Han Cao

Tel: (+86) 18801313786| Email: caoh16@mails.tsinghua.edu.cn

RESEARCH INTERESTS

Data/Text Mining, Natural Language Processing, Machine Learning & Their Applications

EDUCATION

Tsinghua University **09/2016 - 06/2020(expected)**

Bachelor of Electronic Engineering

- **Overall GPA:** 3.86/4.0 (ranked 14/266, top 6% in the department)
- **Major GPA:** 3.90/4.0
- **Computer Science& Mathematics Courses**: Linear Algebra (1) A (4.0/4.0); Probability and Stochastic Processes (1) A (4.0/4.0); Probability and Stochastic Processes (2) A+(4.0/4.0); C++Programme Design and Training A (4.0/4.0);Database A (4.0/4.0); Operating System A (4.0/4.0); Introduction to Auditory-visual Information System A (4.0/4.0); Digital Image Processing A (4.0/4.0)

RESEARCH EXPERIENCE

Transformer-based Models for Knowledge Base Completion **08/2019 – now**

Independent Student Research Project at UC Santa Barbara, Supervised by Assistant Professor William Wang

- Apply BERT to model paths on Knowledge Graphs generated by random walk.
- Our model consists of a two-layer model for pretraining and a one-layer model for downstream tasks including link prediction and triple classification.
- Achieves competitive results compared with current state-of-the-art method: MRR: 0.346, MR: 227, Hits@10: 0.533 on FB15k237 dataset and MRR: 0.455, MR: 3918, Hits@10: 0.521 on WN18RR dataset.

User Identity Linkage via Deep Neural Network from Heterogeneous Mobility Data **02/2019 – 05/2019**

Student Research Project at Tsinghua University, Supervised by Associate Professor Yong Li

- Our task is to find the same users by matching the personal trajectory data from different apps.
- We proposed a model which contains an RNN based encoder and a location encoder to extract spatio-temporal features, after which attention based network will select similar parts for MLP to calculate the similarity score.
- Our model outperforms the state-of-the-art solutions by more than 15% in terms of hit-precision.
- My contribution: obtained results of an important baseline method (ICDE 2018), improved the RNN based trajectory encoder to achieve better performance and conducted extensive hyperparameter experiments.

PUBLICTIONS

Feng, Jie; Li, Yong; Zhang, Mingyang; Yang, Zeyu; Wang, Huandong; **Cao, Han**; Jin, Depeng, "User Identity Linkage via Co-Attentional Neural Network From Heterogeneous Mobility Data" submitted to TKDE 2019

SELECTED COURSE WORKS

Sound Source Separation and Location Based on Audio-visual Information **12/2018 - 01/2019**

The course work of Introduction to Auditory-visual Information System

- The task is to locate the solo sound source in the video and separate the solo audios from duet audios.
- Based on the idea of one paper in *ECCV 2018*, we applied the model Inception_V3 to predict the sound source. Then we used MIML (multi-instance multi-label) model and NMF (non-negative matrix factorization) algorithm to separate the duet audios based on the predicted sound source.
- The model achieved outstanding performance. I'm one of the **7 students** who got **full marks** in this project among **150 students** and I also got an A (4.0/4.0) in this course.

PROFESSIONAL SKILLS

Programming: Proficient in C/C++, Python, MATLAB, Linux, Pytorch, Spark, Hadoop

English: TOEFL IBT: Total 103(Reading 28/Listening 26/Speaking 22/Writing 27)

GRE: 325(V155/Q170) +AW3.5

HONORS& AWARDS

National Inspirational Scholarship (TOP10%) 2019

National Inspirational Scholarship 2018

Friend of Tsinghua University—Huang YiCong Couple Scholarship 2017