Minghao **Wu**

Beijing, P. R. China

+86 176 0044 2356 | minghaowu_2015@163.com | https://github.com/minghao-wu

Education

The University of Melbourne

Melbourne, Australia

Master of Information Technology

Mar. 2016 - Jul. 2018

- Specialization in Computing, Natural Language Processing.

The University of Sydney

Sydney, Australia

B.S. in Statistics & Information Systems

Mar. 2013 - Mar. 2016

- Most of statistics and mathematics subjects are in Advanced stream.

Experience

JD AI Research

Beijing, P. R. China

Research Engineer

Aug. 2018 - Jun. 2019

- Design and develop online system based on business needs, with natural language processing technologies.
- Develop, evaluate and optimize models, both deep learning models and machine learning models, especially for tasks like text classification, named entity recognition and ranking systems.
- Follow up and reproduce latest academic research advances.
- Conduct innovative research in natural language processing and have the outcomes published.

Publications

Evaluating the Utility of Hand-crafted Features in Sequence Labelling

The University of Melbourne

Keywords: Natural Language Processing, PyTorch, Name Entity Recognition

Mar. 2018 - Aug. 2018

- Propose a novel neural architecture for utilizing hand-crafted features by auto-encoders in Deep Learning models.
- Obtain F_1 of 91.89 on CoNLL 2003 NER English shared task, setting a new state of the art, achieving comparative performance against BERT-Base and Bi-LSTM-CNN-CRF+ELMo.
- This work is accepted by EMNLP 2018, the second best conference in NLP.
- https://aclweb.org/anthology/D18-1310
- https://github.com/minghao-wu/CRF-AE

Projects

JD AlphaSales - Smart Chatbot

JD AI Research

Retrieval-based Chatbot, including three major modules, Frequently Asked Questions, Ranking Module and Knowledge-Based Question Answering.

Aug. 2018 - Feb. 2019

- Train text classification model which achieves 79% of accuracy and natural language inference model which achieves 85% of accuracy.
- Introduce click model into Ranking Module and improve the performance on self-defined test set by 6%
- Introduce answer frequency feature into Ranking Module and improve the performance on self-defined test set by 4%
- Combining last two points, improve the overall performance on self-defined test set by 6.9%. Note: Quarterly departmental KPI is to enhance the performance by 8%.
- Based on XGBOOST, design and implement Confidence Model to filter those answers with lower confidence.

AI+Healthcare - Pharmacy Knowledge Graph

JD AI Research

Extract information from unstructured text, namely medicine specification, and build knowledge graph.

Feb. 2019 - Present

- Parse unstructured medicine specification into semi-structured format.
- With Protégé, design and implement ontology for knowledge graph.
- Train character-level Named Entity Recognition model, namely BiLSTM-CRF, on medical corpus, achieving 80.02 of F_1 score.
- With Stanford CoreNLP, develop rule-based templates.
- With Neural NER model and rule-based template, extract structured information at a finer granularity, achieve more than 80% overall accuracy.

Additional Information

Languages Skills

Self assessment - Common European Framework of Reference level

- Chinese: Mother tongue
- English: Listening C1, Reading C2, Spoken Interaction C1, Spoken Production C1, Writing B2 (equivalent to IELTS 7.5 - 8.0)

Postgraduate Academic Supervisors

Complete the final project under supervision

- Prof. Trevor Cohn: Assoicate Professor at The University of Melbourne
  trevor.cohn@unimelb.edu.au