Zichen Wang, PhD

📍 Edgewater, NJ 07020 • ✉ wangzc921@gmail.com • 📱 +1 (917) 826-8235
🏠 https://wangz10.github.io • ⭘ wangz10 • 🔗 zichenwang • Ⓜ wangz10

EMPLOYMENT

Amazon Web Services, New York, NY

- Applied Scientist   Apr 2021 – present
  - ML research and applications.

Sema4, Stamford, CT

- Principal Scientist   Sep 2019 – Apr 2021
  - Performed an observational study using the electronic health records (EHR) from hospitalized COVID-19 patients using survival analysis [paper] .
  - Developed digital phenotyping algorithms to identify the patients with pregnancy complications and neonatal disorders using longitudinal EHR.
  - Created a ML pipeline (gradient boosting machine) to predict the disease risks for patients.
  - Developed clinical NLP pipeline to improve the digital phenotyping algorithm.

Icahn School of Medicine at Mount Sinai, New York, NY

- Research Assistant Professor   Jun 2017 – Sep 2019
  - Developed a novel graph-based dimensionality reduction algorithm with comparable performance to t-SNE to visualize drug-induced gene expression data and implemented a web application L1000FWD to explore the data manifold.
  - Implemented a semi-supervised VAE model to predict biological activities for molecules and to generate novel compounds with desired properties.
  - Integrated gene-centric datasets from 3 distinctive sources (biomedical literature, RNA-seq datasets and ontologies) using topic model (latent Dirichlet allocation) and word embeddings (Word2vec and GloVe).
- Postdoctoral Fellow   Jan 2017 – Jun 2017
  - Developed a deep neural network leveraging clinical variables from the EHR to predict physiological age.

EDUCATION

Icahn School of Medicine at Mount Sinai, New York, NY

- Ph.D. in Computational Biology   Aug 2012 – Dec 2016
  - Research topics: gene expression, systems pharmacology, machine learning, software development

China Agricultural University, Beijing, China

- B.S. in Biochemistry and Molecular Biology   Sep 2008 – Jun 2012

PROJECTS

DEEP LEARNING

- **Contrastive representation learning**: benchmarked various contrastive loss functions used for representation learning.
- **Deep generative models**: performed latent interpolation experiments for VAE and GANs.

SOFTWARE DEVELOPMENT

- **react-scatter-board**: a reusable React library for interactive THREE.js 2d/3d scatter plots.
- **PAEA**: a web-based multivariate gene set enrichment analysis tool.

SKILLS

- Programming languages: Python, R, JavaScript, SQL, php, MATLAB
- ML Frameworks: Tensorflow, Keras, PyTorch, Numpy/Scipy, Pandas, Scikit-learn
- Web Development: Flask/Python, Shiny/R, React.js, d3.js, THREE.js
- Platforms: AWS (EC2, S3, RedShift), Jupyter/SageMaker, Docker, MongoDB, Apache Mesos, Hadoop

SELECTED PUBLICATIONS

[1] Z. Wang, A. Lachmann, A. Keenan and A. Ma'ayan: "L1000FWD: fireworks visualization of drug-induced transcriptomic signatures" *Bioinformatics*, vol. 34, no. 12, pp. 2150–2152, (2018).

[2] Z. Wang, L. Li, B. Glicksberg et al: "Predicting age by mining electronic medical records with deep learning characterizes differences between chronological and physiological age" *Journal of Biomedical Informatics*, vol. 76, pp. 59–68, (2017).

[3] Z. Wang, C. Monteiro, K. Jagodnik et al: "Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd" *Nature Communications*, vol. 7, (2016).

[4] Z. Wang, N. Clark, A. Ma'ayan: "Drug-induced adverse events prediction with the LINCS L1000 data" *Bioinformatics*, vol. 32, no. 15, pp. 2338–2345 (2016).