JIANBO DAI

13.4/1 Meadow Court 13-29 Sciennes ◇ Edinburgh EH9 1NJ ◇ (+44) 07913109707 ◇ (+86) 13486963560
jianbodai819@gmail.com ◇ s1792542@sms.ed.ac.uk ◇ https://www.linkedin.com/in/jianbo-dai-9085a2161/

EDUCATION

The University of Edinburgh, UK Sep. 2019 - Nov. 2020
MSc in Artificial Intelligence, Estimated GPA: 3.85/4.00 (74.0/100), Ranking: About top 10% students out of 250 students
Relevant Courses: Machine Learning and Pattern Recognition, Accelerated Natural Language Processing, Deep Learning, Probabilistic Modelling and Reasoning, Natural Language Understanding, Generation, and Machine Translation, Algorithmic Game Theory and Applications, Programming Skills

The University of Manchester, UK Sep. 2017 - Jun. 2019
BEng in Chemical Engineering, GPA: 3.82/4.00 (75.9/100), Ranking: About top 7% students out of 302 students
Relevant Courses: Numerical Methods for Differential Equations, Discrete Mathematics, Process Optimization

Dalian University of Technology, China Sep. 2015 - Jun. 2019
BEng in Chemical Engineering, GPA: 3.83/4.00 (88.3/100), Ranking: Top 3% students out of 219 students
Relevant Courses: Advanced Mathematics I and II, C language programming, Probability and Statics, Linear Algebra, Fundamentals of Computers, Electrical Engineering, Electronic Engineering
Selected Honors: National 8^{th} of Chinese College Student Chem-E-Car Competition in 2017; 1^{st} Prize of China's National Undergraduate Math Contest in Dalian in 2016; 3^{rd} Prize in International Mathematical Modeling Competition in 2017; Outstanding Graduates of DLUT in 2019

TECHNICAL SKILLS

Computer Languages	Fluent in Python, C, LaTeX; Have experience in HTML, SQL, C++, Qt
Framework	Fluent in Pytorch, OpenNMT, Numpy, Matplotlib, SciPy
Natural Language	Chinese (native), English (professional), French (beginner)

PROJECT

User Question Input Classification and Suggestion *Summer Research, Yale LILY Group* 2020.06 - Present

- Formalized positive data and did negative sampling based on Spider, SparC and CoSQL dataset (some semantic parsing and text-to-SQL datasets). Created SQL templates and NL-SQL templates corpus.
- Pre-trained and ran RatSQL model which is a seq2seq framework augmented with relation-aware self-attention mechanism on formalized dataset, calculated recalls.
- Generated three partial input datasets based on different definitions of partial questions.

Contextualized Neural Lemmatization *Master Project, UoE* 2020.05 - Present

- Trained the Lematus model which is a contextualized neural encoder-decoder model as a baseline, the results are evaluated by lemma accuracy and Levenshtein distance.
- Trained COPYNET on 20 languages which is a seq2seq model incorporated with the copying mechanism, the implementation in the OpenNMT framework if utilized.
- Created templates for lemma-inflected word pairs with edit distance algorithm, used them to train a Course-to-fine model with a structure-aware neural architecture decomposing the semantic parsing process into two stages.

Neural Machine Translation *NLU+ Course, UoE* 2020.02 - 2020.03

- Trained baseline NMT models based on bi-LSTM encoder-decoder architecture, implemented beam search and added layers to enhance model's performance on different datasets by calculating loss, perplexity and BLEU score.
- Implemented the lexcial attention to improve the model as described in paper and investigate its influence by outputting attention heat maps.
- Analysed a basic implementation of the Transformer architecture and implemented the multi-head attention mechanism according to papers.

Exploring Distributional Similarity in Twitter *NLP Course, UoE* 2019.11

- Implemented various vector semantic models to represent words, including basic PPMI and its variants with Laplace smoothing and context distribution smoothing. Investigate different similarity measures such as cosine similarity and Jaccard similarity methods.
- Used data from Twitter to compute relatedness of country names and found that it tends to be higher when they are geographically closer except for when news cause considerable relations between countries.

Trigram Language Model over Characters *NLP Course, UoE* 2019.09 - 2019.10

- Built a trigram language model by collecting counts for character 3-grams and estimating probabilities with smoothing methods such as add-one and add-α.
- Used the model to generate random sentences and to calculate documents' perplexity for language identification task.
- The optimal model is utilized to investigate influence of the genre and size of the training data.