YIFENG TAO

⚲ MI 654F, Carnegie Mellon University, Pittsburgh, PA, 15213

📞 +1 (412) 320-1617 ⋄ ✉ yifengt@cs.cmu.edu ⋄ 👤 http://www.cs.cmu.edu/~yifengt

RESEARCH INTEREST

I am interested in developing machine learning, phylogenetic and text mining methods in cancer genomics by incorporating domain knowledge for precision medicine and personalized treatment.

Keywords: Computational Biology; Machine Learning; Cancer Genomics; Natural Language Processing.

EDUCATION

Carnegie Mellon University (GPA: 4.0/4.0)

- Ph.D. in Computational Biology, School of Computer Science   *Aug 2016 - Aug 2021 (expected)*
  Thesis: Genome-Driven Personalized Medicine of Cancer via Machine Learning and Phylogenetic Models
- M.Sc. in Machine Learning, School of Computer Science   *Aug 2016 - Dec 2018*

Tsinghua University (GPA: 91.3/100)

- B.Eng. in Automation (with honors), School of Information Science and Technology   *Aug 2012 - Aug 2016*
- B.Ec. in Economics (double major), School of Economics and Management   *Aug 2013 - Aug 2016*

RESEARCH EXPERIENCE

Carnegie Mellon University   Aug 2018 - Present
Research Assistant, Advisor: Prof. Russell S. Schwartz   *Pittsburgh, PA*

- Research on developing machine learning and phylogenetic methods for precise prediction of cancer progression.

Carnegie Mellon University   Jul 2016 - Jul 2018
Research Assistant, Advisors: Prof. William W. Cohen and Prof. Xinghua Lu   *Pittsburgh, PA*

- Research on contextual deep learning for gene/tumor embedding representation and phenotype prediction.

Tsinghua University   May 2015 - Aug 2016
Research Assistant, Advisor: Prof. Jianyang Zeng   *Beijing, China*

- Research on the rational protein design of CRISPR-Cas9 system to reduce off-target effect.

Stanford University   Jul 2015 - Sep 2015
Research Assistant, Advisor: Prof. Kerwyn Casey Huang   *Palo Alto, CA*

- Research on the MreB protein in bacterial cell shape determination through molecular dynatic simulation.

University of California San Diego   Jul 2014 - Aug 2014
Research Assistant, Advisor: Prof. Kun Zhang   *La Jolla, CA*

- Programming the fluorescence microscope for automated single-cell imaging and analysis.

EMPLOYMENT

Illumina, Inc.   June 2020 - Aug 2020
Data Scientist Intern, Mentor: Dr. Kimberly Gietzen   *San Diego, CA*

- Research and development of deep learning models for genotyping image anomaly detection and classification.

Roam Analytics, Inc.   May 2018 - Aug 2018
Machine Learning Research Intern, Advisor: Prof. Christopher Potts   *San Mateo, CA*

- Research on hybrid feature representation to improve the clinical text sequence labeling effectively.

FELLOWSHIPS AND AWARDS

- **PSB Travel Award**, *PSB; National Library of Medicine, NIH* — 2020
- **CMLH Fellowship in Digital Health**, *Center for Machine Learning and Health* — 2019 - 2020
  Fellowship awarded annually for $70,000 for tuition and stipend
- **Best Poster Award**, *CPCB Program in Computational Biology* — 2018
- **UGVR (UnderGraduate Visiting Research) Program**, *Stanford University* — 2015
- **Finalist in Interdisciplinary Contest in Modeling**, *COMAP* — 2015
- **Scholarship for Academic Excellence**, *Tsinghua University* — 2014, 2015
- **Spark Program Fellowship for Technology Innovation**, *Tsinghua University* — 2014
- **First Prize in National Physics Olympiad**, *Chinese Physical Society* — 2012
- **Huaying Elite Fellowship**, *Huaying Education Foundation* — 2012

PUBLICATIONS

[1] **Yifeng Tao**, Ashok Rajaraman, Xiaoyue Cui, Ziyi Cui, Haoran Chen, Yuanqi Zhao, Jesse Eaton, Hannah Kim, Jian Ma, and Russell Schwartz. Tumor mutational phenotypes account for a substantial portion of progression risk under various confounding environmental factors. 2020.

[2] **Yifeng Tao**, Haoyun Lei, Adrian V. Lee, Jian Ma, Russell Schwartz. Neural network deconvolution method for resolving pathway-level progression of tumor clonal expression programs with application to breast cancer brain metastases. *Frontiers in Physiology.* 2020.

[3] **Yifeng Tao**, Shuangxia Ren, Michael Q. Ding, Russell Schwartz, and Xinghua Lu. Predicting drug sensitivity of cancer cell lines via collaborative filtering with contextual attention. *Proceedings of the Machine Learning for Healthcare Conference (MLHC).* 2020.

[4] **Yifeng Tao**, Haoyun Lei, Xuecong Fu, Adrian V. Lee, Jian Ma, and Russell Schwartz. Robust and accurate deconvolution of tumor populations uncovers evolutionary mechanisms of breast cancer metastasis. *Proceedings of the Intelligent Systems for Molecular Biology (ISMB). Bioinformatics 36: i407-i416.* 2020.

[5] Haoyun Lei, E. Michael Gertz, Alejandro A. Schäffer, Xuecong Fu, **Yifeng Tao**, Kerstin Heselmeyer-Haddad, Irianna Torres, Xulian Shi, Kui Wu, Guibo Li, Liqin Xu, Yong Hou, Michael Dean, Thomas Ried, and Russell Schwartz. Tumor heterogeneity assessed by sequencing and fluorescence *in situ* hybridization (FISH) data. *bioRxiv 2020.02.29.970392.* 2020.

[6] **Yifeng Tao**, Chunhui Cai, William W. Cohen, and Xinghua Lu. From genome to phenome: Predicting multiple cancer phenotypes based on somatic genomic alterations via the genomic impact transformer. *Proceedings of the Pacific Symposium on Biocomputing 25:79-90 (PSB).* 2020.

[7] **Yifeng Tao**, Ashok Rajaraman, Xiaoyue Cui, Ziyi Cui, Jesse Eaton, Hannah Kim, Jian Ma, and Russell Schwartz. Improving personalized prediction of cancer prognoses with clonal evolution models. *bioRxiv 761510.* 2019.

[8] **Yifeng Tao**, Haoyun Lei, Adrian V. Lee, Jian Ma, and Russell Schwartz. Phylogenies derived from matched transcriptome reveal the evolution of cell populations and temporal order of perturbed pathways in breast cancer brain metastases. *Proceedings of the International Symposium on Mathematical and Computational Oncology 3-28 (ISMCO).* 2019.

[9] **Yifeng Tao**, Bruno Godefroy, Guillaume Genthial, and Christopher Potts. Effective feature representation for clinical text concept extraction. *Proceedings of the Clinical Natural Language Processing Workshop 1-14 (NAACL-ClinicalNLP).* 2019.

[10] Haohan Wang, Xiang Liu, **Yifeng Tao**, Wenting Ye, Qiao Jin, William W. Cohen, and Eric P. Xing. Automatic human-like mining and constructing reliable genetic association database with deep reinforcement learning. *Proceedings of the Pacific Symposium on Biocomputing 24:112-123 (PSB).* 2019.

PATENTS

[1] Xiaoyue Cui, Ziyi Cui, Jian Ma, Ashok Rajaraman, Russell Schwartz, and **Yifeng Tao**. Phylogenetic models for predicting cancer progression. US Patent. Pending.

PROFESSIONAL SERVICE

- Reviewer for Pacific Symposium of Biocomputing (PSB)   2020
- Reviewer for Frontiers of Engineering Management (FEM)   2020
- Reviewer for International Conference on Control, Automation and Systems (ICCAS)   2020
- Secondary reviewer for RECOMB Computational Cancer Biology (RECOMB-CCB)   2020
- Reviewer for PLOS Computational Biology   2019
- Reviewer for International Conference on Research in Computational Molecular Biology (RECOMB)   2019
- Reviewer for International Journal of Production Research (IJPR)   2019
- Secondary reviewer for International Conference on Intelligent Systems for Molecular Biology (ISMB)   2019
- Secondary reviewer for Workshop on Algorithms in Bioinformatics (WABI)   2019
- Secondary reviewer for IEEE International Conference on Bioinformatics and Biomedicine (BIBM)   2018

TEACHING

Instructor

- Introduction to Machine Learning, Northeastern University (China)   2019

Teaching Assistant

- Probabilistic Graphical Models, Carnegie Mellon University   2018
- Convex Optimization, Carnegie Mellon University   2017

COURSEWORK

Machine Learning (Ph.D. level)

- Advanced Introduction to Machine Learning (10-715; A)
- Intermediate Statistics (10-705; A+)
- Statistical Machine Learning (10-702; A+)
- Convex Optimization (10-725; A)
- Probabilistic Graphical Models (10-708; A+)

Computational Biology (Ph.D. level)

- Automation of Biological Research (02-750; A+)
- Computational Genomics (02-710; A)
- Cell and Systems Modeling (02-730; A)
- Introduction to Computational Structural Biology (PT-749; A+)
- Advanced Genetics (03-730; A)
- Laboratory Methods for Computational Biologists (02-760; A)

PROGRAMMING LANGUAGES

Research languages: Python (NumPy, pandas, scikit-learn), Matlab, R
Deep learning packages: PyTorch, TensorFlow
Miscellaneous: LaTeX; Web (HTML, CSS, Bootstrap); bash; Vim; C/C++; SQL; AWS