

Meixi Chen

Email: mc2125@cantab.ac.uk, Tel: +86 19168911813

Personal webpage: <https://meixichentracy.github.io>

Education

M.Eng. & B.A. Engineering, University of Cambridge *Oct. 2018 – Jul. 2023*

2022-2023: Master's degree with Merit.

Class I in Project: *Automatic Children Story Video Generation*

Research Supervisor: Dr. Samuel Albanie

2021-2022: Intermission year in industry; 2020-2021: Class II.1;

2019-2020: Not classed due to COVID; 2018-2019: Class II.1.

Course modules related to research interests:

- Deep learning and structured data, Computer vision, Practical optimisation, Probabilistic machine learning, Signals and systems, Inference.

Harrow International School Hong Kong *Sept. 2016 – Jun. 2018*

A-Levels: Mathematics (A*), Further Mathematics (A*), Physics (A*), Chemistry (A*).

Research Experiences

Automatic Children Story Video Generation

University of Cambridge, M.Eng. project (Oct. 2022 - Jul. 2023)

- Text-to-image diffusion models used to generate images for the video
 - Customisation of text-to-image diffusion models to maintain character consistency, with DreamBooth method as benchmark.
 - Additional manual control to improve DreamBooth via weight interpolation at the inference stage, with no need for further training.
- Automatic generation using: character design images to text description (fine-tuned ViT) + story generation (ChatGPT) + image generation (text prompts by ChatGPT and images by Midjourney) + speech synthesis (play.ht service).
 - Class I overall project grade.
 - For final report please see: https://github.com/MeixiChenTracy/MEng_project.

Few-Shot Text Classification

Tencent, Intern project (Apr. 2022 - Aug. 2022)

- Prompt engineering and prompt tuning methods explored.
- Calibrated in-context learning by retrieving nearest neighbours from datastore
- Manage noise in retrieved neighbours by optional training of two small modules - Adaptive Neighbour Selection and Feature Regularisation.
- Performance with and without training the noise handling modules:
 - Without parameter training: achieved superior results compared to no-training benchmarks, and comparable performance with some training-required benchmarks.
 - With parameter training to handle noise: performance boosted more.
 - For full results please see [arxiv:2212.02216](https://arxiv.org/abs/2212.02216).

Natural Language Tags Clustering

Tencent, Intern project (Jan. 2022 - Apr. 2022)

- Word2vec embeddings trained on game-related domain text resources from: 1) large game-related forum corpus (limited appearance of specific tags) and 2) search engine results specifically for tag words.
- Clustered tags by their embeddings using K-Means and hierarchical clustering.
- Knowledge graph schema construction employed model outputs, saving manual effort.

Intent Classification for Dialogue System

China Merchants Bank (CMB)*, Intern Project (Jun. 2021 - Sept. 2021)

- Aimed to classify user intention in the online banking context, where the boundaries of some classes were very close to each other.
- Modeled by: finBERT + modified triplet network which generates “hard” triplets.
- Best model achieved final accuracy above 98%, average recall and precision above 96%.

Other Internships

Application Research Intern - Tencent (Jan.2022 - Aug 2022)

- Conducted research as listed in Research Experiences section
- Daily SQL request and data analysis support.

Data Analysis Intern - DST Rental (Oct. 2021 - Dec. 2021)

- Predicted order renewals and early-terminations by: binning continuous data by ChiMerge and equal-frequency binning; encoding bins with WOE encoding; modeling data with logic regression, neural network, and regression tree, respectively.
- Achieved total accuracy above 95%, 90% precision and recall on negative samples, 80% recall and 60% precision on positive samples (positive samples contribute to a very small amount of data).

Unity Intern - Dexta Robotics (Jun. 2019 - Sept. 2019)

- 3D bounding boxes automatic generation in Unity from input 3D model.
 - In Python: implemented Graphic calculations.
 - In C#: developed interactions with Unity.
- Attended the SIGGRAPH 2019 Summit as a representative of Dexta Robotics.

Publications

Nie Feng, **Meixi Chen**, Zhirui Zhang, and Xu Cheng. "Improving few-shot performance of language models via nearest neighbor calibration." *arXiv preprint arXiv:2212.02216*, 2022.

Additional Skills and Interests

Language skills: English (fluent), Chinese (native), Japanese (intermediate).

Other Interests: Bird-watching, Badminton (college team captain for one year).

* Officially employed by CMB Network Technology, a sub-company of China Merchants Bank (CMB), but worked with the team in CMB.