Yu Deng

CONTACT INFORMATION

Feinberg School of Medicine, 303 E Superior street, Chicago, IL.

Email:yudeng2015@u.northwestern.edu
Mobil: 3124839916

RESEARCH INTEREST

I am a PhD student majoring in biomedical informatics. My focus is on applying supervised and unsupervised learning methods for clinical research. I have applied various machine learning techniques to find disease subtypes and novel risk factors. I have used deep learning/joint model to improve prediction performance using longitudinal data as well as using Natural Language Processing (NLP) for computational phenotyping.

EDUCATION

Northwestern University, Chicago, IL, USA
● PhD Candidate, biomedical informatics, Feinberg School of Medicine, expected graduation June 2020
● Major GPA: 3.79/4.00
● Coursework: Programming for Big Data, Advanced Biostatistics, Deep Learning from Scratch

INTERNSHIP

Data Science Intern, National Institute of Health (NIH), Bethesda, MD, USA, 2019
● Internship at Dr. Zhiyong Lu's text mining group, National Center for Biotechnology Information, NIH
● Developing/analyzing full text chemical corpus for downstream analysis such as name entity recognition
● Developed customized CNN architectures utilizing longitudinal data for cardiovascular disease prediction

SKILLS

Programming Skills
● Python, R, SQL
Machine Learning Algorithms
● **Deep Learning** (MLP, CNN, RNN), Classical & Penalized Regression Methods (LASSO, Ridge), SVM, Random Forest, K-nearest Neighbors, Adaboosting, Cox regression; **K-means**, hierarchical clustering, Partition Around Medoid, Latent Class Analysis, non-Negative Tensor Factorization (NTF)
● Feature Engineering (e.g. PCA, tensor factorization, forward/backward selection), Regularization

SELECTED PROJECTS

Developed CNN/RNN-survival Model for Cardiovascular Disease (CVD) Prediction, 2019
● Developed CNN and RNN based models to predict CVD outcome by using longitudinal data and considering:
1). Missing data from irregular visits 2) Subject drop-out 3). Correlation within repeated measurements
● Designed 3 convolution layers with convolution kernels convolved with the layer input over the temporal dimension.
● Explore different parameter settings include learning rate, number of hidden layers, epoch, activation function, early stop.
● improved AUC by 3% compared to cox baseline model.

Use of Clinical Phenotypes and Non-negative Tensor Factorization for Heart Failure (HF) Prediction, 2017
● Performed NTF on large scale, sparse medical record data; Generated latent clusters
● Performed dimension reduction including NTF, MTF, PCA on medical record data to get important features
● Used the output of different feature reduction techniques as the input features of logistic regression.

Use of NLP to Improve Systemic Lupus Erythematosus Criteria Identification in Electronic Health Records, 2019
● Used Metamap to extract name entities, used l1/l2 penalized logistic regression used for feature selection and prediction.
● Evaluated model performance using sensitivity and specificity. NLP based algorithm improved renal sensitivity from 0.4 to 0.7.

SELECTED PUBLICATIONS/CONFERENCES

Deep Neural Network Survival Model on Prediction of Cardiovascular Disease Using Longitudinal Measurements
Yu Deng, Yifan Peng, Yishu Wei, Lihui Zhao, Zhiyong Lu. (2019). NIH poster day

Use of Clinical Phenotypes and Non-negative Tensor Factorization for Heart Failure Prediction
Deng Y., Furmanchuk A., Chen R., Ahmad F., Sun J., Kho A.(2017). *AMIA Annual Symposium. Podium Presentation*

Use of Repeated Measurements for Cardiovascular Disease Prediction: the Application of Joint Model
Deng Y., Zhong Y., Kho A., Zhao L.(2019). ENAR 2019. Poster presentation

Identification of Systemic Lupus Erythematosus Subtypes using Latent Class Analysis
Deng Y., Ghosh A., Luo Y., Kho A., Goldman R., Walunas T.(2019). Medinfo. Podium Presentation

Natural Language Processing for EHR-Based Computational Phenotyping
Zeng, Z., **Deng, Y.**, Li, X., Naumann, T., & Luo, Y. (2018). Natural Language Processing for EHR-Based Computational Phenotyping. arXiv preprint arXiv:1806.04820.

Characterizing Design Patterns of EHR-Driven Phenotype Extraction Algorithms
Zhong, Y., Rasmussen L., **Deng Y.**, Pacheco J., Smith M., Starren J., Wei W., Walton N., Hripcsak G., Chute C., Luo Y. Characterizing Design Patterns of EHR-Driven Phenotype Extraction Algorithms. IEEE on Bioinformatics and Biomedicine

Measuring Nascent Transcripts by Nascent-seq. Methods in Molecular Biology
Chen, F. X., Marshall, S. A., **Deng, Y**., & Tianjiao, S. (2018). Measuring Nascent Transcripts by Nascent-seq. In Next Generation Sequencing (pp. 19-26). Humana Press, New York, NY.

HONOURS/ DISTINCTIONS

Honorable Mention in Student Poster Competition, International Chinese Statistical Association (ICSA), 2018
First Prize in Student Poster Competition, Northwestern Biomedical Informatics Day, 2017
Driskill Graduate Program Full Scholarship, Northwestern University 2015 – 2017