Interactive Data-Driven Business Process Simulation (Extended Abstract)

Gerhardus van Hulzen
Research group Business Informatics
Hasselt University
Hasselt, Belgium
0000-0001-8962-9515

I. Introduction

Today, healthcare systems worldwide are under constant pressure. On the one hand, increasing population numbers, ageing populations, lifestyle factors, and new technologies are increasing the yearly expenses on healthcare. On the other hand, budgets are under pressure due to economic austerity [1]. In order to provide high-quality care to all patients, healthcare managers are forced to improve their care processes. Efficient Capacity Management (CM) is one of the key aspects to ensure this. This involves, amongst others, determining the suitable resource levels – i.e. staff size, equipment, and facilities [2].

Business Process Simulation (BPS) can be used to support managers during CM decisions. BPS uses a (computer) model to imitate the behaviour of a business process. This approach allows evaluating the effects of changes before implementing them [3]. For instance, BPS can be used to determine suitable equipment levels, e.g. by simulating the effect of an additional X-ray scanner on patient waiting times, throughput rates, and staff workload.

In Process Mining (PM) the emerging field of data-driven process simulation provides promising first results to generate simulation models from information captured in event logs [4]. These "discovered" models can form the basis to compare the operational effects of various capacity levels. The main advantage of data-driven process simulation over "traditional" simulation model development is the availability and objectivity of event logs compared to information sources, such as interviews, process documentation, and observations [5]. However, some challenges remain in the field of automated BPS discovery. Most importantly, the lack of *domain knowledge* makes it challenging to extract a reliable and usable simulation model. In addition, event logs often suffer from *data quality issues*, which strongly affects the reliability of the simulation results [6]. Therefore, it is imperative to take these problems seriously.

II. Research Objectives

Given the context outlined above, this PhD research pursues the following two objectives:

1) *Extended support for key BPS modelling tasks:* While the field of automated BPS discovery renders promising results; there are still challenges ahead to discover individual BPS model components to make it usable to support CM decisions.
2) *Enabling interactive data-driven process simulation:* Domain knowledge should be closely integrated during the discovery of BPS models to ensure the reliability and usability of the discovered simulation models.

III. Planned Research Activities

The following subsections give an overview of the planned research activities for the two research objectives.

A. Extended Support for Key BPS Modelling Tasks

Based on a systematic literature review, we concluded that defining the control-flow, entity arrival rates, activity execution times, gateway routing logic, entity types, queueing disciplines, resource schedules, resource requirements, and resource roles are the most important modelling tasks to support CM decisions via simulation. These tasks correspond to a subset of modelling tasks given by [7]. Most attention of PM research has been dedicated to control-flow definition [7]. However, for creating a simulation model for supporting CM decisions, we believe that all aforementioned tasks are required – albeit some tasks are more important than others.

In PM, only limited amount of work has been devoted to integrating the various tasks needed to build a simulation model. The authors in [8] were the first to generate an initial simulation model from data. They included the process-flow, gateway routing logic, and resource pools. Later, the authors extended their work with activity durations and entity inter-arrival times [5]. Nevertheless, the authors emphasise that the derived initial model still has to be verified and – if required – augmented by domain experts to ensure validity.

In [9], a PM approach is proposed to generate BPS models for short-term KPI prediction. A similar approach as in [5] is used. However, the resource perspective is left aside, assuming an infinite amount of resources is available [9].

Control-flow, resources, activity durations, and gateway routing logic are supported by the approach in [10]. In addition, they also support inter-arrival times and resource schedules. However, the latter have to be defined manually by the domain expert.

None of the aforementioned studies tried to integrate all elements into a single, simulation-ready model. This is where

Simod [11] extends the work on data-driven process simulation. Simod is a tool which automatically discovers BPS models from event logs. In addition, Simod is also capable of measuring the accuracy of the obtained simulation model and allows to optimise the accuracy using hyper-parameters [11].

While the initial results of data-driven BPS algorithms are promising, there are still challenges to automatically derive a simulation model for supporting CM decisions from event logs. Especially the resource perspective is crucial for CM decisions. Incorrect resource requirements, pools, and schedules make the results of the model unreliable, resulting in inaccurate capacity requirement estimations. The state-of-the-art still has limitations when it comes to defining the resource perspective. Part of this PhD research will be dedicated to improving the support of the resource perspective in data-driven BPS.

B. Enabling Interactive Data-Driven Process Simulation

As mentioned earlier, data quality issues should be taken seriously to ensure the reliability of the data-driven simulation model. Detecting these issues often requires domain knowledge. Therefore, it would be beneficial to involve the domain experts as early as possible to detect and handle data quality issues before integrating everything into a single simulation model. Especially in stochastic models, such as simulation, a problem in one part of the model may have a profound impact on other parts. It is much easier to solve issues at the root, then having to trace back the problem in a full simulation model.

Ideally, domain experts would conduct simulation studies themselves. After all, they know the process best. However, conducting simulation studies requires specific knowledge which domain experts often do not possess. Of course, they could learn more about constructing simulation models, but usually, they are very busy and do not have the time to master the required skills.

Against this background, we propose a framework to interactively involve domain experts during the development of data-driven simulation models. The framework consists of three cycles. The *first cycle* is the initial model construction. In this step, for each required modelling task (e.g. determining the inter-arrival rates, activity durations, resource requirements, the control-flow, etc.) the data requirements are established. If these requirements are fulfilled, the quality of the data is assessed, and a discovery algorithm is applied. The results of this algorithm, together with the detected data quality issues (e.g. missing values, outliers, inconsistencies, etc.), are presented to the domain expert for validation. If needed, the expert can correct these issues and alter the discovery parameters until he or she is satisfied with the results.

In the *second cycle*, all the initial model components from the first cycle are integrated into a single simulation-ready model. The entire model will run for the first time, and the preliminary results will be validated for the first time by the domain expert. By altering parameters, the domain expert can "calibrate" the model until he or she is satisfied with the preliminary results. During this calibration, the domain expert should immediately obtain an estimation of the impact of the changed parameter, instead of having to wait until the simulation has finished running, which could – depending on the complexity of the model – take quite a while.

The *third cycle* of the framework involves the actual model validation. The calibrated model is simulated extensively, and the domain expert validates the simulation results. If needed, the parameters of the simulation model can be altered again to obtain more realistic results. The validated model can be used for further analyses and to evaluate different scenarios.

The goal of this part of the PhD research is to develop a prototype which supports the interactive development of data-driven simulation models.

IV. Concluding Remarks

This PhD will mainly focus on the resource aspect of data-driven BPS and how domain experts can be interactively involved in the discovery of simulation models. This should culminate in the development of a prototype tool which allows interactive data-driven generation of BPS models based on event logs and domain knowledge. The derived simulation model will form the basis for supporting CM decisions in healthcare. Nevertheless, the prototype would also be usable in many other applications in different fields besides healthcare, such as production planning in manufacturing, supply chain logistics, and transportation.

References

[1] C. Hicks, T. McGovern, G. Prior, and I. Smith, "Applying Lean Principles to the Design of Healthcare Facilities," *International Journal of Production Economics*, vol. 170, pp. 677–686, 2015.

[2] F. R. Jacobs and R. B. Chase, "Strategic Capacity Management," in *Operations and Supply Management: The Core*, ser. Operations and Decision Sciences. New York, NY, USA: McGraw Hill/Irwin, 2008, pp. 51–79.

[3] N. Melão and M. Pidd, "Use of Business Process Simulation: A Survey of Practitioners," *Journal of the Operational Research Society*, vol. 54, no. 1, pp. 2–10, 2003.

[4] B. Depaire and N. Martin, "Data-Driven Process Simulation," *Encyclopedia of Big Data Technologies*, 2018.

[5] A. Rozinat, R. S. Mans, M. Song, and W. M. P. van der Aalst, "Discovering Simulation Models," *Information Systems*, vol. 34, no. 3, pp. 305–327, 2009.

[6] L. Vanbrabant, N. Martin, K. Ramaekers, and K. Braekers, "Quality of Input Data in Emergency Department Simulations: Framework and Assessment Techniques," *Simulation Modelling Practice and Theory*, vol. 91, pp. 83–101, 2019.

[7] N. Martin, B. Depaire, and A. Caris, "The Use of Process Mining in Business Process Simulation Model Construction," *Business & Information Systems Engineering*, vol. 58, no. 1, pp. 73–87, 2016.

[8] A. Rozinat, R. S. Mans, and W. M. P. van der Aalst, "Mining CPN Models: Discovering Process Models with Data from Event Logs," in *Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools*, K. Jensen, Ed., Aarhus, Denmark, 2006, pp. 57–76.

[9] I. Khodyrev and S. Popova, "Discrete Modeling and Simulation of Business Processes Using Event Logs," in *Proceedings of the 14th International Conference on Computational Science*, ser. Procedia Computer Science, D. Abramson, M. Lees, V. Krzhizhanovskaya, J. Dongarra, and P. M. A. Sloot, Eds., vol. 29. Cairns, QLD, Australia: Elsevier, 2014, pp. 322–331.

[10] B. Gawin and B. Marcinkowski, "How Close to Reality is the "as-is" Business Process Simulation Model?" *Organizacija*, vol. 48, no. 3, pp. 155–175, 2015.

[11] M. Camargo, M. Dumas, and O. González-Rojas, "Automated Discovery of Business Process Simulation Models from Event Logs," *Decision Support Systems*, vol. 134, 2020.