Wenyan Li

https://wenyanli.org

Email:wenyanl62@gmail.com

EDUCATION

- **University of Maryland** — College Park, MD
  Master of Science in Electrical Engineering — *Aug 2018*
  - Relevant Coursework: Computational Linguistics, Machine Learning, Database Design, Convex Optimization, Computer Processing of Pictorial Information
- **Northwestern Polytechnical University** — Xi'an, China
  Bachelor of Engineering in Electrical Engineering and Automation (Ranked 1/97) — *June 2016*

EMPLOYMENT

- **Comcast Applied AI** — *Senior Research Engineer, Machine Learning* — Washington, D.C.
  NLP & Data Science — *Jan 2019 – July 2021*
  - Built a Spark-based automatic data processing pipeline, which cleans and preprocesses 40M raw user logs daily
  - Designed an **unsupervised auto-annotation and active learning pipeline** which used user behavioral modeling to automatically produce reliable training data, identify errors in speech recognition and NLP systems, and suggest corrections
  - Developed a context-based approach that discovered misclassified user queries in question answering systems by performing **semantic search with Sentence-BERT**
  - Leveraged **subword-level query representation** and adversarial training in customer care dialogue system for misspelled user queries, which improved user intent classification accuracy by 18%
  - Utilized a self-attentive model for automatic entity extraction, linking, and extension to guide human annotation for customer care queries
- **JD Digits AI Lab** — *Research Intern* — Mountain View, CA
  Customer Service Chatbot — *Oct 2018 – Dec 2018*
  - Implemented attention-based CNN and RNN models for query classification in customer care dialogue system

RESEARCH EXPERIENCE

- **CLIP Lab, University of Maryland** — *Master's Thesis Research* — College Park, MD
  Deep Learning for Verb Prediction; Advisor: Jordan Boyd-Graber — *Sep 2017 – Aug 2018*
  - Developed an end-to-end and incremental verb prediction model for reducing translation latency in simultaneous machine translation, and significantly improved prediction accuracy in both German and Japanese over baseline
  - Implemented synonym-aware verb prediction for German and provided interpretable visualization of the prediction process
- **Computational Biology Group, University of Maryland** — College Park, MD
  Predicting Phenotype from Genomic Sequences; Advisor: Max Leiserson — *Sep 2017 – Dec 2017*
  - Experimented with random forest and an attention-based CNN and LSTM model for genotype-phenotype reasoning which predicts genetic interactions directly from DNA/amino-acid sequences

PUBLICATIONS & PATENTS

- R. Tang, K. Kumar, K. Chalkley, J. Xin, L. Zhang, **W. Li**, ..., G. Murray, J. Lin, "Voice Query Auto Completion", in *EMNLP*, 2021
- **W. Li**, A. Grissom II, J. Boyd-Graber, "ANVIIL: An Attentive Recurrent Model for Incremental Prediction of Sentence-final Verbs", in *Findings of EMNLP*, 2020
- **W. Li** and F. Ture, "Auto-annotation for voice-enabled entertainment systems", in *SIGIR, July 2020*
- **W. Li**, F. Ture, J. Casillas, T. Jardins, "Systems and Methods for Training Voice Query Models". U.S. Application Serial No.: 63/056,361. filed July 24, 2020. Patent Pending.

PROGRAMMING SKILLS

- **Languages**: Python, MATLAB, SQL, Java, C++
- **Frameworks and Tools**: PyTorch, Tensorflow, Keras, Scikit-Learn, PySpark, Git, Docker, Snorkel, Latex