Discovering generative models from event logs: data-driven simulation vs deep learning

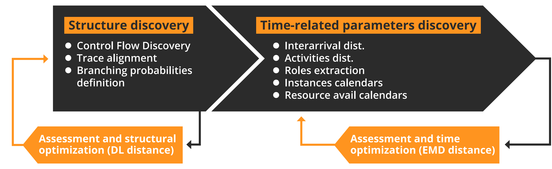
Manuel Camargo Corresp., 1, 2, Marlon Dumas 1, Oscar González-Rojas 2

1 Institute of Computer Science, University of Tartu, Tartu, Tartu, Estonia
2 Computer and Systems Engineering Department, Universidad de Los Andes, Bogotá, Bogotá, Colombia

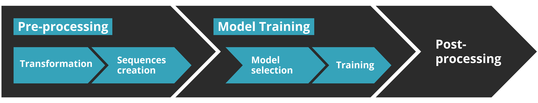
BACKGROUND

A generative model is a statistical model capable of generating new data instances from previously observed ones. In business processes, a generative model creates new execution traces from a set of historical traces, also known as an event log. **Two types of generative business process models have been developed in previous work: data-driven simulation (DDS) models and deep learning (DL) models**.

DDS MODEL



DL MODEL



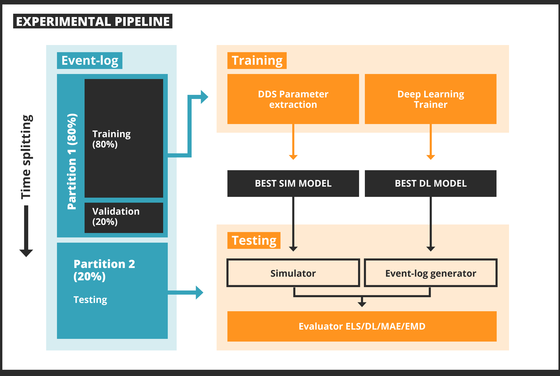
EXPERIMENTAL SETUP

Until now, these two approaches have evolved independently, and their relative performance has not been studied. **This article fills this gap by empirically comparing a data-driven simulation approach with multiple deep learning approaches for building generative business process models.**

EVENT LOGS DESCRIPTION

Size	Type of source	Event log	Num. traces	Num. events
LARGE	REAL	POC	70512	415261
LARGE	REAL	BPI17W	30276	240854
LARGE	REAL	BPI12W	8616	59302
LARGE	REAL	CALL	3885	7548
LARGE	SYNTHETIC	CVS	10000	103906
LARGE	SYNTHETIC	CFM	2000	44373
SMALL	REAL	INS	1182	23141
SMALL	REAL	ACR	954	4962
SMALL	REAL	MP	225	4503
SMALL	SYNTHETIC	CFS	1000	21221
SMALL	SYNTHETIC	P2P	608	9199

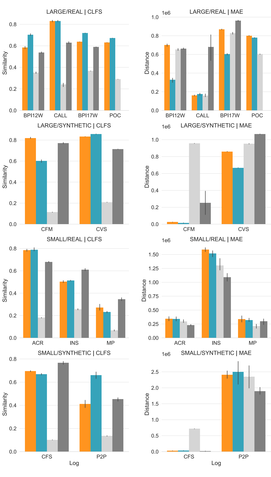
EXPERIMENTAL PIPELINE



RESULTS

The results suggest that DDS models are suitable for capturing the sequence of activities of a process.

On the other hand, **DL models outperform DDS models when predicting the activity timings**, specifically the waiting time between activities. This observation can be explained because DDS approaches do not take into account the multitude of sources of waiting times that may arise in practice, such as waiting times caused by batching, prioritization of some cases relative to others, resources being involved in other business processes, or fatigue effects.



KEY

- GRU
- LSTM
- LSTM (GAN)
- SIMOD

CONCLUSION

A natural direction for future work is to **extend existing DDS approaches to take into account a wider range of mechanisms affecting waiting times to increase their temporal accuracy**.

An alternative approach would be to **combine DDS and DL approaches to take advantage of their relative strengths**. In such a hybrid approach, the DDS model would capture the control-flow perspective, while the DL model would capture the temporal dynamics, particularly waiting times.

UNIVERSITY OF TARTU
Universidad de los Andes Colombia

This research is funded by the European Research Council.