Outcome-Oriented Predictive Process Monitoring to Predict Unplanned ICU Readmission in MIMIC-IV Database

Qifan Chen[1], Yang Lu[1], Charmaine Tam[2], and Simon K. Poon[1]

[1] School of Computer Science, The University of Sydney, Sydney, Australia
{qifan.chen,yang.lu,simon.poon}@sydney.edu.au
[2] Centre for Translational Data Science and Northern Clinical School, University of Sydney, Sydney, Australia
charmaine.tam@sydney.edu.au

1 Introduction

Unplanned readmission entails not only preventable risks for patients but also unnecessary medical resources allocated to increased likelihood of length of stay. Predicting readmission traditionally is performed by applying statistical techniques on clinical information recorded in historical completed patient records. In this paper, we investigate process-oriented event features of in-patient journey (ICU admission to discharge) captured in EHR to predict the risk of ICU readmission in the future. Predictive process monitoring aims to learn from previous completed cases to predict the future outcome of the current (ongoing) cases. In this study, we adopt process monitoring to consider each ICU stay as a continuous process trace. This technique enables us to incorporate process-oriented features (i.e. the sequence of chart events) in conjunction with commonly used clinical features to train our deep learning model to predict future adverse outcome of the patient (the risk of readmission). Our goal is to show that process-oriented features can add value to improve the accuracy of the predictive model, and by incorporating process monitoring technique, we are also able to predict future adverse events by observing ongoing events of existing patients.

2 Methodology

The typical pipeline [2] for predictive process monitoring is followed, as shown in Figure 1. The readmission event log is constructed from the publicly available MIMIC-IV database. First, we filter out patients under 18 years old and patients that died in the ICU. Each ICU stay is treated as a process trace as some patients could have multiple ICU stays over time. We take 12 ICU patients' chart events as activities with timestamps and their observation results as event-level attributes. We also extract demographic features, including gender, age, ethnicity and insurance along with ICD diagnoses as case-level attributes. We encode the event log using one-hot encoding for activities and min-max encoding

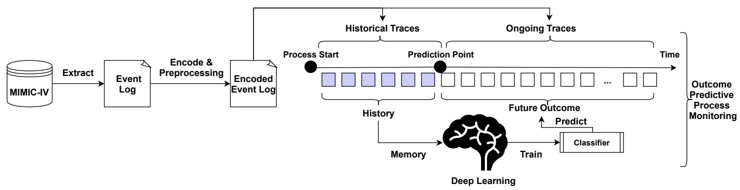

Fig. 1. Overview of the proposed approach

for numerical attribute values. We adopt the Elixhauser comorbidity score to encode the patients with multiple ICD diagnoses. We implement a LSTM+CNN deep learning model to train a classifier from the historical traces and make predictions for ongoing traces.

3 Results and Discussion

We evaluate our approach against baseline [1] using the extracted event logs. To stimulate real-life process, all traces are sorted based on start time. We use first 80 % of the event log for training, and remaining 20 % for testing [2]. Earliness measures how "early" (i.e. number of events completed) of the ongoing trace is adequate to reach a predefined accuracy [2]. We apply the trained model to different prefixes of the trace to explore how "early" we could predict unplanned ICU readmission. The baseline model considered solely the clinical result values of every chart event of completed ICU stay as statistical features in model building. The predictive accuracy of the baseline model was found to be around 60 %. In our proposed process monitoring inspired framework, by adjusting prefixes in the holdout cases for evaluation, as expected the prediction performance will improve as prefix length increases. Our model was able to outperform the baseline after considering the prefix length of 6 events, and progressively increased to 65 % when prefix length extended to 10 events. In conclusion, we show that by incorporating process-oriented features in prediction, we are able to achieve better prediction against baseline after considering 6 events as prefix (in comparison of 43 events used on average) in the testing phase. Furthermore, the usefulness of this framework is the applicability to predict readmission risk of ongoing patient cases. This framework can be further extended for simulating unseen patient journey and online process management of patient care.

References

1. Assaf, R., Jayousi, R.: 30-day hospital readmission prediction using mimic data. In: 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT). pp. 1–6. IEEE (2020)
2. Teinemaa, I., Dumas, M., Rosa, M.L., Maggi, F.M.: Outcome-oriented predictive process monitoring: Review and benchmark. ACM Transactions on Knowledge Discovery from Data (TKDD) **13**(2), 1–57 (2019)