

Event Log Sampling for Predictive Monitoring

---- User Manual

Prepared by Malek Alhelwany, Taekeun Jeong, Xiaoyan Jin

<https://github.com/Malekhy/ws2122-lspm>

Overview

The event log sampling for Predictive Monitoring (LSPM) proposes an instance selection procedure that allows sampling training process instances for prediction models.

We implement 3 different sampling algorithms as a training service in the form of web services, combine it with existing predictive models, and produce prediction results.

On our web service page, the user will be able to import (csv/xes) and export (xes) their files. With choosing the suitable sampling method as they want, they will get the prediction results with reliable levels of prediction accuracy instantly.



Getting started!



Steps (using virtual environment)

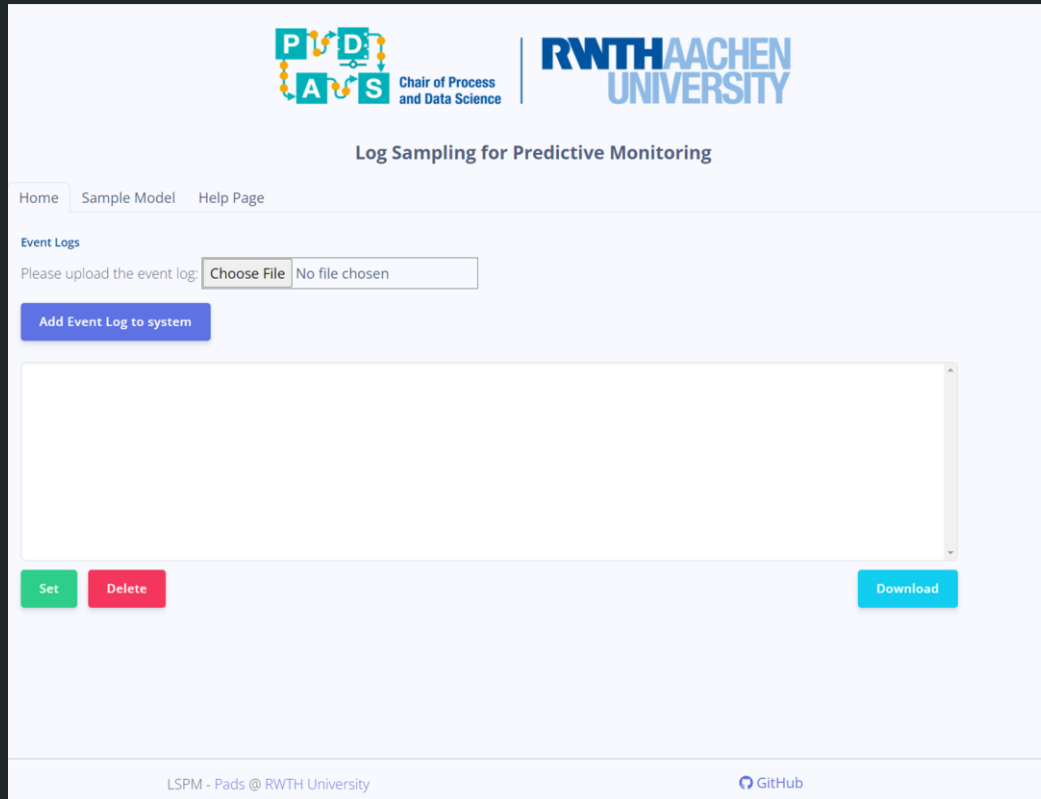
- Clone the project from <https://github.com/Malekhy/ws2122-lspm>
- Open the command terminal 'cmd'
- Move into the directory folder of the project:
[\(project-path\)\ws2122-lspm](#)
- Move into project environment:
 - For Windows users: `.\Scripts\activate`
 - For Mac users: `source Scripts\activate`
- Install requirements: `pip install -r requirements.txt`
- Run the web app: `python manage.py runserver`
- Open the browser and hit the URL: <http://localhost:8000/>



Steps (using docker image)

- Follow the instruction from here to install Docker if you don't have:
<https://docs.docker.com/desktop/>
(Please make sure you run Docker Desktop before start)
- Clone the project from <https://github.com/Malekhy/ws2122-lspm>
- Open the command terminal 'cmd'
- Move into the directory folder of the project:
[\(project-path\)\](#)
- Build the docker image using this command:
`docker build --tag lspm .`
- Run the docker container using this command:
`docker-compose up`
- Open the browser and hit the URL: <http://localhost:8000/>

Then you will get... 👍



The screenshot displays the web application interface for Log Sampling for Predictive Monitoring (LSPM). At the top, the logo for the Chair of Process and Data Science is shown alongside the RWTH Aachen University logo. The main title of the application is "Log Sampling for Predictive Monitoring". Below this, there is a navigation bar with links to "Home", "Sample Model", and "Help Page". The "Event Logs" section is active, showing a file upload interface. It includes a text prompt "Please upload the event log:", a "Choose File" button, and a status indicator "No file chosen". A large blue button labeled "Add Event Log to system" is positioned below the upload area. A large, empty text area for the event log content is visible. At the bottom of this section, there are three buttons: a green "Set" button, a red "Delete" button, and a blue "Download" button. The footer of the application contains the text "LSPM - Pads @ RWTH University" and a GitHub logo.


Chair of Process and Data Science | RWTH AACHEN UNIVERSITY

Log Sampling for Predictive Monitoring

Home Sample Model Help Page

Event Logs

Please upload the event log: No file chosen

LSPM - Pads @ RWTH University  GitHub



Instruction

1. Click here to choose file (.xes,.csv)

2. Click here to add the file into the working directory

*Click here to delete the chosen file

3. Click 'Set' to proceed to the next step

*Click here to download the chosen file in the list

The screenshot shows the 'Log Sampling for Predictive Monitoring' web application. At the top, there are logos for 'PADS Chair of Process and Data Science' and 'RWTH AACHEN UNIVERSITY'. Below the logos, the title 'Log Sampling for Predictive Monitoring' is displayed. A navigation bar contains links for 'Home', 'Sample Model', and 'Help Page'. The main section is titled 'Event Logs' and contains the text 'Please upload the event log'. A 'Choose File' button is circled in red, with an arrow pointing to it from the first instruction. Below this, an 'Add Event Log to system' button is circled in red, with an arrow pointing to it from the second instruction. A list of event logs is shown, with 'roadtraffic100traces.csv' highlighted in blue. An arrow points from the third instruction to the 'Delete' button, which is circled in red. Another arrow points from the fourth instruction to the 'Download' button, which is also circled in red. The footer contains the text 'LSPM - Pads @ RWTH University' and a 'GitHub' link.



Instruction

4. Click here to adjust the preferred characteristics

5. Click here to choose the sampling method

6. Click here to start sampling

Add Event Log to system

CPMExport.csv
ItalianHelpdeskFinal - Concept.csv
ItalianHelpdeskFinal - Concept.xes
ItalianHelpdeskFinal.csv
repairExample.xes
roadtraffic100traces.csv
roadtraffic50traces.xes

Set

Delete

Download

roadtraffic100traces.csv is used as input.

Please choose event log characteristics

Please choose the Case ID:

case:concept.name

Please choose the Activity Name:

concept.name

Please choose the Start Time Stamp:

time:timestamp

Please choose the sampling method

☒ Unique Selection

☐ Logarithmic distribution

☐ Division

Start Sampling


LSPM - Pads @ RWTH University

GitHub



Instruction

Then you will get an alert page showing a notification with your choices.
Click 'OK' to continue.



127.0.0.1:8000 says
Event log sampled successfully using following criterias:
Selected method is : Unique-Selection
Case ID is : case:conceptname
Activity Name is : conceptname
Time Stamp is : time:timestamp

OK

Home Sample Model Help Page

Sample model window

Original eventlog and sampled one some statistics [See all](#)


ATTRIBUTE	ORIGINAL	SAMPLED	BOUNCE RATE
Event log name	roadtraffic100traces.csv	Calculating the results..	-
Number of cases	100	Calculating the results..	
Number of events	390	Calculating the results..	
Average case duration	312d 11h 15m 36s	Calculating the results..	-
Number of variants	10	Calculating the results..	
Total case duration	31246d 22h 0m 0s	Calculating the results..	-

Please navigate to Home tab to choose another sampling method or change event log

LSPM - Pads @ RWTH University [GitHub](#)



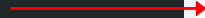
Instruction


results page 




In the Bounce Rate column we can see the reduced percentage after sampling process.

e.g. In this row we can see the significant changes after implementing our sampling methods.





Chair of Process
and Data Science



Home Sample Model Help Page

Original eventlog and sampled one some statistics [See all](#)

ATTRIBUTE	ORIGINAL	SAMPLED	BOUNCE RATE
Event log name	roadtraffic100traces.csv	"LSPM - Unique-Selection - roadtraffic100traces.csv-220249.csv"	-
Number of cases	100	8	-1150.00 %
Number of events	390	34	-1047.06 %
Average case duration	312d 11h 15m 36s	316d 21h 0m 0s	-
Number of variants	10	8	-25.00 %
Total case duration	31246d 22h 0m 0s	2535d 0h 0m 0s	-

LSPM - Pads @ RWTH University [GitHub](#)

Happy process mining :)
