

Predicting ICU Readmission with Context-Enriched Deep Learning

Rafael Zamora-Resendiz¹, Xiang Wang¹, Xinlian Liu¹, Silvia Crivelli¹

¹Lawrence Berkeley National Laboratory

1. Introduction

The explosion of healthcare information stored in Electronic Health Records (EHR) has led to an increasing trend of EHR-based applications in computational biomedicine. Unfortunately, applying deep learning (DL) to medicine is no trivial task as EHR data is extremely complex, usually unbalanced, muddled with missing or invalid values and frequently contains a heterogeneous mixture of data types and structured/unstructured formats. The problem has been compounded by the lack of publicly available datasets that are large enough for the development of deep learning methods as well as by the lack of benchmarking tasks and metrics to compare results. The creation of the MIMIC-III Clinical Database [1] and the recent work of Harutyunyan et al. [2] proposing benchmarking tasks and metrics are accelerating advances in the field.

To deal with the complexity of the EHR data, methods usually focus on expert-defined phenotypes, thus not fully exploiting the richness of EHR data and the power of DL approaches, which do not rely on feature engineering tactics. Rajkomar et al. [3] presented the most relevant DL-based method so far, which uses entire EHRs as input to predict inpatient mortality, unplanned 30-day hospital readmission, length of hospital stay, and discharge diagnoses. Among the advantages of this method, it allows using all the EHR data at large scale, including thousands of predictors for each patient while identifying which data were important for a particular prediction. However, the method relies on a very large hyperparameter search, which makes it very computationally intensive requiring specialized expertise and hardware, and the resulting models have limited “transferability” to other hospitals or cohorts. The lack of transferability is both a limitation of the Fast Healthcare Interoperability Resources (FHIR) format, which does not ensure semantic consistency between sites, and the model’s embedding method, which is tuned to a particular site’s feature codes. Furthermore, their study is based on private data.

We propose a more generalizable approach for analysing structured EHR data using DL-based methods. The approach converts source data to Observational Medical Outcomes Partnership (OMOP) format (<https://ohdsi.org>), which organizes data in a relational database with medical concept tags, and then combines embedded contextual information contained within the schema with corresponding clinical records. OMOP provides a uniform representation of medical records across systems by both using a common data model (CDM) and providing references to standard medical concepts as defined by source vocabularies from the UMLS (<https://nlm.nih.gov/research/umls/>). We use Poincare-space embedding methods to embed OMOP’s medical concept trees as these methods are better suited for hierarchically structured data than Euclidean space embedders such as Word2Vec. Since

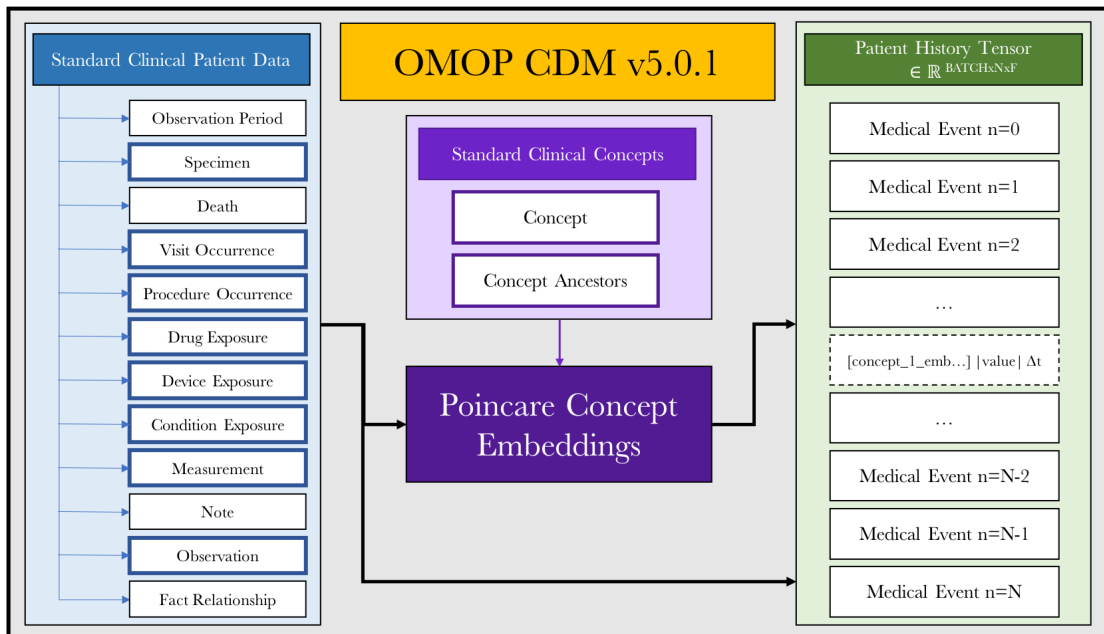


Figure 1. Patient medical history representation gathered from clinical data tables and medical concept tables in OMOP v5.0.1 for multi-head attention longitudinal model. Medical events contain F concatenated features constructed from records’ time of entry, normalized values of the records’ main attribute, and embeddings of associated concept identifiers.

OMOP concepts and relations are independent of the source dataset, the embedded concepts are transferable to other cohorts employing the OMOP schema.

We focus on predicting the likelihood that a patient will die or have an unplanned readmission to the ICU 30 days after been released from the ICU. We compare our results against those in [4], which also use MIMIC-III data. Our results show that our method is able to outperform current DL methods on readmission prediction in precision, accuracy and area-under-ROC (AUC).

2. Data

Our retrospective studies focus on the MIMIC-III Clinical Database [1], which contains around 60,000 intensive care units (ICU) admissions of patients to a Boston hospital between 2001 and 2012. The MIMIC-III database includes demographics, vital sign measurements made at the bedside (~ 1 data point per hour), laboratory test results, procedures, medications, nurse and physician notes, imaging reports, and out-of-hospital mortality.

We worked on three variations of the 30-day unplanned readmission task. This prediction task was casted as a binary classification problem at the time of discharge. Each variant defines the positive condition as follows:

- *Readmission + Transfer*: We consider the positive labels used in [4] which include unplanned readmission within 30-days after discharge, death within 30-days after discharge, and transfers from lower wards back to the ICU.
- *General Readmission*: For this set, we consider positive labels to be admissions which result in unplanned 30-day readmission or death. Transfers were not included in order to provide a task that can be compared to a general readmission prediction task that is not specific to ICUs.

Method	Features	Labels	Precision	Recall	F1	Kappa	AUC	Acc.
LSTM	Lin et al.	Readmission + Transfer	0.3630	0.7330	0.4855	~	0.787	0.695
LSTM + CNN	Lin et al.	Readmission + Transfer	0.3670	0.7420	0.4910	~	0.791	0.698
Multi-Head Attention	OMOP Schema	Readmission + Transfer	0.8912	0.4877	0.6304	0.5475	0.8157	0.8519
Multi-Head Attention	OMOP Schema	General Readmission	0.9419	0.5878	0.7239	0.6774	0.8321	0.9159
Multi-Head Attention	OMOP Schema	DoD Readmission	0.9412	0.5449	0.6902	0.6431	0.8203	0.9120

Table 1. Results for multi-head attention model on the three 30-day unplanned readmission variants. We include the results published in [4] for comparison. Metrics include precision, recall, and F1 w.r.t. the positive class. Cohen’s kappa coefficient, area-under-ROC, and average accuracy are also included.

- “Diseases of Despair” (DoD) Readmission: We consider a subset of the general readmission task composed of a cohort of patients containing International Classification of Disease (ICD9) codes related to substance-use disorders, suicides, and alcohol-related diseases.

3. Methods

Our proposed methodology follows the idea of creating ‘patient state vectors’ using the full EHR data as presented in [3], but differs in the dataset representation, embeddings, and longitudinal analysis model. Prior to processing, the MIMIC-III database was converted to OMOP format. We define ‘patient state vectors’ to be medical events based off records from OMOP’s standardized clinical data tables. We focused on tables that include medical observations, measurements, conditions, procedures, device-exposures and drug-exposures. We then create medical event vectors using records’ time of entry, normalized values of the records’ main attribute, and embeddings of associated concept identifiers. These are used to construct medical history tensors from a series of 1D medical event vectors (see Fig. 1.)

OMOP’s standardized vocabulary contains a medical concept schema based on a number of UMLS source vocabularies including RX-Norm, ICD, and SNOMED. Within OMOP’s standard library, there is a table of concept relations that contains approximately 68 million hierarchical relationships for 3.3 million unique concepts. Embedding of these concepts allows for the organization of concepts in a way that preserves their similarity between closely related concepts in the dependency graph and embedding spaces. However, embeddings of large, complex trees may be computationally infeasible. To increase the representation capacity of embeddings so that they can be used for large trees, we use embeddings in hyperbolic space, in particular a Poincare ball model. The final embeddings provide a 10-dimension vector for each of the 3.3 million concepts.

We implemented a multi-head attention longitudinal model to classify medical history tensors into positive and negative examples of n -day unplanned-readmissions. Multi-head attention has shown to better capture long-term dependencies in series data than RNNs, and

greatly decrease the sparsity of the data representation without the need of averaging along equal time intervals. While multi-head attention has been explored prior to this study, our method’s main contribution is the design of the input data representation which is better formatted to take advantage of multi-head attention. In addition, the model processes patient’s full medical histories, which is significantly more data than the 48-hour windows used in [4]. The final model was trained for 100 epochs on a Tesla-V100 graphics processor using ADAM for gradient-descent optimization and categorical cross entropy as loss.

4. Results and Future Work

Each dataset was split randomly into training, validation, and testing set according to a 80%/10%/10% distribution. Table 1 details the results for trained models evaluated on test sets for 30-day unplanned readmission for the three task’s variants. We measure model performance using positive class precision, recall, F1, area-under-ROC (AUC) and overall accuracy. The reliability of the models was also compared using Cohen’s kappa coefficient, which accounts for the possibility of classification occurring by chance. The multi-head attention longitudinal model outperforms current deep learning methods as presented in [4] in precision, F1, AUC and overall accuracy, while underperforming on recall. The results for general and DoD readmission show higher performance for all defined metrics.

We believe the improved performance is due in part to the contribution of contextual information provided by the Poincare hierarchical embeddings. Since these embeddings are trained using data common among all OMOP databases, they promote “transfer learning” application among different medical cohorts. We show that multi-head attention architectures allow for dense representations of patient histories, which allow us to incorporate patients’ full medical histories during prediction at the time of discharge. More work needs to be conducted looking into the effects of missing data points on prediction performance. Due to the conversion of schema between native MIMIC-III to OMOP, many concept tags are missing and require manual cleaning though the current model configuration still outperforms other methods. Future work will also be conducted to benchmark the model on tasks proposed in [2], which include predicting risk of mortality, forecasting length of stay, detecting physiologic decline, and classifying different phenotypes.

5. References

- [1] Johnson, Alistair EW, et al. "MIMIC-III, a freely accessible critical care database." *Scientific data* 3 (2016): 160035.
- [2] Harutyunyan, Hrayr, et al. "Multitask learning and benchmarking with clinical time series data." *Scientific data* 6.1 (2019): 96.
- [3] Rajkomar, Alvin, et al. "Scalable and accurate deep learning with electronic health records." *npj Digital Medicine* 1, no. 1 (2018): 18.
- [4] Lin, Yu-Wei, et al. "Analysis and Prediction of Unplanned Intensive Care Unit Readmission using Recurrent Neural Networks with Long Short-Term Memory." *bioRxiv* (2018): 385518.