Sicheng Wang

Room 524B, Zijing 2, Tsinghua University, Beijing, China, 100084

 (+86)18810701659 • wang-sc17@mails.tsinghua.edu.cn
 https://github.com/Molv1659

Education

Tsinghua University <div align="right">**Beijing, China**
Aug. 2017 — Jul. 2021</div>

- Senior undergraduate, Department of Automation
- GPA: Overall: 3.82/4.00 | Major: 3.86/4.00 | Ranked 24/183
- Standard Test: TOFEL:98 (R:29,L:27,S:21,W:21), GRE: V:155, Q:169, AW:3.5
- Advisor: Prof. Linxuan Zhang

University of California San Diego <div align="right">**San Diego, USA**
Feb. 2020 — Oct. 2020</div>

- Research Intern, Department of Electrical and Computer Engineering
- Advisor: Prof. Pengtao Xie

Publications

CERT: Contrastive Self-supervised Learning for Language Understanding
- Hongchao Fang, **Sicheng Wang**, Meng Zhou, Jiayuan Ding, Pengtao Xie
- In preparation for Transactions of the Association for Computational Linguistics. TACL 2020.

Identify Risk Factors of COVID-19 from Medical Literature
- **Sicheng Wang**, Pengtao Xie
- Technical report available in: report

MedDialog: Two Large-scale Medical Dialogue Datasets
- Xuehai He, Shu Chen, Zeqian Ju, Xiangyu Dong, Hongchao Fang, **Sicheng Wang**, Yue Yang, Jiaqi Zeng, Ruisi Zhang, Ruoyu Zhang, Meng Zhou, Penghui Zhu, Pengtao Xie
- Accepted by the 2020 Conference on Empirical Methods in Natural Language Processing. EMNLP 2020.

Honors and Awards

- **Philobiblion Scholarship**(Top 2 students in Dept. of Automation), Tsinghua University. 2019
- **Best Rookie Award in the 37th Challenge Cup**(Top 16 out of 784), Tsinghua University. 2019
- **First Prize** in the 35th National Physics Competition of College Students, Non Physical Professional Group. 2018
- **Guangyao Scolarship**(Top 5 students in Dept. of Automation), Tsinghua University. 2018

Research Experiences

Identify Risk Factors of COVID-19 from Medical Literature

Advisor: Prof. Pengtao Xie, ECE, UCSD <div align="right">*Feb. 2020 — May. 2020*</div>

- Further pretrained the BioBERT model on CORD-19 Dataset and finetuned a sentence classifier which can choose the sentences that contain the risk factors (e.g., leukocytosis) of the COVID-19 from unstructured medical literature.
- Designed various negative sentence selection schemes and validate the importance of them with experiments.
- Proposed to use Question Answering model to avoid the overfitting problem of NER model while extracting the risk factor from positive sentences.
- Achieved significant performance gains in terms of F1 & Exact Match metrics on the mission of risk factor extraction.
- Report available in: report

CERT: Contrastive Self-supervised Learning for Language Understanding

Advisor: Prof. Pengtao Xie, ECE, UCSD   *May. 2020 — Jul. 2020*

- Proposed a novel language data augmentation technique by back translation (re-translate a document that was previously translated into another language, back to the original language).
- Designed and conducted a comprehensive set of experiments independently using the BERT model and the structure of SimCLR for Self-supervised Learning to achieve better performance on SST-2 task.
- Conducted the CERT experiments on SQuAD v1.1, SST-2 and MRPC and achieved significant performance gains on MRPC task.
- In preparation for Transactions of the Association for Computational Linguistics. TACL 2020.

Progressive Text Generation

Advisor: Prof. Pengtao Xie, ECE, UCSD   *Aug. 2020 — Present*

- Proposed a progressive text generation method generating a summary first and then generating a long document according to the summary.
- Designed a novel metric Summary FBD to evaluate the extent to which the document is focused on the topic.

Prediction of Remaining Life of Trains

Advisor : Prof. Linxuan Zhang, DA, THU   *May. 2019 — Jul. 2019*

- Experimented different ways to sample the data and decided which way is effective by using PCA to reduce the dimensions and comparing the clustering effect of K-means algorithm.
- Introduced the element-wise attention gate to the LSTM model to improve the regression performance.

Semi-supervised Sentiment Analysis of Danmu (video comments)

Advisor : Prof. Linxuan Zhang, DA, THU   *Dec. 2018 — Mar. 2019*

- Segmented the Chinese video comments into words and obtained emotion labels based on a Chinese common words emotion database.
- Calculated the TF-IDF of each word and selected the key words of Danmu which tend to be new network terms.
- Applied the LSTM model and implemented self-training algorithm as a semi-supervised approach to predict the emotion of the extracted keywords.

Leadership Activities

The Publicity Department of the Student Union

Member   *Aug. 2018 — Jul. 2019*

- Managed the interview, data arrangement and the countdown poster making of the Student's Day
- Designed the key chain for the freshmen
- Edited and produced the promotion video for the sports meeting held by three departments

"Robots in my heart" Service Robot Survey Team

Team Leader   *Aug. 2018 — Jun. 2018*

- Created and organized the team to investigate the service robot technology and business model of CloudMinds Company in Beijing
- Investigated people's understanding and acceptance of service robot through street interviews and questionnaires

Skills

- Solid expertise in Python, C++, C#, MATLAB
- Experienced in Verilog HDL, Assenmbly Language
- Languages: Mandarin Chinese, English Proficient in English listening, reading, speaking and writing