朱逸伦

135-2142-5306 | allen.yl.zhu@gmail.com | GitHub: y-l-zhu | 微信: zhudalun1553

教育背景

乔治城大学 2017 年 8 月 - 2022 年 5 月(预计)
计算语言学 博士，GPA: 3.93/4.0 华盛顿
- 相关课程：自然语言处理, 统计机器翻译, 算法与数据结构, 语料库语言学

南京大学 2013 年 9 月 - 2017 年 6 月
英语语言学 / 计算机科学(辅修) 本科，GPA: 3.6/4.0 | 荣誉/奖项：南京大学优秀毕业生，2013-16 人民奖学金 南京
- 相关课程：微积分, 线性代数, 离散数学, 程序设计, 计算机系统

专业技能

- 编程/脚本语言：Python, C++, Java, Haskell (入门), Bash
- 框架 & 工具：Numpy, Pandas, Keras, Scikit-learn, TensorFlow, StanfordNLP, NLTK, 基于 Linux 编程, Git, MySQL, LaTeX
- 语言能力：汉语, 英语, 法语 (中级)

文章发表

- (即将发表) **Yilun Zhu**, Yue Yu (2019) "Evaluation on UCCA and USim for Paraphrase Detection."
- (收录) Yue Yu, **Yilun Zhu**, Yang Liu, Yan Liu, Siyao Peng, Mackenzie Gong and Amir Zeldes (2019) "GumDrop at the DISRPT2019 Shared Task: A Model Stacking Approach to Discourse Unit Segmentation and Connective Detection." In: *Proceedings of Discourse Relation Parsing and Treebanking (DISRPT 2019)* at NAACL-HLT 2019, Minneapolis, MN.
- **Yilun Zhu**, Yang Liu, Siyao Peng, Austin Blodgett, Yushi Zhao and Nathan Schneider (2019) "Adpositional Super-senses for Mandarin Chinese." In: *Proceedings of the Society for Computation in Linguistics (SCiL 2019)* at LSA 2019 Annual Meeting, New York, NY.
- **Yilun, Zhu** (2018) "Extreme Predicative Adjectives in Mandarin Chinese." Presentation at *8th International Conference on Formal Linguistics (ICFL-8)*, Hangzhou, China.

实习经历

中国电信 2018 年 5 月 - 2018 年 7 月
自然语言处理实习生 | 软件服务部，大数据研发中心 北京
- 创建一个用户情感分析系统，该系统可以根据用户对于某一车型的评论 (外观、动力等) 来预测该车型在用户中的口碑
- 使用 python 语言编写网络爬虫脚本，并爬取超过 11 万条、覆盖 513 种车型的用户评论；用 StanfordNLP 对用户评论进行分词处理，建立包含汽车相关术语的词库
- 使用 TensorFlow 框架建立 RNN 模型，使用预训练词向量和词库作为输入特征。该模型的 F1 达到 86.74，帮助准确筛选出好评率高的厂家和车型

科研&项目经历

评估 UCCA 和 USim 对于语义理解的帮助 2019 年 3 月 - 至今
- 使用 TUPA parser 处理 MSRP 语料库 (Microsoft Research Paraphrase Corpus) 中每一对语义相似的句子，从而生成相应的 UCCA 结构，以便于分析 UCCA 结构和 USim 分数能否准确反映出两个句子的语义相似度
- 搭建 Tree RNN 模型，并把 UCCA 结构和预训练词向量作为特征加入模型，来评估语法树对语义理解准确率的帮助

篇章自动划分系统 (GUMDROP) 2019 年 1 月 - 2019 年 3 月
- 搭建包含 wide(语言学特征) & deep(预训练词向量) 特征的 DNN 模型，帮助系统在大语料库中准确划分句子和篇章单元
- 搭建基于 Gradient boosting 的 ensemble 模型，整合基于词频统计和 bi-LSTM/CNN-CRF 框架，提高了 4%的预测准确率

语义共指消解与(非)命名体识别系统 (xrenner) 2018 年 2 月 - 2018 年 11 月
助研 | 指导老师：Amir Zeldes *博士* 华盛顿
- 创建了中文子系统的含有人名、地理知识、常见命名体、常见名词的列表，并将其整合到现有 xrenner 系统当中
- 使用 model stacking 方法搭建逻辑回归模型，用于选择当前预测最准确的基于规则的分类器，提高了 7%的准确率

中文介词的语义理解 2018 年 8 月 - 2018 年 10 月
助研 | 指导老师：Nathan Schneider *博士* 华盛顿
- 将 SNACS 拓展到中文，标注了 20 章节、11000+个词的《小王子》，并建立了可用于介词自动识别系统的标注标准

中文量词预测系统 2018 年 4 月
- 带领小组设计了一个利用语言学特征 (词类, 语境等) 和 SVM 模型准确预测中文量词的系统

实践活动

康平学社 *社长* | 举办"大文大理知识竞赛"，创办时事访谈节目"康平有话说" 2014 年 9 月 - 2015 年 6 月
南京大学外国语学院青年志愿者协会 *副会长* 2014 年 9 月 - 2015 年 6 月
守望者支教 英语教师 2014 年 7 月 - 2015 年 8 月

Yilun Zhu

135-2142-5306 | allen.yl.zhu@gmail.com | GitHub: y-l-zhu | WeChat: zhudalun1553

EDUCATION

Georgetown University Aug 2017 - May 2022 (expected)
PhD in Computational Linguistics, GPA: 3.93/4 Washington DC

- Courses: Empirical method in NLP, Data structs & algorithms, Statistical machine translation, Comp discourse modelling

Nanjing University Sept 2013 - June 2017
BA in English Linguistics and Literature (Hons.), Minor in Computer Science, GPA: 3.6/4.0 Nanjing, China

- Awards: 2013-16 Renmin Scholarship
- Courses: Calculus, Linear algebra, Discrete mathematics, Basics of C++, Introduction to computer systems

SKILLS

- Programming/scripting languages：Python, C++, Java, Haskell (入门), Bash
- Frameworks & tools：Numpy, Pandas, Keras, Scikit-learn, TensorFlow, StanfordNLP, NLTK, Linux, Git, SQL, LaTeX
- Natural languages：Mandarin (native), English (proficient), French (intermediate)

PUBLICATIONS & PRESENTATIONS

- (preparing) **Yilun Zhu**, Yue Yu (2019) "Evaluation on UCCA and USim for Paraphrase Detection."
- (to appear) Yue Yu, **Yilun Zhu**, Yang Liu, Yan Liu, Siyao Peng, Mackenzie Gong and Amir Zeldes (2019) "GumDrop at the DISRPT2019 Shared Task: A Model Stacking Approach to Discourse Unit Segmentation and Connective Detection." In: *Proceedings of Discourse Relation Parsing and Treebanking (DISRPT 2019)* at NAACL-HLT 2019, Minneapolis, MN.
- **Yilun Zhu**, Yang Liu, Siyao Peng, Austin Blodgett, Yushi Zhao and Nathan Schneider (2019) "Adpositional Super-senses for Mandarin Chinese." In: *Proceedings of the Society for Computation in Linguistics (SCiL 2019)* at LSA 2019 Annual Meeting, New York, NY.
- **Yilun, Zhu** (2018) "Extreme Predicative Adjectives in Mandarin Chinese." Presentation at *8th International Conference on Formal Linguistics (ICFL-8)*, Hangzhou, China.

INTERNSHIP

China Telecom May 2018 - July 2018
NLP Intern | Department of Software Service, Big Data R&D Center Beijing, China

- Built a sentiment analysis system, which assists clients to know the feedback of each attributes of car models from customers
- Prompted an automatic web crawler in Python to collect 110,000+ pieces of positive and negative evaluations of 513 types of cars from online comments; Parsed comments by using StanfordNLP and built a corpus about terminologies in motor vehicles
- Trained RNN by using TensorFlow, with pre-trained word embeddings and the corpus as input, to predict the sentimental polarity of each comment. The decent F-score 86.74 assists to filter out those car models with high qualities

RESEARCH

Evaluation on UCCA and USim for Paraphrase Detection Mar 2019 - present

- Use TUPA parser parsing each pair of paraphrases in MSRP corpus (Microsoft Research Paraphrase Corpus) to generate UCCA structures for analyzing whether UCCA and USim accurately reflect semantic similarities
- Build UCCA structure via Tree RNN and add pertained word embeddings as input, evaluating whether semantic structure contributes to paraphrase detection

Georgetown University Multilingual Discourse Region Partitioner (GUMDROP) Jan 2019 - Mar 2019

- Developed a DNN model with wide (linguistic features) and deep (pre-trained word embeddings) programming for sentence splitting and discourse unit segmentation, assisting the system to generalize and have better predictions in large corpora
- Prompted an ensemble by using a Gradient Boosting classifier to blend lookup Wikipedia frequency and a bi-LSTM/CNN-CRF sequence labelling framework for connective detection, increasing the predicting accuracy by .04 in 15 datasets in average

Externally configurable reference and non-named entity recognizer (xrenner) Feb 2018 - Nov 2018
Research Assistant | Supervisor: Dr. Amir Zeldes Washington DC

- Established benchmark entities (names, gazetteer, etc.) for a rule-based model in the Chinese subsystem
- Developed a Logistic Regression classifier with model stacking to predict named entities that are unseen the corpus by blending rule-based and CRF models, increasing the accuracy by .07 for coreference prediction

Semantic network of adposition and case supersenses (SNACS) for Chinese Aug 2018 - Oct 2018
Research Assistant | Supervisor: Dr. Nathan Schneider Washington DC

- Annotated 20 Chapter of The Little Prince and demonstrated the adaptability for SNACS annotation to Chinese, which can further support automatic disambiguation of adpositions in Chinese

The Prediction of Chinese Classifiers Apr 2018

- Led a team to develop a system to predict the correct classifier by using a SVM model with linguistic features (e.g. POS) as input

LEADERSHIP

KANG PING Academic Club, President, Nanjing University Sept 2014 - June 2015
Young Volunteers Association, Vice president, Nanjing University, School of Foreign Studies Sept 2014 - June 2015
English teacher (voluntary teaching), Yunnan, China July 2014 - Aug 2014