Zeyu Li

zyli93.github.io | lizeyu_cs@foxmail.com | (424) 386-8519

EDUCATION

University of California, Los Angeles (UCLA) **Los Angeles, CA**
Ph.D. Student in Computer Science *Sept. 2016 - Present*
Advisor: Prof. Wei Wang
Research Interests: *Recommender Systems, Social Network Analysis, Natural Language Processing, Deep Learning*

Harbin Institute of Technology (HIT) **Harbin, China**
B.S. in Computer Science *June 2016*

PUBLICATIONS

Zeyu Li, Jyun-Yu Jiang, Yizhou Sun, Wei Wang. *Personalized Question Routing via Network Embedding. Thirty-Third AAAI Conference on Artificial Intelligence* **[AAAI'19]** (pdf, code) (Oral presentation, Acceptance rate: 16%)

Jieyu Zhao, Yichao Zhou, **Zeyu Li**, Wei Wang, Kai-Wei Chang. *Learning Gender-Neutral Word Embedding. 2018 Conference on Empirical Methods in Natural Language Processing* **[EMNLP'18]** (pdf, code)

Zeyu Li, Hongzhi Wang, Wei Shao, Jianzhong Li, Hong Gao. *Repairing Data through Regular Expressions. 42nd International Conference on Very Large Data Bases* **[VLDB'16]** (pdf) (Acceptance rate: 16%)

SKILLS

Programming Languages: Python, Scala, SQL, Java, C/C++, R

Tools and Framework: PyTorch, TensorFlow, Keras, Numpy, Pandas, Sci-kit Learn, Spark, AWS SageMaker/EMR

EXPERIENCE

Intuit Inc. **Woodland Hills, CA**
Data Scientist Intern | Spark, Scala, Tensorflow, Keras, AWS *June 2018 – Sept. 2018*

• Preprocessed (cleaning, labeling, and down-sampling) and featurized (normalizing and embedding) the transaction data of year of 2017 (800GB) from *Quickbooks Online* using Spark/Scala and EMR on Amazon Web Service (AWS)
• Designed a Bidirectional LSTM + Attention model to predict fraud transactions
• Implemented the model via TensorFlow and Keras that can handle variable lengths of input time segments and deployed the model on Multiple GPUs (model parallelism) on Amazon SageMaker

SELECTED PROJECTS

Personalized Question Routing via Network Embedding **UCLA |** *Aug. 2017 - Sept. 2018*
Graduate Student Researcher | Numpy, PyTorch

- Conducted Question Routing using both question content and question raiser information
- Learnt the embeddings of entities using network structural information and 'accepted-answer' ranking information to provide personalized answerer recommendation given a new *(question content, question raiser)* query
- Used LSTM-equipped Metapath-based heterogeneous network embedding model for embedding learning and Convolutional Neural Network ranking model
- Jointly optimized network embedding loss and ranking loss among answers
- Achieved MRR at 0.56, Hit@5 at 0.83, and Precision@1 at 0.37 on two real-world datasets and outperformed state-of-the-art models

Neutralizing Gender-bias in Word Embedding Models **UCLA |** *Jan. 2017 - Feb. 2018*
Graduate Student Researcher | C, Python

- The first work of neutralizing the gender stereotype in the training time of a word embedding model
- Enforced specific dimensions of the word vectors to capture the desired attributes, which ameliorates the interpretability of word representations
- Learnt other dimensions as neutralized word embeddings, which benefits bias-free downstream applications
- Reduced the stereotype by 35% and effectively preserved the word proximity

Research of Basic Theory and Critical Technique on Big Data Usability **HIT |** *Mar. 2015 - Jun. 2016*
Undergraduate Student Researcher | C++

- Proposed Regular-expression-based Structural Repair (RSR) algorithm which is a dynamic programming method with time complexity $O(nm^2)$ and space complexity $O(nm)$
- Combined the RSR algorithm (for structural repair) with the value repair algorithm, a selection between association rule mining and edit distance
- Constructed whole repairing methodology and conducted excessive experiment to prove the effectiveness and efficiency