Jun Yu

Redmond, WA | (541) 678-8666 | yujunnokia@gmail.com | zariable.github.io | linkedin.com/in/zariable

---

QUALIFICATION A proven track record of leading teams of scientists and engineers to build end-to-end machine learning solutions to solve real world problems and deliver business value.

- ❏ 5+ years experience as tech lead and work in several domains, including recommendation system and ranking, causal model, forecasting, advertising and trust.
- ❏ 10+ years experience in developing machine learning algorithms and applying machine learning techniques to solve large scale real world problems.

EXPERIENCE

Amazon - Senior Applied Scientist and Tech Lead October 2017 – Now

- ❏ **[Recommendation]** built ML models to rank recommendations in Sell Central Homepage to optimize user engagement and suppress low quality recommendations.
- ❏ **[Downstream Impact]** use causal model to estimate the incremental long-term economic impact of a single seller-initiated action to guide recommendations and resource allocation.
- ❏ **[Forecasting]** build ML models to predict the product sales velocity to identify high-quality deals and optimize the deal selection and scheduling to maximize GMS.

University of Washington - Adjunct Professor January 2019 – Now

I have a passion for teaching and helping more people into the field of machine learning and data science. I currently teach several courses in the Foster School of Business, including Advanced Machine Learning, Deep Learning, and Natural Language Processing.

eBay - Applied Science Tech Lead and Engineer December 2013 – October 2017

- ❏ **[Marketing]** Led a team of 8 applied researchers to power eBay's paid internet marketing by improving the bidding strategies on Google and Facebook to maximize ROI. Our models improved ROI by 27% in 2016, leading to millions more GMB annually.
- ❏ **[Trust]** Built a large-scale machine learning model to predict seller risk and reduce the number of defective transactions on eBay via search ranking demotion, which translates to 100+ million GMB lift annually. [4]
- ❏ **[Search]** Implemented a topic model based approach to retrieve diverse items based on user buying intents and improved the user satisfaction by 6+%. [8]
- ❏ **Award:** eBay Critical Talent Award & eBay Seattle Technical Achievement Award.

Oregon State University - PhD Researcher September 2006 – December 2013

- ❏ Developed probabilistic graphical models to predict species distribution with citizen science data, which significantly improved the accuracy of species distribution. [3, 5, 10, 16, 6, 14].
- ❏ Developed a generative mixture model in the eBird human/computer learning network to quantify the skill level of citizen scientists in the eBird project. [1, 2, 7, 9, 11, 12]

EDUCATION

Ph.D. in Computer Science September 2006 – December 2013
Oregon State University, OR GPA: 3.98
Thesis: *Machine Learning For Improving The Quality of Citizen Science Data.*

B.S. in Computer Science September 2002 – June 2006
Wuhan University, China

SKILL SETS

Amazon Web Service for building ML products using various AWS products.
Python and R for training machine learning models, data analytics and rapid prototyping.
Scala and Java for building and deploying end-to-end application in production.
Spark and Scalding for building data pipelines and training larges-scale ML models.

PUBLICATION

[1] *Can observation skills of citizen scientists be estimated using species accumulation curves?* PLOS-ONE, 2015.

[2] *The eBird enterprise: an integrated approach to the development and application of citizen science*. Biological Conservation, 2014.

[3] *eBird: a human/computer learning network for biodiversity conservation and research*. AI magazine, 2013.

[4] *Bayesian and empirical Bayesian forests*. ICML. 2015.

[5] *Modeling misidentification of bird species by citizen scientists*. AAAI. 2014.

[6] *HC-Search for Multi-Label Prediction: An Empirical Study*. AAAI. 2014.

[7] *Clustering species accumulation curves to identify groups of citizen scientists with similar skill levels*. IAAI. 2014.

[8] *Latent Dirichlet Allocation based diversified retrieval for e-commerce search*. WSDM. 2014.

[9] *Clustering species accumulation curves to identify groups of citizen scientists with similar skill levels*. NIPS workshop. 2013.

[10] *Modeling misidentification of bird species by citizen scientists*. NIPS workshop, 2013.

[11] *Automated data verification in a large-scale citizen science project: a case study*. eScience, 2013.

[12] *eBird: a human/computer learning network for biodiversity conservation and research*. IAAI. 2012.

[13] *Crowdsourcing citizen science data quality with a human-computer learning network*. NIPS workshop. 2013.

[14] *Multi-label classification for species distribution modeling*. ICML workshop. 2011.

[15] *The implementation of automated data verification processes in a large-scale citizen science project*. eScience workshop. 2012.

[16] *Modeling experts and novices in citizen science data for species distribution modeling*. ICDM. 2010.

[17] *Learning algorithms for link prediction based on chance-constraints*. ECML. 2010.

[18] *Chance-constrained programs for link prediction*. NIPS workshop. 2009.