Name: Océane SOULIER

Name of your Level 1: Morjane SAIDANI

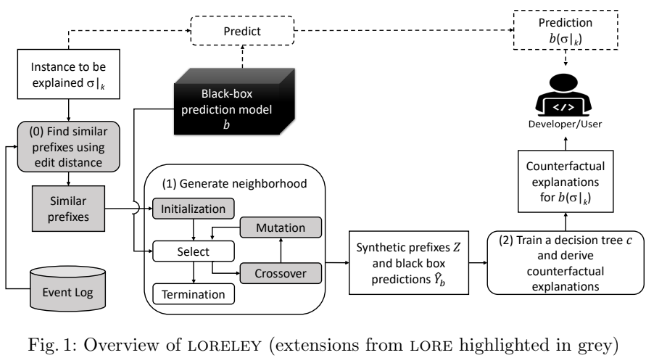
Paper title: Counterfactual Explanations for Predictive Business Process Monitoring

Source: Google scholars

Keywords specific to the paper: business process monitoring, deep learning, counterfactual explanation, explainable AI

Summary of the main contributions (use text paragraphs, tables and if necessary, figures):

The paper proposes a new technique to generate counterfactual explanations for predictive process monitoring models. Called LORELEY, it is actually an extension of LORE, an algorithm a decision tree is built in the neighbourhood of an instance to be explained, to acquire a single rule for the decision and a set of counterfactual rules for the inverse decision[[1]](#footnote-0). The raised issue of the paper is the lack of interpretability of AI models, from the simplest to the more sophisticated. In other words, the lack of interpretability can lead to potential risks especially if black box models (meaning impenetrable systems arriving at conclusions without providing explanations[[2]](#footnote-1)) are used. The technique of LORE is extended by incorporating process constraints during the generation of neighboring examples: LORELEY now generates high-quality explanations (counterfactual explanations) while being accurate from the black-box prediction model.



LORE generates synthetic data for training decision trees. Decision trees have two main advantages: decisions can be easily interpretable and their explanations are either factual or counterfactual. However, both Local Interpretable Model-Agnostic Explanations (LIME) and LORE techniques do not take process constraints into account, thus not the order in which business activities have to be done. It faces the risk of generating unrealistic instances. Another problem is that LORE is limited to binary classification, which excludes some applications that often face multi-class predictions (like Business Process Model).

The approach was evaluated on a public event log dataset using an LSTM prediction model. Researchers have then two questions: how faithful are the explanations generated by LORELEY to the underlying black-box prediction model? and what is their quality?

The generated models reach 97.69% of fidelity, that is a good score. Overall, the variance is less than 0.5%: they reach the conclusion that LORELEY can successfully predict the black-box model’s decisions accurately.

The explanations generated by LORELEY also match the domain knowledge about the process, demonstrating high-quality explanations.

Future experiments may evaluate LORELEY with further benchmark datasets and different black-box predictions models.

AI model used (e.g. Neural network, etc.):

Related works used LSTM to extract attribute importance from each process instance. LSTM is also used in other work and trained to solve a binary classification problem.

Researchers also use LSTM in this very paper, which is a two-layer model (a black-box prediction model).

Supported by a software application? (If yes, provide more details):

eXtreme Gradient Boosting is a software supporting a related work from Sindhgatta et al. where LIME (Local Interpretable Model-Agnostic Explanations) is used to interpret the results of the process.

1. [LORE explanation](https://arxiv.org/pdf/1908.03840.pdf#:~:text=LOcal%20Rule%2Dbased%20Explanations%20) [↑](#footnote-ref-0)
2. [What is black box AI ?](https://www.techtarget.com/whatis/definition/black-box-AI) [↑](#footnote-ref-1)