# Capturing Process Behavior with Log-Based Process Metrics

Marijke Swennen[1], Gert Janssenswillen[1,2], Mieke Jans[1], Benoît Depaire[1], Koen Vanhoof[1]

[1]Hasselt University, Agoralaan Gebouw D, 3590 Diepenbeek, Belgium
[2]Research Foundation Flanders (FWO), Egmontstraat 5, 1000 Brussels, Belgium
{marijke.swennen, gert.janssenswillen, mieke.jans, benoit.depaire, koen.vanhoof}@uhasselt.be

**Abstract.** Currently, process mining literature is primarily focused on the discovery of comprehensible process models that best capture the underlying behavior in event logs. Consequently, the resulting models a) aggregate information, based on algorithm-specific assumptions, and b) transform information into a simplified representation. Both characteristics, which are valuable in certain, different contexts, suffer from the inability to describe objectively the behavior that is inherent to the event log at hand. In this paper, we present the need for log-based process metrics to capture the process behavior in an event log, without the need to first discover a model. The metrics provide a process owner with unbiased, algorithm-agnostic information of the event log, as a starting point of the process analysis. The constructed metrics also serve as a mean to objectively compare different event logs in terms of time-related and variance aspects.

**Keywords:** Process mining • Operational excellence • Process behavior • Log-based process metrics

## 1 Introduction

Process mining is intended to detect strategic insight from business processes by extracting valuable information from event logs. Next to discovering process models from event logs, process mining is also used to check the conformance between a process model and reality and to extend process models with extra information [12]. Starting point for performing a process mining task is an event log. When performing a discovery task on an event log, a process model is extracted without using any additional information [13]. Many discovery algorithms have been introduced [3], [15] and each has its specific assumptions resulting in models not suited for or aimed at describing the behavior that is inherent to the event log objectively and in a detailed fashion.

Looking from a Business Process Management perspective to models, business processes should be modified and improved continuously driven by the continuous improvement concept. This concept is related to methodologies such as lean management, Six Sigma and business process improvement and reengineering [1], [4].

In literature, it has been suggested that process mining can be used to support operational excellence in companies [12,13,14]. However, if process mining is used to implement methodologies such as lean management or Six Sigma, it can be cumbersome to decide which discovery algorithms and assumptions to choose. Moreover, process models discovered from an event log are not always perfect representations of reality.

Therefore, the goal of this paper is to present the need for log-based process metrics, which are measures indicating how the current process is running, without the need of a process model. In contrast to traditionally used KPIs (for measuring performance), the proposed metrics are constructed on the level of the event log or the activities executed in the event log instead of the output level. The metrics provide an unbiased picture of the present process behavior.

# 2 Log-Based Process Metrics

Building on the idea to calculate the distance between two event logs, which was presented in [11], the goal of this research is providing process metrics to identify and quantify the behavior of a process. Four categories of process performance indicators - quality, time, costs and flexibility- have been defined in [6]. Quality and costs can be seen as derivatives of process behavior. Time and flexibility, however, are inherent to the way in which a process is carried out. In this paper, we will focus on the dimensions *time* and *structuredness.* Structuredness is chosen because we want to measure how structured -and not how flexible- the behavior in the event log is. Although structuredness is defined in [15] as a quality metric to measure the ease of interpretation of a process model, we define structuredness as the level of variation in the event log.

According to the study on model-log evaluation metrics in [2], only one dimension or level of analysis should be measured by each metric, in order to remain comprehensible. Building on the different feature scopes presented in [11], possible levels of analysis are: the *log level*, which represents the complete event log, the *trace level*, representing characteristics of sequences of activities, and the *activity level*, representing characteristics of the activity types, aggregated over the entire log.

## 2.1 Time

Interesting concepts for the time dimension are, among others, the duration, the actual processing time and the waiting time of cases or activities. If we, for example, have a look at the actual processing time, or service time, of an activity in the event log, a list of summary statistics can be interesting to get a notion of the duration of each activity in the process. Building from this, a *bottleneck activity* can be revealed. In a process, a bottleneck is an activity that obstructs other activities to be executed properly and determines the continuation of the whole process [10]. According to the theory of constraints [5], bottlenecks or constraints should be eliminated from a process because ‘a process is only as strong as its weakest link’. A bottleneck indicator could be calculated by searching for the activity in the process that has the longest duration compared to the duration of the other activities in the process.

### 2.2 Structuredness

Concepts to analyze the structuredness of an event log can be variance, self-loops, repetitions and the presence of batch processing. A first notion of the structuredness or variance in an event log is the number of patterns, or distinct traces, that are recorded. Next to this, the minimum number of traces that is required to cover for example 80 % of the cases can also be of interest to a company. Moreover, the frequency of specific traces or specific activity types can help a company to get an insight in which activities or sequences of activities should be paid the most attention to. An overview of which activities are usually the last activity in a case or the amount of different end activities in an event log can be an indication of the number of pending cases.

The key goal of lean management is avoiding non-value adding activities or waste [16]. Activity instances of the same activity type that are executed more than once immediately after each other are in a self-loop (length-1-loop), what might be an indication of not adding value to the process. Next to this, repetitions of activities in a case, not immediately after each other, might also be an indication of waste.

Another form of waste is batch processing, which can be defined as activities piled and handled simultaneously by the same resource [9], [16]. This results in cases waiting to be handled while other cases are handled immediately. The importance of identifying batch processing in event logs is put forward in [8]. For example, a comparison between the duration of activities executed in a batch and the same activities executed not in a batch can provide the company with an overview of activities for which it is more beneficial to be handled together instead of handling them immediately at arrival time.

## 3 Implementation and Evaluation

All metrics will be implemented in the R-package edeaR [7], which stands for Exploratory and Descriptive Event-based data Analysis in R. To evaluate the added value of the metrics, all metrics will be applied to both artificial and real event logs.

## 4 Conclusions and Future Work

From literature, we can infer that plenty of metrics exist for checking the conformance of process models with reality or for measuring the performance of discovery algorithms. However, choosing the right process discovery technique and its specific assumptions can be cumbersome for companies that have dynamic and rapidly changing processes. Moreover, the resulting process models are not suited for or aimed at describing objectively the behavior that is inherent to the event log. Therefore, log-based process metrics are needed, which provide business people with an objective start to look at their processes. All metrics will be discussed with people from industry, implemented in the edeaR-package in R [7] and applied in a real life case study.

However, some challenges and different perspectives can provide an even better indication of the process behavior observed in an event log. First, the resources level of analysis can be of interest to see which worker is executing a task. Next to this, an alternative metric on relevance of a trace, other than frequency, would add incremental value to the current metrics. Moreover, other dimensions of behavior can be taken into account to provide business people with an overall view of their business processes. Finally, metrics should not be considered to be independent from each other. The results of one metric can be the input of or complement other metrics as stated in [6].

## References


1. Bigelow, M.: How To Achieve Operational Excellence. Quality Progress, 35(10), 70-75 (2002)
2. De Weerdt, J., De Backer, M., Vanthienen, J.,Baesens, B.: A Critical Evaluation Study of Model-Log Metrics in Process Discovery. In: zur Muehlen, M., Su, J. (eds.) Business Process Management Workshops, pp. 158-169. Springer, Heidelberg (2011)
3. De Weerdt, J., De Backer, M., Vanthienen, J., Baesens, B.: A Multi-Dimensional Quality Assessment of State-of-the-art Process Discovery Algorithms Using Real-Life Event Logs. Information Systems, 37(7), 654-676 (2012)
4. Drohomeretski, E., Gouvea da Costa, S.E., Pinheiro de Lima, E., Garbuio, P.A.D.R.: Lean, Six Sigma and Lean Six Sigma: an Analysis Based on Operations Strategy. International Journal of Production Research, 52(3), 804-824 (2014)
5. Goldratt, E.M., Cox, J.: The Goal – A Process of ongoing improvement. North River Press Inc., New York (1984)
6. Heckl, D., Moormann, J.: Process Performance Management. In: Handbook on Business Process Management 2, pp. 115-135. Springer, Heidelberg (2010)
7. Janssenswillen, G., Swennen, M., Depaire B., Jans, M., Vanhoof, K.: Enabling Event-data Analysis in R - Demonstration. Proceedings of the 5th International Symposium on Data-driven Process Discovery and Analysis (SIMPDA), Vienna (2015)
8. Martin, N., Depaire, B., Caris, A.: The use of process mining in business process simulation model construction: structuring the field. Business & Information Systems Engineering (forthcoming)
9. Martin, N, Swennen, M., Depaire, B., Jans, M., Caris, A., Vanhoof, K.: Batch Processing: Definition and Event Log Identification. Proceedings of the 5th International Symposium on Data-driven Process Discovery and Analysis (SIMPDA), Vienna (2015)
10. Melton, T.: The Benefits of Lean Manufacturing: What Lean Thinking has to Offer the Process Industries. Chemical Engineering Research and Design, 83(6), 662-673 (2005)
11. Ribeiro, J., Carmona, J., Mısır, M., Sebag, M.: A Recommender System for Process Discovery. In: Sadiq, S., Soffer, P. (eds.) International Conference on Business Process Management 2014, LNCS, vol. 8659, pp. 67-83, Springer, Heidelberg (2005)
12. van der Aalst, W.M.P.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Heidelberg (2011)
13. van der Aalst, W.M.P. et al.: Process Mining Manifesto. In: Business Process Management Workshops, LNBIP, vol. 99, pp. 169-194, Springer, Heidelberg (2011)
14. van der Aalst, W.M.P., Adriansyah, A., van Dongen, B.: Replaying History on Process Models for Conformance Checking and Performance Analysis. WIREs Data Mining and Knowledge Discovery, 2(2), 182-192 (2012)
15. vanden Broucke, S.: Advances in Process Mining: Artificial Negative Events and Other Techniques. Ph.D. thesis, KU Leuven (2014)
16. Womack, J., Jones, D.T.: Lean Thinking: Banish Waste and Create Wealth in Your Corporation. Simon and Schuster, London (1996)