Junwen Bai

Curriculum Vitae

Research Interests

I am interested in the general areas of machine learning and language technology, with research focuses on sequence representation learning and probabilistic modeling, often under scenarios with low-supervision. I have developed scalable and general machine learning methods for real-world problems including automatic speech recognition, climate change and scientific discovery.

Education

2017–present **Ph.D. Student, Computer Information and Science**.
Cornell University, Ithaca, NY
Advisors: Carla Gomes, Bart Selman
GPA: 4.00/4.30

2013–2017 **Bachelor of Science, Computer Science and Technology**.
ACM Class, Shanghai Jiao Tong University
Advisors: Hongtao Lu, Yong Yu
GPA: 4.00/4.30

Publications

* for equal contribution.

Conferences:

2022 **Gaussian Mixture Variational Autoencoder with Contrastive Learning for Multi-Label Classification**.
Junwen Bai, Shufeng Kong, Carla Gomes.
International Conference on Machine Learning (ICML), 2022.
a workshop version was presented at NeurIPS Workshop on Deep Generative Models and Downstream Applications, 2021

2022 **Joint Unsupervised and Supervised Training for Multilingual ASR**.
Junwen Bai, Bo Li, Yu Zhang, Ankur Bapna, Nikhil Siddhartha, Khe Chai Sim, Tara N. Sainath.
International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2022

2022 **A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction**.
Joshua Fan*, **Junwen Bai***, Zhiyun Li*, Ariel Ortiz-Bobea, Carla Gomes.
AAAI Conference on Artificial Intelligence (AAAI), 2022
a workshop version won **Best ML Innovation Paper** award at NeurIPS workshop on Tackling Climate Change with Machine Learning, 2021

344 Gates Hall, 107 Hoy Rd, Cornell University – Ithaca, NY 14853
☎ (+1) 607-262-1268 • ✉ jb2467@cornell.edu

2021 **Scaling End-to-End Models for Large-Scale Multilingual ASR**.
Bo Li, Ruoming Pang, Tara N. Sainath, Anmol Gulati, Yu Zhang, James Qin, Parisa Haghani, W. Ronny Huang, Min Ma, **Junwen Bai**.
IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2021

2021 **Contrastively Disentangled Sequential Variational Audoencoder**.
Junwen Bai, Weiran Wang, Carla Gomes.
Advances In Neural Information Processing Systems (NeurIPS), 2021

2021 **Representation Learning for Sequence Data with Deep Autoencoding Predictive Components**.
Junwen Bai, Weiran Wang, Yingbo Zhou, Caiming Xiong.
International Conference on Learning Representations (ICLR), 2021.

2021 **HOT-VAE: Learning High-Order Label Correlation for Multi-Label Classification via Attention-Based Variational Autoencoders**.
Wenting Zhao, Shufeng Kong, **Junwen Bai**, Daniel Fink, Carla Gomes
AAAI Conference on Artificial Intelligence (AAAI), 2021

2020 **Disentangled Variational Autoencoder based Multi-Label Classification with Covariance-Aware Multivariate Probit Model**.
Junwen Bai, Shufeng Kong, Carla Gomes.
International Joint Conference on Artificial Intelligence - Pacific Rim International Conference on Artificial Intelligence (IJCAI-PRICAI), 2020 **(Acceptance rate: 12.6%)**

2020 **Deep Hurdle Networks for Zero-Inflated Multi-Target Regression: Application to Multiple Species Abundance Estimation**.
Shufeng Kong, **Junwen Bai**, Jae Hee Lee, Di Chen, Andrew Allyn, Michell Stuart, Malin Pinsky, Kathy Mills, Carla Gomes.
International Joint Conference on Artificial Intelligence - Pacific Rim International Conference on Artificial Intelligence (IJCAI-PRICAI), 2020 **(Acceptance rate: 12.6%)**

2019 **SWALP: Stochastic Weight Averaging in Low Precision Training**.
Guandao Yang, Tianyi Zhang, Polina Kirichenko, **Junwen Bai**, Andrew Wilson, Chris De Sa
International Conference on Machine Learning (ICML), 2019

2019 **Imitation Refinement For X-Ray Diffraction Signal Processing**.
Junwen Bai, Zihang Lai, Runzhe Yang, Yexiang Xue, John Gregoire, Carla P. Gomes
International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2019

2018 **An Efficient Relaxed Projection Method for Constrained Non-negative Matrix Factorization with Application to the Phase-Mapping Problem in Materials Science**.
Junwen Bai, Sebastian Ament, Guillaume Perez, John M. Gregoire, Carla P. Gomes
International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR), 2018

2017 **Relaxation Methods for Constrained Matrix Factorization Problems: Solving the Phase Mapping Problem in Materials Discovery**.
Junwen Bai, Johan Bjorck, Yexiang Xue, Santosh K. Suram, John M. Gregoire, Carla P. Gomes
International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR), 2017

2017 **Phase-Mapper: An AI Platform to Accelerate High Throughput Materials Discovery**.
Yexiang Xue, **Junwen Bai**, Ronan Le Bras, Brendan Rappazo, Richard Bernstein, Johan Bjorck, Liane Longpre, Santosh K. Suram, Robert B. van Dover, John Gregoire, Carla P. Gomes
AAAI Conference on Artificial Intelligence (AAAI), 2017 **(Innovative Application Award)**

***Journals*:**

2019 **CRYSTAL: a Multi-agent AI System for Automated Mapping of Materials' Crystal Structures**.
Carla P. Gomes, **Junwen Bai**, Yexiang Xue, Johan Björck, Brendan Rappazzo, Sebastian Ament, Richard Bernstein, Shufeng Kong, Santosh K Suram, Robert Bruce van Dover, John M Gregoire. Materials Research Society (MRS) Communications, 2019 **(#4 of Top 10 Coolest Army Science and Technology Advances of 2019)**

2018 **Phase Mapper: Accelerating Materials Discovery with AI**.
Junwen Bai, Yexiang Xue, Johan Bjorck, Ronan Le Bras, Brendan Rappazzo, Richard Bernstein, Santosh K. Suram, Robert Bruce van Dover, John M. Gregoire, Carla P. Gomes
AI Magazine 39(1): 15-26 2018 **(cover story)**

2016 **Automated Phase Mapping with AgileFD and its Application to Light Absorber Discovery in the V-Mn-Nb Oxide System**.
Santosh K. Suram, Yexiang Xue, **Junwen Bai**, Ronan Le Bras, Brendan Rappazzo, Richard Bernstein, Johan Bjorck, Lan Zhou, Robert B.van Dover, Carla P. Gomes, John M. Gregoire
ACS Combinatorial Science 19.1 (2016): 37-46 **(Editor's choice and the cover story)**

Professional Services

SPC: IJCAI '21

PC/reviewer: IJCAI '20, AAAI '21, IJCAI '21, ICML '21, NeurIPS '21 (Outstanding Reviewer), AAAI '22, SAS@AAAI '22, ICLR '22, ICASSP '22, IJCAI '22, AIforGood@IJCAI '22, ICML '22, NeurIPS '22

Journal reviewer: Journal of Chemometrics and Intelligent Laboratory Systems, Computational Materials Science, Transactions on Image Processing (TIP), Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Journal of Selected Topics in Signal Processing (JSTSP), Transactions on Machine Learning Research (TMLR), GeoInformatica

Session Chair: IJCAI '20

Judge: ENVISION WiSTEM competition

Mentor: Cornell University High School Programming Contest

Experience

Jun 2021 - Nov 2021 **Research Intern**, Google.
- Mentor: Bo Li, Yu Zhang
- Developed a massive multilingual automatic speech recognition (MMASR) system for multiple languages. The proposed model greatly outperformed the prior works and achieved the state-of-the-art performance on a widely adopted public multilingual ASR dataset.

May 2020 - Aug 2020 **Research Intern**, Salesforce Research.
- Mentor: Weiran Wang, Caiming Xiong
- Proposed a novel self-supervised learning method based on predictive information and masked reconstruction, which boosted the interpretability and predictability of the learnt representations and significantly improved automatic speech recognition performances.

June 2019 - Aug 2019 **PhD Intern**, *Ads ranking team*, Facebook.
- Developed an LSTM-based framework for a multi-task learning problem on click through rate, click through conversion rate and view through conversion rate. The performance w.r.t. normalized entropy (NE) improved by $\sim 0.5\%$.

Jul 2016 - June 2017 **Research Intern**, *Institute for Computational Sustainability*, Cornell University.
- Advisor: Carla Gomes
- Developed machine learning methods and the corresponding public online materials analysis platform "phase-mapper" to promote materials discovery.

Oct 2015 - Jun 2016 **Research Intern**, *Center for Brain-like Computing and Machine Intelligence*, Shanghai Jiao Tong University, China.
- Advisor: Hongtao Lu
- Improved deep learning model for computer vision tasks, such as scene classification and understanding.

Oct 2015 **Research Intern**, *National Institute of Informatics*, Tokyo, Japan.
- Advisor: Yuan Sun
- Statistical analysis for behavioral and emotional engagement and task performances of university students.

Talks

2017 **Junwen Bai**, Relaxation Methods for Constrained Matrix Factorization Problems. Presented at 2017 Doctoral Consortium on Computational Sustainability.

Skills

PyTorch, TensorFlow, Babelfish, Python, C++, Java

Teaching Experience

Aug 2021 - Dec 2021 **Head Teaching Assistant**, *CS 4700/5700: Foundations of Artificial Intelligence*, Cornell University.

Sept 2020 - Dec 2020 **Teaching Assistant**, *CS 4740/5740: Introduction to Natural Language Processing*, Cornell University.

Jul 2015 - Aug 2015 **Teaching Assistant**, *Programming Practice*, Shanghai Jiao Tong University.

Mar 2015 - Jun 2015 **Teaching Assistant**, *Data Structure*, Shanghai Jiao Tong University.

Awards

2022 Outstanding TA award

2022 AAAI 2022 student scholarship

2022 AI for Earth Microsoft Azure Grant

2021 NeurIPS 2021 Outstanding Reviewer Award

2017 Outstanding Graduate of Shanghai Jiao Tong University

2017 Outstanding Graduate of Shanghai

2017 National Scholarship, China

2016 National Scholarship, China