Xinyu Ma

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

Research Interest

I am passionate about natural language processing and machine translation :

- low-resource machine translation
- multilingual translation
- parameter-efficiency tuning

Education

Harbin Institute of Technology, ShenZhen Sept. 2020 – present

- third year of studying Computer Science and Technology
- average credit score of 89.51 and a GPA of 3.734/4

Research experience

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

EMNLP 2023 Conference Submission 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 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

Complete eight directions from Chinese to French, French to Chinese, Chinese to Russian, Russian to Chinese, Chinese to Thai, Thai to Chinese, Chinese to Arabic and Arabic to Chinese.

- 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-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

- I have experience in writing conference papers and I am proficient in using LaTeX as well as data visualization tools such as Pyplot.
- I am knowledgeable about the current state-of-the-art large-scale models and have a strong understanding of multilingual machine translation techniques.
- I have accumulated over one year of research experience in HITSZ-ICI-哈翻深圳 team.
- I am well-versed in the utilization of Linux operating systems and possess fundamental knowledge in cluster maintenance.
- Language : CET6 (Satisfying the demands of academic literature review and daily life, will take IELTS soon)

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