

ZHENHAO LI

(+44) (0) 7422538781 \diamond zhenhao.li18@imperial.ac.uk

EDUCATION

Imperial College London, UK

October 2019 - Present

Supervisor: Prof. Lucia Specia

PhD in Natural Language Processing

Department of Computing

Imperial College London, UK

September 2018 - September 2019

Supervisor: Prof. Lucia Specia

MSc in Computing(Artificial Intelligence)

Department of Computing

degree: Distinction

Nanjing University, China

September 2014 - June 2018

Supervisor: Prof. Jidong Ge

BEng in Software Engineering

Overall Percentage: 87.8

RECENT RESEARCH

Improving Neural Machine Translation Robustness via Data Augmentation (Master Thesis) *May 2019 - Sept 2019*

Explored effect of data augmentation methods (back-translation, forward translation, fuzzy match) and improved model robustness. Proposed to use data from ASR/human transcripts from audios in improving NMT model performance on noisy texts.

Study in ZH-EN NMT with Different Granularity Levels

Jan 2019 - June 2019

Participate in WMT19 zh-en news translation shared task and compare NMT models with different granularity levels (word, subword, character and subcharacter). Showed the benefit of fine-grained word embedding under limited training data.

PageRank-Based Chinese Word Segmentation on Legal Documents (Bachelor Thesis)

Nov 2017 - June 2018

Presented an unsupervised Chinese Word Segmentation method based on the PageRank algorithm and applied on legal documents. Achieved 93/100 in the final presentation.

PUBLICATIONS

Zhenhao Li, Lucia Specia. "Improving Neural Machine Translation Robustness via Data Augmentation: Beyond Back-Translation" The 5th Workshop on Noisy User-generated Text (W-NUT 2019).

Zhenhao Li, Lucia Specia. "A Comparison on Fine-grained Pre-trained Embeddings for the WMT19 Chinese-English News Translation Task." The 4th Conference on Machine Translation (WMT2019).

PATENT

Jidong Ge, Chuanyi Li, **Zhenhao Li**, Miaomiao Lei, Linxia Yao, Xiaoyu Zhou, Bin Luo, "A Word Segmentation Method on Judicial Document Based on PageRank and Information Entropy.", CN Patent, Application Number:201810534689.4, 30 May 2018

INTERESTS

Neural Machine Translation, NLP on noisy texts