HUIMING WANG

huiming_wang@mymail.sutd.edu.sg | (+65) 80395663

EDUCATION

Zhejiang University (ZJU) *Sep. 2016 – Jul. 2021*

Double Major: B.Eng., Computer Science and B.Eng., Biosystems Engineering

- **GPA:** (Overall) 3.79/4.00, (CS) 3.93/4.00

Singapore University of Technology and Design (SUTD) *Sep. 2021 – Present*

PhD of Computer Science

- **Supervisor:** Prof. Soh De Wen

PREPRINTS & PUBLICATIONS

1. Guoshun Nan*, **Huiming Wang***, Wei Lu. "Multi-Level Contrastive Distillation for Semi-Supervised Relation Extraction." *In Findings of ACL. 2021*. (Withdraw)
2. Zhanming Jie*, **Huiming Wang***, Wei Lu, Lidong Bing, Luo Si. "Low-Resource NER with Latent Dependency Trees." (Draft for *EMNLP. 2020.* Short.)
3. **Huiming Wang***, Zhanming Jie*, Wei Lu, Lidong Bing, Luo Si. "Learning to Recognize New Entities with Shared-Latent Structures." (Draft for *ACL. 2020.*)

RESEARCH EXPERIENCE

FUTURE TOPICS

- **Data Augmentation for Various Autoregressive-based Generation Tasks**: a new data augmentation approach for generative frameworks, including at least aspect-based sentiment analysis and NER
- **Coarse-to-fine Token Classification**: a method designed to combine the advantages of both autoregressive and non-autoregressive generation strategies
- **Planned Generation**: due to the property that already generated tags can serve as guidance to other parts, we can design a latent variable to learn a generation order, rather than left-to-right or simultaneously
- **Sequence-to-sequence Learning with Transferred Language-specific Structures:** a hierarchical approach for sequence-to-sequence learning, where each node in the target structure is transduced by a node in the source structure
- **Directed Acyclic Graph for Span-based Extraction:** a method that can extract span-based terms (entity, aspect, or opinion) in linear time complexity, and can also deal with nested or discontinuous scenarios

ONGOING AND SELECTED FORMER TOPICS

StatNLP | ISTD Pillar | Singapore University of Technology and Design (SUTD)

*Advised by Professor **Wei Lu***

- **Weighted Left-to-right HMM for Non-autoregressive Machine Translation**
  - Based on the observations that existing directed acyclic graph-based non-autoregressive machine translation systems will achieve bad performance with last few tokens of each sentence, we proposed a global weighted HMM as an alternative
  - Rather than greedy or beam search, we further applied a Viterbi-style decoding algorithm to better fit the training procedure, and can find the best graph and output translation sequence simultaneously
- **Learning Better Representations for Semi-Supervised Relation Extraction**
  - Proposed a novel multi-level contrastive learning method that was able to learn better three-levels representations for Semi-Supervised Relation Extraction
  - Proposed a novel iterative knowledge distillation method to enhance the model's capability in representation learning from unlabeled data. It was demonstrated that such a joint and interactive learning approach was capable of yielding more meaningful representations
  - Conducted extensively quantitative and qualitative experiments on several standard benchmark datasets and confirmed the effectiveness of our approach under various SSRE settings
- **Learning New Entity Type Recognition with Fine-Grained Latent LSTM-CRF**

- Introduced shared encoding and shared negation structures to facilitate the determination of the relationship between rare type and other entity types.
- Proposed a shared-attribute CRF model in which the shared structures were represented as latent attributes. The proposed model was able to capture the interactions among the shared structures by using hyperedges to connect the latent attributes. And experimental results demonstrated the capability of the proposed model for improving the performance of rare entity types without any external knowledge.
- Established a framework to deal with low-resource tasks under most scenarios.

OTHERS

- **Learning Span-Level Interactions for Various NER and ABSA Subtasks:** proposed a span-level generation approach that each span can be determined by a single index, while still holding the ability to tackle various subtasks of both fields
- **Leveraging Label Semantic Information for NER:** conventional token-level NER systems are more like clustering rather than classification, as they use a fully randomly initialized classifier and it serves as measuring the similarities of representations (nodes in a high-dimension space) and classifier weights (clustering centroids)
- **Masked Entity Language Modeling:** proposed a new language modeling (fine-tuning objective) which randomly masks some tokens in a sentence and forces the model to predict both the entity tags and the tokens
- **Autoregressive Generation Based on RoBERTa**
- **Dynamic Programming Guided Language Modeling with LSTM:** proposed a new dynamic programming guided language modeling with LSTM, based on Markov assumption. In this modeling, we focused only on local feature window and merge equivalent states to keep a certain label (window) size, and decode with Viterbi
- **Low-Resource NER with Latent Dependency Trees:** proposed an end-to-end model that regards the syntactic structure as a latent variable and employed a latent biaffine dependency parsing module that allowed us to incorporate the loss of dependencies and perform multi-task learning during the training phase

Digital Media Computing & Design Lab | Dept. of Computer Science | ZJU *Jan. 2019 – May. 2020*

*Advised by Professor **Siliang Tang***

- **CoNLL Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies**
  - The focus of the task is learning syntactic dependency parsers that can work in a real-world setting, starting from raw text, and that can work over many typologically different languages, even low-resource languages for which there was little or no training data, by exploiting a common syntactic annotation standard
- **EMNLP Workshop FEVER: Fact Extraction and Verification**
  - Following the setting of *"Build it, Break it, Fix it"* from adversarial training, existing baseline models were fooled by the generated adversarial examples. Then new data from the *Breakers* was used to further improve their classification performance

Digital Media Computing & Design Lab | Dept. of Computer Science | ZJU *Mar. 2018 – May. 2019*

*Advised by Professor **Yueting Zhuang***

- **Speech Text Alignment Based on Neural Machine Translation Model**
  - Baseline: Trained an acoustic model with the audio data. Trained a LM with the text data and then decoded the audio to obtain time information of words in the sentences. Measured the distance of the original text with generated text and did the alignment. The accuracy of this baseline and pipeline system was 53%.
  - Attention-based Model: used a machine translation-based model with encoder-decoder architecture. A sequence of an input audio was projected into a continuous low dimensional space and the output sequence was generated from this representation. The output sequence was formed by the start and end indices of certain texts, so alignment could be executed soon after the output sequence was generated. This end-to-end speech-to-text alignment system could achieve 75% accuracy.

WORK EXPERIENCE

Eigen Technology Co., Ltd. *Jul. 2018 – Sep. 2018*

Natural Language Processing Algorithms Intern

Oracle (China) Software Systems Co., Ltd. *Apr. 2018 – Jul. 2018*

Big Data Development Intern & Solutions Specialist Intern

Selected Open-Source Projects

Operating System (MIPS SDK on C) [code]

- Developed an Operating System on FPGA-based hardware (supporting MIPS instructions).
- Designed the complete fair schedule (CFS) for process scheduling, system calls, and loading user programs.

Mini Database Design [code]

- Implemented a SQL database engine using B+ trees and supported storing and loading of the data.
- Simulated the MySQL database using minimizing codes (less than 2,000 lines).

Selected Honors

- **Zhejiang Provincial Government Scholarship,** *Top 2% Provincewide* *2017*
- **Tang Zhongying Scholarship,** *Top40 out of 6673 undergraduates, Lifetime honor* *2017*
- **First-Class Scholarship for Excellence in Research and Innovation** *2018 & 2019*
- **National Innovation Training Research Project,** *Top 20 out of 1000+ SRTP projects of ZJU* *2018*
- **Second-Class Scholarship for Outstanding Students,** *Top12% Uni-wide* *2017*
- **Second-Class Scholarship for Academic Excellence,** *Top12% Uni-wide* *2017*
- **Outstanding Student Leader Awards** *2017 & 2018*
- **Meritorious Winner in MCM/ICM, USA** *2019*
- **Honorable Mention in MCM/ICM, USA** *2018*