Digital ICU : Data Processing and Visualization Pipeline for Patient Data in Intensive Care Unit

General Info

Contact Person: Ee Heng Chen
Contact Email: eeheng.chen@tum.de

Project Abstract

Digitalization in healthcare has led to the increasing use of digital medical systems in the Intensive Care Unit (ICU). They generate a large amount of data, such as the vital signs of patients, the blood gas analysis results, and the medication that a patient receives. This data can be analyzed using machine learning and data analytics techniques to help clinicians identify clinical deterioration in patients earlier and determine if a patient's treatment is working. However, due to the lack of industry standards, the data format from the medical systems differs between different manufacturers. This issue makes it hard to employ machine learning and data analytics techniques on the raw data since they often require the input data to be structured. A solution would be to create a generic data processing pipeline capable of producing structured data from data in different formats in terms of the data type, data value unit, and data frequency. This project aims to build such a pipeline by focusing on data from medical systems in the ICU. Moreover, the project also strives to implement a visualization system capable of visualizing both the raw unstructured data and the final structured data to better inform the clinicians on how the data is processed.

Task Description

- Literature review on SOTA works working on ICU data [8,9,10,11,12,13], with a focus on the strategies used to preprocess raw ICU data.
- Develop a pipeline concept to process unstructured raw ICU data [1,2] (opt. [3]) into structured ones. The works in [6,7,12,13] can be used as references.
- Implement the concept pipeline in Python. The pipeline should ideally be compatible with the framework in [7].
- Implement a visualization tool to intuitively visualize the processed data.
- Write a final technical report/documentation.
- Present the concept and implementations.

Technical Prerequisites

- Intermediate or advanced programming experience with Python3.
- Basic knowledge on database (SQL).
- (Optional) Experience with the following python libraries: Matplotlib, Plotly, OpenCV, scikit-learn.

References

[1] Johnson, A., Pollard, T., Badawi, O., & Raffa, J. (2021). eICU Collaborative Research Database Demo (version 2.0.1). PhysioNet. https://doi.org/10.13026/4mxk-na84.
[2] Kallfelz, M., Tsvetkova, A., Pollard, T., Kwong, M., Lipori, G., Huser, V., Osborn, J., Hao, S., & Williams, A. (2021). MIMIC-IV demo data in the OMOP Common Data Model (version 0.9). PhysioNet. https://doi.org/10.13026/p1f5-7x35.
[3] Yèche, H., Kuznetsova, R., Zimmermann, M., Hüser, M., Lyu, X., Faltys, M., & Ratsch, G. (2021). HiRID-ICU-Benchmark---A Comprehensive Machine Learning Benchmark on High-resolution ICU Data.
[6] https://ohdsi.github.io/CommonDataModel/index.html

[7] Jarrett, D., Yoon, J., Bica, I., Qian, Z., Ercole, A., & van der Schaar, M. (2020, September). Clairvoyance: A pipeline toolkit for medical time series. In International Conference on Learning Representations.
[8] Sato, N., Uchino, E., Kojima, R., Hiragi, S., Yanagita, M., & Okuno, Y. (2021). Prediction and visualization of acute kidney injury in intensive care unit using one-dimensional convolutional neural networks based on routinely collected data. Computer Methods and Programs in Biomedicine, 206, 106129.
[9] Qian, Q., Wu, J., Wang, J., Sun, H., & Yang, L. (2021). Prediction Models for AKI in ICU: A Comparative Study. International Journal of General Medicine, 14, 623.
[10] Purushotham, S., Meng, C., Che, Z., & Liu, Y. (2018). Benchmarking deep learning models on large healthcare datasets. Journal of biomedical informatics, 83, 112-134.
[11] Harutyunyan, H., Khachatrian, H., Kale, D. C., Ver Steeg, G., & Galstyan, A. (2019). Multitask learning and benchmarking with clinical time series data. Scientific data, 6(1), 1-18.
[12] Tang, S., Davarmanesh, P., Song, Y., Koutra, D., Sjoding, M. W., & Wiens, J. (2020). Democratizing EHR analyses with FIDDLE: a flexible data-driven preprocessing pipeline for structured clinical data. Journal of the American Medical Informatics Association, 27(12), 1921-1934.
[13] Wang, S., McDermott, M. B., Chauhan, G., Ghassemi, M., Hughes, M. C., & Naumann, T. (2020, April). Mimic-extract: A data extraction, preprocessing, and representation pipeline for mimic-iii. In Proceedings of the ACM Conference on Health, Inference, and Learning (pp. 222-235).