Introduction

You work in a hospital that has agreed to provide sepsis cases data to a private company for analysis. The hospital has recorded sepsis cases as a set of activities that were performed for a specific patient. Personal information of the patients has already been removed from the dataset. However, the management board of the hospital has decided that this alone is not enough to protect the privacy of their patients.

It was agreed that the hospital will create an artificial dataset of sepsis cases that is representative of the actual real-life treatment process. You have been given the task of creating that dataset.

To achieve this you have been given access to the anonymized dataset of sepsis cases (separated into training and test set) and a process mining tool called RuM.

Task I

We will begin by discovering a model of the sepsis cases treatment process. We will use the training set as input for discovery.

1. Discover an initial model from "1_Training_Set_Sepsis.xes" using the following parameters
    a. Method: "Declare Miner"
    b. Templates: Existence, Precedence; Response; Chain Precedence; Chain Response
    c. Minimum constraint support: 80%
    d. Pruning type: All Reductions
    e. Vacuity detection: on
2. Filter out all activities with support less than 90%
3. Navigate the different views available to inspect the model
4. Play around with the different labels that can be attached to constraints in the "Declare" view
5. Export the model as a Declare (*.decl) file

Task II

Before using the model to generate the artificial dataset we want to make sure that the model was discovered correctly. We can do this by checking if the model is conformant with the data contained in the test set.

1. Check if the model from the previous task conforms to the "2_Test_Set_Sepsis.xes" using the Declare Analyzer.
2. Navigate the output of the Analyzer. What information can you get from it?
3. Check if the model from the previous task conforms to the "2_Test_Set_Sepsis.xes" using the Declare Replayer using the following parameters:
   a. Checking method: Declare replayer
   b. Default costs
      i. Insertion cost: 5.0
      ii. Deletion cost: 10.0
   c. Specific costs for activity ER Registration
      i. Insertion cost 2.0
      ii. Deletion cost 15.0
4. Try to play a bit by switching between original trace views and alignment views
5. Try to find some activities that are have been inserted in some of the traces to align the model with the test set
6. Try to find the constraint that causes the highest number of deletions

Task III

It seems that the discovered model needs some modifications before it can be used to generate the artificial dataset. We will do these modifications manually using the MP-Declare editor.

1. Edit the model from Task I
   a. Remove the constraint that caused the most deletions
   b. Add the following attribute to activity ER Triage
      i. Name: Treatment
      ii. Type: Enumeration
      iii. Possible values: T1, T2
   c. Change constraint Chain Response [ER Triage, ER Sepsis Triage]
      i. Template to: Response
      ii. Activation condition: A.Treatment is T1
2. Use this model for the log generation task

Task IV

Now that the model of the process is completed we can use it for generating the artificial dataset.

1. Generate a log based on the model from Task III with the following parameters
   a. Generation Method: AlloyLogGenerator
   b. Minimum events per trace: 5
   c. Maximum events per trace: 10
   d. Number of traces: 20
   e. Negative traces: 20%
   f. Non-Vacuous traces: 50%
2. Check if activity ER_Triage was generated with the attribute Treatment and that some of the constraints in the model are actually satisfied in the generated traces
3. Export the generated log

This concludes the tasks for this evaluation. Next we will continue with a small interview discussing your impression of RuM - what has been done well and what could be improved?