REMEMBER: Uncovering Complex Temporal Dependencies in Process Logs with Variable Length Markov Chains

Emilio Incerto[1] and Andrea Vandin[2,3]

[1] IMT School for Advanced Studies, Lucca
emilio.incerto@imtlucca.it
[2] L'EMbeDS and Institute of Economics,
Sant'Anna School for Advanced Studies
a.vandin@santannapisa.it
[3] DTU Technical University of Denmark

Abstract. We present REMEMBER, a Python tool for the extraction[AV: mining?learning?] of Variable-Length Markov Chains from event logs. The tool can uncover complex temporal dependencies between activities. [AV: Inspired by techniques for indexing large bodies of text,[non mi ricordo cosa si diceva]], REMEMBER constructs a probabilistic suffix tree to efficiently encode path dependencies of variable length by using variable-length memory. Delivered as a set of user-friendly Python notebooks, also available online on Google COLAB without requiring any installation, the tool facilitates the computation of trace likelihoods for advanced Process Mining tasks like Stochastic Conformance Checking and Anomaly Detection. As shown in recent literature, the techniques implemented in REMEMBER outperform the state-of-the-art in Stochastic Conformance Checking. REMEMBER offers a practical solution for deriving precise, memory-aware stochastic models from event logs.

Keywords: Probabilistic Process Minining,· Stochastic conformance checking, · Higher-order Markov chains

1 Introduction

Stochastic process mining [**?**] has become increasingly significant within the Business Process Management (BPM) community, as it enables the analysis and understanding of processes under uncertainty by extracting probabilistic models from event logs. These models are crucial for tasks such as stochastic conformance checking [**?**], anomaly detection [**?**], and predictive monitoring [**?**], as they provide insights into the likelihood of various process executions.

A central challenge in this domain is accurately modeling complex path dependencies between activities within process traces. Traditional fixed-order Markov models often fall short in capturing the variable-length dependencies observed in real-world processes, leading to oversimplified representations that may not reflect the true stochastic nature of the underlying processes.

To address this challenge, we present REMEMBER, a Python-based tool designed to extract Variable-Length Markov Chains (VLMCs) from event logs. REMEMBER constructs probabilistic suffix trees that efficiently encode path dependencies of variable lengths, allowing for a more nuanced representation of process behaviors. The tool facilitates the computation of trace likelihoods, supporting advanced process mining tasks such as stochastic conformance checking and anomaly detection. Delivered as a set of user-friendly Python notebooks, REMEMBER is accessible via Google Colab, requiring no installation.

The techniques implemented in REMEMBER have been demonstrated to outperform state-of-the-art methods in stochastic conformance checking, as detailed in our previous paper [1]. By providing a practical solution for deriving precise, memory-aware stochastic models from event logs, REMEMBER contributes to advancing the capabilities of stochastic process mining within the BPM field.

The remainder of this paper is structured as follows: Section 2 provides an overview of the tool's architecture and features and discusses its application in a specific use case; Section 3 reports the information about tool's availability; and Section 5 concludes.

2 REMEMBER

2.1 Overview

2.2 A Simple Case Study

3 Availability

4 Conclusion

References

1. Incerto, E., Vandin, A., Ahrabi, S.S.: Stochastic conformance checking based on variable-length markov chains. Information Systems p. 102561 (2025)