Deep Learning in Health Informatics

Fei Wang

Department of Healthcare Policy and Research. Weill Cornell Medicine. New York City. NY. USA.,

Abstract

The Precision Medicine Initiative brings tremendous opportunities to speedup scientific discovery and promote quality improvement in medicine. However, it also raises big challenges in dealing with large and massive data from heterogeneous sources. Traditional data mining methods tend to favor clean and structured data, which may not be able to effectively utilize the rich information in biomedical data. The latest breakthrough in deep learning technologies provides a unique opportunity to retrieve information from complex and heterogeneous sources in an effective way. In this tutorial I will review the recent literature on applying deep learning models to healthcare data. Based on the analyzed work, I suggest that deep learning approaches could be the vehicle for translating big biomedical data into improved human health. But deep learning methods also have limitations, especially in terms of ease-of-understanding for domain experts and citizen scientists. I will discuss such challenges and suggest developing meaningful contextualized representations to bridge deep learning models and human interpretability.

Keywords:

Deep Learning, Health Informatics

Introduction

Health is closely related with everyone's daily life. Recently, with the rapid development of computer related technologies, more and more health related data are becoming readily available. Artificial intelligence related methods and algorithms are gradually being applied in helping analyze those healthcare data and improve healthcare quality. Deep learning is one popular AI technology that has been proven to be useful in a lot of different application areas. In this tutorial I will review recent and forthcoming applications of deep learning in medicine, highlighting the key aspects to significantly impact healthcare. I will start with general introductions on health data, deep learning architectures that could be useful for analyzing health data. Then I will introduce with detailed examples on how deep learning models can be applied in analyzing health data including electronic health records, genomic data and mobile and sensor data, with detailed explanations on what are the typical health problems associated with those data. I will conclude the whole tutorial with discussions and future directions.

Duration

3 Hours

Outline

- Introduction: deep learning and health data (60 min)
  - The current status of healthcare
  - What is health informatics
  - Examples of health informatics problems
    - Predictive Modeling
    - Cohort Identification
    - Pharmacovigilance
  - Health Data
    - Electronic Health Records
    - Genetic Data
    - Mobile and Sensor Data
  - What is deep learning
  - Typical Architectures of Deep Learning Models
    - Deep Belief Network
    - Restricted Boltzman Machine
    - Autoencoders
    - Convolutional Neural Network
    - Recurrent Neural Network
- Deep Learning with Electronic Health Records (EHR) (40 min)
  - Deep Representation Learning with EHR
    - Word2Vec, Doc2Vec, Med2Vec
    - Autoencoders

  - Deep Predictive Modeling with EHR
    - Recurrent Neural Network
    - Long Short-Term Memory
  - Deep Patient Similarity Learning with EHR
    - Structured Recurrent Neural Network
- Deep Learning with Genetic Data (40 min)
  - Prediction the Splicing Activity of Individual Exons with Deep Neural Network
  - DNA Sequencing Study with Convolutional Neural Network
    - DeepBind
    - Basset
  - Gene Expression Analysis with Autoencoders
- Deep Learning with Mobile and Sensor Data (30 min)
  - Deep Learning with EEG Data
  - DeepEar: Deep Learning with Audio Sensing Data
  - Deep Learning with Physiological Signals in ICU
- Discussions and Future Directions (10 min)

Presenter

Fei Wang is an Assistant Professor in Division of Health Informatics, Department of Healthcare Policy and Research, Cornell University. His major research interest is data analytics and its applications in health informatics. His papers have received over 4,200 citations so far with an H-index 35. His papers won paper award runner-up in ICDM 2016, best short paper award in ICHI 2016, best student paper in ICDM 2015, best research paper nomination in ICDM 2010, Marco Romani Best paper nomination in AMIA TBI 2014, and his paper was selected into the best paper finalist in SDM 2011 and 2015. Dr. Wang is an action editor of the journal Data Mining and Knowledge Discovery, an associate editor of Journal of Health Informatics Research and Smart Health, and an editorial board member of Pattern Recognition and International Journal of Big Data and Analytics in Healthcare. Dr. Wang has successfully organized numerous workshop series, including "Data Mining for Medical Informatics" in conjunction with AMIA annual symposium from 2014 to 2016 and "Data Mining for Medicine and Healthcare" in conjunction with SDM from 2013 to 2017. Dr. Wang is the vice chair of the KDD working group in AMIA.

Dr. Wang has done lots of research on deep learning and health [1-4]. Moreover, in one of his recent effort, Dr. Wang won the Parkinson's Progression Markers Initiative (PPMI) data challenge on subtyping Parkinson's disease organized by Michael J. Fox Foundation. The press release can be found at https://www.michaeljfox.org/foundation/publication-detail.html?id=625&category=7. The model Dr. Wang used is Long-Short Term Memory (LSTM) [5], which is a popular variant of Recurrent Neural Network (RNN) [6], that has been applied successfully in many real world applications, expecially Natural Language Processing (NLP).

Dr. Wang has extensive experience on presenting tutorials. The tutorials he has given include "Computational Phenotyping" in AMIA Annual Symposium 2016 (3 hours), "Computational Phenotyping from Massive Clinical Data" in IEEE International Conference on Health Informatics 2016 (6 hours), "Healthcare Data Mining with Matrix Models" in ACM KDD 2016 (3 hours) and SIAM Data Mining Conference (2 hours), "Feature Engineering in Healthcare Analytics" in AAAI 2014 (4 hours) and CIKM (3 hours), etc. All tutorials are well received.

Audience

All the researchers and practitioners engaged in data analytics and health informatics are welcome. Basic knowledge on deep learning is preferred.

Acknowledgements

Part of Fei Wang's work is supported by National Science Foundation under number IIS-1650723.

References

[1] Z. Zhu, C. Yin, B. Qian, Y. Cheng, J. Wei, F. Wang. Measuring Patient Similarities via A Deep Architecture with Medical Concept Embedding. In *Proceedings of 2016 IEEE International Conference on Data Mining (ICDM)*. 2016.

[2] Y. Cheng, F. Wang, P. Zhang, J. Hu. Risk Prediction with Electronic Health Records: A Deep Learning Approach. In *Proceedings of 2016 SIAM International Conference on Data Mining (SDM)*. 2016.

[3] C. Che, C. Xiao, J. Liang, B. Jin, J. Zhou, F. Wang. An RNN Architecture with Dynamic Temporal Matching for Personalized Predictions of Parkinson's Disease. In *Proceedings of 2016 SIAM International Conference on Data Mining (SDM)*. 2016.

[4] J. Liang, R. Lu, C. Zhang, F. Wang. Predicting Seizures from Electroencephalography Recordings: A Knowledge Transfer Strategy. In *Healthcare Informatics (ICHI), 2016 IEEE International Conference on*. 2016. pp. 184-191. IEEE.

[5] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural computation, 9(8), pp.1735-1780. 1997.

[6] L. R. Medsker and L. C. Jain. "Recurrent neural networks." Design and Applications. 2001.

Address for correspondence

Fei Wang is with 425 East 61 Street, New York City, NY 10065. USA. Email: few2001@med.cornell.edu