Sanxing Chen

Website: *www.cs.virginia.edu/~sc3hn/* E-mail: *sc3hn@virginia.edu* Tel: 434-466-9037

Research Interest

My long-term research interests lie primarily in helping computers to learn natural languages like humans. Currently, I focus on grounded language learning, especially in finding intuitive learning methods to facilitate language learning and reasoning with other structured knowledge.

Education

University of Virginia (Charlottesville) Aug. 2019 - Exp. Dec. 2020
M.S. in Computer Science
GPA **3.87**/4
Course: Machine Learning, Deep Learning for Visual Recognition, Natural Language Processing

China University of Geosciences (Beijing) Sept. 2015 - Jun. 2019
Bachelor of Computer Science
Senior GPA **3.66**/4

Experience

Natural Language Processing Group, University of Virginia
Student Researcher *supervised by Prof. Yangfeng Ji* Aug. 2019 - Now

- Researched in cross-domain Text-to-SQL semantic parsing, conducted experiments on the Spider dataset using different neural models, analyzed current state of the art models for their performance bottleneck.
- Explored the idea of using graph neural network models, e.g. RNN-based GNNs and the relational-aware transformer, to facilitate semantic parsing with grounding to relational databases.
- Explored the entity generation procedure in Text-to-SQL, formulated two linking processes from it, further strengthened their cooperation. We got one paper submitted to ACL 2020.

Bing Ads & Microsoft Research
Research Intern *supervised by Dr. Xiaodong Liu and Dr. Jian Jiao* May. 2020 - Now

- Researched in pretrained graph embeddings and NLP.

Big Data Mining Group, Microsoft Research Asia
Research Intern *supervised by Dr. Börje F. Karlsson* Mar. 2019 - Jun. 2019

- Researched in word representations for time expression recognition, built the first neural system which reaches the SOTA results in this task, submitted one paper to EMNLP 2019.
- Enhanced a rule-based time expression extraction system by adding an neural classifier to the pipeline to filter out false positive cases causing by ambiguous word senses.

Big Data Mining Group, Microsoft Research Asia
Research Intern *supervised by Guoxin Wang* Feb. 2018 - Sep. 2018

- Developed *Microsoft.Recognizers.Text* which is a toolkit provides robust multilingual support for recognition and resolution of entities like numbers, date/time, and units for downstream NLP applications. It powers pre-built entities in both *LUIS* and *Microsoft Bot Framework*, and is shipped to many Microsoft products including PowerPoint.
https://github.com/Microsoft/Recognizers-Text
- Built a learning-based statistical Time Expression Recognition model, which leverages segment-level features by a conditional random field model: responsible for cleaning and pre-processing of the datasets and create baselines for benchmarking.
- Researched broadly in NLP, mainly focus on Time Expression Recognition, shared papers in group hosted seminars, and participated in internal paper reviews.

Projects

Research in Domain Adaptation for Low-Resource Neural Machine Translation
Supervised by Prof. Xiaohui Ji Jan. 2019 - Mar. 2019

- Researched in transfer learning based methods for NMT, especially for low-resource domains and language pairs.

- Performed experiments on the WMT 2017 Chinese-English dataset using a Transformer model, analyzed experimental observations before and after adapting to the IWSLT 2017 TED dataset using fine-tune and mix fine-tune.
- Release the first open-sourced pre-processing code implementation of WMT 2017 Chinese-English dataset on GitHub.

Honors and Awards

Academic Excellence Fellowship
Department of Computer Science, University of Virginia Sept. 2019

Award of Technological Innovation Model
CUGB Annual Outstanding Students (only one undergraduate in the department) Nov. 2018

Award of Excellence
Stars of Tomorrow Internship Program in Microsoft Research Asia Sept. 2018

First Prize
Lanqiao Cup National Software and Information Technology Talents Competition Apr. 2018

Honorable Mention
Interdisciplinary Contest in Modeling Mar. 2018

Bronze Medal
the ACM-ICPC Asia Regional Contest Qingdao Site Nov. 2017

Bronze Medal
the ACM-ICPC Asia Regional Contest Nanning Site Nov. 2017

Bronze Medal
China Collegiate Programming Contest Harbin Site Oct. 2017

Qualifications

Programming Languages: Python, JavaScript, C#, C/C++, Java

Libraries and Tools: PyTorch, Linux, Git, Vim, HTML5, LaTeX, Adobe Suite

Music: Amateur Grade 10 Certificate in the Violin