Denghui Zhang

Institute of Computing Technology, Chinese Academy of Sciences, No.6 Kexueyuan South Road, Haidian, Beijng, China

(+86) 13269365613 | zhangdenghui@ict.ac.cn | zhangdenghui.tech

Educations

Insititute of Computing Technology, Chinese Academy of Sciences, Beijing, China *Sep. 2015 – June.2018*

- M.E. 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 - June.2015*

- B.E. in the Communication Engineering, at School of Computer and Communication Engineering
- Major GPA: 3.85/4.0, Overall GPA: 3.57/4.0 (Top 5%)
- National Scholarship (Top 2%), National Motivational Scholarship (Top5%), USTB Excellent Student Awards (Top 5%)

Publications

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

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, minor revision)*

RESEARCH INTEREST

Natural Language Processing, Knowledge Graph and Machine Learning

Research Projects

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

- Problem: Fine-grained entity typing employs distant supervision to automatically generate training data, which labels the entities with types in knowledge bases, but it ignores the certain context of entities and thus introduce noises in training data.
- Came up with the idea to use attention mechanism to dynamically reduce the weights of wrong labeled sentences for each type.
- 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, which can improve the fine-grained typing accuracy for 4 percent.
- The paper was submitted to **AAAI 2018** (Lead Author), and is under review now.

Efficient Parallel Framework for Translating Embedding of Knowledge Graphs *Jan. 2017 - Apr. 2017*
Natural Science Foundation of China (NSFC) No. 61572473 *Advisor: Yuanzhuo Wang*

- Problem: Knowledge graph translating embedding aims to embed entities and relations into low-dimensional vector spaces. However, a major limitation of these methods is the time consuming training process, which may take several days or even weeks for large knowledge graphs, and result in great difficulty in practical applications.
- 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.
- ParTrans-X can speed up the training process by more than an order of magnitude, with scaling performance along with increasing number of processors.
- The paper was accepted as a regular paper in **WI 2017** (Lead Author).

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

- Problem: 50% triples in knowledge graphs are organized in hierarchical structures, which also contain rich inference patterns to predict links but do not be fully used.
- Participated in dividing the hierarchical structures into two categories, i.e., single-step hierarchical structures and multi-step hierarchical structures.
- Participated in proposing a link prediction method based on knowledge graph embedding, called **hTransM**, which can separate negative and positive examples by using single-step and multi-step specific margin.
- Conducted the experiment on three datasets to demonstrate the effectiveness of hTransM.
- The paper was submitted to **IEEE Transaction on Big Data** (Second Author) and is in a revision now.

Big Data Analysis Platform   *Sep. 2016 - Dec. 2016*
KEY PROGRAM OF CAS KEY LABORATORY OF NETWORK DATA SCIENCE AND TECHNOLOGY   *Advisor: Cao Lei, Jun Xu*

- Objective: Providing a GUI web tool and Machine Learning library for data scientists to make workflows of various data mining applications fast and easily. The demo paper was accepted by CIKM 2016. (online demo: http://159.226.40.104:18080/dev/)
- Implemented GBDT (Gradient Boosting Decision Tree) algorithm in Spark.
- Optimized the GBDT algorithm in Spark, and it runs faster than Spark MLlib, e.g. 2.4 times faster with 100GB data.
- Optimized the IO cost and storage cost of the workflow using parquet file format.

Projects

Knowledge Graph Construction and Analysis for Videos   *Jan. 2017 - Apr. 2017*
COLLABORATIVE PROJECT OF CHINESE ACADEMY OF SCIENCES AND HUAWEI INC.   *Advisor: Yuanzhuo Wang*

- Objective: Constructing a knowledge graph for videos, including movies, series, etc. The knowledge graph can be enriched automatically from different data source, and supports a plenty of applications such like relation extraction, tag inference.
- Implemented the parallel version of key knowledge graph construction modules, including ontology alignment algorithm, relation extraction algorithm and tag inference algorithm, which are based on knowledge graph embedding.
- Shorted the whole updating procedure to 5-6 hours from 3 days with 60k video entities, which was regarded as one of the main innovations by Huawei Inc.

Commodity Demand Prediction Program   *Apr. 2016 - Jun. 2016*
ALIBABA COMMODITY SUPPLY CHAIN PREDICTION COMPETITION

- Objective: Predicting the future sales of certain products in TaoBao, using the historical sales data and user behavior data.
- Designed and implemented the algorithm individually, which **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 lead to a great promotion in accuracy.

Connect Six: A Computer Game   *Jul. 2014 - Sep. 2014*
NATIONAL COMPUTER GAME TOURNAMENT   *Advisor: Ke Zhou*

- Objective: Developing a computer program to play the board game Connect Six.
- Lead a team of three people and our team 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

Oct. 2016	CCF Agricultural Product Price Prediction Competition (**Rank: 2/547**)
Jun. 2016	AliBaba Commodity Supply Chain Prediction Competition (**Rank: 13/2807**)
Aug. 2014	National Computer Game Tournament (Connect Six) (**First prize**)
Oct. 2013	National College Competition on Internet of Things (**Second prize**)
Scholar	Chinese Academy of Sciences Scholarships (**Top 10%**, 2015)
-ship	National Scholarship (**Top 2%**, 2015), National Motivational Scholarship (**Top 5%**, 2014)
Awards	Excellent Student Awards (**Top 5%**, 2012, 2013, 2016), Excellent Student Leader Awards (**Top 10%**, 2013)

Skills

Language	Python, C/C++, Java, Scala, MySql, Shell
Tools	Sklearn, Spark MLlib, PyTorch, Tensorflow
Courses	Machine Learning, Natural Language Processing, Data Mining, Information Retrieval, Knowledge Graph, etc.
English	Tofel: 102, GRE: 320 (Verbal 153, Quantities 167) + AW 3.5