Denghui Zhang

Institute of Computing Technology, Chinese Academy of Sciences, No.6 Kexueyuan South Road, Beijng, China
(+86) 13269365613 | zhangdenghui@ict.ac.cn | zhangdenghui.tech

Educations

Insititute 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 Excellent Student Awards (**Top 5%**), Chinese Academy of Sciences Scholarships (Top 10%)

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
- Major GPA: 3.83/4.0 (90.9/100), Overall GPA: 3.57/4.0 (86.7/100), WES GPA: 3.61/4.0
- National Scholarship (**Top 2%**), National Motivational Scholarship (Top5%), USTB Excellent Student Awards (Top 5%)

Research Interest

Natural Language Processing and Data Mining

Publications

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)*

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.

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

Research Projects

Path-Based Attention Neural Model for Fine-Grained Entity Typing *Jun. 2017 - Sep. 2017*
Natural Science Foundation of China (NSFC) No.61572469 *PI: Yuanzhuo Wang*

- Problem: Fine-grained entity typing employs distant supervision to automatically generate training data. It labels entities with types in knowledge bases without considering the certain context of entities, and thus introduces noises.
- Came up with the idea to use attention mechanism to dynamically reduce the weights of wrong labeled sentences for each type.
- Designed path-based attention mechanism using the paths in type hierarchy of the knowledge base, so that the attention learned for a parent type could assist the learning of the attention for its subtype.
- Proposed an end-to-end typing model, called the path-based attention neural model (PAN), by leveraging the distinct hierarchical structure of types to learn a noise robust performance.
- Experimented PAN by PyTorch, and improved the fine-grained typing accuracy for 4 percent.
- Wrote a poster paper which was accepted by **AAAI 2018** (Lead Author).

Efficient Parallel Framework for Knowledge Graph Embedding *Jan. 2017 - Apr. 2017*
Natural Science Foundation of China (NSFC) No. 61402442 *PI: Yantao Jia*

- Problem: Knowledge graph embedding aims to embed entities and relations into low dimensional spaces. However, a major limitation of these methods is 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.
- Speeded up the training process by more than an order of magnitude, with scaling performance along with increasing processors.
- Wrote a regular paper which was accepted by **WI 2017** (Lead Author).

Link Prediction Using Hierarchical Structures in Knowledge Graphs *Dec. 2016 - Apr. 2017*
National Grand Fundamental Research 973 Program of China No.2014CB340401 *Co-PI: Yuanzhuo Wang*

- Problem: Hierarchical structures contain rich inference patterns to predict links, but do not be fully used.
- Participated in designing a link prediction method based on knowledge graph embedding, called hTransM, which separated negative and positive examples by optimal single-step and multi-step specific margin.
- Came up with the idea of optimal margin calculation delay strategy, and speeded up the training process for ten times.
- Implemented and experimented hTransM on three datasets to demonstrate the effectiveness of hTransM.
- Coauthored a regular paper submitted to **IEEE Transactions on Big Data** (Second Author).

Engineering Projects

Knowledge Graph Construction and Analysis for Videos *Jan. 2017 - Apr. 2017*

Collaborative Project of Chinese Academy of Sciences and Huawei Inc. *Advisor: Yantao Jia*

- Objective: Constructed a knowledge graph for videos, including movies, series, etc. The knowledge graph could be enriched automatically from different data source, and supported a plenty of applications.
- Implemented the parallel version of key modules for knowledge graph construction, including ontology alignment, relation extraction, relation inference and tag inference algorithms, which are based on knowledge graph embedding.
- Shortened the incremental procedure to 5-6 hours from 3 days given 60k video entities, which was regarded as one of the main innovations by Huawei Inc.

Big Data Analysis Platform *Sep. 2016 - Dec. 2016*

Key Program of CAS Key Laboratory of Network Data Science and Technology *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 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.

Agricultural Product Price Prediction Program *Jan. 2017 - Apr. 2017*

CCF Big Data and Computer Intelligence (BDCI) Agricultural Product Price Prediction Competition *Advisor: Xueqi Cheng*

- Objective: Predicting the future price of certain agricultural products in China, based on the historical price data.
- Worked in 5 member team and our algorithm **ranked 2 out of 547 teams**, with the first prize vacancy.
- Implemented GBDT Model and Random Forest for ensemble.
- Proposed the loss function by weighting more to the data with lower price according to this specific problem.

Commodity Demand Prediction *Apr. 2016 - Jun. 2016*

AliBaba Commodity Supply Chain Prediction Competition *Individual Work*

- Objective: Predicted the future sales of certain products in TaoBao, using historical sales data and user behavior data.
- Designed and implemented the algorithm individually, and **ranked 13 out of 2807 teams**.
- Implemented GBDT algorithm using Java in MapReduce framework to make it run on the appointed platform ODPS.
- Designed the loss function to assign different loss to lower and higher prediction, which promoted the rank by 100+.

Connect Six: A Computer Game *Jul. 2014 - Sep. 2014*

National Computer Game Tournament *Advisor: Ke Zhou*

- Objective: Developed a computer program to play the board game Connect Six.
- Lead a team of three people and won the **national first prize**.
- Implemented Alpha-Beta Search algorithm to search the game tree of Connect Six.
- Optimized the search algorithm by adding VCF(Victory of Continuous Four) strategy.

Honors

2016.10	CCF BDCI Agricultural Product Price Prediction Competition (**Rank: 2/547, with the first prize vacancy**)	
2016.06	AliBaba Tianchi Commodity Supply Chain Prediction Competition (**Rank: 13/2807**)	
2014.08	National Computer Game Tournament (Connect Six) (**First prize**)	
2014.08	National Computer Game Tournament (Nogo) (**Second prize**)	
2013.10	National Competition of International Contest of Innovation (iCAN'13) (**Second prize**)	
Scholarship	National Scholarship (**Top 2%**, 2015)	Chinese
	National Motivational Scholarship (**Top 5%**, 2014)	
	Academy of Sciences Scholarships (**Top 10%**, 2015)	
	USTB Excellent Student Awards (**Top 5%**, 2012, 2013)	
Awards	USTB Excellent Student Leader Awards (**Top 10%**, 2013)	
	Chinese Academy of Sciences Excellent Student Awards (**Top 5%**, 2016)	

Skills

Language	Python, C/C++, Java, Scala, MySql, Shell
Tools	Sklearn, Spark MLlib, PyTorch, Tensorflow
English	Tofel: 102, GRE: Verbal 153, Quantities 167, AW 3.5

Extracurricular Activities

2012-2013	Vice-President and Publicity Department Director of the Source Technology Association
2015-2017	Haidian Hospital Volunteer, Yanqi Lake Forest Ranger Water Delivery Volunteer