Guanzhong Liu

Email : guanzhol@andrew.cmu.edu
Linkedin: https://www.linkedin.com/in/guanzhong-liu-653b73229//
Mobile : +1-724-734-6971
Github: https://github.com/guanzhongliu

Education

- **Carnegie Mellon University, Language Technologies Institute** — Pittsburgh, Pennsylvania
  Masters of Science in Intelligent Information Systems — *Aug 2022 - Dec 2023*
- **Tianjin University, College of Intelligence and Computing** — Tianjin, China
  Bachelor of Engineering in Computer Science and Technology — *Sep 2018 - Jun 2022*

Work Experience

- **Tencent** — Beijing, China
  Machine Learning Intern — *Apr 2021 - May 2022*
  - **Document-based Question Answering System**: Designed model architecture which, based on the dense retrieval model (DPR), increasing the F1 score and MRR score by 0.152 and 0.145, respectively
  - **Document-grounded Dialogue System**: Implemented a three-stage pipeline model, which predicts groundings from retrieved passages to guide the generation of the final response *(ranked 2nd in DialDoc@ACL2022 (SEEN subtask))*
  - **Similar Case Retrieval**: Leveraged text features (BM25, keywords and statistical features) to enhance Longformer's modeling ability of language *(ranked 5th in the Challenge of AI in law (CAIL2021)*
  - **Lead Scoring Model**: Leveraged incremental learning to smoothly iterate the model, increasing conversion rate by 3.75x
- **TWT Studio** — Tianjin, China
  Software Development — *Sep 2018 - Apr 2021*
  - **Modularization**: Reconstructed the campus app, reorganizing the codebase into loosely coupled and self-contained modules
  - **Basic Module**: Supported cache-based responsive asynchronous network requests based on LiveData, Retrofit and Kotlin coroutines, handling network requests with dynamic proxy

Research Experience

- **Research Assistant at Tianjin University**
  Machine Learning Research with Prof. Liu Yang — *Apr 2021 - May 2022*
  - **Low Data Resource Federated Learning**: Designed a federated adversarial data augmentation technique, achieving comparable performance using only 50% the data under extremely unbalanced data distribution settings
  - **Classifier as Discriminator**: Combined the classifier with GAN's discriminator by 2k-Loss, finding the unified decision boundaries of both real and fake samples with same label, improved accuracy in the case of low data regime by 7%+
  - **Knowledge Distillation**: Implemented knowledge distillation in the global update phase to sync the augmenter with stronger privacy protection, lower communication cost, and acceptable performance loss
- **Research Assistant at Tianjin University**
  Natural Language Processing Research with Prof. Xin Wang — *Apr 2020 - Apr 2021*
  - **Pre-training BERT**: Trained BERT model with ancient Chinese corpus, and enabled the model to learn from the special syntactic features and sentence segmentations of ancient Chinese, improving the F1-score in NER by 10%+
  - **Relation Extraction**: Designed a pipeline to leverage the tags predicted by NER model, improving the F1-score by 2.5%+
  - **Boosting**: Implemented a vote mechanism of multiple models for NER predictions, improving the F1-score by 5%+
- **Research Assistant at Tianjin University**
  Computer Vision Research with Prof. Wenyu Qu — *Apr 2019 - Apr 2020*
  - **Low latency**: Used one-stage detection based on CNN to achieve high inference speed on underwater robots for detection
  - **Dimension Clusters**: Generated the anchor boxes based on training data, improving the recall by 2%+
  - **Convolutional Block Attention Module**: Integrated CBAM seamlessly into CNN architectures with negligible overheads, improving mAP by 0.8+

Skills Summary

- **Languages**: Python, Java, C++, C, Kotlin, Matlab, Shell Script, SQL
- **Tools & Framework**: Git, Pytorch, Tensorflow, PostgreSQL, Neo4j, Vim, Latex, Pandas, Numpy
- **AI Models**: Regression, Boosting, Bagging, Decision Tree, SVM, CNN, RNN, LSTM, GAN, Autoencoder, Transformer

Publications

- **G4: Grounding-guided Goal-oriented Dialogues Generation with Multiple Documents**: (Shiwei Zhang, Yiyang Du, Guanzhong Liu, Zhao Yan, Yunbo Cao) *DialDoc Workshop at ACL 2022*
- **Constructing Chinese Historical Literature Knowledge Graph Based on BERT**: (Qingyan Guo, Yang Sun, Guanzhong Liu, Zijun Wang, Zijing Ji, Yuxin Shen, Xin Wang) *WISA 2021: Web Information Systems and Applications*