Pan Xie

New main building E813, Beihang University, Beijing, 100191, China

Email: ftdpanxie@gmail.com

Homepage: panxiaoxie.cn

Github: https://github.com/PanXiebit

Research Interest: machine translation, natural language generation and lifelong learning

EDUCATION

Beihang University, Beijing, China

Doctor of philosophy in Computer Science

- **Core course:** Electronic Image Processing, Artificial Intelligence, Machine learning**,** Speech and language processing, Algorithm and design analysis

Beijing Institute of Technology, Beijing, China

M.S in mechanical engineering and Automation, GPA: 3.6/4 (no rank)

Wuhan University of Technology, Wuhan, China

B.E in mechanical engineering and Automation, GPA: 3.34/4 (top 20%)

Research Experience

Adversarial Neural Machine Translation with pair-wise discriminator *Jan, 2019 – Present*

Mentor: Xiaohui Hu, Institute of Software, Chinese Academy of Sciences (ISCAS)

- The discriminator and generator are parameter-shared transformer models, and REINFORCE is used to solve the problem that gradient cannot be returned when the text data is discrete.
- Utilizing the consistency traits of bidirectional translation of machine translation, build rewards with margin loss between fake target sentence and real target sentence back translate to source sentence.

Improving machine translation with monolingual data *Mar, 2019 – Aug,2019*

Research Intern, XiaoMi AI Lab

- Constructing pseudo-parallel corpora with monolingual and specific dictionary, improving the quality of translation, and to some extent solve the translation problem of translation pairs that do not appear in parallel corpus.
- Pseudo-parallel corpus is constructed by replacing certain words in a monolingual data, especially low-frequency words that never appear in parallel corpora.

PROJECT EXPERIENCE

Machine reading comprehension *Aug, 2018 – Jan, 2019*

Intern, Intelligent understanding Lab, Samsung China Research Institute

- Apply related techniques of machine reading comprehension to children's educational robots. Automatically match the most appropriate answer based on the reading document and the question.
- Combine the pre-trained language model BERT and bi-attention structure BIDAF, fine-tuning on the specific task
- Deploy the model on the server through tensorflow serving

Text cluster and classification *Jan, 2018 – Mar, 2018*

Intern, Sina Advertising Research and Development Center

- Based on the clustering algorithm, the sina weibo blog text is coarsely classified and tagged. Then, the Text-CNN classification model is trained according to the existing tagged data, and the newly added microblog posts are classified. The goal is to increase the recall rate of targeted blog posts for targeted ad serving

opinion Question Machine Reading Comprehension, AI challenger 2018 *Aug, 2018 – Jan, 2019*

Rank 16th/260

- *implement with the combination of multiway attention and BIDAF model using pytorch.*

AWARDS

Second Prize of Scholarship, Beijing Institute of Technology 2016, 2017 (Top 20%)

Second Prize of Scholarship, Wuhan University of Technology 2011-2012 (Top 15%)

Excellent Student Leaders, Wuhan University of Technology 2011-2012 (Top 15%)

Skills & Other Information

- Programming Languages: Python (Expert), C++(Proficient), C (Proficient), SQL (Proficient)
- Machine Learning Package: NLTK, Tensorflow, Pytorch, Gensim, Skicit, Numpy, Scipy,
- Programming Experience: Machine Learning, Deep Learning, GPU Programming with Nvidia