This zip file contains a collection of 800 synthesized models and their corresponding event logs. They are used in the experiments for the paper "Handling Duplicated Tasks in Process Discovery by Refining Event Labels". The results of the experiments are summarized in the four .xlsx files.

How to make the link between .xlsx and the models and logs?

In each .xlsx file, in the "allN" sheet, for each record, you find a letter at the beginning and a date at the end, which indicate the file and the model. For example, in "noImprInLoop_default_OD.xlsx", the first entry has "A" and "feb16-1625", as shown in Figure 1. This means that it is obtained by using the model in the file "noImprInLoop_default_OD\feb16-1625\models\A_ModelGenfeb16-1625.ptml" (see Figure 2) and the logs in "noImprInLoop_default_OD\feb16-1625\logs\A_..." (see Figure 3).

The .xlsx file also indicates the duplicated task: the tasks in the “Source” column have the same label as the task in the “Target” column. For example, in Figure 1, model A has four task nodes “C”, “B”, “E” and “D” that all have the same activity label “D”.

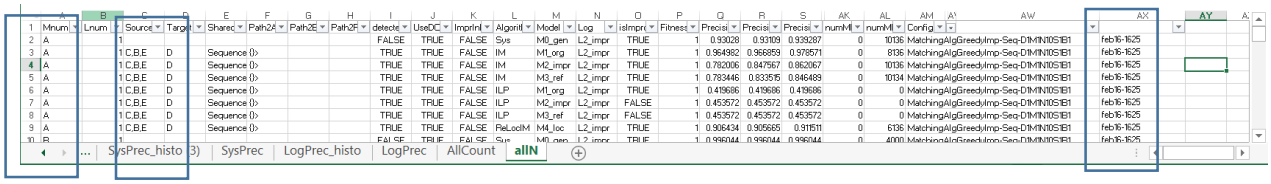

Figure 1: A snapshot of noImprInLoop_default_OD.xlsx

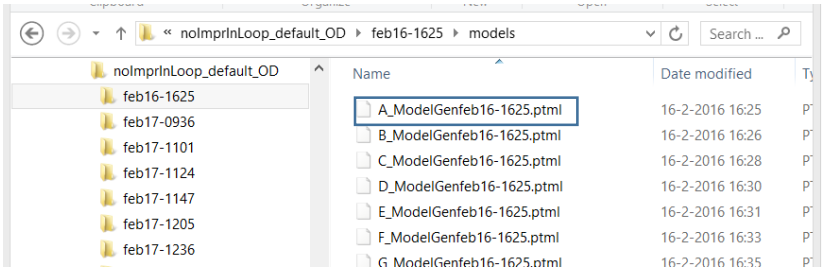

Figure 2: Models used in the experiment “noImprInLoop_default_OD”.

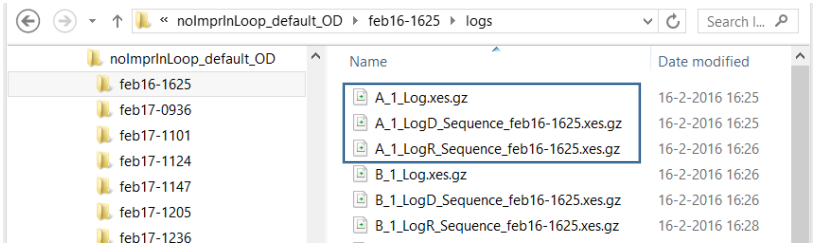

Figure 3: There are three logs for each model: (1) the original log named “... Log.xes.gz”; (2) the log with imprecise labels, name “... LogD...xes.gz”; (3) the log with refined labels, named “...LogR...xes.gz”, which is obtained using the approach proposed in the paper.