Review of a Doctoral Thesis at FIT BUT

Doctoral thesis: MODELLING AND ANALYSIS OF LOGISTICS PROCESSES BY APPLYING PROCESS AND DATA MINING TECHNIQUES

Name of the doctoral student: Mgr. Julia Rudnitckaia

Name and institution of the reviewer:

Jaroslav Zendulka, Brno University of Technology, Faculty of Information Technology, Czech Republic

---

I. Doctoral Thesis

Appropriateness and Relevance

Process mining has been an integral part of the data mining and data science disciplines since the beginning of this century. Due to the availability of data related to processes in many areas, it has high potential for process management. One of the areas for which the process mining application can be very beneficial are logistics systems, whose process model is usually very complex, often non-structured, and as a result difficult to describe and understand. Therefore, I consider the topic and content of this doctoral thesis to be topical and relevant to the Computer Science and Engineering doctoral study programme.

A summary of the Contributions of the Thesis, Novelty and Significance

The goal of the thesis is stated on page 4 as "to extract knowledge and models from real-world raw event logs from interconnected processes and to identify problems to further improve the complex non-structured processes". The expected result was to be "a consistent set of methods and techniques, which are strongly interconnected".

I consider the main result of the thesis to be the proposal of a methodology for the use of process mining and other data science methods and techniques in the field of logistics. It represents a systematic approach of the development of advanced tools supporting planning and process management. Context-awareness is used here to deal with complex and non-structured processes.

The methodology is presented as a framework covering the individual steps of the methodology. As part of a more detailed description of the steps, the available approaches and techniques that can be used for subtasks that need to be performed are recommended. In two cases (bottleneck analysis and predictive modeling), the author proposes her own methods. For the last step of the methodology – the creation of a process map, only basic idea and possible services provided are outlined on the example of the Google Maps tool. I agree that a more detailed proposal is beyond the scope of this thesis.

The case study in Chap. 4, in which real data obtained from the ship handling part of a seaport system are used. It can be considered as a proof of concept of some aspects of the proposed framework. I consider the use of real data to be very valuable, as it allowed the evaluation of the proposed approach in conditions close to the target ones.

I consider the declared goal of the thesis to be fulfilled and its results to be original and usable in further research and in practice.

Evaluation of the Formal Aspects of the Thesis:

The thesis is structured to five chapters. The first chapter is an introduction and focuses mainly on the motivation of solving the given topic and on the goals of the thesis. In the second chapter, the background and state-of-the-art in the main areas related to the thesis topic and the areas in which the original results were achieved are presented. There is also an introduction to the seaport logistics environment here. The third and fourth chapters represent the core of the thesis. In the third, the above-mentioned methodology and the framework, and in the fourth, a case study are presented. The fifth chapter is the conclusion with possible directions for future research.

Overall, the thesis is mostly well written, has a logical structure and the content is understandable. There are several places where the presentation could have been better (see my defense questions below).

The number of errors and inaccuracies in the work is not high, except for the description of the Confidence interval method on pages 69 – 71. There are inaccuracies here, which make understanding more difficult, although the main idea of the method is simple. Specifically:

- If the median of EQT for operation A12800 is 86s, then the threshold for confidence interval with respect to the experts' recommendation to extend the interval by 6 – 7s is not 25s. Where does Algorithm 1 take this recommendation into account? Is the value *nb* (no bottleneck) of the 3rd operation in the *direct* sequence on p. 71 the result of such extending? According to Algorithm 1, if *diff* < 0, result should be *Slower,* and a bottleneck is present.
- Formally, variables *diff* and *results* are declared in Algorithm 1 as arrays/vectors but used as simple variables. Similarly *direct* in Algorithm 2.
- Some members of the *EQT* sequence on page 70 and the *diff* sequence on page 71 are missing a decimal point. In addition, the mean times for the EQTs of the A12800 and A13500 operations differ from those in Fig. 3.7 (both in Fig. 3.10 and in the sequences below it).

Of the other inaccuracies in the thesis, I only mention the fact that equation 3.4 on p.73 is not for MSE, but MAE, and that the right parenthesis is missing there.

The thesis is written in English, there are occasional typos and minor grammatical errors, but their number is not high.

Quality of Publications

The thesis contains references to 5 publications of the author ([22] and [67] are duplicates). Four of them are part of the author's attached publications (Time Series Analysis and Predictive Statistical Models for Ship Handling Time at an Oil Terminal are attached as an abstract in the Proceedings and in the full version). The referenced publication [64] is not attached. Of these cited publications, one ([68]), is a journal publication, 3 ([64], [67], [71]) are conference papers and one ([54]) was published in a university journal. Julia Rudnitckaia is the first author of all these publications.

A substantial part of the original results of the thesis was published in the journal publication. It is a publication in a quality journal IEEE Access (Web of Science: JIF 2021 is 3.48, JCR Q2 (76/246) in Computer Science, Information Systems; Scopus: SCR Q1 (34/231) in General Computer Science). In the conference papers, some partial results were presented (creation of a process model, prediction of time parameters). The corresponding conferences are not among those included in the CORE rating, but their Proceedings are indexed in the Web of Science (3) and in Scopus (4). The number of citations of these publications is one in Web of Science and 3 in Scopus (all self-citations), and 6 (4 without self-citations) in Google Scholar. I consider the requirement to publish the core of the dissertation fulfilled.

II. Candidate's Overall Achievements

From the author's CV, it is evident that she completed internships at two European universities and another one at Honeywell company in Germany during her doctoral studies. She was involved in solving research projects related to the topic of her doctoral thesis there. This collaboration also resulted in several joint publications. In addition, she worked as a business and database analyst in two companies in the Czech Republic. She, therefore, has experience of involvement in research and development teams outside of her university.

III. Defense Questions

1. Your description of the Two-step predictive model in Chapters 3 and 4 is very brief. Explain it in more detail. Focus mainly on the connection between the two steps (what from the first step and how is be used in the second step) and how context-aware predictors are applied when using the SARIMA method. Is the corresponding time sequence a multivariate one with exogenous variables?
2. When trying to deal with non-structured (spaghetti) models, you suggest using various methods. To what extent can data mining methods be applied here and to what extent are manual approaches using expert knowledge applied?
3. In Fig. 4.4 on page 81 you show three different notations for visualizing the process model. In text you state that "the loop in the process model is visible on all three notations". Where is the loop in the model visualized as a DFG?
4. In Fig. 4.5 on page 82, you compare the variants of the obtained descriptive model with a normative process model. How was the normative model defined?

IV. Conclusion

Overall, I evaluate the doctoral thesis of Mgr. Julia Rudnitckaia positively. Although I have some comments about her thesis, they are not fundamental. In my opinion, she has demonstrated the ability of independent creative work in the field of her study. Her contribution, publications, and potential usefulness of the thesis results for practice and research community meet the standard requirements imposed on a doctoral thesis. Therefore, **I recommend her thesis for defense**.

Brno 19. 3. 2023

Jaroslav Zendulka