Master Thesis Proposal

Kefang Ding

August 7, 2018

Title:	Repair Model by Incorporating Negative KPI Outcomes
Author:	Kefang Ding
Supervisor:	Bastian von der Zelst Professor Wil van der Aalst
Degree:	Master of Science

Introduction

Process mining techniques aim to support the analysis of business processes based on event logs. These techniques focus on three tasks: process discovery, conformance checking and model enhancement. Process discovery creates a new model from event log to understand the business process. Conformance checking analyze the deviation of process model with observed behavior after model execution. Enhancement improves the existing model based on event log by extending model with more data perspectives or repairing existing model to better reflect observed behavior.

Between those techniques, repair model stands between process discovery and conformance checking. It improves the existing model by adding loops, subprocess, and removing infrequent transitions at last step[paper from Dirk]. Later in [paper Dees and ANja], focuses are not only on fitness improvement but also on good performance from model. However, the state-of-the-art model repair methods only consider the positive examples which benefits the good performance, but leave out the impact from negative examples. It results in a less precise model, like shown in the next graph.

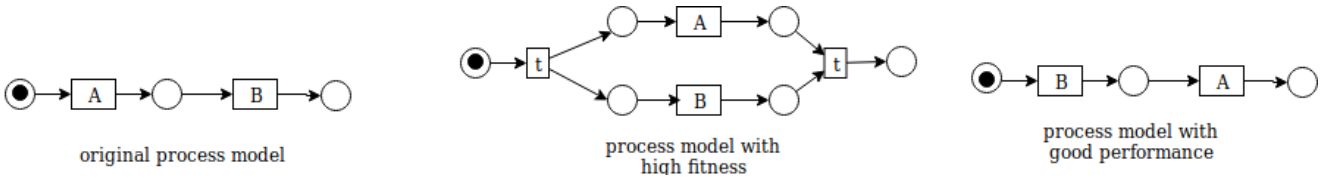

Figure 1: example for current repair model techniques

The original model is defined in the left part of figure above, where A is followed directly by B. During its execution in real life, event log is generated:

$$< A, B >^{30}, < B, A >^{105}$$

After repair model for high fitness, process model in the middle has A and B in parallel relation. With analysis the impact on performance, we have divided event log into this set:

$$Positive\ examples : < A, B >^{10}, < B, A >^{100}$$

$$Negative\ examples : < A, B >^{20}, < B, A >^{5}$$

The wished model in the right part of figure provides a better way to execute business processes, in the way it reinforces positive examples and avoids negative examples. Yet, the current methods can't achieve it.

On the other side, the current methods are not capable to combine multiple KPI at the same time, which leads the balance of good model difficult.

Aim

The inputs for model repair include the following data.

- Existing process model in Petri net
- Event Log in real execution of process
- Key Performance Indicators(KPI)

My aim is to improve repair model techniques by incorporating negative KPI outcomes, and at end provide a better business process with higher fitness, precision and better KPIs.

To evaluate the improvement, I will build a confusion matrix in the table 1. The columns represent model allowed behavior and not allowed behavior; the rows means the positive examples and negative ones in the event log. Good improvement should increase the diagonal ratio in the table, namely the AP and NN, while lowering the percentage of reverse diagonal, which is PN and AN.

Table 1: confusion matrix for evaluation

	Allowed behavior	Not allowed behavior
Positive Examples	AP	PN
Negative Examples	AN	NN

But how to define the real evaluation of it ??

Under the current techniques for repair model, only positive example are considered and the model in the middle is generated. In the view of its confusion matrix in

Table 2: confusion matrix for given example in current approach

	Allowed behavior	Not allowed behavior
Positive Examples	110	0
Negative Examples	25	0

Table 3: confusion matrix for given example in my approach

	Allowed behavior	Not allowed behavior
Positive Examples	100	10
Negative Examples	5	20

Table 2, it has less precision by allowing all behavior in model. The evaluation result is ???

In my approach, confusion matrix is built to use both negative and positive examples. The evaluation result is ??

Also, given multiple KPIs, it should be able to combine them together and provide a balanced model with good performance. But this suggestion should be based on the hypothesis:

Method

Given input of one existing process model and event log, my approach produces a repaired model with good KPIs outcomes. As shown in the following graph, the first phase accepts the existing process and event log as input, then repair model w.r.t. fitness by using framework in [paper dirk]. Next, the original repaired model is passed to the second phase, repair model w.r.t. KPI, where model is modified to reinforce the good examples while avoiding the negative examples.

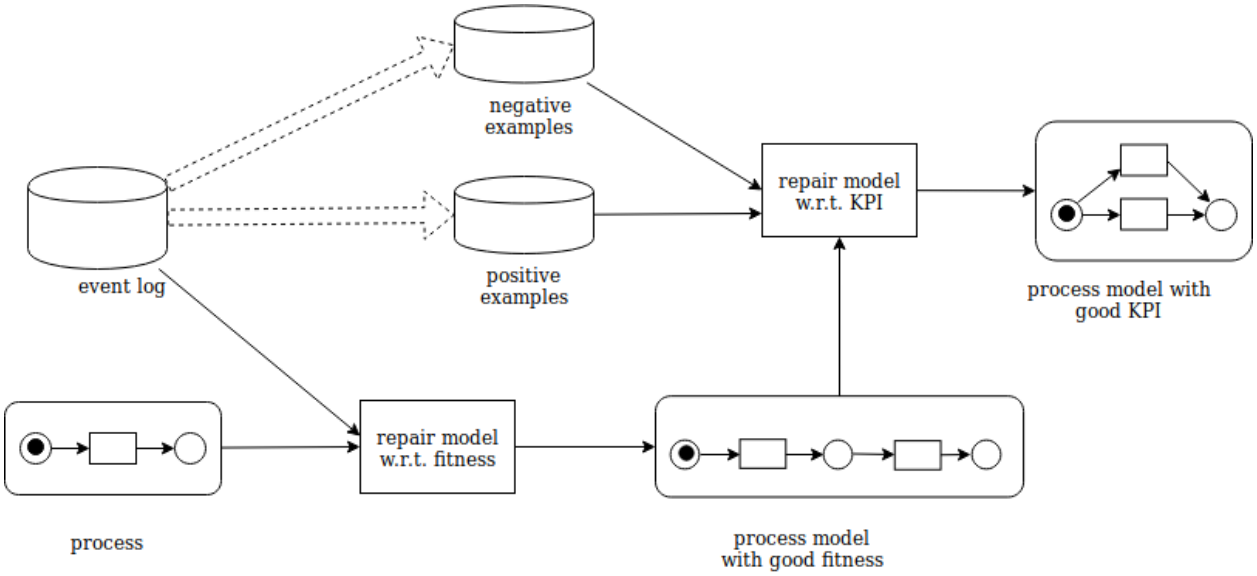

Figure 2: my approach to repair model

The process repair model w.r.t. KPI is motivated from Artificial Generated Negative Events(AGNEs) in [paper AGNEs]. (AGNEs) is used to discover process model. At first, it uses TILDE to learn rules from both positive and artificial generated negative traces. Later it transforms the classification rules to Petri net.

Now, I combine it into my approach to deal with positive and negative impact on model. It works in such way:

- change Petri net into rules patterns.
- train rules based on the positive and negative examples and get modified rules
- transform the modified rules into Petri net

To make sure this method above feasible, several hypotheses are needed:

1. Petri net patterns and the rules generated from TILDE are bijective, such that the reverse transformation is feasible.
2. It considers the impact from positive and negative examples but still keep the similarity to the original model.

The next primary work should test on the hypotheses.

Another improvement of my approach is to combine multiple KPIs, and quantify the goodness of event trace, but based on the hypothesis.

- By implying the weighted impact on model repair, we could get better result than by only using the classification method.

Software and Hardware Requirements

The platform is ProM6, a open source project for process mining. To implement my approach, those plugins are in demand.

- repair w.r.t. fitness

  Repair Model in Uma package

- repair w.r.t. KPI

  Process Discovery in AGNEs

  Decision Tree

  Associated Rules

  Regression function

Timeline

My original plan for my master thesis is in the table listed below.

Table 4: My time line for my work

time	taks
August	Test the proposed hypotheses by creating small demo
September	Register master thesis and design the architecture
October	Coding and debug
November	Test and Write Thesis
December	Write thesis and representation

References

[1] Fahland, Dirk, and Wil MP van der Aalst. "Model repairaligning process models to reality." Information Systems 47 (2015): 220-243.

[2] Dees, Marcus, Massimiliano de Leoni, and Felix Mannhardt. "Enhancing process models to improve business performance: a methodology and case studies." OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Springer, Cham, 2017.

[3] Adriansyah, Arya, et al. "Alignment based precision checking." International Conference on Business Process Management. Springer, Berlin, Heidelberg, 2012.

[4] vanden Broucke, Seppe KLM, et al. "Determining process model precision and generalization with weighted artificial negative events." IEEE Transactions on Knowledge & Data Engineering 1 (2013): 1.

[5] Buijs, Joos CAM, et al. "Improving business process models using observed behavior." International Symposium on Data-Driven Process Discovery and Analysis. Springer, Berlin, Heidelberg, 2012.