Zetian (Neal) Wu

+1-(443)-630-2430 | ztwu.nil@gmail.com | neal-ztwu.github.io

Education

Oregon State University — Oregon, United States
PhD in Computer Science & Artificial Intelligence, Advisor: Professor Liang Huang — *Sept. 2022 –*

Johns Hopkins University — Maryland, United States
MSE in Data Science — *Jan. 2020 – Dec. 2021*

Zhejiang University — Zhejiang, China
BS in Physics | Minor in Finance — *Sept. 2015 – Jun. 2019*

Research Experience

Inducing Generalizable and Interpretable Lexica
(*Paper accepted by EMNLP 2022 Findings*) — Jan. 2021 – Jan. 2022

- **Data Cleaning**: Removed non-English words and HTML strings in Yelp Reviews, Amazon Reviews and NRC dataset, and conducted down sampling to balance labels.
- **Implementation**: Built FFN, SVM, RoBERTa, DistilBERT models and implemented several lexica generation methods to create lexicon for both generalization ability and human annotation evaluation.

MultiBench: Multiscale Benchmarks for Multimodal Representation Learning
(*Paper accepted by NeurIPS 2021 Track Datasets and Benchmarks*) — Apr. 2021 – Jun. 2021

- **Data Processing**: Implemented data reader for MOSI, MOSEI and MM-IMDB dataset.
- **Implementation**: Implemented several multimodal fusion methods including early/late fusion, CCA, and etc., and built the benchmark pipeline from training to evaluation for all the datasets and methods.

Anti-fraud Model for New Financial Leasing Services
(*Top Prize in China Collegiate Computing Contest-AI Innovation Contest*) — Jul. 2018 – Nov. 2018

- **Feature Extraction**: Constructed two kinds of features: one is obtained from Bipartite Graph as statistical features and the other is extracted from Unipartite Graph using DeepWalk model as node embeddings.
- **Implementation**: Built supervised learning model (DeepFM), increasing the anti-fraud ability of the new financial leasing services by 6% on AUC.

Interactive Rare-Category-of-Interest Mining from Large Datasets
(*Paper accepted by AAAI 2020*) — Apr. 2018 – Jul. 2018

- **Feature Extraction**: Built a web crawler for data collecting and a CNN-based feature extractor to construct a real audio dataset (Birdcall) along with a numerical dataset (Medicine) for performance evaluation.
- **Implementation**: Built a Rare Category Detection (RCD) model, reducing the time complexity of query answering from quadratic to logarithmic; built a Rare Category Exploration (RCE) model, and compared our model with baseline algorithms, resulting in at least 11.75% improvement in accuracy.

Work Experience

Applied Scientist Intern — May. 2022 – Aug. 2022
Amazon — *Washington, United States*

- Fine-grained Multi-lingual Disentangled Autoencoder for Language-agnostic Representation Learning, accepted by EMNLP 2022 MMNLU Workshop

Machine Learning Engineer — Aug. 2019 – Mar. 2020
Hangzhou Enjoymusic Technology Co. Ltd. — *Zhejiang, China*

- Built a sequence-to-sequence model for music style transferring using TransformerXL and Discriminator.
- Formulated automatic music piece generation problem as a conditional sequence generation task that decodes MIDI sequence from drum beats, and modelled with VAE architecture.
- Refactored Typescript Midi-me codes using Python for integration with our own platform and application.

Skills and Additional Information

Programming/Framework: Python, PyTorch, TensorFlow, AllenNLP, Linux, C/C++, MATLAB, R, SQL
Awards: Top Prize in China Collegiate Computing Contest-AI Innovation Contest, Honorable Award in COMAP
Honors: 2nd Level in Training Plan of the National Basic Subject Top-notch Talent Scholarship