Zhiyu Xie

Tsinghua University – Beijing, China
+86 18030308900 • xiezhiyu32768@gmail.com
xiezhiyu01.github.io

Education

Tsinghua University **Beijing, China**
2019.8 - Present

- Bachelor of Engineering in Computer Science and Technology
- GPA: 3.96/4.00 **(Ranking: 3/202)**
- Core Courses: Linear Algebra(**A+**), Probability and Statistics(**A**), Data Structures(**A**), Operating Systems(**A**), Introduction of Theory of Computation(**A**), Introduction to Artificial Intelligence(**A**)

Honors and Awards

- Academic Excellence Scholarship in Tsinghua University (**Top 1%**) ***3 times*** *in 2020,2021,2022*
- Google Women Techmakers Scholarship (34 winners in China) *2020*
- Freshman Scholarship in Tsinghua University *2019*
- FuGuang Scholarship in Fujian Province *2019*
- The Bronze Medal in National Olympiad in Informatics (**NOI**, Category D) *2017*
- The Bronze Medal in the 11th Asia-Pacific Informatics Olympiad (**APIO**) *2017*

Internship

Model Distillation in Model Maker **Tensorflow Lite, Google Beijing**
STEP Intern, Mentor: Tian Lin *2021.7 - 2021.9*

- To further facilitate TensorFlow Lite on-device training, this work introduced model distillation method for model compression in *TensorFlow Lite Model Maker library*.
- Followed Object-oriented programming principles to implement model distillation for image, audio and text classification tasks in Model Maker Framework, enabling users to create a fine-tuned end-to-end model on a customized dataset in just **6 lines of code**.
- Carried out comprehensive experiments to show that the work can lead to **90% reduce of parameters** while maintaining a competitive and sometimes better accuracy performance compared to the teacher network.

Research Experience

Prefix Generator for Low-resource Event Extraction **PlusLab, UCLA**
Advisor: Nanyun Peng *2022.6 - Present*

- Generative models have been important for event extraction tasks because of its flexibility and efficiency. However, it's difficult to incorporate useful external information (e.g. syntax tree) into the framework.
- Proposed a novel method called Prefix Generator, which encodes external information by pre-training and mapping its representations into prefixes in encoder-decoder models.
- Experimental results of incorporating information such as AMR(Abstract Meaning Representation) graph or Optimus robust representation show that the method of Prefix Generator is generally applicable and especially effective in low-resource settings.

Domain Relabeling for Subpopulation Shift **IRIS Lab, Stanford**
Advisor: Huaxiu Yao *2022.6 - Present*

- Currently, many methods leveraging subpopulation shifts requires manually specifying only one feature to annotate as domain label, which cannot eliminate the problem of learning spurious correlations.
- We realized that real-world datasets (e.g. medical, weather forecasting) usually contain many candidate features in the metadata table, providing an opportunity to integrate them to get higher-quality domain labels.
- Through Reinforcement Learning, we evaluate the importance of each metadata feature using the feedback such as error rate variance from downstream tasks, and iteratively update the domain label, leading to improved worst-group performance.

Reviewing Test Protocols of Distantly Supervised Relation Extraction **THUNLP, Tsinghua**
Advisor: Zhiyuan Liu *2020.9 - 2021.3*

- Since the distant supervision adopted by the two popular datasets (NYT10 and Wiki20) for relation extraction doesn't consider context information, the annotation errors it bring make the evaluation of the models inaccurate.
- Thus, we proposed a improved relation ontology and adopted data-cleaning to build manually-annotated test sets for NYT10 and Wiki20, correcting 53% wrong labels in NYT10.
- We thoroughly evaluated competitive models on manually-annotated and distantly supervised datasets, the results indicated very different conclusions between the two datasets, which sheds light on the importance of more accurate evaluation procedures for relation extraction research.

Publication

Manual Evaluation Matters: Reviewing Test Protocols of Distantly Supervised Relation Extraction

- Tianyu Gao, Xu Han, Keyue Qiu, Yuzhuo Bai, **Zhiyu Xie**, Yankai Lin, Zhiyuan Liu, Peng Li, Maosong Sun, Jie Zhou.
- In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*. **ACL Findings 2021.**

Skills

- Programming Language: C++, Python, Java, HTML+CSS, RISC-V, Verilog
- Language Skills: Mandarin(Native), English(Fluent, TOEFL: 112, GRE: V158+Q170+4.0)