

Liwei Cai

liweicai@andrew.cmu.edu • (412) 608-5497
linkedin.com/in/liwei-cai • github.com/cai-lw • liweicai.me - personal homepage

Education

Carnegie Mellon University (CMU) • School of Computer Science PITTSBURGH, PA
Master of Science in Intelligent Information Systems Aug 2018 – Dec 2019
Selected Coursework (in progress): Machine Learning for Text Mining, Search Engines, Design and Engineering of Intelligent Information Systems.

Tsinghua University (THU) • Department of Electronic Engineering BEIJING, CHINA
Bachelor of Engineering in Electronic Information Science and Technology Aug 2014 – Jul 2018
GPA: 89/100 (top 20% in the department)
Selected Coursework: Data and Algorithms, Modern Computer Architecture, Operating Systems, Functional Programming, Pattern Recognition, Advanced Applications of Machine Learning.

Work Experience

Mobvoi BEIJING, CHINA
NLP Intern Apr 2018 - Jul 2018

- Designed architectures of knowledge base ontology and automatized pipeline for knowledge extraction with teammates.
 - Investigated and developed the knowledge storage and query module, which represented facts as typed RDF triples and stored them in MySQL, and translated queries into SQL with type check.
 - Migrated the existing unannotated knowledge base to the new type-safe one, increasing the robustness of the question answering (QA) service and the whole voice assistant ecosystem it supports.
-

Research Experience

UC Santa Barbara • Natural Language Processing Lab SANTA BARBARA, CA
Research Assistant (Advisor: William Wang) Jun 2017 – Sep 2017

- Proposed a generative adversarial model that can adaptively generate better negative training data according to the behavior of the knowledge graph embedding (KGE) model being trained.
- Implemented the model in PyTorch and conducted experiments on various datasets to demonstrate its ability to improve the Hit@10 performance metric of existing KGE models by up to 3%.
- Authored the paper *KBGAN: Adversarial Learning for Knowledge Graph Embedding* as the first author and published it in NAACL 2018. Paper (bit.ly/cai18kbgan) and code (bit.ly/kbgan) are available online.

Projects

Content-based Recommender System for Electronic Medical Records (EMRs) *Thesis project at THU, Mar 2018 - Jun 2018*

- Developed a two-stage content-based text recommender system, in which candidate documents are first retrieved by approximate nearest neighbor search, and then ranked by various methods based on word embeddings and RNN models.
- Conducted comprehensive experiments on an EMR dataset to demonstrate its improvement in multiple precision metrics by up to 6%, compared to traditional vector space models.

Image Captioning in Chinese (code available online: bit.ly/imgcapzh) *Course project at THU, May 2017 - Jun 2017*

- Implemented and tuned recurrent neural network (RNN) with and without attention mechanism from scratch in Tensorflow. The model ranked top 10% in the class in performance on concealed test set.
- Incorporated dropout, skip connection, beam search width limitation and short sequence punishment to control overfitting.

Interpreter of Lisp-like Language (code available online: bit.ly/lispintpr) *Course project at THU, Dec 2016 - Jan 2017*

- Developed an interpreter of a Lisp-like mini language in Haskell, using a modified “finally tagless” mechanism for computation and type check, and parser combinator library for parsing source code.
- Supported lambda calculus as well as primitive numeric/boolean/string operations with dynamic type check.

Skills

Programming Languages: Python, C++, MATLAB, Java, JavaScript, Haskell, Rust
Softwares and Tools: PyTorch, Tensorflow, Git, MongoDB, MySQL
Development Environment: Windows and Linux