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 11, 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

LCM Least Common Multiple. 8

ML Machine Learning. 3

One-Way Delay The time a packet uses through a network from one host to another. 2

OWD One-Way Delay. 2, *Glossary:* One-Way Delay

XAI eXplanable AI. 3, 5

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 trajec-

tory 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 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[2]. 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"[3].

2.1.1 Related Literature

Rationality - Counterfactual thinking play a crucial role in planning actions

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

2.1.2 Definition in the context of this Thesis

2.1.3 Formal Definition

[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 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.

2.1.4 Process Mining

2.2 Research Question

2.3 General Approach

2.4 What is Process Mining?

2.5 Challenges of Processing Process Data

Chapter 3

Related Papers

Chapter 4

Methods

Least Common Multiple (LCM)

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

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/