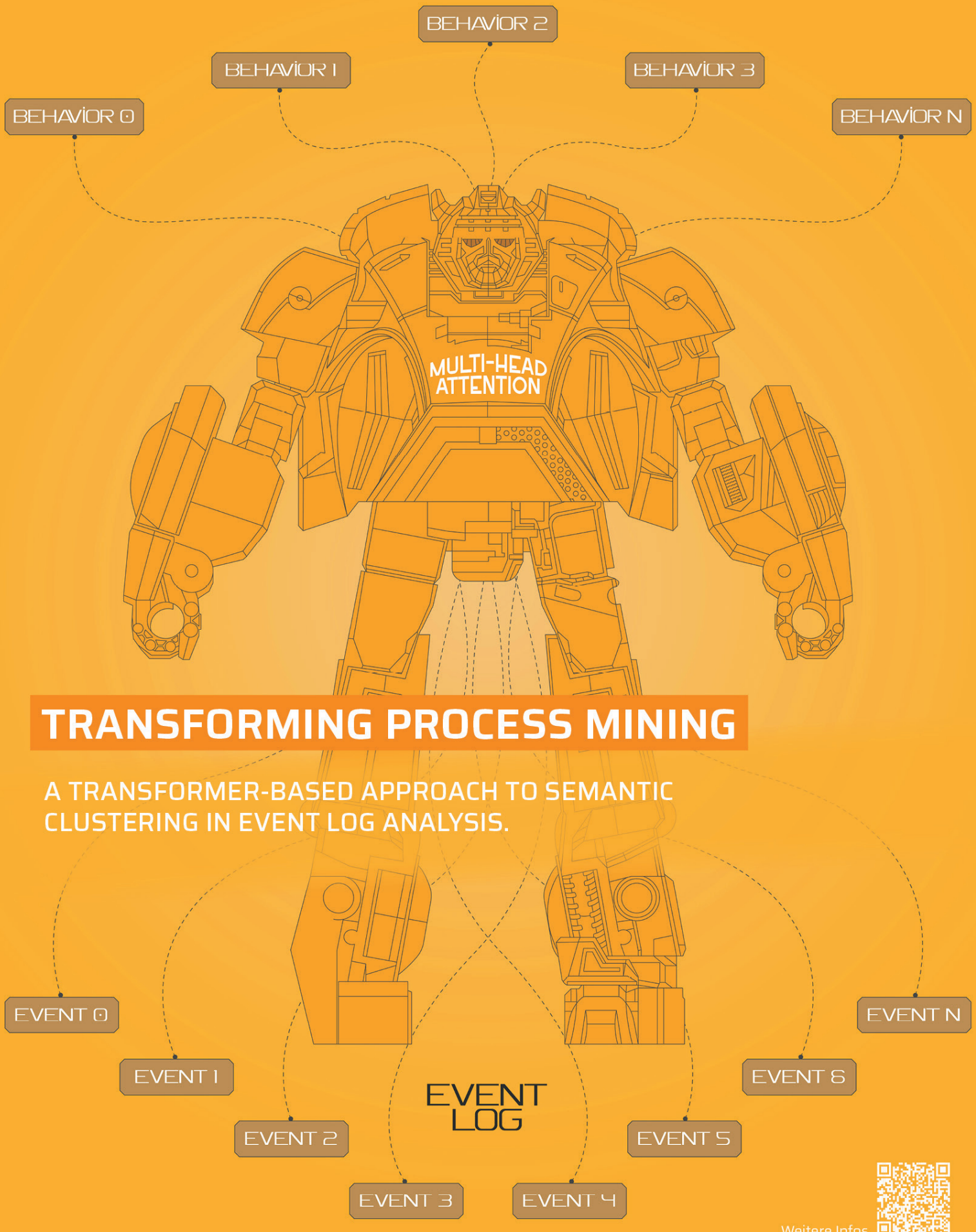
TECHNISCHE HOCHSCHULE DEGGENDORF



TRANSFORMING PROCESS MINING

A TRANSFORMER-BASED APPROACH TO SEMANTIC CLUSTERING IN EVENT LOG ANALYSIS.

Weitere Infos

Abstract

Projekttitel/ Project title:

Transforming Process Mining: A Transformer-Based Approach to Semantic Clustering in Event Log Analysis

Kurztitel/ Short title:

Semantic Clustering of Transformed Process Mining

Einleitung/ Introduction:

Process mining is integral for business optimization, focusing on analyzing digital footprints through event log inspection. Event log analysis identifies patterns, irregularities, and potential bottlenecks. Transformer models, known for success in tasks like context comprehension, present a promising avenue for enhancing process understanding.

Ziel/ Aim:

This research explores the potential of transformer models and clustering algorithms to enhance event log analysis in process mining. The primary objective is to foster the creation of clusters that are semantically meaningful, providing a refined representation of distinct behavioral patterns within the logs.

Methode/ Method:

Adopting a two-step approach, the study begins with training a transformer model as a masked language model. This process equips the encoder to learn contextual and semantic relationships. The subsequent step involves utilizing the trained model to transform event logs into high-dimensional vectors. These vectors are then subjected to a clustering algorithm to group meaningful clusters. Careful consideration is given to aspects like tokenizer selection, model configuration, and the introduction of novel evaluation measures.

The choice of tokenizer is crucial, balancing model compatibility, text granularity, and domain specificity. Efficiency impacts computational costs and sequence lengths, aligning with practical constraints. Configuring the transformer model, including the number of encoder layers and attention heads, is guided by dataset size, task complexity, and computational resources. Initial configurations draw inspiration from established architectures like BERT or RoBERTa, refined through a hyperparameter search based on validation performance.

Clustering high-dimensional vectors generated by transformers for structured data involves dimensionality reduction techniques like t-SNE or PCA. Alternatively, density-based algorithms like DBSCAN or HDBSCAN, accommodating varying densities without pre-specified cluster numbers, are explored for clustering.

Ergebnis/ Result:

The anticipated outcomes include the formation of cohesive case clusters through the application of transformer models and clustering algorithms in event log analysis. These clusters are expected to exhibit meaningful semantic coherence, offering a representation of specific behavioral patterns within the logs. The research aims to evaluate the model's versatility through fine-tuning tasks, such as anomaly detection, employing metrics like AUC-ROC and AUC-PR. Novel evaluation metrics will be incorporated to assess the model's generalization capacity and its performance in real-world scenarios, providing a comprehensive framework for analysis.

Projektbeteiligte/ Project participants:

TH Deggendorf, dab: Daten - Analysen & Beratung GmbH, Otto-Friedrich-Universität Bamberg.

Projektpartner/ Project partners:

- TH Deggendorf: Prof. Andreas Fischer, Zineddine Bettouche, Johannes Reisinger, Nicki Bodenschatz
- Uni Bamberg: Prof. Stephan Scheele, Christian Dormagen
- dab:GmbH: Marco Kretschmann, Martin Riedl, Ledia Leka

Gefördert durch/ Funded by:

Project KIGA: KI-basierte Geschäftsprozess-Analyse

Logos/ Logos:

insgesamt maximal 450 Wörter/ limit of 450 words in total