XIAOMIN YU

Yu-xm ✉ yuxm02@gmail.com

Education

QingDao Univeristy of Science and Technology   Sep. 2020 – Jun. 2024
Bachelor of Data Science and Big Data Technology   *Qingdao, Shandong, China*

Research Experience

Visiting Student at Greate Bay University   **April 2024 – August 2024**

Supervised by GKD-Lab Prof. Ziyue Qiao

The research area has been Model-efficient Machine Learning. The goal is to minimize the need for training new models by focusing on utilization maximization and sustainability of models/knowledge through techniques such as transfer learning, continual learning, and data-centric learning.

MLE Intern at Xreal   **August 2024 – Present**

Supervised by FAITA-Lab Prof. Zhaoxin Fan and Xreal Dr. Kejian Wu

The research area has been Embodied AI and Multimodal Large Language Models (MLLMs). The focus is on enhancing agent environment interaction by integrating vision, language, and control, leveraging pretrained models to enable complex task performance and adaptive learning.

Paper List

Similarity-Aware Multimodal Prompt Learning for Fake News Detection.
[Published] [Journal] [Second Author] Information Science (IF=8.1)   *September 2022 – August 2023*

- Challenges in practical applications of multimodal fake news detection due to large training data requirements and inconvenient parameter updating.
- By introducing the SAMPLE framework, utilizing prompt learning and a similarity-aware multimodal feature fusion approach, we achieved leading high F1 scores and accuracy rates in benchmark tests, showcasing outstanding performance in both low-data and data-rich environments.

ArcSin: Adaptive ranged cosine Similarity injected noise for Language-Driven Visual Tasks.
[Under Review] [Conference] [Second Author] IJCV (IF=11.6)   *March 2023 – Present*

- In our study, we explore how to bridge the modality gap between learning from language and inference for visual tasks in vision-language tasks, while preserving the semantic content of the original modality embedding.
- We propose a novel method called "Adaptive ranged cosine Similarity injected noise" (ArcSin), which employs an adaptive noise injection and a similarity pool strategy to effectively narrow the domain gap while maintaining the integrity of the semantic content.

Adapting Graph Models via Target Integrity Assessment and Source Distribution Hypothesis.
[Under Review] [Journal] [Co-first Author] TNNLS (IF=10.2)   *March 2023 – Present*

- This paper studies domain adaptation in graph transfer learning when the source graph cannot be used due to privacy or security concerns. We propose a source-free domain adaptation method that enhances the model's discriminative ability through a weighted information maximization loss and posterior integrity assessment and improves alignment between graph distributions by integrating a reconstruction decoder and adversarial learning techniques.
- Experimental results validate the superiority of our method in handling unlabeled target graphs.

Fake Artificial Intelligence Generated Contents (FAIGC): A Survey of Theories, Detection Methods, and Opportunities.
[Under Review] [Survey] [Journal] [First Author] PR (IF=8.0)   *January 2024 – Present*

- We make a comprehensive analysis of Fake AI-Generated Content (FAIGC), proposing novel categorization methods and detailing the main technologies behind the generation of FAIGC in text, visual, audio, and multimodal.
- We conducted a thorough investigation into the detection techniques for FAIGC, compiled an extensive list of relevant datasets, and offered insights into future research directions.

SpikEmo: Enhancing Emotion Recognition With Spiking Temporal Dynamics In Multimodal Conversations.
[Under Review] [conference] [First Author] ICASSP   *December 2023 – Present*

- We propose a two-stage representation method based on semantic and dynamic encoding using spiking neurons to more accurately capture the complex modality and temporal characteristics in multimodal emotion data.
- Additionally, we designed an innovative combination of loss functions to effectively address class imbalance and emotional semantic similarity issues. Extensive experiments on multiple benchmark datasets demonstrate the effectiveness of this framework in the ERC task.

CubeRobot: Grounding Language in Rubik's Cube Manipulation via Vision-Language Model.
[Under Review] [conference] [Co-first Author] ICASSP   *December 2023 – Present*

- We developed CubeRobot, a vision-language model (VLM) designed for solving 3x3 Rubik's Cubes. It utilizes the CubeCoT image dataset, which includes 43 challenging subtasks, and incorporates a Dual-loop VisionCoT architecture and Memory Stream for task planning and decision-making.
- CubeRobot independently manages Rubik's Cube tasks at various levels, achieving 100% accuracy in low- and medium-level tasks and 80% accuracy in high-level tasks.

Deep Learning Competition

SemEval-2023 (The 17th International Workshop on Semantic Evaluation)
*Task 3 Subtask 1: **News Genre Categorisation***

- The submitted system1 achieves **the second best** in Italian and Spanish (zero-shot) in subtask-1.
- Automate the detection of news genres within a multilingual framework, addressing the challenge of imbalanced and insufficient training data.
- Fine-tuned the XLM-RoBERTa model using a combination of sample and class weights, effectively managing multilingual data imbalance and surpassing traditional methods with a task-dependent fine-tuning approach.

Honors and Awards

- Contemporary Undergraduate Mathematical Contest in Modeling — **National Second Prize** 2022
- National College Students' Innovation and Entrepreneurship Training Program — **National Level Project** 2022
- National College Students' Innovation and Entrepreneurship Training Program — **National Level Project** 2021
- Chinese Collegiate Computing Competition — **Provincial First Prize** 2022
- The Mathematical Contest in Modeling — **Honorable Mentions** 2021

Entrepreneurship Program

Project 1.: Teaching platform based on RGB human-computer interaction
*National College Students' Innovation and Entrepreneurship Training Program **National Level Project 2021.***

- The project is designed to reduce the teacher's dependence on input equipment when using multimedia teaching.
- The system collects depth images through Kinect to obtain real-time three-dimensional skeletal images of the controller, and identifies the controller's movements through the data of bone joints to control the use of the teaching system.

Project 2.: Intelligent visual aid system based on multimodal deep learning algorithm.
*National College Students' Innovation and Entrepreneurship Training Program **National Level Project 2022.***

- The project was designed to address the problem of people with visual impairments lacking visual perception of the outside world.
- The core algorithm of the project is the image caption module, which converts the scene image obtained by gesture interaction into a complete text description of the scene.

Leadership

Data Science Research Association — Autumn 2021 – June 2024
Founder, Association President — *QingDao University of Science and Technology*

- The Data Science Research Association is dedicated to advancing **AI** technology, organizing **AI-related competitions** and **research activities**, and has been honored with **multiple national and corporate awards**.
- Collaborating with companies including **Huawei**, **Baidu PaddlePaddle**, and **the open-source organization**.
- The association has over **140** members, and the cumulative number of participants in association activities exceeds **500**.

QUST PaddlePaddle Developer Group — April 2022 – June 2024
Founder — *QingDao University of Science and Technology*

- Organized learning sessions on deep learning technologies with PaddlePaddle as the core framework.
- Organized learning sessions on deep learning technologies with PaddlePaddle as the core framework.
- Led team members in implementing projects and generating outputs around the PaddlePaddle framework.

DataWhale — Spring 2022 – 2023
Open Source Organization Member — *China*

- Served multiple times as a **professional teaching assistant** for DataWhale team learning, **assisting in the review and correction of open-source projects and courses**.
- Participate in **university collaborations** to promote undergraduate involvement in artificial intelligence-related studies and competitions.

Academic service

- Conference: Coling 24, Coling 25, ICME 24, ICASSP 25
- Journey: Computer Speech & Language