Type: Master thesis

Title: Model: Analysis of token flows in processes in order to predict interesting performance patterns

Description:

In this thesis, we assume that an event log and a Petri net model are given. Replaying the event log on the model provides useful information about how the tokens flow through the process. Assume that we first partition the process model into some zones (each place belongs to one zone, but a zone can have more places) and a predefined time window (e.g., 1 hour, 1 day, or 1 week) is given. It is easily possible to extract and store a wide range of descriptive features for each zone by replaying the event log. Some examples of descriptive features for each zone are:

- Number of cases entering the zone
- Number of cases leaving the zone
- Number of cases staying in the zone
- Time-related features (average/standard deviation of waiting times) in the zone

Many other features can be defined as descriptive features. Our final goal is to use these descriptive features to predict interesting performance patterns. The idea is originated from the performance spectrum tool that is recently proposed to observe the token flows better. Please refer to the pointers for more information. Some performance patterns such as overtaking, batching and queuing might not be captured in the aggregated level but using the performance spectrum, we can analyze them.

We expect the algorithm to predict the target features for a specific zone and time period in future as an output. We divide the event log into test/training event log and calculate standard supervised learning evaluation metrics to evaluate the final predictions.

In summary, the selected candidate for this thesis is supposed to:

1. Define a wide range of meaningful descriptive features. Later you should find the set of features that could predict the target features better.
2. Get to know interesting performance patterns and work on automatically detecting them based on token flows.
3. Implement the ideas and extract tables that illustrate the descriptive and target features for a specific zone and different time periods
4. Search for causality relations between the descriptive and target features and perform interpretable classification techniques to classify the instances

5. Evaluate the techniques using standard evaluation metrics and perform many experiments using synthetic and public real-life event logs

Prerequisites:

Python programming (experience in PM4Py is a plus)

Process mining knowledge (especially conformance checking and performance analysis techniques)

Data science knowledge (supervised learning algorithms and evaluation techniques)

Pointers:

- Please take a look at: https://youtu.be/-ysPFFZsWFY and https://youtu.be/MkBQ_JXyiVs
- van der Aalst, W.M., Tacke Genannt Unterberg, D., Denisov, V. and Fahland, D., 2020, June. Visualizing token flows using interactive performance spectra. In International Conference on Applications and Theory of Petri Nets and Concurrency (pp. 369-380). Springer, Cham.
- Denisov, V., Fahland, D. and van der Aalst, W.M., 2018, September. Unbiased, fine-grained description of processes performance from event data. In International Conference on Business Process Management (pp. 139-157). Springer, Cham.
- Pourbafrani, M. and van der Aalst, W.M., 2021, June. Extracting process features from event logs to learn coarse-grained simulation models. In International Conference on Advanced Information Systems Engineering (pp. 125-140). Springer, Cham.

Supervisor:
Prof.dr.ir. Wil van der Aalst

Advisor:
Ali Norouzifar

For more Information:
Send an e-mail to ali.norouzifar@pads.rwth-aachen.de. Make sure to include detailed information about your background and scores for completed courses.

Chair of Process and Data Science
Ahornstr. 55 (Eingang Mies-van-der-Rohe-Str.), Erweiterungsbau E2
52074 Aachen
Phone: +49 241 80 21 901
http://www.pads.rwth-aachen.de