Binwei YAO

https://bigbinnie.github.io | binwei.yao@wisc.edu

Education

University of Wisconsin-Madison, Madison, WI   Sept. 2023 – Present
Ph.D. in Computer Sciences
Shanghai Jiao Tong University (SJTU), Shanghai, China   Sept. 2018 – Jun. 2022
B.Eng in Software Engineering

Publications

No Preference Left Behind: Group Distributional Preference Optimization
Binwei Yao, Zefan Cai, Yun-Shiuan Chuang, Shanglin Yang, Ming Jiang, Diyi Yang, Junjie Hu
Under Review

LLM in Professional Email Communication: Exploring the Usage, Disclosure, and Perceptual Impacts among Non-native English Speakers
Kexin Zhang*, Hanxiu Zhu*, Hailun Zhang, **Binwei Yao**
Under Review

Benchmarking Machine Translation with Cultural Awareness
Binwei Yao, Ming Jiang, Tara Bobinac, Diyi Yang, Junjie Hu
*Accepted by **EMNLP 2024** (Findings)*

Towards Reliable and Empathetic Depression-Diagnosis-Oriented Chats
Kunyao Lan, Cong Ming, **Binwei Yao**, Lu Chen, Mengyue Wu
Preprint

D^4: a Chinese Dialogue Dataset for Depression-Diagnosis-Oriented Chat
Binwei Yao, Chao Shi, Likai Zou, Lingfeng Dai, Lu Chen, Mengyue Wu, Zhen Wang, Kai Yu.
*Accepted by **EMNLP 2022** (Oral)*

MSDWild: Multi-modal Speaker Diarization Dataset in the Wild
Tao Liu*, Shuai Fan*, Xu Xiang, Hongbo Song, Shaoxiong Lin, Jiaqi Sun, Tianyuan Han, Siyuan Chen, **Binwei Yao**, Sen Liu, Yifei Wu, Yanmin Qian and Kai Yu.
*Accepted by **Interspeech 2022***

Research Experiences

HuLab, **University of Wisconsin-Madison**   Madison, WI   Sept. 2023 – Present
Group Distributional Alignment for LLMs
Advisor: Prof. Junjie Hu

- *How can a probabilistic LLM align with distributional preferences within a group?* To address this issue, we propose **Group Distribution Preference Optimization (GDPO)**, a novel framework that aligns language models with the distribution of preferences within a group by incorporating the concept of beliefs that shape individual preferences.
- Our experiments show that DPO fails to align with the targeted belief distributions, while GDPO consistently reduces this alignment gap during training. Additionally, our evaluation metrics demonstrate that GDPO outperforms existing approaches in aligning with group distributional preferences, marking a significant advance in pluralistic alignment.

SALT Lab, **Stanford University**　　　　　　Remote　　　　　　Oct. 2022 – June. 2023
Culture-Aware Machine Translation System
Advisors: Prof. Diyi Yang, Prof. Junjie Hu

- *How to evaluate machine translation systems' cultural-awareness?* To address this gap, we introduce a novel parallel corpus, enriched with CSI annotations in 6 language pairs for investigating Culturally-Aware Machine Translation—CAMT.
- We design two evaluation metrics to assess CSI translations, focusing on their pragmatic translation quality. Our findings show the superior ability of LLMs over neural MTs in leveraging external cultural knowledge for translating CSIs.

X-LANCE Lab, **Shanghai Jiao Tong University**　　Shanghai, China　　Dec. 2020 – June. 2023
The Dialogue System for Depression Diagnosis
Advisors: Prof. Mengyue Wu, Prof. Lu Chen, Prof. Kai Yu

- *How to create a reliable and empathetic dialogue system for depression diagnosis?* We design a 3-phase procedure to construct D4: the first Chinese Dialogue Dataset for Depression Diagnosis-Oriented Chat, including 1,339 simulated multi-turn conversations, which is a brand new type of dialogue - Task-Oriented Chat (task-oriented dialogue + chit-chat).
- Upon the newly-constructed dataset, four tasks mirroring the depression diagnosis process are established: response generation, topic prediction, dialog summary, and severity classification of depressive episode and suicide risk. Multi-scale evaluation results demonstrate that a more empathy-driven and diagnostic-accurate consultation dialogue system trained on our dataset can be achieved compared to rule-based bots.

Internship

AISpeech Information Technology Co., Ltd. Suzhou, China　　　Jul. 2021 – Sept. 2021
Summer Intern, Dialogue and Multimodal Group

- Constructed the ruled-based depression-screening agent Android App and WeChat miniprogram to collect users' depressive symptoms and provide a diagnostic report.
- Implemented the language generation model based on the state machine and the language understanding model based on similarity calculation and regular matching.
- Extracted core depressive symptoms into semantic slots and stored collected symptoms in the dialogue state to control the dialogue flow in a user-specific way.
- Converted the WeChat miniprogram to the first mental health support platform - Jiao Wo Xin in SJTU, which is still serving students on campus.

Academic Service

Reviewer: ACL Rolling Review　　　　　　　　　　　　　　　　　2023–Present

Teaching Experiences

CS400 programming III　　　　　　　　　　　　　　Aug. 2023 – Dec. 2023
CS400 programming III　　　　　　　　　　　　　　Jan. 2024 – May. 2024

Technical Skills

Programming Languages: Proficient in *Python*, *C++*, *Java* , *Javascript*; Capable of *Cuda*, *Go* and *Shell*.
Tools: Proficient in *PyTorch*, *Spring* , *Flask*, *React* and *Vue*; Familar with *Linux* and *git*.