cost predictions on the other hand leverage a process model extended with costs, considering information about production, volume and time [1].

Categorial Predictions are further split into risk and outcome predictions. Former are outcome-oriented, including approaches that minimize process risks by applying process models. These approaches utilise decision trees generated from the logs of past executions, which are then processed, resulting in probabilities returned to the user. The other group are categorial outcome predictions, which consider the fulfilment of predicates. In contrast to outcome predictions, most approaches evaluated in the paper of NAME, NAME und NAME do not apply any explicit model. Instead, frameworks that leverage the sequence of events and data payload of the last activity are applied to make predictions [1].

Next Activities Predictions try to predict a sequence of future activities and their corresponding data payloads, by studying the previously observed activities [1].

The results of the paper are condensed into the tables 1 and 2. Read from left to right, the tables distinguish approaches utilizing the four criteria deduced from the literature review, guiding readers to one of the 55 reviewed approaches.

Pred. type	Det. Pred. type	Input 1	Input 2	Input 3	Tool	Domain	Family of algorithms 1	Family of algorithms 2	Family of algorithms 3	Refer.
time	maint. time	event log (with timestamps)			N	automotive	time series	probabilistic model		[58]
	activity delays					financial telecomm.	queueing theory	transition system		[61,62]
	rem. time					telecomm.	stat. analysis			[6]
					ProM plugin	public admin.	transition system			[3,2]
						customer supp.	stochastic Petri net			[55]
			process model			financial customer supp.	regression	classification		[69]
		event log (with timestamps) with data			N	unspecified	pattern mining			[10]
					Y but unavail.	unspecified	classification	time series		[9]
					Y	financial public admin.	regression	classification		[68]
						healthcare	stochastic Petri net			[60]
					ProM plugin	public admin.	regression			[20]
						public admin.	transition system	regression	classification	[53]
						customer supp. public admin. financial	transition system	regression		[54]
						financial logistics	stochastic Petri net			[56,57]
			inter-case metrics		Y	healthcare manufacturing	regression			[59]
		event log (with timestamps) with data and contextual information			Y but unavail.	logistics	clustering	regression		[31]
					Y	logistics	clustering	pattern mining		[5]
								transition system		[30]
					ProM plugin	logistics	clustering	transition system		[29]
			labeling funct.	proc. model	ProM plugin	no validation	classification			[38]
categorical outcome	outcome	act. durations and routing probab.	threshold(s)	proc. model	N	synthetic	simulation	stat. analysis		[63]
		event log	labeling funct.		Y implem.	financial automotive	prob. automata			[7]
		event log (with timestamps) with data			N	logistics	neural network	constraint-sat.	QoS aggregation	[44]
			labeling funct.			healthcare	probab. automata	classification		[39]
						logistics	classification			[8]
						synthetic	classification			[34]
					Y but unavail.	synthetic	probab. automata			[37]
						synthetic	classification	neural network		[42]
						healthcare	clustering	classification		[18]
					Y	no valid.	stat. analysis			[24]
						logistics	stat. analysis			[46]
						financial public admin.	classification			[68]
					ProM pl.	healthcare	classification			[41]
						healthcare	clustering	classification		[17]
					ProM and Camunda pl.	automotive healthcare	evol. algorithm			[48]
			threshold(s)		Y but unavail.	unspecified	classification			[9]
		event log (with timestamps) with data and contextual information			Y	domotic	stat. analysis			[25]
			labeling funct.		Y but unavail.	healthcare	clustering	classification		[16]
				proc. model	ProM plugin	no validation	classification			[38]
			threshold(s)		ProM plugin	logistics	clustering	transition system		[28]
			clusters of behav.		N	logistics manufacturing	clustering	classification		[27]
		event log (with timestamps) with data and unstructured text	labeling funct.		N	financial	classification	text mining		[65]
	next activity	event log (with timestamps) with data			N	unspecified	pattern mining			[10]
					Y	domotic	stat. analysis			[26]
		event log (with timestamps) with data and contextual information			Y	domotic	stat. analysis			[25]
			labeling funct.	proc. model	ProM plugin	no valid.	classification			[38]
	last value of an attribute	event log (with timestamps) with data and contextual information	labeling funct.	proc. model	ProM plugin	no valid.	classification			[38]

Table 1. Predictive process monitoring framework: time and categorical outcome predictions [1]

Pred. type	Det. Pred. type	Input 1	Input 2	Input 3	Tool	Domain	Family of algorithms 1	Family of algorithms 2	Family of algorithms 3	Refer.
sequence of outcomes/values	sequence of future activities	event log (with timestamps)			Y implem.	customer supp. financial public admin.	neural network			[64]
					Y but unavail.	financial automotive customer supp.	neural network			[43]
					Y	financial automotive	neural network			[22]
			backgr. knowledge		Y impl.	healthcare automotive financial public admin. customer supp.	neural network			[19]
					Y	financial automotive	neural network			[23]
		event log (with timestamps) with data			ProM plugin	customer supp. public admin. financial	neural network			[54]
		event log (with timestamps) with data and contextual information			Y but unavail.	logistics	clustering	regression		[31]
	sequence of future activity timestamps	event log (with timestamps)			N	customer supp. financial public admin.	neural network			[64]
			backgr. knowledge		Y impl.	healthcare automotive financial public admin. customer supp.	neural network			[19]
					Y	financial automotive	neural network			[23]
risk	risk	event log (with timestamps)			Y but unavail.	logistics	clustering	classification		[11]
			labeling funct.		Camunda pl.	financial	similarity-weight. graph	stat. analysis		[13]
		event log (with timestamps) with data	threshold(s)		N	transport logistics	neural network			[45]
					ProM plugin	financial logistics	stochastic Petri net			[57]
			labeling funct.	proc. model	Yawl pugin	no valid.	classification			[15]
						logistics unspecified	evol. algorithm			[14]
inter-case metr.	inter-case metrics	event log (with timestamps)			Y but unavail.	logistics	clustering	regression		[11]
		event log (with timestamps) with data	labeling funct.		ProM plugin	unspec.	regression			[52]
			threshold(s)		N	transport logistics	neural net.			[45]
						no valid.	classification	regression	time series	[71]
					Y but unavail.	unspec.	classification	time series		[71]
	workload	event log (with timestamps) with data and contextual information	labeling funct.		ProM plugin	no valid.	classification			[9]
cost	cost	event log (with timestamps) with data	threshold(s)		N	transport logistics	neural net.			[45]
		event log (with timestamps) with resources	cost schema		ProM plugin	no valid.	trans. system	stat.analysis		[70]

Table 2. Predictive process monitoring framework: sequence of outcomes/values, risk, inter-case metrics, cost [1]

3 Application Example

In this passage the resulting tables 1 and 2 are applied to the simple example of a three step production line for self aware robots. In activity A, body and limbs are assembled. This step is dependent on the availability of parts. Furthermore an event log containing assembly and waiting times is laid out. Activity B is the quality assurance step for consciousness-independent functions. Basic motor functions are tested. The testing is dependent on the previous assembly of robot bodies. An event log containing testing and waiting times is laid out. In the last Activity C, the artificial consciousness is simply uploaded to the robot body. Due to the software architecture and capacity of the factory the upload can only be implemented in parallel for ten soon to be self aware robots at a time. The factory operators have noticed great fluctuations in the time it takes to fill up the ten spots for the consciousness upload in activity C, descending from delays in assembly or product testing. They decide to apply predictive process monitoring to proactively identify and counteract delays. To find a suitable approach they apply their problem to the tables 1 and 2.

Going through the table from left to right, first the prediction type has to be determined. In the case of the self aware robot factory, it is time, more precisely activity delays. The second category are the available inputs. For this example, there are time stamped event logs and a process model describing the connection between activities A, B and C available. The third relevant category for the operators is the domain for which the approaches are most suitable. In this case, manufacturing. Lastly the family of applied algorithms is considered, since the developers of the company already have experience with regression and stochastic Petri nets. Considering the prediction and input type, as well as the family of algorithms the companies' developers are familiar with, the papers [2] and [4], are chosen to be further reviewed. Since the paper [3] is more suitable for the manufacturing domain and also applies regression, the operators decide to also consider it, even though it would require additional input, in the form of inter-case metrics, to be generated.

4 Strong and Weak Points

4.1 Strong Points

well formulation what the paper tries to achieve −¿ formulation of a research question and splitting it top down into smaller points of interest −¿ well structured Convincing modus operandi for the selection and review of the subject. Overseeable presentation of results in the form of tables. Easily implementable for businesses to find fitting approaches.

4.2 Weak Points

Some passages are though to read, because the sentences get very long A lot of repetition. review process is described in some detail within the introduction and then again within a separate section

References

1. Di Francescomarino, C., Ghidini, C., Maggi, F. M., and Milani, F. Predictive process monitoring methods: Which one suits me best? *arXiv preprint arXiv:1804.02422* (2018).
2. Rogge-Solti, A., Vana, L., and Mendling, J. Time series petri net models-enrichment and prediction. CEUR Workshop Proceedings.
3. Senderovich, A., Di Francescomarino, C., Ghidini, C., Jorbina, K., and Maggi, F. M. Intra and inter-case features in predictive process monitoring: A tale of two dimensions. In *International Conference on Business Process Management* (2017), Springer, pp. 306–323.
4. Verenich, I., Nguyen, H., La Rosa, M., and Dumas, M. White-box prediction of process performance indicators via flow analysis. In *Proceedings of the 2017 International Conference on Software and System Process Pages* (2017), ACM, pp. 85–94.