Wenbo Cui
669-230-2324 | wenbo.c95@gmail.com

Education:

Carnegie Mellon University **M.S. Machine Learning** (GPA: **3.89**) 2020 - present
- **Research Assistant** supervised by **Ameet Talwalkar** in **Interpretable ML**

University of Washington **B.S. Computer Science/B.S. Mathematics** (GPA: **3.80**) 2013 - 2017
- **Teaching Assistant** for **Data Structure** and **Algorithm** classes
- **Research Assistant** supervised by **Anna Karlin** in **Game Theory**

Professional Experience:

Yahoo Inc. (Yahoo Email - Information Extraction) 2017 - 2020

Software Dev Engineer
- Designed an end-to-end real-time email classification workflow, including:
  - A **Hadoop** based email sampling and pre-clustering workflow
  - A **PHP/JavaScript** based front end tagging tool, and corresponding data cleaning/collection back end
  - Assisted data scientist to train an **XGboost** model
  - Designed an offline processing **Spark** workflow to aggregate post-classification data
  - Designed an **JAVA** back end to serve real time email classification **<10 ms** on average
  - ***Email Classification at Scale*** *(first author)* *Yahoo Tech Pulse[1] 2019*
- Designed a **JAVA RESTful** content serving backend system
  - Refactored from a monolith server into **microservices** with different qps requirements
  - Served with **Jenkins/Jetty**
  - Support various data backends: **Redis/Hive/Yahoo Vespa/Yahoo Sherpa**
  - **CI/CD** using **Screwdriver** and **Docker** integration
- Designed a series of **Hadoop** data preparation workflows for email sampling and clustering
  - Process ~4 billion data per day
  - Mainly developed with **Oozie**, **Pig** and **Java**
- Designed an email clustering algorithm utilizing **machine learning** and graph-based clustering
  - Designed an algorithm to find the (approximate) shortest XML path query, which covers selected XML components
  - ***Graph Based Mail Clustering*** *Yahoo Tech Pulse 2018*
- A front-end tool that helps editors create extraction rules using **PHP** and **JavaScript**.

Other Experience:

Personalized Language Modeling via Meta Learning

Utilizing graphical models in meta-learning set-up for personalized next-word prediction. An extension of existing work, PLATIPUS. **Co-first author**, class project.

Best Practices for Interpretable Machine Learning in Computational Biology

Interpretable machine learning research aims to provide guidance for computational biology researchers to utilize IML tools in their projects, set up templates for further experiments to follow, and highlight pitfalls found in existing research. **Co-first author**, in progress

Evaluating the Real-World Utility of Explainable ML Methods

IML research aims to evaluate various IML methods in a real-world set-up, where it was used to assist domain experts to make judgments, by performing real world experiments on real data. A collaboration between CMU and Feedzai. In progress

---

[1] Yahoo Tech Pulse is an internal tech conference, all works are peer (engineer) reviewed.