Shengqiang Zhang

Email: sq.zhang@pku.edu.cn
https://shengqiang-zhang.github.io/

Education

- **Peking University** — Beijing, China
  Master of Engineering in Computer Technology — *Sept. 2018 – Jul. 2021*
- **Harbin Institute of Technology** — Harbin, China
  Exchange student — *Sept. 2015 – Jul. 2016*
- **Jilin University** — Changchun, China
  Bachelor of Science in Computer Science and Technology — *Sept. 2014 – Jul. 2018*
  Overall GPA: 85/100 (top 5%)

Papers

* indicates equal contribution (co-first author).

- **Attention Temperature Matters in Abstractive Summarization Distillation.**
  Shengqiang Zhang*, Xingxing Zhang*, Hangbo Bao, Furu Wei.
  In ***Preprint.*** Under review.

- **Pre-trained Language Model based Ranking in Baidu Search.**
  Lixin Zou*, **Shengqiang Zhang***, Hengyi Cai, Dehong Ma, Suqi Cheng, Daiting Shi, Shuaiqiang Wang, Zhicong Cheng, Dawei Yin.
  In ***KDD 2021.***

- **CUNY-PKU Parser at SemEval-2019 Task 1: Cross-Lingual Semantic Parsing with UCCA**
  Weimin Lyu, Sheng Huang, Abdul Rafae Khan, **Shengqiang Zhang**, Weiwei Sun, Jia Xu.
  In ***SemEval 2019.***

Research Experience

- **Microsoft Research Asia (MSRA)** — Beijing, China
  Research Intern, Natural Language Computing (NLC) Group — *Jan. 2021 – Jun. 2021*
  - **Mentor**: Xingxing Zhang
  - **Topic**: Document Summarization Distillation
  - **Paper**: Our work was submitted to ACL 2022.
  - We propose a simple but effective extension of pseudo-labeling method for summarization distillation. Experiments on three summarization datasets show our proposed method consistently outperforms the vanilla pseudo-labeling method. Further empirical analysis shows that both pseudo labels and summaries produced by the students are shorter and more abstractive.

- **Baidu** — Beijing, China
  Intern, Learning to Rank (LTR) Group in Baidu Search — *Sept. 2020 – Jan. 2021*
  - **Mentor**: Dehong Ma
  - **Topic**: Pre-trained Language Model based Ranking in Search Engines
  - **Paper**: Our work was accepted by KDD 2021.
  - We propose several practical solutions to employ the state-of-the-art Chinese pre-trained language model–ERNIE–in the large-scale online ranking system. Firstly, we propose a novel practice to summarize the lengthy document and then capture the query-document relevance efficiently through a Pyramid-ERNIE architecture. Secondly, we design an innovative relevance-oriented pre-training paradigm to finely exploit the large-scale post-click behavioral data. Lastly, we propose a human-anchored fine-tuning strategy tailored for the online ranking system. Extensive offline and online experimental results show that the proposed techniques significantly boost the search engine's performance.

- **Microsoft Research Asia (MSRA)** Beijing, China
  Research Intern, Speech Group *Dec. 2019 – Jun. 2020*
  - **Mentor**: Wenping Hu
  - **Topic**: Information Extraction from 2-D Visually Rich Documents
  - We propose several methods to improve current methods of extracting information from two-dimensional visually rich documents. Firstly, we propose a new label set for the unique discontinuous entity problem when we model the problem as a named entity recognition task. Secondly, we try to apply the machine reading comprehension model to solve the information extraction problem. Lastly, we propose a graph neural network combining both image features and text features well.
- **Wangxuan Institute of Computer Technology, Peking University** Beijing, China
  Research Assistant *2017 – 2019*
  - **Advisor**: Weiwei Sun
  - **Topic**: Sub-word based Named Entity Recognition, Semantic Parsing
  - **Paper**: One work was accepted by SemEval 2019.
  - We propose a sub-word based named entity recognition method for Chinese. Experiments demonstrate our method can outperform the previous state-of-the-art baseline on two datasets.
  - We participated in the SemEval-2019 cross-lingual semantic parsing with UCCA task. We introduce a novel method by applying a cascaded MLP and BiLSTM model. Then, we ensemble multiple system outputs by reparsing. Our system won second place in the German-20K-Closed track, and third place in the English-20K-Closed track.

Honors and Awards

- **Award of Excellence**, Microsoft Research Asia, 2021
- **Outstanding Graduates**, Jilin University, 2018

Skills

- **Programming Languages**: Python, C/C++, Bash, SQL
- **Deep Learning Framework**: PyTorch, PaddlePaddle
- **Toolkit**: Vim, LaTeX