Xinyu Ma

Email : mxinyuma@gmail.com · Homepage :https://ecoli-hit.github.io/ ·

Research Interest

I am passionate about natural language processing and machine translation :

- Large Model Transfer Learning and Data Synthesis
- Large Model Task-specific Instruction Fine-tuning
- Low-resource Machine Translation

Education

Harbin Institute of Technology, ShenZhen   Sept. 2020 – July 2024

- average credit score of 89.51/100 and a GPA of 3.734/4

Research experience

Clustering Pseudo Language Family in Multilingual Translation Models with Fisher Information Matrix   Feb. 2023 – Dec. 2023

EMNLP 2023 Conference in process   Advised by Xuebo Liu

We propose an approach that leverages Fisher Information to search the task-specific parameters of pretrained multilingual model to represent language pairs and propose Pseudo Family for language clustering. Our main contributions is showing below :

- We propose a novel method for clustering language families without accessing the data or modifying the architecture
- We introduce a simple and effective fisher information matrix method for pseudo language family clustering.
- The results show that the pseudo family clustered by our method can yield more promising results than the vanilla linguistic language family.

The 4th IKCEST The Belt and Road International Big Data Competition and the 8th Baidu & Xian Jiaotong University Big Data Competition   Aug. 2022 – Nov. 2022

We complete eight translation directions between Chinese $\leftrightarrow$ French, Chinese $\leftrightarrow$ Russian, Chinese $\leftrightarrow$ Thai, and Chinese $\leftrightarrow$ Arabic.

- Utilizing the m2m multilingual pretrained model to establish the model.
- Employing backtranslation in conjunction with data augmentation.
- Enhancing the model's performance by implementing the rdrop algorithm.
- Training an r2l model for reranking, which partially resolves the issue of poor right-side generation in autoregressive models that generate from left to right.
- Utilizing a fine-tuned model with a small amount of data for data selection, resulting in a significant improvement in translation quality (with an average increase of approximately 2.0 BLEU score).
- Attaining the third prize in the final competition as an undergraduate team, achieving an average BLEU score of 32.965 for Chinese $\leftrightarrow$ Arabic translation.

ASC22 Student Supercomputer Challenge   Oct. 2021 – Apr. 2022

I was responsible for optimizing the training process of the deepmp-kit machine learning molecular dynamics tool:

- Utilized knowledge of OpenMP, assembly language, and SIMD (Single Instruction, Multiple Data) to perform loop unrolling, memory optimization, and parallel processing on the algorithm.
- Optimized the training environment by adjusting hardware parameters.
- The team was awarded the second prize.

Teaching

Teaching Assistant, Harbin Institute of Technology , ShenZhen
COMP2008 Principles of Computer Composition

⚙ Skills

- Python, C++, JAVA , Latex.
- I am well-versed in the utilization of Linux operating systems and possess fundamental knowledge in cluster maintenance.
- Language : IELTS 6.5(L:6.5 R:7.5 W:6.5 S:6.0)

♡ Awards and Honors

8th Baidu & Xian Jiaotong University Big Data Competition, International Third Prize		Nov. 2022
ASC22 Student Supercomputer Challenge, Second Prize		Apr. 2022
China Undergraduate Mathematical Contest in Modeling, First Prize of Guangdong Province		Sept. 2022
Mathematical Contest In Modeling, Meritorious Winner		May 2023