北京理工大学

Xiao Zhang

13883159686 xiaoz987@gmail.com Beijing
24 Male https://github.com/Yottaxx

EDUCATION

Beijing Institute of Technology Sep 2020 - Jun 2023

Computer Science Master

- Key Laboratory of the Ministry of Industry and Information Technology
- Supervisor: Huanghe Yan and Xianling Mao
- Research interests: Natural Language Processing, Machine Reading Comprehension

Chongqing University Sep 2016 - Jun 2020

Computer Science Bachelor

- GPA:3.63 / 4.0
- Honors/awards: Mathematical Contest in Modeling & Interdisciplinary Contest in Modeling (M)(2018, 2019), Chinese University Student Computer Design Competition Second Prize, National Mathematical Modeling Competition Chongqing Division First Prize, Chongqing Computer Design Competition Second Prize, Chongqing University Outstanding Graduates, Chongqing University Outstanding Students, Chongqing University Outstanding Student Cadres.

PROFESSIONAL EXPERIENCE

ByteDance May 2022 - Jul 2022

NLP algorithm intern Data-Douyin Beijing

- Language Model Pretraining
  - Cantonese pretrained models: Cant-Roberta, Cant-Electra, Cant-Dlectra, Cant-Deberta:
  - Multilingual pretrained models: XLM-Deberta, XLM-Roberta, XLM-Dlectra, XLM-Electra
- Language Model Finetune stategies: RDrop, FreeLB, EKD
- Online Improvement:
  - Cantonese dialect risk classification: precision +65%
  - Mandarin recruit risk classification: precision +23%

ByteDance Jun 2021 - Sep 2021

NLP algorithm intern (secret project) Data-EDU Beijing

- Rule model enhanced -> model accuracy increased by 15%
- Pre-trained model domain transfer-> model accuracy increased by 5%
- Data cleaning -> model recognition accuracy increased by 13.5%
- Model pre-research-> lightweight, reverse condition generation.

RESEARCH EXPERIENCE

ET5: A Novel End-to-end Framework for Conversational Machine Reading Comprehension Jan 2022
（COLING2022 first author）

A new end-to-end CMRC simplified framework

- The traditional CMRC framework is a pipeline structure and is mainly divided into three processes: decision classification based on entailment reasoning, span extraction based on missing information, and question generation based on extraction.
- This architecture **cannot make full use of fine-grained supervision information** containing entailment reasoning in the problem generation stage and suffers from error propagation problems.
- This paper proposes **a novel and simple end-to-end CMRC framework**, which makes decisions or generates questions at the same time and shares the effective signals of fine-grained entailment reasoning. The framework consists of Text-to-Text Transformer and an additional fine-grained entailment reasoning decoder.
- Experimental results show that the proposed SimCMRC has achieved the **SOTA** performance of BLEU scores in the ShARC dataset.

Cross-lingual Phrase Retrieval (ACL 2022 co-first author) Nov 2021

Phrase representation learning based on unlabeled example sentences

- **A cross-lingual phrase retrieval task** is proposed, and phrase pairs containing 420M example sentences 65K are constructed through Wiki.
- The sentence-based phrase contrast representation learning method **(XPR)** proposed in this paper has achieved **the most advanced performance** in supervised, unsupervised, zero-shot transfer, multilingual transfer four experiments.

Key Technologies and Systems of Human-like Intelligence Based on Big Data May 2021

National key R&D program (863)

- Responsible for the machine reading comprehension part of gaokao examination in Chinese.
- 《A Knowledge Enhanced Chinese GaoKao Reading Comprehension Method》 **(2021 ICBK first author)**.
- Haihua gaokao machine reading comprehension competition:
  - **3rd (0.2%)**. The participating teams include Peking University, Shanghai Jiaotong University, Fudan University, Ant Group, Tencent, and other academic and industrial teams.
- FreeLB-EKD-MacBERT MultichoicesModel based on Transformers and PyTorch.

HammerScholar Academic Search Engines Feb 2021

Website: https://hammerscholar.net/

- DBLP, SemantiScholar, and other academic websites are used for data collection and processing based on the Scrapy framework.

COMPETITION EXPERIENCE

Mathematical Contest in Modeling & Interdisciplinary Contest in Modeling (M) Jan 2019

- Responsible for modeling and proposing an environmental degradation model based on expert knowledge and physical process environmental assessment.

Mathematical Contest in Modeling & Interdisciplinary Contest in Modeling (M) Jan 2018

- Responsible for modeling and model solving: Practical application of Markov chain model in national vulnerability assessment

Chinese University Student Computer Design Competition (Second Prize) Jul 2019

- Responsible for building a front-end and real-time automatic lip translation website by using React

LEADERSHIP EXPERIENCE

Class Monitor; President of the Basketball Club; Student Representative

SKILLS, CERTIFICATIONS & OTHERS

- **Skills:** Familiar with C, Java, Python, PyTorch, and Tensorflow.
- **Languages:** English (CET-6)