Lesheng **Jin**

Undergraduate in Computer Science

(+86) 18817550378 | king18817550378@sjtu.edu.cn | leshengjin.github.io | LeshengJin

Education

Shanghai Jiao Tong University *Shanghai, China*

Bachelor of Science in Computer Science *Sept. 2017 June 2021*

- Member of **ACM Honors Class**, which is an elite CS program for top 5% talented students.

Honors & Scholarships

Xingcai Scholarship , Award for **Top 1** student in the ACM class	2018
Zhiyuan Honorary Scholarship	2017, 2018, 2019
Academic Excellence Scholarship	2018, 2019, 2020
Zhiyuan Leadership Scholarship	2019

Publication

Retrieve And Rewrite With Automated Labeler For Chest X-ray Report Generation NAACL 2021

Lesheng Jin, Zilong Wang, Jingbo Shang *Under Review*

Addressing The Polysemy Problem In Language Modeling With Attentional Multi-Sense Embeddings ICASSP 2020

Rao Ma, **Lesheng Jin**, Qi Liu, Lu Chen, Kai Yu

Research Experience

University of California, San Diego, advised by Prof. Jingbo Shang *La Jolla, CA, USA*

Research Intern *Jun. 2020 - Present*

- Disease detection on chest X-rays using computer version
  - Developed multi-label classifiers based on different deep convolutional neural networks like ResNet and VGG. The model achieves a mean AUC of 0.85, better than all three other individual radiologists in the CheXpert dataset.
- Retrieve And Rewrite With Automated Labeler For Chest X-ray Report Generation
  - Build an effective pipeline to generate text reports from chest radiograph with commonly used techniques in Data Mining, Natural Language Processing, and Computer Version.
  - Design an unsupervised labeler based on Automated Phrase Mining System. The labeler automatically highlights the keywords in the text reports, then chooses high-frequency keywords to build the label set.
  - Implement a developed beam search to retrieve sentences of different labels based on the prediction of multi-label classification. Retrieving on different labels will get different templates. The result of retrieval outperforms traditional end2end frameworks like CNN-RNN, LRCN, and TieNet by 5%-10%.
  - Develop an image-text-co-attention rewrite framework to remove duplicate sentences and rewrite the template into detailed reports. Generated reports cover most of observations in ground truth and make few mistakes.

Shanghai Jiao Tong University, advised by Prof. Kai Yu *Shanghai, China*

Undergraduate Researcher *Jun. 2019 - Jun. 2020*

- Addressing the polysemy problem in language modeling with attentional multi-sense embeddings
  - Built a model to discriminate among different senses of a word with attention mechanism in an unsupervised manner.
  - Improved the performance of the language model on disambiguation tasks and word similarity tasks.
- U-BERT: syntax and semantic knowledge enhanced pre-trained language model for NLI
  - Proposed a novel model, U-BERT, incorporating two kinds of knowledge graph, dependency tree and abstract meaning representation, into BERT to improve the performance of pre-trained model on natural language inference.
  - We learned from the architecture of widely used U-Net in computer vision. In U-BERT, there are two paths: *contracting* path and *expansive* path. In the *contracting* path, the model obtains contextual information form BERT, then extracts syntax knowledge from the dependency tree and semantic knowledge from the AMR graph. In the *expansive* path, U-Bert sequentially incorporates syntax and semantic features from the *contracting* path into contextual word representation.
  - U-BERT achieved the state of the art on two NLI datasets, MNLI and SNLI, and showed great potential on other NLP tasks like Question Answering.
- Experiment text matching models on Chinese short paragraphs
  - Built a consistent framework to run experiments using text matching models like BiMPM and ESIM on two Chinese semantic matching datasets, BQ and LCQMC.

Selected projects

M* Compiler *Java, 2019*

- Developed a Java-and-C-like language compiler from scratch and implemented optimizations like graph coloring register allocator, dead code elimination and inline expansion.

Variational Conditional Information Bottleneck *Python C++, 2019*

- Implemented Variational Conditional Information Bottleneck(VCIB), a latent variable model, on the ImageNet dataset.

Toy Music Recommended System for Dancing Video *Python, 2019*

- Extracted features of music by traditional method(MFCCs) and pre-trained network, also developed a siamese network to match the video and the music.

RISC-V CPU *Verilog, 2018*

- Designed and implemented a FPGA-supported RISC-V CPU with 5-stage pipeline.
- Achieved the highest clock rate by applying simplification on the whole pipeline.

Train Ticket Reservation System *C++, 2017*

- Designed and implemented the backend of the train ticket reservation system based on self-made B+ Tree.

Teaching Experience

Lead Teaching Assistant, Data Structures(SJTU CS149) Spring 2019

Lead Teaching Assistant, Principle and Practice of Computer Algorithms(SJTU MS125) Summer 2019

Activities

Student Council Vice President 2019

Vice-minister of the Department of Culture and Sports, Student Union 2018

Class Monitor, ACM Honors Class 2017, 2018