Deep Learning for Electronic Health Records

David Chang
April 11, 2018

Abstract

The amount of data stored in electronic health records (EHR) systems has grown tremendously in the past decade, leading to rapid advancement in the development of a wide variety of methods for performing clinical informatics tasks. The concurrent progress in machine learning, particularly deep learning, has motivated researchers to apply deep learning methods to EHR data on tasks including patient trajectory modeling, clinical decision support systems, medical concept extraction, outcome prediction, computational phenotyping, etc.

Several properties of EHR data make the application of deep learning methods challenging, such as data heterogeneity (many different data types), irregular measures (different time scales), and the information contained in unstructured data (i.e. free text, which requires NLP techniques). For uncommon diseases or conditions, the sample sizes are often insufficient for deep learning models. Furthermore, the issue of interpretability and model transparency are extremely important in the clinical domain, since predictions can be used to make decisions that affect patients. Most applications of deep learning for EHR data so far involve modifying preexisting deep learning models to address task-specific concerns and the more general EHR-related challenges mentioned above.

In this presentation, I will be discussing two models: ehrGAN (Che et al. 2017) and deep counterfactual network with propensity dropout (DCN-PD) (Alaa et al. 2017). ehrGAN is a modified GAN that can learn to generate plausible EHR data in order to augment the training dataset in a semi-supervised learning approach and used in combination with a CNN-based prediction model to improve prediction performance and generalization. DCN-PN is a novel approach for inferring individualized causal effects of treatment from observational data by modeling a subject's potential outcomes using a deep multitask network and using a propensity-dropout regularization scheme to account for selection bias.

Readings

A. Alaa, M. Weisz, and M. van der Schaar, "Deep counterfactual networks with propensity-dropout," *ICML Workshop on Principled Approaches to Deep Learning*, 2017. https://arxiv.org/abs/1706.05966

Che, Zhengping, Yu Cheng, Shuangfei Zhai, Zhaonan Sun, and Yan Liu, "Boosting Deep Learning Risk Prediction with Generative Adversarial Networks for Electronic Health Records," *arXiv preprint arXiv:1709.01648 [cs.LG]* https://arxiv.org/abs/1709.01648

Spotlight question

What are some ways you might approach dealing with the main challenges associated with applying deep learning on EHR?