Humboldt-Universität zu Berlin
Mathematisch-Naturwissenschaftliche Fakultät
Institut für Informatik

Neural Networks for Partially Ordered Trace Resolution

Bachelorarbeit

zur Erlangung des akademischen Grades
Bachelor of Science (B. Sc.)

eingereicht von: Glenn Dittmann
geboren am: 13.06.1993
geboren in: Berlin

Gutachter/innen: Prof. Dr. Matthias Weidlich
Dr. Wolfgang Koessler

eingereicht am: .... verteidigt am: ....

Contents

Purpose and scope of your entire report. The purpose of your entire report is to make a **scientific argument using the scientific method**.A scientific argument always has the following steps that all must come in this order.

SM1 **Explicate the assumptions and state of the art** on which you are going to conduct your research to investigate your research problem / test the hypothesis.

SM2 Clearly and precisely **formulate a research problem or hypothesis**.

SM3 **Describe the (research) method** that you followed to investigate the problem / to test the hypothesis in a way that **allows someone else to reproduce your steps**. The method must include steps and criteria for evaluating whether you answered your question successfully or not.

SM4 **Provide execution details** on how you followed the method in the given, specific situation.

SM5 **Report your results** by describing and summarizing your measurements. You must not interpret your results.

SM6 **Now interpret your results** by contextualizing your measurements and drawing conclusion that lead to answering your research problem or defining further follow-up research problems.

1 Introduction

Purpose and scope of Section 1. The introduction is a summary of your work and your scientific argument that shall be understandable to anyone in your scientific field, e.g., anyone in Data Science. A reader must be able to comprehend the problem, method, relevant execution details, results, and their interpretation by reading the introduction and the introduction alone. Section 1.1 introduces the general topic of your research Section 1.2 discusses the state of the art and identifies a research. Section 1.3 then states the research problem to investigate. Section 1.4 explains the research method that was followed, possibly with execution details. Section 1.5 then presents the results and their interpretation. Only if a reader thinks they are not convinced or they need more details to reproduce your study, they shall have to read further. The individual chapters and sections provide the details for each of the steps in your scientific argument.

You usually write the introduction chapter *after* you wrote all other chapters, but you should keep on making notes for each of the subsections as you write the later chapters.

1.1 Context and Topic (SM1)

1.2 State of the Art (SM1)

1.3 Research Question (SM2)

1.4 Method or Approach (SM3, SM4)

1.5 Findings (SM5, SM6)

2 Background

2.1 Preliminaries

In the field of process mining, a subfield of (see process mining book) we deliberately investigate the happenings of processes of the "real world". It is about modeling, checking and improving such processes. Furthermore we can apply the area of conformance, checking on the area of process mining, which is basically a more in depth investigation how well a model of a process and the actual execution of such processes are conforming. This chapter is therefore used to gently introduce typical concepts that are used in process mining and more precisely defining the attributes we need in order to answer the research question. Additionally our (main) method to solve the problem of finding resolutions for partially ordered traces will be using neural networks, i.e. a special method from the wide field of machine learning. Thus we will give a short overview over typical machine learning ideas and more concisely introduce artificial neural networks, all of their properties and capabilities as well as the different network architectures we will be have used.
Most of the plain mathematical concepts should be of knowledge to the reader, and therefore instead of defined mostly explained in text. There might be some exceptions relating to the fields of either conformance checking or neural network such as conformance checking measures or algorithms, gradient descent functions for error minimizing in ANN's or the way of encoding input data in an ANN.

Definition 1. Activity Activities are ... captured in the set of all activities.

Definition 2. Event An event is... mapping from events to activities; events are unique in a log

Definition 3. Trace A trace is...

Definition 4. Event Log An event log is...

Based on the timestamps associated with events in a given log, we define a partial ordering of events for a certain trace. Thus an event $e1 < e2$, if the timestamp of e1, i.e. time(e1), is smaller than/happens before, the timestamp of e2, i.e time(e2). Otherwise the timestamps of e1 and e2 are the same, leading to $e1 = e2$, i.e. $time(e1) = time(e2)$.

Definition 5. Uncertain Trace A trace is called uncertain if there is a trace for which for any two events ei, ej $time(ei) = time(ej)$

Definition 6. Certain Trace A trace is called certain if for all events e1 to en for all traces, ei < ei+1.

Based on this, if all traces in a log are certain, we call it a certain log and an uncertain log otherwise.
Event-Name, Event-Alias, Event-Properties

Definition 7. Artifical Neural Network Artifial Neural Network (ANN)

Definition 8. feedforward-ANN feedforward-ANN

Definition 9. Recurrent Nerual Network Recurrent Neural Network (RNN)

Definition 10. One-Hot-Encoding One-Hot-Encoding

Definition 11. Embed-Encoding Embed-Encoding

Partial Order, Total Order

2.2 Related Work

The field of conformance checking and process mining is very broad, so a lot of research has been done there up to today. Furthermore the field of machine learning has reached another peak of high interest in business applications, as well as media, teaching and research. The particular problem of solving partially ordered event logs from real-life process applications though, has been investigated in a fairly small amount compared to the above. Different techniques have been used to solve the problem differently and machine learning was only used once for predicting information of event logs.
To my best knowledge and also stated in previous work [LMFvdA14] so far there has been little research done addressing the problem of only partially ordered event logs.
M. de Leoni et al. presented a technique to abstract the problem of aligning partially ordered traces into a PDDL-encoded planning problem. This approach will either report, that there exists no solution, i.e. optimal alignment, for an explicit trace and petri net. Or it will, in finite time, find an optimal alignment for a trace and petri net, whether or not the trace is sequential, i.e. totally ordered or a trace containing concurrent events, i.e. partially ordered. (note that by definition every totally ordered trace it also a partially ordered trace).[dLLM18]
Van der Aalst et al. took another approach in defining partially ordered traces and from those compute partially ordered alignments, with the aim to provide a model that can express concurrently running events and from there getting insight in the meaning of those. [LFvdA14] They researched the usefulness of those partially-ordered alignments with case studies relying on real-world data from a Dutch hospital. [LMFvdA14]
In [TVLRD17] Niek Tax et. al. introduce an approach to tackle three question not yet visited in the field of process mining. They use RNN's, LSTM's specifically, to answer

three questions 1),2),3). However they let the order certainity of traces unvisited and thus, while exploring an answer for the next activity, omit the information of certain parts of a trace already being order / they only give answer on how future traces could end most possibly.
Weidlich et. al. have sought three algorithmic approaches for resolving partially ordered traces in giving a probability distribution over all possible resolutions and from there on efficiently compute the conformance of partially ordered traces with a given process model.[vdALW19]
As of my best knowledge no research has yet been done, to resolve partial ordering of traces in conformance checking / process mining. We expect to find valuable solutions for resolving the partial ordering of events happening at the same time, exploiting the field of deep learning, neural networks respectively.
Trying to cite the books read so far:
Process Mining [Aal16]
Conformance Checking [CvDSW18]
Deep Learning [GBC16]
Hands-on ML [Gér19]

3 Problem Exposition (optional)

3.1 Context / Business Understanding (SM1)

3.2 Data Understanding (SM1)

3.3 Detailed Research Questions (SM2)

3.4 Detailed Method (SM3)

4 First Real Chapter addressing first Research Problem

4.1 First Sub-Problem

4.2 Second Sub-Problem

5 Second Real Chapter

6 Evaluation

6.1 Objective (SM2)

6.2 Setup (SM3)

6.3 Execution (SM4)

6.4 Results (SM5)

6.5 Discussion (SM6)

7 Conclusion

Your conclusions are not just a factual summary of your work, but they position, interpret and defend your findings against the state of the art that you discussed in Sect. 1.2. You specifically outline which concrete findings or methodological contributions advance our knowledge towards the general objective you introduced in Sect. 1.1. Objectively discuss which parts you solved and in which parts you failed.
You should explicitly discuss limitations and shortcomings of your work and detail what kind of future studies are needed to overcome these limitations. Be specific in the sense that your arguments for future work should be based on concrete findings and insights you obtained in your report.

References

[Aal16] Wil van der Aalst. Process mining : data science in action / by wil van der aalst, 2016.

[CvDSW18] Josep Carmona, Boudewijn van Dongen, Andreas Solti, and Matthias Weidlich. *Conformance Checking*. Springer, 2018.

[dLLM18] Massimiliano de Leoni, Giacomo Lanciano, and Andrea Marrella. Aligning partially-ordered process-execution traces and models using automated planning. In *Twenty-Eighth International Conference on Automated Planning and Scheduling*, 2018.

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[Gér19] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.

[LFvdA14] Xixi Lu, Dirk Fahland, and Wil MP van der Aalst. Conformance checking based on partially ordered event data. In *International conference on business process management*, pages 75–88. Springer, 2014.

[LMFvdA14] Xixi Lu, Ronny S Mans, Dirk Fahland, and Wil MP van der Aalst. Conformance checking in healthcare based on partially ordered event data. In *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, pages 1–8. IEEE, 2014.

[TVLRD17] Niek Tax, Ilya Verenich, Marcello La Rosa, and Marlon Dumas. Predictive business process monitoring with lstm neural networks. In *International Conference on Advanced Information Systems Engineering*, pages 477–492. Springer, 2017.

[vdALW19] Han van der Aa, Henrik Leopold, and Matthias Weidlich. Partial order resolution of event logs for process conformance checking. 2019.

Selbständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und noch nicht für andere Prüfungen eingereicht habe. Sämtliche Quellen einschließlich Internetquellen, die unverändert oder abgewandelt wiedergegeben werden, insbesondere Quellen für Texte, Grafiken, Tabellen und Bilder, sind als solche kenntlich gemacht. Mir ist bekannt, dass bei Verstößen gegen diese Grundsätze ein Verfahren wegen Täuschungsversuchs bzw. Täuschung eingeleitet wird.

Berlin, den 11/08/2020 ....