Improving Medical Predictions by Irregular Multimodal Electronic Health Records Modeling

Xinlu Zhang [*1] Shiyang Li [*1] Zhiyu Chen [1] Xifeng Yan [1] Linda Ruth Petzold [1]

Abstract

Health conditions among patients in intensive care units (ICUs) are monitored via electronic health records (EHRs), composed of numerical time series and lengthy clinical note sequences, both taken at *irregular* time intervals. Dealing with such irregularity in every modality, and integrating irregularity into multimodal representations to improve medical predictions, is a challenging problem. Our method first addresses irregularity in each single modality by (1) modeling irregular time series by dynamically incorporating hand-crafted imputation embeddings into learned interpolation embeddings via a gating mechanism, and (2) casting a series of clinical note representations as multivariate irregular time series and tackling irregularity via a time attention mechanism. We further integrate irregularity in multimodal fusion with an interleaved attention mechanism across temporal steps. To the best of our knowledge, this is the first work to thoroughly model irregularity in multimodalities for improving medical predictions. Our proposed methods for two medical prediction tasks consistently outperforms state-of-the-art (SOTA) baselines in each single modality and multimodal fusion scenarios. Specifically, we observe relative improvements of 6.5%, 3.6%, and 4.3% in F1 for time series, clinical notes, and multimodal fusion, respectively. These results demonstrate the effectiveness of our methods and the importance of considering irregularity in multimodal EHRs. [1].

[*]Equal contribution [1]University of California, Santa Barbara. Correspondence to: Xinlu Zhang <xinluzhang@ucsb.edu>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

[1]Our code is released at https://github.com/XZhang97666/MultimodalMIMIC

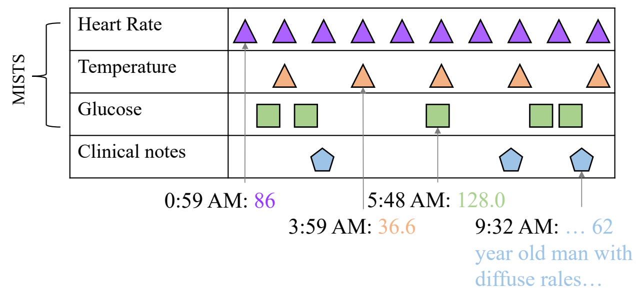

Figure 1: An example of a patient's ICU stay includes MISTS with three features and a series of clinical notes. For MISTS, heart rate and temperature are monitored regularly with different frequencies, and glucose is a laboratory test ordered at irregular time intervals based on doctors' decisions. Clinical notes are free text, collected with much sparser irregular time points than clinical measurements.

1. Introduction

ICUs admit patients with life-threatening conditions, e.g. trauma (Tisherman & Stein, 2018), sepsis (Alberti et al., 2002), and organ failure (Afessa et al., 2007). Care in the first few hours after admission is critical to patient outcomes. This period is also more prone to medical decision errors than later times (Otero-López et al., 2006). Automated tools with effective and real-time predictions can be much beneficial in assisting clinicians in providing appropriate treatments. Recently, the health conditions of patients in ICUs have been recorded in EHRs (Adler-Milstein et al., 2015), bringing the possibility of applying deep neural networks to healthcare (Xiao et al., 2018; Shickel et al., 2017), e.g. mortality prediction (Zhang et al., 2021a) and phenotype classification (Harutyunyan et al., 2019). EHRs contain multivariate irregularly sampled time series (MISTS) and irregular clinical note sequences, as shown in Figure 1. The multimodal structure and complex irregular temporal nature of the data present challenges for prediction. This leads us to formulate two research objectives:

1. *Tackling irregularity in both time series and clinical notes*
2. *Integrating irregularity into multimodal representation learning*

To the best of our knowledge, none of the existing works has fully considered irregularity in multimodal representation learning.

We observed three major drawbacks for irregular multimodal EHRs modeling in existing works. 1) *MISTS models perform diversely.* While the numerous MISTS models have been proposed to tackle irregularity (Lipton et al., 2016; Shukla & Marlin, 2019; 2021; Zhang et al., 2021b; Horn et al., 2020; Rubanova et al., 2019), none of the approaches consistently outperforms the others. Even among ***T**emporal **d**iscretization-based **e**mbedding* (TDE) methods, including hand-crafted imputation (Lipton et al., 2016) and learned interpolation (Shukla & Marlin, 2019; 2021), which transform MISTS into regular time representations to interface with deep neural networks for regular time series, there is no clear superior approach. 2) *Irregularity in clinical notes is not well tackled.* Most existing works (Golmaei & Luo, 2021; Mahbub et al., 2022) directly concatenate all clinical notes of each patient but ignore the note-taking time information. Although Zhang et al. (2020) proposes an LSTM variant to model time decay among clinical notes, this approach utilizes only a few trainable parameters, which could be less powerful. 3) *Exiting works ignore irregularity in multimodal fusion.* Deznabi et al. (2021); Yang et al. (2021) have demonstrated the effectiveness of combining time series and clinical notes for medical prediction tasks, however these works are deployed only on multimodal data without considering irregularity. Their fusion strategies may not be able to fully integrate irregular time information into multimodal representations, which can be essential for prediction performance in real-world scenarios.

Our Contributions. To tackle the aforementioned issues, we separately model irregularity in MISTS and irregular clinical notes, and further integrate multimodalities across temporal steps, so as to provide powerful medical predictions based on the complicated irregular time pattern and multimodal structure of EHRs. Specifically, we first show that different TDE methods of tackling MISTS are complementary for medical predictions, by introducing a gating mechanism that incorporates different TDE embeddings specific to each patient. Secondly, we cast note representations and note-taking time as MISTS, and leverage a time attention mechanism (Shukla & Marlin, 2021) to model the irregularity in each dimension of note representations. Finally, we incorporate irregularity into multimodal representations by adopting a fusion method that interleaves self-attentions and cross-attentions (Vaswani et al., 2017) to integrate multimodal knowledge across temporal steps. To the best of our knowledge, this is the first work for a unified system that fully considers irregularity to improve medical predictions, not only in every single modality but also in multimodal fusion scenarios. Our approach demonstrates superior performance compared to baselines in both single modality and multimodal fusion scenarios, with notable relative improvements of 6.5%, 3.6%, and 4.3% in terms of F1 for MISTS, clinical notes, and multimodal fusion, respectively. Our comprehensive ablation study demonstrates that tackling irregularity in every single modality benefits not only their own modality but also multimodal fusion. We also show that modeling long sequential clinical notes further improves medical prediction performance.

2. Related Work

Multivariate irregularly sampled time series (MISTS). MISTS refer to observations of each variable that are acquired at irregular time intervals and can have misaligned observation times across different variables (Zerveas et al., 2021). GRU-D (Che et al., 2018) captures temporal dependencies by decaying the hidden states in gated recurrent units. SeFT (Horn et al., 2020) represents the MISTS to a set of observations based on differentiable set function learning. ODE-RNN (Rubanova et al., 2019) uses latent neural ordinary differential equations (Chen et al., 2018) to specify hidden state dynamics and update RNN hidden states with a new observation. RAINDROP (Zhang et al., 2021b) models MISTS as separate sensor graphs and leverages graph neural networks to learn the dependencies among variables. These approaches model irregular temporal dependencies in MISTS from different perspectives through specialized design. TDE methods are a subset of methods for handling MISTS, converting them to fixed-dimensional feature spaces, and feeding regular time representations into deep neural models for regular time series. Imputation methods (Lipton et al., 2016; Harutyunyan et al., 2019; McDermott et al., 2021) are straightforward TDE methods to discretize MISTS into regular time series with manual missing values imputation, but these ignore the irregularity in the raw data. To fill this gap, Shukla & Marlin (2019) presents interpolation-prediction networks (IP-Nets) to interpolate MISTS at a set of regular reference points via a kernel function with learned parameters. Shukla & Marlin (2021) further presents a time attention mechanism with time embeddings to learn interpolation representations. However, learned interpolation strategies do not always outperform simple imputation methods. This may be due to complicated data sampling patterns (Horn et al., 2020). Inspired by Mixture-of-Experts (MoE) (Shazeer et al., 2017; Jacobs et al., 1991), which maintains a set of experts (neural networks) and seeks a combination of the experts specific to each input via a gating mechanism, we leverage different TDE methods as submodules and integrate hand-crafted imputation embeddings into learned interpolation embeddings to improve medical predictions.

Irregular clinical notes modeling. (Golmaei & Luo, 2021; Mahbub et al., 2022) concatenate each patient's clinical