Transformer-Based Unsupervised Learning for Early Detection of Sepsis (Student Abstract)

Yutao Dou, Wei Li, Albert Y. Zomaya

School of Computer Science, The University of Sydney, Darlington, NSW 2008, Australia
ydou5134@uni.sydney.edu.au, weiwilson.li@sydney.edu.au, albert.zomaya@sydney.edu.au

Abstract

A 6-hour early detection of sepsis leads to a significant increase in the chance of surviving it. Previous sepsis early detection studies have focused on improving the performance of supervised learning algorithms while ignoring the potential correlation in data mining, and there was no reliable method to deal with the problem of incomplete data. In this paper, we proposed the Denoising Transformer AutoEncoder (DTAE) for the first time combining transformer and unsupervised learning. DTAE can learn the correlation of the features required for early detection of sepsis without the label. This method can effectively solve the problems of data sparsity and noise and discover the potential correlation of features by adding DTAE enhancement module without modifying the existing algorithms. Finally, the experimental results show that the proposed method improves the existing algorithms and achieves the best results of early detection.

Introduction

Sepsis is the systemic inflammatory response syndrome caused by infection with significant mortality and morbidity, making an average of 20% of all deaths worldwide but can cause more than 50% mortality in some countries (Rhee et al. 2017). Previous studies showed that a 6-hour early detection could effectively prevent 80% of deaths caused by sepsis, and the mortality rate is increased by 4-7% for each hour delay (Rhee et al. 2017). Although sustainable efforts have been made, the 6-hour early detection is still severely constrained. This situation is mainly caused by the required data for early detection which are not systematically recorded in electronic medical records.

To address the data sparsity and imbalance problem, (Vaswani et al. 2017) first proposed transformer that using the self-attention mechanism can capture the complex relationship between features and has demonstrated its value in natural language processing. (Arik and Pfister 2021) designed a transformer-based solution to process tabular data and demonstrated its superiority. Because the transformer is a variant of the neural network, it is difficult to produce performance without labeled data. Relatively, unsupervised learning can alleviate the problem by learning a large number of unlabeled data sets to generate a general representation of nonlinear features and find the potential relationship between features (Wang, Reimers, and Gurevych 2021).

Inspired by the above works, we proposed a data enhancement module, Denoising Transformer AutoEncoder (DTAE), an unsupervised learning approach to address the above issues. DTAE aims to learn low-dimensional general representations by unifying the transformer and deep auto-encoder (DAE). Combining deep auto-encoder and transformer strengths, DTAE can learn better representations with more effective and stable performance. With this module, the performance of existing algorithms for the 6-hour early sepsis detection task can be notably improved.

Methodology

Problem Definition

We aim to improve the survival rate of patients with sepsis by issuing accurate and reliable warnings 6 hours ahead of the clinical diagnosis of sepsis. However, we have encountered several problems, a large part of which comes from the data. Because electronic medical data are not systematically recorded, noisy and sparse data are common causes for degrading the performance of machine learning algorithms in clinical applications. In addition, missing medical data values are generally processed by data interpolation, such as forward filling, etc. However, these methods cannot reconstruct the original data or even cause further deviation since the data source is noisy and sparse. Therefore, a reliable data enhancement method that works closely with the existing machine learning algorithms is urgently needed.

Denoising Transformer AutoEncoder

To endow the unsupervised learning model with the ability to interact with higher-order features, we explore incorporating the transformer module into the unsupervised learning model and propose a data enhancement method as Denoising Transformer AutoEncoder. Our DTAE structure is related to that of denoising auto-encoders, but the intermediate layer is based on transformer.

Firstly, we incorporate a certain probability distribution and random masking into the original input X to make the learning mode robust to noise and missing data. In this way, the network can learn from the corrupted data and to make it as close as possible to the actual data.

Secondly, the Transformer framework, which performs encoding/decoding, was used to train the processed data $\widetilde{X}$. We feed the unlabeled data $\widetilde{X}$ into the encoder, which contains two main parts: a multi-head self-attention layer and a fully connected feed-forward network. The core component of Multi-Head Self-Attention is the attention mechanism, which is obtained by Q(uery), K(ey) and V(alue) through the scaled dot product calculation. Specifically, Q, K, and V are derived from the embedding of input data, and they are generated by three different linear layers. To control the output size, it is divided by a scaling factor of $\sqrt{d_k}$, where d_k is the dimensions of K. The formula is shown below:

$$attention\_output = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

The specific approach is to map the same Q, K and V to different subspaces of the original high-dimensional space for Attention calculation under the condition that the total number of parameters remains unchanged, and then merge the Attention information in different subspaces in the last step. We aim to project the outputs back to the embedding of dimension, using a fully connected layer.

In short, when the data are fitted into the model, they will first be encoded by the encoder module and converted into an abstract feature. Furthermore, the converted feature will be sent to the decoder module for decoding. Once decoded, the reconstructed data can help discover the hidden potential feature correlation and improve the robustness.

Experiments

Datasets and Evaluation

We employed the PhysioNet Computing in Cardiology Challenge 2019 dataset(Reyna et al. 2020) to evaluate the performance of DTAE on sepsis early detection. The dataset comprises a total of 40,336 patients, where 2932 of them are sepsis patients. Patient data includes demographics, vital signs, laboratory data, and hourly patient status labeled. In our experiment, the dataset was randomly split into training (85%) and testing (15%) sets.

A Bayesian-based method was employed to find the optimal hyper-parameters for DTAE. We also implemented multiple baseline approaches for performance comparison, including Lasso, decision trees, random forest, multi-layer perceptron (MLP), gradient boost trees, and deep neural network. Our experiments used AUC and accuracy as the performance metrics, and all the results were the average values after five runs.

Analysis of Results

The results of the 6-hours early sepsis detection are shown in Table 1. LASSO algorithm had the worst performance in terms of AUC and accuracy. Random forest and MLP showed similar performance to LASSO. In contrast, the more advanced algorithms LightGBM and RNN showed better results, especially LightGBM outperformed the other baseline algorithms. We also evaluated the performance of DTAE combined with some baseline approaches. As can be seen in the Table, the DTAE with the Lasso classifier can provide nearly the same AUC as LightGBM and achieve better accuracy. Such combination brings more than 17% performance improvements since DTAE recovers the correlations between features that are needed for LASSO. Moreover, the combination DTAE and LightGBM achieve the best performance in both AUC and accuracy. It has at least 1.5% performance improvement than the original LightGBM, but not as significant as the previous case. This is because LightGBM is a sophisticated algorithm and has the embedded mechanism to handle missing data.

Methods	**AUC**	**Accuracy**
Lasso	0.619	0.649
Random Forest	0.676	0.673
MLP	0.651	0.657
RNN	0.764	0.767
LightGBM	0.798	0.785
DTAE + Lasso	0.792	0.798
DTAE + Raw Features + LightGBM	**0.813**	**0.819**

Table 1: The results of different methods for the 6-hour sepsis early detection.

Conclusion

We proposed a Transformer-based unsupervised learning algorithm that can learn the feature correlation required for sepsis early detection tasks without labeling. The experiment illustrates that the DTAE module can improve existing models' performance and achieve the best results of early detection.

References

Arik, S. Ö.; and Pfister, T. 2021. TabNet: Attentive Interpretable Tabular Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 6679–6687.

Reyna, M. A.; Josef, C. S.; Jeter, R.; Shashikumar, S. P.; Westover, M. B.; Nemati, S.; Clifford, G. D.; and Sharma, A. 2020. Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. *Critical care medicine*, 48(2): 210–217.

Rhee, C.; Dantes, R.; Epstein, L.; Murphy, D. J.; Seymour, C. W.; Iwashyna, T. J.; Kadri, S. S.; Angus, D. C.; Danner, R. L.; Fiore, A. E.; Jernigan, J. A.; Martin, G. S.; Septimus, E.; Warren, D. K.; Karcz, A.; Chan, C.; Menchaca, J. T.; Wang, R.; Gruber, S.; and Klompas, M. 2017. Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014. *JAMA : the journal of the American Medical Association*, 318(13): 1241–1249.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is All you Need. *Advances in Neural Information Processing Systems*, 30: 5998–6008.

Wang, K.; Reimers, N.; and Gurevych, I. 2021. TSDAE: Using Transformer-based Sequential Denoising Auto-Encoderfor Unsupervised Sentence Embedding Learning. *arXiv preprint arXiv:2104.06979*.