Utrecht University

Department of Mathematics and Computer Science
Process Analytics

The generation of interpretable counterfactual examples by finding minimal edit sequences using event data in complex processes

Master Thesis

Olusanmi Hundogan

Supervisors:
Xixi Lu
Yupei Du

February 18, 2022

Abstract

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Contents

List of terms

BI Business Intelligence. 3, *see:* Business Intelligence

BPM Business Process Management. 3, 9, *see:* Business Process Management

Business Intelligence XXX. 3

Business Process Management XXX. 3, 9

Comma Seperated Values A structured data format to store information. Every line relates to a data point and every feature is seperated by a seperator. The seperator is commonly a comma but other characters like tabs or semicolons are valid as well.. 3, 9

Continuous Process Improvement XXX. 3, 9

Corporate Performance Management XXX. 3

CPI Continuous Process Improvement. 3, 9, *see:* Continuous Process Improvement

CPM Corporate Performance Management. 3, *see:* Corporate Performance Management

CSV Comma Seperated Values. 3, 9, *see:* Comma Seperated Values

Data Mining XXX. 9

Event Log A collection of event data, that's produced by the process. They are the main input of every process mining venture.. 9

eXtensible Event Stream An XML-based data format to store event logs. The format was developed and adopted by the IEEE Task Force on Process Mining.. 4, 9

Information System XXX. 4, 9

IS Information System. 4, 9, *see:* Information System

ML Machine Learning. 5

Process Event Also called activities. A discrete step in the process.. 8

Process Instance Also called case. A collection of activities that belong to a common entity that is produced by the process.. 8–10

Total Quality Management XXX. 4

TQM Total Quality Management. 4, *see:* Total Quality Management

XAI eXplanable AI. 5, 10

XES eXtensible Event Stream. 4, 9, *see:* eXtensible Event Stream

Chapter 1

Introduction

1.1 Context of this Thesis

Many processes, often medical, economical, or administrative in nature, are governed by sequential events and their contextual environment. Many of these events and their order of appearance play a crucial part in the determination of every possible outcome. With the rise of AI and the increased abundance of data in recent years several techqniques emerged that help to predict the outcomes of complex processes in the real world.**[Expand the domain application.]**

For instance, research in the Process Mining discipline has shown that is possible to predict the outcome of a particular process fairly well [CITE] .**[However, while many prediction models can easily certain outcomes, it remains a difficult challenge to understand what led to a particular outcome. This obstacle is undesirable, as knowing the main factors to an outcome can help understand how to steer a process to a desired outcome with minimal effort.]** In other words, we want to change the outcome of a particular event, by making it maximally likely, with as little interventions as possible [CITE TEST].

One-way to better understand the Machine Learning (ML) models lies within the eXplanable AI (XAI) discipline. XAI dedicates its research to the **research and** development of so-called *black-box models* that are difficult to interpret. Most of the discipline's techniques produce explanations that guide our understanding.

A prominent and human-friendly approach uses the generation of counterfactuals as primary explanation tool. Counterfactuals within the AI framework help us to answer hypothetical "what-if" questions. In this thesis, we will raise the question, how we can use counterfactuals to change the trajectory of a models' prediction towards a desired outcome. Knowing the answers will help us further understand what to do to avoid or enforce the outcome of a process.**[WHY]**

1.2 Problem Space

In this paper, we will approach the problem of generating counterfactuals for processes. The literature has provided a multitude of techniques to generate counterfactuals for AI models, that are derived from static data[1]. However, little research has focussed on counterfactuals for dynamic data[2]. A major reason, emerges from a **[multitude – better #]** of challenges, when dealing with counterfactuals and sequences. First, counterfactuals within AI attempt to explain outcomes, that did not happen. Therefore, there is no evidence data, from which one can infer predictions. Subsequently, this lack of evidence further complicates the evaluation of generated counterfactuals. In other words, you cannot validate the correctness of a theoretical outcome that has never occured. Second, sequential data is not only has a highly variable form, too **CITE** . The sequential nature of the data impedes the tractability of many problems due to the combinatorial explosion of possible sequences which depends on the length of the sequence. Third, process data of requires knowledge of the underlying and often hidden causal structures that produce the data in the first place. However, these structures are often hidden and it is a NP-hard problem to elicit them **CITE Check process discovery literature**. Furthermore, the data generated is seldomly one-dimensional or discrete. Henceforth, each dimension's contribution can vary in dependance of its context, the time and magnitude. Hence, the field in which we can contribute to this open challenge is vast. As a result, we have to restrict the solution space by imposing limitations and assumptions. Therefore, the result of this paper will describe a framework that will only apply to a subset of problems. In the following sections, we will explore these restrictions by describing the most important concepts in chapter 2.

[1] With static data, we refer to data that does not change over a time dimension.

[2] With dynamic data, we refer to data that has time as a major component, which is also inherently sequential

Chapter 2

Background

This chapter will explore the most important concepts for this work. Most of the concepts can have several meanings depending on the varying context in which they are applied. For this purpose, we will provide an intuitive understanding, the ensuing challenges, a concrete definition for this work and lastly and a mathematically formal description. The concepts we will cover encompase **sequence modelling**, process mining and counterfactual explanations.

2.1 Process Mining

2.1.1 A definition for Business Processes

Before elaborating on Process Mining, we have to establsih the meaning of the term *process* in the context of this paper. The term is broadly used in many contexts and therefore has a rich semantic volume. A process generally refers to something that advances and changes over time[2]. Although, legal or biological processes may be valid understandings, we focus on processes *business processes*.

An example is a loan application process in which an applicant may request a loan at a specific point in time. The case is then assessed and reviewed by multiple approvers and ends in a final decision. The loan may be granted or denied. The *business* part may be misleading as these processes are not confined to commercial settings. For instance, a medical business process may cover a patients admission to a hospital, followed by a series of diagnostics and treatments and ending with the recovery or death of a patient. Another example from a human-computer-interaction**[Add to glossary]** perspective would be an order process for an online retail service like Amazon. The buyer might start the process by adding articles to the shopping cart and proceeding with specifying their bank account details. This order process would end with the submission or receival of the order.

All of these examples have a number of common characteristics. They have a clear starting point which is followed by numerous intermediary steps and end in one of the possible sets of outcomes.For this paper we will refer to each step, including start and end points, as Process Event. Each Process Event may contain additional information in the form of event attributes. A collection of these Process Events refer to a Process Instance, if they all relate to a single run of a process. In line with the aforementioned examples, these Process Instances could be understood as a single loan application, a medical case or a buy order. We can also attach Process Instance related information to each instance. Examples would be the applicants race, a patients age or the buyers budget. In its entirety, a business process can be summarised as a *graph* or *flowchart*, in which every node represents an event and each arc the path to another event. This graphical representation is referred to as *process map*. Figure 2.1 shows an example of such a representation.

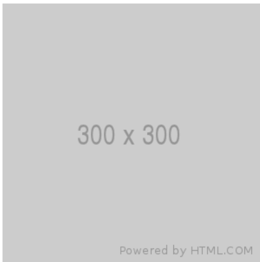

Figure 2.1: This graph shows an example of various process maps.

In conclusion, in this thesis a *business process* refers to

> *A finite series of discrete events with one or more starting points, intermediary steps and end points.*

However, we have to address a number of issues with this definition. First, this definition excludes infinite processes like [XXX] or continuous processes such as [XXX]. There may be valid arguments to include processes with these characteristics, but they are not relevant for this thesis. Second, in each example we deliberately used words that accentuate modality such as *may*, *can* or *would*. It is important to understand that each process anchors its definition in its application context. Hence, what defines a business process is indisputably subjective. For instance, while an online marketplace like Amazon might be interested in the process from the customers first click to the successful shipment, an Amazon vendor might be interested in the delivery process of a product only. Third, the example provided in Figure 2.1 may not relate to the reality of a data generating process. In line with the second point, these examples subjective models of a process. They may or may not be accurate. The *true* process is often unknown to every actor. Therefore, we will distinguish between the *true process model* and a *process model*. The *true process model* is a hypothetical concept whose *true* structure remains unknown.

2.1.2 What is Process Mining

Having established our understanding of a process, we can turn towards *Process Mining*. This young discipline has many connections to other fields that focus on the modeling and analysis of processes such as Continuous Process Improvement (CPI) or Business Process Management (BPM). However, its data-centric approaches originate in Data Mining. The authors van der Aalst et al. describe this field as a discipline "to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's (information) systems"[6]. The discipline revolves around the analysis of Event Logs. An Event Log is a collection of Process Instances. These logs are retrievable from various sources like an Information Systems (ISs) or database. Those logs are often stored in data formats such as Comma Seperated Values (CSV) or eXtensible Event Stream (XES).

2.1.3 The Difficulties of Process Mining

As mentioned in chapter 1, process data modelling and analysis is a challenging task. van der Aalst et al. mentions a number of issues that arise from processes[6].

The first issue arises from the quality of the data set. Process logs are seldomly collected with the primary goal of mining information and hence, often appear to be of subpar quality. The information is often in complete due to a lack of context information, the ommision of logged process steps or wrong levels of granularity.

This issue is exacerbated by the second major issue with process data. Mainly, its complexity. Not only does a process logs complexity arise from the variety of data sources and differing levels of complexity, but also from the datas' characteristics. The data can often be viewed as multivariate sequence with discrete and continuous features and variable length. This characteristic alone creates problems explored in section 2.2. **[Also refer to**

variability in sequence section.] However, the data is also just a *sample* of the process. Hence, it may not reflect the real process in its entirety. In fact, mining techniques need to incorporate the *open world assumption* as the original process may generate unseen Process Instances.

A third issue which contributes to the datasets incompleteness and complexity is a phenomenon called *concept drift*. This phenomenon relates possibility of a change in the *true* process. The change may occur suddenly or gradually and can appear in isolation or periodically. An expression of such a drift may be a sudden inclusion of a new process step or domain changes of certain features. These changes are not uncommon and their likelihood increases with the temporal coverage and level of granularity of the dataset **CITE** .In other words, the more *time* the dataset covers and the higher its detail, the more likely a change might have occured over the time.

All three issues relate to the *representativeness* of the data with regards to the unknown *true* process that generated the data. However, they also represent open challenges that require research on their own. For our purpose, we have to assume that the data is representative and its underlying process is static. These assumptions are widely applied in the body of process mining literature **CITE** .

2.2 Multivariate Time-Series Modelling

2.3 What are Counterfactuals?

Counterfactuals have various definitions. However, their semantic meaning refers to "a conditional whose antecedent is false"[1]. A simpler definition from Starr states, counterfactual modality concerns itself with *what is not, but could or would have been.* Both definitions are related to linguistics and philosophy. Within AI and the mathematical framework various formal definitions can be found within causal inference[3]. However, for this paper, we will use the understanding established within the eXplanable AI (XAI) context[1]. Within XAI, counterfactuals act as a prediction which "describes the smallest change to the feature values that changes the prediction to a predefined output"[4].

[Causal inference definition], **[XAI definition]**. One can understand this as prediction of "what" happens "if" a precursing event would have been different. **[They all share the question of "what if", which is always**

[1]**[XAI is a discipline which seeks to develop techniques to better understand machine learning models.]**

highly subjective with regards to the assumptions made. This will seep into the remainder of the paper.]

[What are counterfactuals?]

Counterfactuals are commonly to relate to questions about the outcomes of situations that

Hence, we want to minimally edit a process to understand the changes necessary to achieve an alternative outcome.

2.4 Related Literature

Rationality - Counterfactual thinking play a crucial role in planning actions

2.5 Formal Definitions

To formalise the log and a process model, we will use the formalisation established by **???** CITE .

2.6 Research Question

2.7 General Approach

2.8 What is Process Mining?

2.9 Challenges of Processing Process Data

Chapter 3

Related Papers

Chapter 4

Methods

4.1 Datasets

4.2 Preprocessing

4.3 Framework

Chapter 5

Results

5.1 Evaluation

Chapter 6

Discussion

Chapter 7

Conclusion

Bibliography

Counterfactual. (n.d.). doi:10.1093/oi/authority.20110803095642948

Definition of PROCESS. (n.d.). Retrieved February 17, 2022, from https://www.merriam-webster.com/dictionary/process

Hitchcock, C. (2020). Causal Models. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020). Metaphysics Research Lab, Stanford University. Retrieved February 10, 2022, from https://plato.stanford.edu/archives/sum2020/entries/causal-models/

Molnar, C. (2019). *Interpretable machine learning. A Guide for Making Black Box Models Explainable.*

Starr, W. (2021). Counterfactuals. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. Retrieved February 9, 2022, from https://plato.stanford.edu/archives/sum2021/entries/counterfactuals/

van der Aalst, W., Adriansyah, A., de Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., … Wynn, M. (2012). Process Mining Manifesto. In F. Daniel, K. Barkaoui, & S. Dustdar (Eds.), *Business Process Management Workshops* (pp. 169–194). doi:10.1007/978-3-642-28108-2_19