Qisheng Liao

Phone: +971 585933506, +1 908 793 9748, +86 135 0935 0177
E-mail: Qisheng.Liao@mbzuai.ac.ae
Address: Masdar City, Abu Dhabi, UAE

Education

08/2022-Present **Mohamed bin Zayed University of Artificial Intelligence (MBZUAI)**, United Arab Emirates
Supervised by Tim Baldwin, and Muhammad Abdul-Mageed,
- Master of Science in Natural Language Processing

09/2020-01/2022 **New York University (NYU)**, United States
- Master of Science in Computer Science
- Member of Music X Lab in NYU Shanghai supervised by Gus Xia

09/2016-03/2020 **University of California Santa Cruz (UCSC)**, United States
- Bachelor of Science in Computer Science (Highest Honor)

Publication

Qisheng Liao, Gus Xia, Zhinuo Wang
Calliffusion: Chinese Calligraphy Generation and Style Transfer with Diffusion Modeling
14th International Conference on Computational Creativity (ICCC), 2023

Qisheng Liao, Meiting Lai, and Preslav Nakov
MarsEclipse at SemEval-2023 Task 3: Multi-lingual and Multi-label Framing Detection with Contrastive Learning
Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval), 2023

Teaching Experience

MBZUAI, Teaching Assistant, Lab Session Instructor
- Advanced Natural Language Processing. Spring 2023

New York University Shanghai, Teaching Assistant, Recitation Session Instructor
- Network Analytics. Fall 2021
- Data Science for Soc. and Info. Networks. Spring 2022

Research Experience

Language Identification
Unknown Language Identification.
- Finetuned pretrained model for language identification problem and the model also could identify unknown language in testing.

An Investigation into ChatGPT's Language Identification Ability
- Evaluated more than 500 languages with ChatGPT and investigated the abilities of ChatGPT for low resources languages.

Contrastive Learning for Unseen Language Identification & Arabic Dialect Identification.
- Finetuned RoBERTa model in flore-101 & NADI 2021 datasets with contrastive loss

Semantic Matching

Syntax Relations in Semantic Matching Tasks

- Modified BERT structure and added information from syntax trees to the model.
  - A Graph Convolution Network, cross-attention and gates module are added into BERT.
  - Experimental results on 10 standard benchmarks demonstrate that our model performs better in semantic matching tasks.

Large Language Models

A Study of Sociopragmatic Understanding in LLMs

- we present an extensive multilingual benchmark specifically designed for sociopragmatic meaning understanding and tested on popular large language models.

Deep Learning and Diffusion Models for Chinese Calligraphy and Handwriting Texts

Calliffusion: Chinese Calligraphy Generation and Style Transfer with Diffusion Modeling

- Designed diffusion models for Chinese Calligraphy generation conditioned with characters, styles, and authors.
  - Based on Denoising Diffusion Probabilistic Models, we use natural language text to describe the calligraphy we want including, characters, scripts, and styles.
  - Any symbols from other languages or untrained characters can be adapted to Chinese Calligraphy with Low-Rank Adaptation by one-shot or few-shots.
  - The result of subjective survey showed that people cannot distinguish our generated characters with real samples.

Chinese Calligraphy Scripts and Characters Recognition.

- Designed a multitask model to predict styles and characters jointly.
- Designed algorithms to recognize a whole artwork not a single character.
  - Both the script and character accuracy can be improved to 0.99
- Designed a supervised contrastive model for artwork generation based on existing data.
  - An artwork can be generated based on the contrastive loss between each character.

Propaganda Framing

Propaganda Framing for Data from European Union

- Try to do propaganda framing to 2M data from European Union

SemEval 2023 Shared Task, Task3, Sub-Task 2: Framing Detection

- Designed a multilabel multitask contrastive learning model.
  - Based on SimCLR and SimCSE, we modified the loss function to make it fit for multilabel task setting.
  - Our system was ranked first on the official test set and on the leaderboard for five of the six languages.

Other Projects

Chinese Classifier-Noun Agreement

Diffusion model for Text Simplification

Diffusion model for Text to Speech System

Contrastive Learning for Check-Worthiness Detection

Arabic Dialect Identification with Contrastive Federated Learning