Project Initiation
Label Refinement by Behavioral Similarity

Document owner:
Bianka
Christopher Beine
Nicole
Juan

Last edited: April 17, 2019

1 Overview

Many processes involve carrying out an action multiple times. An example for this would be an online shop in which you first have to pay a registration fee before ordering an item and paying it. This process contains the event "payment" twice, but in different contexts, so that the payments are actually two different tasks. In the context of analysing processes, the event logs usually only contain the event names, so that the "payment" actions would be treated as the same task and loops would be induced in the resulting models. However, these loops do not match the actual process, which is the issue this project addresses.

These imprecise logs should be refined based on the structural contexts of the events. We want to refine the logs without any filtering. Moreover, we want to allow an interactive change of the thresholds used to refine the labels since this can differ for every log and we have no knowledge of the correctness of the refined log in general.

By designing an interface that allows the users to upload an event log, to set the thresholds and to download the modified event log, carrying out this project will save the data analysts a lot of time which is needed to refine the log and will make their results more accurate. Using this approach, the process logs can be refined to reach a higher precision in the subsequent analysis of up to 89 %, which highly increases the quality of the process discovery results. These better results can lead to more optimized processes and that way reduce the company's expenses while increasing its efficiency.

2 Business Case

2.1 Intial situation

2.2 Scope

During the project, we will create both code and documentations. Thus, the scope will be divided into these two aspects:
Documentation:

- develop and describe the design of the interface we will use and picture how users can set the thresholds in the interface

- design the algorithm structure by stating the usage of classes as well as the inputs and outputs for the functions that are used

Implementation:

- set up a Web Service based on Python that uses the label refinement algorithm proposed by Xixi Lu, et al.

- create a user interface that allows the users to upload the original event log, set thresholds and imprecise label scope and finally download the refined event log

2.3 Key Benefits

2.4 Assumptions

In this project we will assume that an event log is given by the user, i.e., data that contains at least the attributes "id", "time stamp" and "activity name". Moreover, we will assume that these event logs are given in the standard XES format.

3 Feasibility study

3.1 Theoretical point of view

3.2 Technical point of view

3.3 Risks and migiations

3.3.1 Project management risks

3.3.2 Technical risks

4 Project Plan

4.1 Milestones

The project starts on the 09/04/2019 and ends on the 08/07/2019 and is divided into nine milestones.

Table 1: Overview Milestones

Milestone	Description	Deadline
Project Initiation document	The Project Initiation Document provides all of the key information required to start and run the project. This includes the project description, business case, feasibility study and a project team presentation.	19/04/2019
Requirements Specification document	The Requirements Specification document contains functional and none functional requirements such as a set of use cases to describe the system interactions.	29/04/2019
Design Analysis and dummy P.o.C.	The final document is a description about the planned software system architectural background and a proof of concept visualizing the main UI components.	13/05/2019
Sprint 1 code and documentation	TODO: sprint description depending on GANTT chart	24/05/2019
Sprint 2 code and documentation	TODO: sprint description depending on GANTT chart	07/06/2019
Sprint 3 code and documentation	TODO: sprint description depending on GANTT chart	21/06/2019
Testing, assessment and deployment	The application is checked for accuracy and should be aviable for use.	01/07/2019
Final report on the project	The final report provides an overview about the project course and the result.	08/07/2019

4.2 Deliverables

4.3 Timetable

5 Project Team

5.1 Competences

5.1.1 Nicole Ventsch

I am a Master student studying Mathematics and Data Science in parallel. Moreover, I work as a student assistant in the field of Data Analytics / Business Intelligence. Due to my background in mathematics, I have a good understanding of theoretical foundations. Moreover, I am very interested in Data Science and already took many courses in that area. Since I took the course "Introduction to Data Science", I also worked with Python before, so that I should be able to implement an algorithm in Python.
Though I have a strong theoretical background, I never worked on user interfaces or with web services, so that this aspect of the project could be challenging for me.

5.2 Roles

6 Project office

6.1 Description