Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230110

DeepTSE: A Time-Sensitive Deep Embedding of ICU Data for Patient Modeling and Missing Data Imputation

Michael FUJARSKI[a,1,2], Christian PORSCHEN [b,2], Lucas PLAGWITZ[a], Daniel STROTH[a], Catharina Marie VAN ALEN[a], Mahan SADJADI[b], Raphael WEISS[b], Alexander Zarbock[b], Thilo VON GROOTE[b,3] and Julian VARGHESE[a,3]
[a] *Institute of Medical Informatics, University of Münster, Germany*
[b] *Department of Anaesthesiology, Intensive Care and Pain Medicine, University Hospital Münster, Münster, Germany*

Abstract. Missing data is a common problem in the intensive care unit as a variety of factors contribute to incomplete data collection in this clinical setting. This missing data has a significant impact on the accuracy and validity of statistical analyses and prognostic models. Several imputation methods can be used to estimate the missing values based on the available data. Although simple imputations with mean or median generate reasonable results in terms of mean absolute error, they do not account for the currentness of the data. Furthermore, heterogeneous time span of data records adds to this complexity, especially in high-frequency intensive care unit datasets. Therefore, we present DeepTSE, a deep model that is able to cope with both, missing data and heterogeneous time spans. We achieved promising results on the MIMIC-IV dataset that can compete with and even outperform established imputation methods.

Keywords. MIMIC IV, ICU, Machine Learning, Deep Embedding, Time-Sensitive Data Imputation, Patient Modeling

1. Introduction

In the intensive care unit (ICU), patient data, such as vital signs or lab values, are often collected at irregular time intervals, resulting in sparse data with many missing values. This is further complicated by time intervals in which data collection is not possible due to the patient not being present in the ICU, due to interventional or diagnostic procedures, or because measuring devices (such as invasive monitoring devices or catheters) are not yet placed. This sparsity of data can make it challenging to accurately assess patient status and make informed treatment decisions. Finally, incomplete datasets prohibit the use of machine learning methods that assume a dataset with no missing values. In addition, traditional

[1] Corresponding Author: Michael Fujarski, E-mail: michael.fujarski@uni-muenster.de
[2] the authors contributed equally to this work and share the first authorship
[3] the authors contributed equally to this work and share the coordinating authorship

statistical methods may not be well-suited to analyze sparse data, as they often rely on the assumption of a sufficient number of observations. To address this data sparsity in the ICU, advanced data imputation techniques or machine learning methods may be necessary to effectively utilize the available data and make accurate predictions. While some methods have been developed in this regards, many of them have significant limitations, such as relying on randomness of missing values, correlations between features, or equidistant time spans. As shown by Sharafoddini et al. missing values do not necessarily occur at random and can even have a predictive potential for patients' 30-day mortality [1].

Time-sensitive deep embeddings and auto-encoders can directly learn from missing values and may therefore be well-suited methods to impute data. They address the issue of data sparsity, as well as feature correlation and relation of features that are not missing by chance. Therefore, we investigated whether and how a variational autoencoder with time-sensitive imputation competes with established methods of data imputation and may therefore yield opportunity for use in clinical data science in the ICU setting.

2. Methods

2.1. Dataset

We used the Medical Information Mart for Intensive Care (MIMIC)-IV database for training and evaluation of the different imputation methods and models2]. MIMIC is a large database of electronic health records (EHRs) of ICU patients at the Beth Israel Deaconess Medical Center in Boston, Massachusetts. It contains a wide range of data including demographics, laboratory test results, medication administrations, and vital sign measurements for more than 40,000 patients. The parameters we considered for this analysis are split into three categories:

The demographics features consist of age, weight, height, and gender of patients at admission. These features are considered as constant during the stay in the ICU.

The vital signs consist of the parameters temperature, respiratory rate, heart rate, mean blood pressure (MBP), urine output as well as the laboratory parameters creatinine, glucose, hemoglobin (Hb), lactate, pH of blood, potassium, oxygen saturation (SpO2), sodium, and the white blood cell count (WBC).

The events are currently limited to the administration of epinephrine, fentanyl, metronidazole, midazolam, morphine, norepinephrine, paracetamol, and propofol. The procedures currently focus on the artificial ventilation of the patients and include the positive end-expiratory pressure (PEEP), the plateau pressure, and the tidal volume.

All EHRs were split into a training (27,000), validation (7,000), and test (8700) set. For every patient, a single time point was defined based on the duration of the ICU stay. On average, the used EHRs were at 71 hours into intensive care for the training set (validation: 71.25, test: 71.5) with a standard deviation of 91.5 (96.5 and 98.25 respectively). The data were used up to the defined point in time per patient. Missing values were replaced by a token depending on the used imputer.

2.2. Imputation

The performance of an imputation method is measured on the test set by masking each parameter separately and calculating the mean absolute error of the available data.
We compare a set of imputation methods with different underlying assumptions and complexity. Simple imputation methods with mean and median assume the data to be normally distributed and to be missing completely at random. The replacement value can be calculated once on a training set resulting in a fast imputation method. Mean and median imputations were done using pandas (1.5.1) directly. It solely depends on populations distribution of the missing value and does not consider any other parameters. More sophisticated approaches like the k-nearest neighbors (KNN) imputer consider all available data to impute values based on a neighborhood. Since KNN strongly depends on the absolute distances, we scaled the data using a standard scaler. Both the standard scaler, and the KNN imputer, were used from scikit-learn (1.1.1). Therefore, it assumes that similar patients, measured as a distance on the available data, can be used to calculate a (weighted) mean of the found neighborhood. Nevertheless, calculating the nearest neighbors on a large dataset can be time-consuming. The multiple imputation by chained equation (MICE) calculates various imputations by modeling the features as functions of other features in a round robin fashion [3]. Further, we propose a Deep Model as another possible imputer.

2.3. Deep Variational Autoencoder

We developed a Variational Autoencoder (VAE) for the ICU simliar to Pinheiro Cinelli et al. [4]. VAEs encode statistical variance into the latent space by modeling learned dimensions as probability functions. However, we expanded the architecture, as shown in Figure 1, to directly learn from missing values as well as encode heterogeneous time spans. Demographic features are encoded as two-dimensional vectors with their actual value and a flag indicating a masked value. Unavailable or masked data are replaced by [0, 1] per feature per patient. Other parameters are three-dimensional vectors, consisting of their value, a timestamp in minutes since the measurements, and a mask. Therefore, the model is capable of handling differently spaced time series. Additionally, our model contains a projection from the latent space into the latent space that predicts the vital signs of a patient for a given positive delta t. The model was trained using a reconstruction loss, predictive loss, and the Kullback-Leibler divergence as implemented by torch (1.12.1). In order to estimate and improve the reconstruction loss on the masked data, we artificially masked the input data. For each epoch, 70% of the input samples were randomly chosen and one random demographic or vital feature was masked. The reconstruction loss was calculated on both the available data as well as the artificially masked input. The predictive loss was calculated during each epoch with 14 out of 18 randomly chosen features to avoid overfitting the model. We further introduced a local Kullback-Leibler divergence on the learned distribution of the input data in order to force an overall lower variance on similar samples. The model was trained for a total of 45,000 epochs. Hyperparameter optimization and model selection were performed on a validation set.

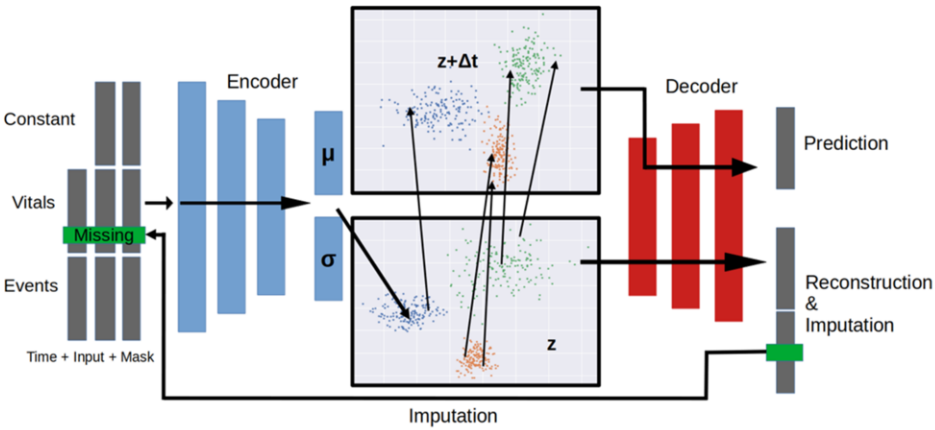

Figure 1. DeepTSE. A Variational Autoencoder with time-sensitive imputation capability. Training losses are calculated on reconstruction, prediction, and local and global Kullback-Leibler divergence.

Performances were evaluated using the separate test set.

3. Results

DeepTSE outperformed other imputation methods in 9 out of 18 categories. Iterative imputation with MICE performed best in 6 categories. Imputation with the train median still performed best in 3 categories, while mean and KNN imputations did not improve the imputation in any case. DeepTSE improved the imputation for the height, weight, gender, pH, creatinine, lactate, urine output, sodium, and WBC. MICE performed best for age, respiratory rate, MBP, heart rate, Hb, and potassium. Median imputation best performed on the body temperature, SpO2, and glucose. Figure 2 depicts the critical differences diagram calculated on the MAE ranking of the different methods. In comparison to the other imputation methods, DeepTSE demonstrated superior performance in 9 out of 18 categories. KNN imputation was outperformed in 14 out of 18 categories, but overall produced only slightly less accurate results to those of MICE. MICE consistently outperformed KNN in the majority of categories and outperformed all other imputation methods in 6 categories. Moreover, MICE is a computationally efficient method, as it is an eager learner, resulting in faster transformations after an initial training. In contrast to the other methods, mean and median imputation methods yielded consistent results across categories, with median imputation performing slightly better than mean imputation.

4. Discussion and Conclusion

DeepTSE has proven to be a well-suited imputer for missing data in the ICU. The combination of time encoding and masking of features enabled the model to learn feature correlations as well as the systematically missing values. Nevertheless, median imputation performed better in three categories indicating that models tend to overfit to features that are missing at random [5]. The imputation of laboratory

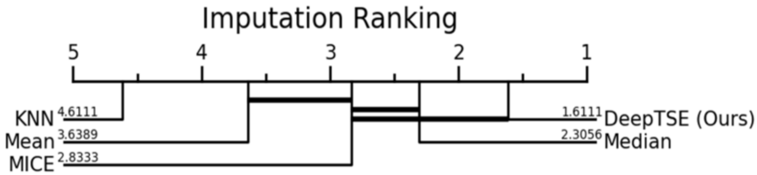

Figure 2. Critical Differences Diagram of the imputation methods and their performances.

values such as lactate and creatinine are particularly of high value for deep learning models, as well as medical professionals. These values are often missing or taken at irregular time points, due to the patient not being present at the ICU or undergoing interventional procedures. Nevertheless, since such values are measured less frequently compared to e.g. vital parameters in an EHR, the loss of information for a potential model is high. If these imputations could be used for estimations of future values, these laboratory measurements could be targets of high interest as well. The early detection of clinical deterioration, as indicated by parameters of lactate or creatinine, are of high importance to clinicians to initiate adequate treatment. For example, serum creatinine is an indicator of acute kidney injury (AKI), the most frequent complication in the critically ill patients, associated with high morbidity and mortality [6]. Similarly, lactate, among other aspects, is an indicator of critical illness as it suggests microvascular dysfunction and lack of peripheral oxygen supply, which is often a sign of hemodynamic or septic shock.

DeepTSE improved and outperformed various imputation methods in 9 out of 18 categories. On average, DeepTSE improved the imputation resulting in a robust and consistent model. Further analyses regarding possible classification and prediction tasks with the learned embedding are now required. Nevertheless, DeepTSE needs to be evaluated on a larger set of parameters, as well as other datasets.

References

[1] Sharafoddini A, Dubin JA, Maslove DM, Lee J. A New Insight Into Missing Data in Intensive Care Unit Patient Profiles: Observational Study. JMIR Med Info. 2019 Jan.

[2] Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV. Physionet. 2022 Nov.

[3] van Buuren S, Groothuls-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011 Dec.

[4] Cinelli LP, Marins MA, da Silva EAB, Netto SL. Variational Autoenvoder. In Variational Methods for Machine Learning wit hApplications to Deep Networks.; 2021. p. 111-149.

[5] Carpenito T, Manjourides J. MISL: Multiple imputation by super learning. Stat Methods Med Res. 2022 Oct.

[6] Hoste EAJ, Bagshaw SM, Webb S, Kellum JA. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med. 2015 Jul.