

Patient Electronic Health Record as Temporal Graphs for Health Monitoring

Hugo LE BAHER^{a,b,c,1}, Jérôme AZÉ^b, Sandra BRINGAY^{a,d}, Pascal PONCELET^a,
Nancy RODRIGUEZ^a and Caroline DUNOYER^{c,e}

^a LIRMM, UMR 5506, Université de Montpellier, CNRS, Montpellier, France

^b 5 DEGRÉS, Paris, France

^c Département d'Information Médicale, CHU Montpellier, Montpellier, France

^d AMIS, Université Paul-Valéry, Montpellier, France

^e IDESP, UMR UA11, INSERM - Université de Montpellier, Montpellier, France

Abstract. Machine learning methods are becoming increasingly popular to anticipate critical risks in patients under surveillance reducing the burden on caregivers. In this paper, we propose an original modeling that benefits of recent developments in Graph Convolutional Networks: a patient's journey is seen as a graph, where each node is an event and temporal proximities are represented by weighted directed edges. We evaluated this model to predict death at 24 hours on a real dataset and successfully compared our results with the state of the art.

Keywords. Graph neural networks, MIMIC-III, health monitoring, mortality prediction, electronic health record

1. Introduction

Supervision of patient status in medical emergency services is a difficult problem. Jung et al. [1] show that the implementation of decision support tools, which are able to anticipate critical risks in patients under surveillance, would improve the capacity of services to react as fast as possible to unexpected episodes and thus preventing the deaths of their patients. Different approaches have been proposed to predict medical outcomes of a patient from his history. Typically, all data are transformed into vectors in order to use traditional learning techniques such as partitioning, k-nearest neighbor [2], logistic regression, or deep learning models [3]. Under the assumption that the variations over time of the patient's descriptors are essential for predicting his future state, other techniques consider a patient journey as a sequence, consisting of a succession of vectors. This representation allows the use of recursive models like LSTM [4]. The implementation of these techniques requires slicing the patient's history, according to time. Finding the cut protocol that best represents the whole data captured over time is an expensive process and the success of this research influences the predictive ability of the model. For example, if we consider the granularity of the hour, we lose the evolution of the heart rate second by second. On the contrary, if we consider a breakdown to the second, the representation produced is too voluminous to be inputted into most models.

¹ Corresponding Author: Hugo LE BAHER, E-mail: hugo.le-baher@lirmm.fr

Another underlying problem with this representation is that the different patient sequences have varying lengths which require adaptations in the learning model. Recently, GNN (*Graph Neural Network*) have shown to be very suitable for making predictions [5]. Considering the aforementioned objectives and the identified limits of the approaches in the literature, we propose in this article a modeling of patient records as temporal graphs, input into GNNs.

The contributions of the article are as follows: 1) We propose a new original modeling of temporal and heterogeneous data in the form of a graph where each node corresponds to a measurement and where temporal proximities are represented by weighted arcs. 2) We input this representation into a GCN (*Graph Convolutional Network*) that we evaluate on the MIMIC-III dataset described by [6], to predict the patient death at 24 hours, from real-world data collected during their journeys. We show through experiment that our approach offers results comparable to those of the state of the art.

The rest of the article is organized as follows. Section 2 describes the modeling proposed to directly integrate the time dimension. Section 3 describes the experiments carried out. Finally, we conclude by presenting our future work in Section 4.

2. Method

We propose a modeling of data from patient files in the form of graphs which aims to avoid a temporal slicing. Formally, records of a given patient are represented as a graph $G = (E, P, X)$ where $E = \{e_1, e_2, \dots, e_n\}$ represents the events, $P = \{(e_i, e_j, p)\}, 1 \leq i \leq n, 1 \leq j \leq n, p \in \mathbb{R}^+$ represents the set of weighted arcs with the temporal proximity p between two events e_i and e_j and $X \in \mathbb{R}^{n \times m}$ represents the characteristics of events with m number of distinct categories measured. Our objective is to predict the medical outcome of interest, i.e., death at 24 hours. This is a binary classification problem, namely finding a mapping f such that $f: G \rightarrow Y \in \{0,1\}$, Y matches the binary label.

The particularity of the modeling resides in the constitution of the arcs. Information propagates through the convolution process within the final model. The proposed hypothesis is that events close in time have comparable properties. This translates into the graph by calculating a temporal proximity p for each couple (e_i, e_j) from their temporal coordinates $T = \{t_1, t_2, \dots, t_n\}$ which acts as a weight for the arcs. Thus, the closer two events are in time, the greater the weight of the arc that connects them. The gradual decrease of the weight as the temporal distance increases calls for the study of forgetting from cognitive sciences. We compared several forgetting functions by experimentation and retained the function developed in [7] Π , which is defined by:

$$\Pi(b, \delta_{i,j}) = \exp(-b \cdot \delta_{i,j}) * (\delta_{i,j} \geq 0), b > 0, \delta_{i,j} = t_i - t_j$$

δ corresponds to the temporal distance computed for each pair of events. b is a modeling parameter, which influences the intensity of the decrease in the function. The value of this parameter has a direct effect on the structure of the generated graphs.

Figure 1 summarizes the different stages of the proposed modeling, which starts from the data to produce a temporal proximity graph: on the one hand, the centered value and the modality of each measurement are transformed into a matrix where each column

corresponds to a measured modality (1). Each line corresponds to the vector representation of a given node. On the other hand, the temporal coordinates of events are transformed pairwise into distances (2). The distances are then transformed into proximity by the application of the function Π (3). The result is an adjacency matrix that describes the weight of the edges of the final graph. The adjacency matrix and node representation vector put together describe the graph of a patient journey (4), inputted in a supervised model to predict the binary label, graph-wise.

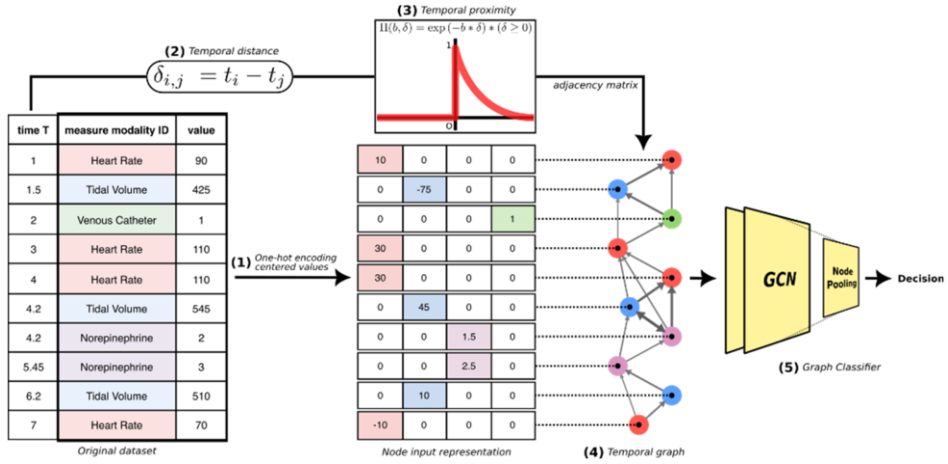


Figure 1. Transformation process of the original data into a temporal graph of measurements. Illustration based on fictional data. (1) Transformation of the values through centered one-hot encoding. (2) Pairwise distance of temporal coordinates. (3) Computation of the temporal proximity. (4) Patient representation: Node features and adjacency matrix which describe the graph. (5) Graph classification model.

3. Experiments and Results

To evaluate our approach, we focused on the 24-hour death prediction task described in Harutyunyan et al. [6]. Only patient journeys that last at least 4 hours are selected. For each of them, we seek to predict for each one-hour span whether a patient dies within the next 24 hours, based on data prior to the observation. The original data is available on Physionet [8] in the public MIMIC-III Clinical Database [9, 10]. This database is often used in the literature to compare approaches for many medical supervision tasks because it is voluminous, heterogeneous and anonymized. Additionally, the data made available describe actual patient journeys that occurred during Beth Israel Deaconess Medical Center activity from 2001 to 2021. For their experiments, Harutyunyan et al. [6] select 17 descriptors (e.g., Heart Rate, Blood Pressure) according to medical criteria. Values are filtered, pre-processed and ordered categorical data are transformed into continuous. Data volumes are described in Table 1. To obtain a fair comparison with our own experiments, we will rely exactly on these transformed data, which are publicly available.

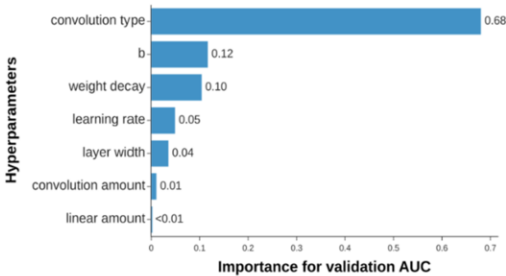
Table 1. Characteristics of data extracted from MIMIC-III.

Amount of patients in training set	28 620
Amount of patients in test set	5 058
Percentage of patients from the minority class	11.0%
Average length of stay	86.6h
Maximum stay duration	2 103h

The model presented is chosen after the search for optimal hyperparameters and structure thanks to the Optuna library [11]. The results presented through the application of our method come from a cross-validation with 5 folds, sliced in the training set then evaluated on the test set. The model referred as "Convolution of Graphs" in the Figure 2 consists of a succession of the following layers: Three layers of GraphSAGE [12] of the DGL library (DeepGraph Library) with the average as aggregation function, then the concatenation of three aggregation functions *min*, *max* and *average* on the representation of all nodes, two linear layers and a *softmax* output. Each trained layer is followed by an activation function *tanh*, except the last. The chosen optimizer is *Adam* with the learning rate set to 0.003 and the weight decay to 0.05. The temporal proximity parameter is fixed at 4.5 and the width of the successive layers is fixed at 15. Learning is computed on a machine equipped with 40 computing cores, 126GB of RAM and 4 NVIDIA GeForce GTX 1080 graphics cards, each with 11GB of VRAM.

Figure 2. On the left, predictive 24-hour mortality results are shown. The different versions of the LSTM are extracted from the original paper corresponding respectively to "S": Standard, "C": Channel-wise and "DS": Deep Supervision. On the right, the importance of hyperparameters on performance following optimization by Optuna for the Convolution of Graphs is shown

Model	AUC
Harutyunyan et al. [6]:	
Logistic Regression	0.870
LSTM - S	0.892
LSTM - C	0.906
LSTM - C + DS	0.911
Our method:	
Convolution of Graphs	0.897



To compare the approaches, we measure the area under the sensitivity/specificity curve, referred to as *AUC*, on the test set. This score is well suited for datasets with unbalanced label proportions. As seen in the Figure 2, our modeling offers results similar to those of the state of the art, observed in Harutyunyan et al. [6], namely logistic regression and the different variants of LSTM. Our objective in this article is to validate the representation of patient data in the form of a graph. The result obtained by our method exceeds some versions of LSTM.

It is questionable to what extent the predictions of the presented model are influenced by the propagation of information from the GNNs through the arcs resulting from the modeling as expected rather than by other effects. The Figure 2 shows the contributing value of different hyperparameters during Optuna's hyperparameter optimization. The results come from a functional analysis of variance, i.e., fANOVA

[13]. We observe on these results that the most important hyperparameters are the type of convolution and the parameter b of the modeling. We deduce that the constitution of the arcs and the way in which the information is propagated within the graph in the model have a strong impact on the measured performances.

4. Conclusions

In this article, we have proposed a new modeling of patient data in the form of a graph. This alternative representation of the data enables to consider time. We use weighted arcs to represent the temporal proximity between two events. We experimented this modeling by taking advantage of graph convolutions. Nodes characteristics are propagated along arcs. The main limitation is the studied population (ICU) and task (mortality). Towards generalization, other population and tasks have to be studied. Considering classification accuracy, our results approach those of the state of the art. Other approaches based on node-level prediction should be considered. Moreover, due to the modularity offered by graph structures and the many recent contributions in the field of GNNs, it is possible to imagine various areas to improve explainability. This project is funded by a CIFRE grant, funded by 5 DEGRÉS, established in collaboration with LIRMM and CHU Montpellier.

References

- [1] Jung B, Daurat A, De Jong A, Chanques G, Mahul M, Monnin M, et al. Rapid response team and hospital mortality in hospitalized patients. *Intensive Care Med.* 2016 Apr;42(4):494-504
- [2] Poongodi T, Sumathi D, Suresh P, Balusamy B. Deep Learning Techniques for Electronic Health Record (EHR) Analysis. In: Bhoi AK, Mallick PK, Liu CM, Balas VE, editors. *Bio-inspired Neurocomputing. Studies in Computational Intelligence.* Springer; 2021. p. 73-103.
- [3] El-Rashidy N, El-Sappagh S, Abuhmed T, Abdelrazek S, El-Bakry HM. Intensive Care Unit Mortality Prediction: An Improved Patient-Specific Stacking Ensemble Model. *IEEE Access.* 2020;8:133541-64.
- [4] Ashfaq A, Sant'Anna A, Lingman M, Nowaczyk S. Readmission prediction using deep learning on electronic health records. *Journal of Biomedical Informatics.* 2019 Sep;97:103256.
- [5] Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. Graph neural networks: A review of methods and applications. *AI Open.* 2020;1:57-81.
- [6] Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. *Scientific Data.* 2019 Dec;6(1):96.
- [7] White KG. Forgetting functions. *Animal Learning & Behavior.* 2001 Aug;29(3):193-207.
- [8] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation.* 2000 Jun;101(23):E215-20. Place: United States.
- [9] Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Scientific Data.* 2016 May;3:160035.
- [10] Johnson A, Pollard T, Mark R. MIMIC-III Clinical Database. PhysioNet; 2016. Version Number: 1.4. Type: dataset. Available from: <https://physionet.org/content/mimiciii/1.4/>.
- [11] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. In: *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2019.
- [12] Hamilton W, Ying Z, Leskovec J. Inductive Representation Learning on Large Graphs. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. *Advances in Neural Information Processing Systems.* vol. 30. Curran Associates, Inc.; 2017.
- [13] Hutter F, Hoos H, Leyton-Brown K. An Efficient Approach for Assessing Hyperparameter Importance. In: *Proceedings of International Conference on Machine Learning 2014 (ICML 2014)*; 2014.