We develop a framework for predicting how an ongoing instance of a business process will unfold up to its completion. Specifically, for a given point during the process case execution, we extract all possible completion scenarios how this case may unfold up to its completion and estimate probabilities of various scenarios. In this project we report only the most likely scenario.

A transition system is a triplet \( TS = (S, A, T) \), where \( S \) is the set of states, \( A \subseteq S \times A \times S \) is the set of activities and \( T \subseteq S \times A \times S \) is the set of transitions. \( S_{\text{init}} \subseteq S \) is the set of initial states, and \( S_{\text{fin}} \subseteq S \) is the set of final (accepting) states.

- We build transition system from the adjacency matrix (based on the trainin set)
- For each test sample, we annotate edges based on the transition probabilities learned from the training samples

Damerau-Levenshtein distance is a distance (string metric) between two strings, i.e., finite sequence of symbols, given by counting the minimum number of operations needed to transform one string into the other, where an operation is defined as an insertion, deletion, or substitution of a single character, or a transposition of two adjacent characters.

\[
f_{DL}^{\text{Damerau-Levenshtein}}(x_1, x_2) = 1 - \frac{f_{S}^{\text{S}}(x_1, x_2)}{\max(|x_1|, |x_2|)}
\]

- KNN outperforms ATS-based approach, but the latter is more stable
- Both are better than random walk
- Performance plateaus with longer suffixes

The project was supervised by Fabrizio Maggi. The studies of Andrii Rozumnyi and Dmytro Chasovskyi are supported by the Estonian Foreign Ministry’s Development Cooperation and Humanitarian Aid funds.

References

Repo: https://github.com/JaakTree/predict_sequence_completion