Denghui Zhang

Rutgers University, NJ, US

☐ 8623158748 | ✉ ustbdenghui@gmail.com | ⌂ zhangdenghui.site

Educations

Rutgers University, New Jersey *Sep. 2018 - Now*

- Ph.D. in Information Technology, at Rutgers Data Mining Group
- Advisor: Prof. Hui Xiong

Insitite of Computing Technology, Chinese Academy of Sciences, Beijing, China *Sep. 2015 - Jun. 2018*

- M.S. in Computer Science and Technology, at CAS Key Laboratory of Network Data Science and Technology
- Advisor: Prof. Jun Xu, and Prof. Yuanzhuo Wang
- Chinese Academy of Sciences Freshman Scholarships (Top 10%), CAS Excellent Student Awards (Top 15%)

University of Science and Technology Beijing, Beijing, China *Sep. 2011 - Jun. 2015*

- B.E. in the Communication Engineering, at School of Computer and Communication Engineering
- National Scholarship (**Top 2%**), National Motivational Scholarship (Top5%), USTB Excellent Student Awards (Top 5%)

Publications

CIKM'19 ***Job2Vec: Job Title Benchmarking with Collective Multi-View Representation Learning***
Denghui Zhang, Junming Liu, Hengshu Zhu, Yanchi Liu, Lichen Wang, Pengyang Wang, Hui Xiong
The 28th ACM International Conference on Information and Knowledge Management (CIKM), 2019.
(regular paper, accepted)

AAAI'18 ***Path-Based Attention Neural Model for Fine-Grained Entity Typing***
Denghui Zhang, Manling Li, Pengshan Cai, Yantao Jia, Yuanzhuo Wang
The Thirty-Second AAAI Conference on Artificial Intelligence, 2018. *(poster, accepted)*

WI'17 ***Efficient Parallel Translating Embedding For Knowledge Graphs***
Denghui Zhang, Manling Li, Yantao Jia, Yuanzhuo Wang, Xueqi Cheng
The IEEE/WIC/ACM International Conference on Web Intelligence, 2017. *(regular paper, accepted)*

IEEE Trans. ***Link Prediction in Knowledge Graphs: A Hierarchy-Constrained Approach***
Manling Li, Denghui Zhang, Yantao Jia, Yuanzhuo Wang, Xueqi Cheng
IEEE Transaction on Big Data Special Issue on Knowledge Graphs: Techniques and Applications, 2017.
(regular paper, under 2nd review)

RESEARCH INTEREST

General: Data Mining, Natural Language Processing, Knowledge Graph, Representation Learning
Applications: Talent Intelligence, Spatiotemporal Prediction

Interns

Data Science Department, NEC Laboratories America, Research Intern *May. 2019 - August. 2019*
Supervisor: Yanchi Liu

- Spatiotemporal traffic volume prediction: Proposed Spatial-Temporal Multi-Scale Multi-Granularity Network (ST-MSGN) for site-level traffic volume prediction, by modeling the complex spatial and temporal dependencies and their interactions. Conduct experiments on real-world datasets (NYC-Bike and NYC-Taxi) to validate the effectiveness of the model.

Baidu Talent Intelligence Center, Research Intern *June. 2018 - August. 2018*
Supervisor: Hui Xiong

- Resume distribution: Developed resume recommender system for Baidu HR department, using NLP techniques to obtain the resume-job similarities and distribute resumes according to similarity scores and the headcount of different departments.
- Job title benchmarking: Developed a data-driven approach to match job titles with similar expertise levels across various companies.

Research Projects

Job2Vec: Job Title Benchmarking with Collective Multi-View Representation Learning
Natural Science Foundation of China (NSFC) No.91746301

Jun. 2018 - Dec. 2018

Supervisor: Hui Xiong

- Problem: Proposed a data-driven approach to match job titles with similar expertise levels across various companies.
- Built Job Graph based on collected online career trajectories.
- Eliminating redundancy of Job Graph by filtering noisy words.
- Proposed a multi-view representation learning method, named Job2Vec, and performed link prediction using the embeddings to enrich Job Graph, extract job title benchmarking insights from refined Job Graph using voting to match.
- Conducted extensive experiments to validate the effectiveness of Job2Vec.

Path-Based Attention Neural Model for Fine-Grained Entity Typing
Natural Science Foundation of China (NSFC) No.61572469

Jun. 2017 - Sep. 2017

PI: Yuanzhuo Wang

- Problem: Fine-grained entity typing employed distant supervision to automatically generate training data. It labeled entities with types in knowledge bases without considering the certain context of entities, and thus introduced noises.
- Came up with the idea to use attention mechanism to dynamically reduce the weights of wrong labeled sentences for each type during training.
- Proposed an end-to-end typing model, named the path-based attention neural model (PAN), by leveraging the distinct hierarchical structure of types to learn a noise robust performance. In this way, the attention learned for a type could assist the learning of the attention for its subtype.
- Experimented PAN by PyTorch, and improved the fine-grained typing accuracy for 4 percent.

Efficient Parallel Framework for Knowledge Graph Embedding
Natural Science Foundation of China (NSFC) No. 61402442

Jan. 2017 - Apr. 2017

PI: Yantao Jia

- Problem: Knowledge graph embedding aimed to embed entities and relations into low dimensional spaces. However, a major limitation of these methods was the time consuming training process, especially for large scale knowledge graphs.
- Proposed a lock free framework for training knowledge graph embedding in parallel, called ParTrans-X.
- Proved the validity of ParTrans-X by formulating the training data of knowledge graph into hypergraphs, and analyzing the collision probability of different threads.
- Speeded up the training process by more than an order of magnitude, with scaling performance along with increasing number of processors.

Link Prediction Using Hierarchical Information in Knowledge Graphs
National Grand Fundamental Research 973 Program of China No.2014CB340401

Dec. 2016 - Apr. 2017

Co-PI: Yuanzhuo Wang

- Problem: 50% triples in knowledge graphs were organized in hierarchical structures, which also contained rich inference patterns to predict links but do not be fully used.
- Divided the hierarchical structures into two categories, i.e., single-step and multi-step hierarchical structures.
- Participated in designing a link prediction method based on knowledge graph embedding, called **hTransM**. It can separate negative and positive examples by optimal margin, which is based on single-step and multi-step specific margin.
- Conducted the experiment on three datasets to demonstrate the effectiveness of hTransM.
- Coauthored a regular paper submitted to **IEEE Transaction on Big Data** (Second Author, under 2nd review).

Big Data Analysis Platform
Key Program of CAS Key Laboratory of Network Data Science and Technology

Sep. 2016 - Dec. 2016

Advisor: Jun Xu

- Objective: Provided a general-purpose dataflow-based system to ease the process of applying machine learning algorithms to real world tasks (online demo: http://159.226.40.104:18080/dev/). It was accepted by **CIKM 2016 demo**.
- Implemented GBDT (Gradient Boosting Decision Tree) algorithm on Spark.
- Optimized the GBDT algorithm on Spark, which ran faster than the official version of Spark MLlib, e.g. 2.4 times faster given 100GB data.
- Optimized the IO cost and storage cost of the workflow using Parquet file format.

Skills

Language	Python, C/C++, Java, Scala, MySql, Shell
Tools	Keras, PyTorch, Tensorflow, Spark MLlib, Sklearn, Pandas