Project Proposal
CS598 DL4H Spring 2023

Shiyu (Sherry) Li and Wei-Lun (Will) Tsai
{shiyuli2, wltsai2}@illinois.edu

Group ID: 53
Paper ID: 28
Presentation link: #TODO
Code link: https://github.com/willtsai/dlh-sp23-team53

1 Cite the original paper.

The paper we have selected for our reproducibility study is: *Context-aware Health Event Prediction via Transition Functions on Dynamic Disease Graphs* (Lu et al., 2022)

2 State the general problem the paper aims to solve. Do not use the same language as the paper.

The authors of the paper propose a new deep learning model called Chet (**c**ontext-aware **h**ealth **e**vent prediction via **t**ransition functions on dynamic disease graphs) that leverages the relationship between diseases and how they develop over time to predict future outcomes and diagnoses. Existing research on deep learning models for classification and prediction of diseases based on longitudinal EHR data have modeled disease diagnoses as independent events in their respective visits. However, intuition and data indicate that there are in fact hidden patterns within the combinations of disease diagnoses that may be useful for predicting future outcomes for patients, but yet have not been leveraged in existing best-in-class healthcare deep learning models. The Chet model is able to learn how diagnosed diseases develop over the course of each patient's doctor visits and then utilize this learned disease combination context to predict future outcomes and diagnoses.

3 Describe the new and specific approach taken by the paper. Discuss why it is interesting or innovative.

TODO: Sherry

4 Identify the specific hypotheses you plan to verify in your reproduction study.

In our reproduction study, we will use the same methodology as proposed by the authors for data selection, cleaning, and preprocessing as the authors of the paper. Specifically, we will join the MIMIC-III and MIMIC-IV datasets along the same overlapping time ranges that the authors describe and split training/validation/test sets randomly using the same ratios they used. We will build the diagnosis graphs and calculate the adjacency matrices for their corresponding subgraphs using the same methodology described by the authors. Leveraging the same parameter settings specified, we will train the model and then compare the performance of the Chet model to the listed baseline performance of best-in-class models for diagnosis prediction and heart failure prediction. We will not attempt to reproduce the baseline performance of best-in-class models.

5 Outline any additional ablations you plan to do and explain why they are interesting.

TODO: Sherry

6 Explain how you have access to the necessary data.

For their experiment, the authors use the MIMIC-III (Johnson et al., 2023b) and MIMIC-IV (Johnson et al., 2023a) datasets for training, validation, and testing. Since we have completed the required training and are credentialed users, we are able to download both the MIMIC-III and MIMIC-IV datasets from PhysioNet (Goldberger et al., 2000 (June 13).

7 Discuss the computational feasibility of your proposed work.

TODO: Sherry (Will to also experiment with the code)

8 Specify if you will be re-using existing code and provide a link to it, or if you will implement the code yourself.

We plan to re-use the existing code provided by the authors of the paper for data preprocessing, building the needed input graphs/matrices, building the model architecture, and training the model for the experiment. We will author our own shell scripts to download, unzip, and move datasets into the appropriate paths. We also plan to evelop our own code to compute and evaluate model performance based on the test dataset following the metrics and methodology detailed in the paper, as well as make the required modifications to the existing model architecture and training code to perform the ablation studies mentioned in the paper as well as our own extra ablations. The code provided by the authors is available at `www.github.com/LuChang-CS/Chet` (Lu et al., 2022), we will clone this as reference for our implementations described above.

References

A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. 2000 (June 13). Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220. Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. 2023a. Mimic-iv.

Alistair Johnson, Tom Pollard, and Roger Mark. 2023b. Mimic-iii clinical database.

Chang Lu, Tian Han, and Yue Ning. 2022. Context-aware health event prediction via transition functions on dynamic disease graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(4):4567–4574.