

JUE WANG

Email: zjuwangjue@gmail.com

EDUCATION

Zhejiang University, Master in Computer Science Sep 2018 - Mar 2021(Expected)
Université Paris Saclay (CentraleSupélec), Double Degree Sep 2016 - Jun 2018
Excellent Eiffel Scholarship 2016;
Zhejiang University, Bachelor in Electrical Engineering Sep 2014 - Jun 2018
GPA: 3.90/4.00 (top 10%), The First Prize Scholarship, Excellent Student Scholarship;

RESEARCH EXPERIENCE

Pyramid: A Layered Model for Nested Named Entity Recognition 2019

- Submission to ACL 2020. First Author.
- A novel layered model for nested NER. It achieves SOTA F1 scores on ACE-2004, ACE-2005, GENIA, NNE, which are 80.27, 79.42, 77.78, and 93.70 with conventional embeddings, and 87.74, 86.34, 79.31, 94.68 with pre-trained contextualized embeddings.

Slot-Filling via Question Answering: A Semi-Supervised Few-Shot Learning Approach 2019

- Submission to ACL 2020. First Author.
- A semi-supervised method for slot-filling. In particular, we formulate the task as slot-question answering and propose SAMIE, which learns to both select and answer slot questions itself. Supplied with very little labeled data (e.g. less than 100 sentences), SAMIE outperforms conventional supervised methods, increasing F1 scores by more than 10%.

EXACT: Attributed Entity Extraction by Annotating Texts 2019

- Accepted by SIGIR 2019. Assistant Developer and Presenter.
- Propose EXACT for extracting attributed entities from textual documents by performing explorative annotation tasks.

INTERNSHIP

Rokid - NLP Department, NLP Intern Jun 2018 - Dec 2018

- Use TextCNN to replace the original word2vec based classification module. Moreover, by pre-computing the convolution hidden states of common adjacent words, we avoid most online computation and dramatically improve the response speed. Pre-computed hidden states are stored in the hash table and updated weekly by the script.
- Implement BERT+CRF to replace the original NER module, with F1 0.91 => 0.96;
- Involved in IntentTree Project, which aims to build a domain-independent tree (with intents to be its branches) to address the problem of domain transferability. Domains are constantly changing while intents are relatively stable.
- Maintain the word embedding module (word2vec and ELMo).

ByteDance - ByteCamp, Camper in Algorithm Track Aug 2019 - Sep 2019

- Multimodal Classification. Achieve 3rd place in Algorithm Track.
- A hybrid model of tensor fusion (Tensor fusion network for multimodal sentiment analysis), multimodal attention (Hybrid Attention based Multimodal Network for Spoken Language Classification), and our implemented author gate.

PROJECTS

Court Document IE SDK, NLP July 2019 - Aug 2019

- Apply SAMIE and BERT to extract key information from court documents;

Event Extraction, NLP Jun 2017 - Aug 2017

- Reproduce and improve paper "On summarization and timeline generation for evolutionary tweet stream".

Lurne : voting platform for CentraleSupélec, Back-End/Node.js Dec 2016 - Feb 2017

- Participate in the design of the storage structure and APIs; implement functionalities in line with JSON-RPC.

TrelloGement : plugin for renting houses, Front-End/JavaScript Oct 2016 - Apr 2017

- Agile development: Work with Theodo to discuss product requirements, evaluate recent work and determine sprint goals;
- The idea is to manage information of different rental websites in a unified style on the kanban system, e.g. Trello.

SKILLS

- **Languages:** English, French and Chinese.
- **Frameworks:** PyTorch, Tensorflow, Keras.