

CHENWEI ZHANG SINCE 1991

CONTACT INFORMATION

500 9th Ave N, Seattle, WA 98109

✉ cwzhang910@gmail.com

✉ cwzhang@amazon.com

🌐 www.cwzhang.com

RESEARCH INTERESTS

My research interests lie in the field of Data Mining and Natural Language Processing. In particular, I am interested in text mining and mining structured information from heterogeneous information sources.

EDUCATION

Ph.D., Computer Science

University of Illinois at Chicago

Advisor: Prof. Philip S. Yu

PhD Thesis: Structured Knowledge Discovery from Massive Text Corpus

2014 - 2019

Chicago, IL

B.Eng., Computer Science and Technology

Southwest University

GPA: 3.9/4.0 (ranking 1/63)

2010 - 2014

Chongqing, China

WORK EXPERIENCE

Amazon

Applied Scientist at Amazon, Product Graph

- Building an authoritative knowledge graph for every product at Amazon, with the goal of answering any question about products and related knowledge.
- Conducted research on information extraction, text/graph mining and representation learning.

Aug. 2019 - Now

Seattle, WA

University of Illinois at Chicago

Research Assistant at Big Data and Social Computing (BDSC) Lab

- Mining user-generated text, such as text queries, question-answering corpus.

- Mining structured information from heterogeneous information sources, or multi-modality data.

Aug. 2017 - May. 2019

Chicago, IL

Tencent America

Research Intern at Tencent Medical AI Lab

Project: Question Intent Detection, Slot Filling and Entity Synonym.

- Keywords: Natural Language Processing, Entity Synonym.
- Complexity and Technology: 5K lines of Python code.
- Accomplishments: 1) Developed neural network models for joint intent detection and slot filling. 2) Constructed an end-to-end training/serving pipeline for large-scale synonym discovery built upon Tensorflow and Elasticsearch. 3) Developed neural network models for entity synonym detection based on contexts from massive corpus.

May 2018 - Aug. 2018

Palo Alto, CA

Baidu Research

Research Intern at Big Data Lab

Project: Generative Knowledge Graph Completion.

- Keywords: Knowledge Graph, Entity Relation Extraction, Deep Generative Models.
- Complexity and Technology: 3K lines of Python code.
- Accomplishments: 1) Developed deep generative models for knowledge graph completion. 2) Built a model to infer the relationship between entities in a context. 3) Augmented the scale of the existing knowledge graph used for production by 61.93% with 92.91% accuracy, with no additional human annotation.

May 2017 - Aug. 2017

Sunnyvale, CA

Baidu Research

May 2016 - Jul. 2016

Research Intern at Big Data Lab

Sunnyvale, CA

Project: Inferring User Intents in Community Question Answering.

· Keywords: Question Answering, Big Data, MapReduce.

· Complexity and Technology: 1K lines of Python & Bash Scripts (Spark), 4K lines of Python code.

· Accomplishments: 1) Collected and cleaned a large amount of data from multiple data sources (Online question-answering forums, etc). 2) Built a model to infer multiple pieces of user intents from a corpus on community question-answering. 3) Improved the intent detection accuracy in multi-round question-answering by 8% on AUC.

Baidu Research

May 2015 - Aug. 2015

Research Intern at Big Data Lab

Sunnyvale, CA

· Proposed deep neural network based intention detection models for medical text queries.

· Applied extensive data analysis on medical question answering text corpus.

Southwest University

Apr. 2013 - Jul. 2014

Undergraduate Research Assistant at Information Fusion Lab

Chongqing, China

· Developed uncertainty-aware methods to ensemble classifiers.

· Fused multiple information sources to detect spam emails.

PUBLICATIONS

- Nasser Zalmout, **Chenwei Zhang**, Xian Li, Yan Liang, and Xin Luna Dong. All You Need to Know to Build a Product Knowledge Graph. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery Data Mining (KDD), 2021.
- Xuming Hu, **Chenwei Zhang**, Yawen Yang, Xiaohe Li, Li Lin, Lijie Wen, and Philip S. Yu. Gradient Imitation Reinforcement Learning for Low Resource Relation Extraction. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.
- Xuming Hu, **Chenwei Zhang**, Fukun Ma, Chenyao Liu, Lijie Wen, and Philip S. Yu. Semi-supervised Relation Extraction via Incremental Meta Self-Training. In Findings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.
- Liqiang Xiao, Jun Ma, Xin Luna Dong, Pascual Martínez-Gómez, Nasser Zalmout, **Chenwei Zhang**, Tong Zhao, Hao He, and Yaohui Jin. End-to-End Conversational Search for Online Shopping with Utterance Transfer. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.
- Xinyang Zhang, **Chenwei Zhang**, Xin Luna Dong, Jingbo Shang, and Jiawei Han. Minimally-Supervised Structure-Rich Text Categorization via Learning on Text-Rich Networks. In Proceedings of the Web Conference (TheWebConf), 2021.
- Congying Xia, **Chenwei Zhang**, Jiawei Zhang, Tingting Liang, Hao Peng, Philip S. Yu. Low-shot Learning in Natural Language Processing. In Proceedings of the Second IEEE International Conference on Cognitive Machine Intelligence (CogMI), 2020.
- Xuming Hu, **Chenwei Zhang**, Yusong Xu, Lijie Wen and Philip S. Yu. SelfORE: Self-supervised Relational Feature Learning for Open Relation Extraction. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
- Hoang Nyugen, **Chenwei Zhang**, Congying Xia, Philip S. Yu. Semantic Matching and Aggregation Network for Few-shot Intent Detection. In Findings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
- Hao Chen, **Chenwei Zhang**, Jun Li, Philip Yu, Ning Jing. KGen: A Generative Approach for Incipient Knowledge Graph Population. In IEEE Transactions on Knowledge and Data Engineering (TKDE), 2020.

- **Chenwei Zhang**, Yaliang Li, Nan Du, Wei Fan, Philip S. Yu. Entity Synonym Discovery via Multi-piece Bilateral Context Matching. In Proceedings of the 29th International Joint Conference on Artificial Intelligence and the 17th Pacific Rim International Conference on Artificial Intelligence (IJCAI-PRICAI), 2020.
- Yuning Mao, Tong Zhao, Andrey Kan, **Chenwei Zhang**, Xin Luna Dong, Christos Faloutsos and Jiawei Han. Octet: Online Catalog Taxonomy Enrichment with Self-Supervision. In Proceedings of the 2020 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2020.
- Xin Luna Dong, Xiang He, Andrey Kan, Xian Li, Yan Liang, Jun Ma, Yifan Ethan Xu, **Chenwei Zhang**, Tong Zhao, Gabriel Blanco Saldana, Saurabh Deshpande, Alexandre Michetti Manduca, Jay Ren, Surender Pal Singh, Fan Xiao, Haw-Shiuan Chang, Giannis Karamanolakis, Yuning Mao, Yaqing Wang, Christos Faloutsos, Andrew McCallum, Jiawei Han. AutoKnow: Self-Driving Knowledge Collection for Products of Thousands of Types. In Proceedings of the 2020 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2020.
- Yue Wang, **Chenwei Zhang**, Shen Wang, Philip S. Yu, Lu Bai, Lixin Cui, Guandong Xu. Generative Temporal Link Prediction via Self-tokenized Sequence Modeling. World Wide Web Journal (WWWJ), 2020.
- Ye Liu, Shaika Chowdhury, **Chenwei Zhang**, Cornelia Caragea, Philip S. Yu. Interpretable Multi-Step Reasoning with Knowledge Extraction on Complex Healthcare Question Answering. arXiv, 2020.
- Congying Xia, **Chenwei Zhang**, Hoang Nguyen, Jiawei Zhang, Philip Yu. CG-BERT: Conditional Text Generation with BERT for Generalized Few-shot Intent Detection. arXiv, 2020.
- Xuming Hu, Fukun Ma, Chenyao Liu, **Chenwei Zhang**, Lijie Wen, and Philip S. Yu. Semi-supervised Relation Extraction via Incremental Meta Self-Training. arXiv, 2020.
- Shaika Chowdhury, **Chenwei Zhang**, Philip S. Yu, Yuan Luo. Med2Meta: Learning Representations of Medical Concepts with Meta-Embeddings. In Proceeding of the 13th International Conference on Health Informatics (HEALTHINF), 2020.
- **Chenwei Zhang**, Structured Knowledge Discovery from Massive Text Corpus. PhD Thesis. University of Illinois at Chicago, 2019.
- **Chenwei Zhang**, Yaliang Li, Nan Du, Wei Fan, Philip S. Yu. Joint Slot Filling and Intent Detection via Capsule Neural Networks. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019.
- Congying Xia, **Chenwei Zhang**, Tao Yang, Yaliang Li, Nan Du, Xian Wu, Wei Fan, Fenglong Ma, Philip S. Yu. Multi-Grained Named Entity Recognition. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019.
- Ye Liu, **Chenwei Zhang**, Xiaohui Yan, Yi Chang, Philip S. Yu. Generative Question Refinement with Deep Reinforce Learning. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM), 2019.
- Jiawei Zhang, **Chenwei Zhang**, Bowen Dong, Yang Yang, Philip S. Yu. Missing Movie Synergistic Completion across Multiple Isomeric Online Movie Knowledge Libraries. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), 2019.
- Yue Wang, Yao Wan, **Chenwei Zhang**, Philip S. Yu. Competitive Multi-Agent Deep Reinforcement Learning with Counterfactual Thinking. In Proceeding of the IEEE International Conference on Data Mining (ICDM), 2019.
- Fenglong Ma, Yaliang Li, **Chenwei Zhang**, Jing Gao, Nan Du, Wei Fan. MCVAE: Margin-based Conditional Variational Autoencoder for Relation Classification and Pattern Generation. In Proceedings of the 2019 World Wide Web Conference (WWW), 2019.

- Shaika Chowdhury, **Chenwei Zhang**, Philip S. Yu., Yuan Luo. Mixed Pooling Multi-View Attention Autoencoder for Representation Learning in Healthcare. arXiv, 2019.
- Shaika Chowdhury, **Chenwei Zhang**, Philip S. Yu., Yuan Luo. Hierarchical Semantic Correspondence Learning for Post-Discharge Patient Mortality Prediction. arXiv, 2019.
- **Chenwei Zhang**, Yaliang Li, Nan Du, Wei Fan, Philip S. Yu. On the Generative Discovery of Structured Medical Knowledge. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2018.
- Congying Xia*, **Chenwei Zhang***, Xiaohui Yan, Yi Chang, Philip S. Yu. Zero-shot User Intent Detection via Capsule Neural Networks. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018. (* equally contributed)
- Shaika Chowdhury, **Chenwei Zhang**, Philip S. Yu. Multi-Task Pharmacovigilance Mining from Social Media Posts. In Proceedings of the 27th edition of The Web Conference (WWW), 2018.
- Zhang-Meng Liu, **Chenwei Zhang**, Philip S. Yu. Direction-of-Arrival Estimation based on Deep Neural Networks with Robustness to Array Imperfections. In the IEEE Transactions on Antennas and Propagation, 2018.
- Yue Wang, **Chenwei Zhang**, Shen Wang, Philip S. Yu, Lu Bai, Lixin Cui. Market Abnormality Period Detection via Co-movement Attention Model. In Proceedings of the IEEE International Conference on Big Data (Big Data), 2018.
- Ye Liu, Jiawei Zhang, **Chenwei Zhang**, Philip S. Yu. Data-driven Blockbuster Planning on Online Movie Knowledge Library. In Proceedings of the IEEE International Conference on Big Data (Big Data), 2018.
- Yue Wang, **Chenwei Zhang**, Shen Wang, Philip S. Yu, Lu Bai, Lixin Cui. Deep Co-investment Network Learning for Financial Assets. arXiv, 2018.
- Yaliang Li, Liuyi Yao, Nan Du, Jing Gao, Qi Li, Chuishi Meng, **Chenwei Zhang**, Wei Fan. Finding Similar Medical Questions from Question Answering Websites. arXiv, 2018.
- **Chenwei Zhang**, Wei Fan, Nan Du, Yaliang Li, Chun-Ta Lu, and Philip S. Yu. Bringing Semantic Structures to User Intent Detection in Online Medical Queries. In Proceedings of the IEEE International Conference on Big Data (Big Data), 2017.
- Bokai Cao, Lei Zheng, **Chenwei Zhang**, Philip S. Yu, Andrea Piscitello, John Zulueta, Olu Ajilore, Kelly Ryan and Alex Leow. DeepMood: Modeling Mobile Phone Typing Dynamics for Mood Detection. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2017.
- Jiawei Zhang, Congying Xia, **Chenwei Zhang**, Limeng Cui, Yanjie Fu, and Philip S. Yu. BL-MNE: Emerging Heterogeneous Social Network Embedding through Broad Learning with Aligned Autoencoder. In Proceeding of the IEEE International Conference on Data Mining (ICDM), 2017.
- Junxing Zhu, Jiawei Zhang, Lifang He, Quanyuan Wu, Bin Zhou, **Chenwei Zhang** and Philip S. Yu. Broad Learning based Multi-Source Collaborative Recommendation. In Proceedings of the 26th ACM International Conference on Information and Knowledge Management (CIKM), 2017.
- Junxing Zhu, Jiawei Zhang, **Chenwei Zhang**, Quanyuan Wu, Yan Jia, Bin Zhou, and Philip S. Yu. CHRS: Cold Start Recommendation across Multiple Heterogeneous Information Networks. IEEE Access (2017).
- **Chenwei Zhang**, Wei Fan, Nan Du and Philip S. Yu. Mining User Intentions from Medical Queries: A Neural Network Based Heterogeneous Jointly Modeling Approach. In Proceedings of the 25th International World Wide Web Conference (WWW), 2016.

- **Chenwei Zhang**, Sihong Xie, Yaliang Li, Jing Gao, Wei Fan and Philip S. Yu. Multi-source Hierarchical Prediction Consolidation. In Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM), 2016.
- Chaochun Liu, Huan Sun, Nan Du, Shulong Tan, Hongliang Fei, Wei Fan, Tao Yang, Hao Wu, Yaliang Li, and **Chenwei Zhang**. An Augmented LSTM Framework to Construct Medical Self-diagnosis Android. In Proceeding of the IEEE International Conference on Data Mining (ICDM), 2016.
- Yaliang Li, Chaochun Liu, Nan Du, Wei Fan, Qi Li, Jing Gao, **Chenwei Zhang**, and Hao Wu. Extracting Medical Knowledge from Crowdsourced Question Answering Website. IEEE Transactions on Big Data (2016).
- **Chenwei Zhang**, Xiaoyan Su, Yong Hu, Zili Zhang, Yong Deng. An Evidential Spam-Filtering Framework. Cybernetics and Systems (2016): 1-18.
- Hongping Wang, Xi Lu, Yuxian Du, **Chenwei Zhang**, Rehan Sadiq, Yong Deng. Fault Tree Analysis Based on TOPSIS and Triangular Fuzzy Number. International Journal of System Assurance Engineering and Management (2014): 1-7.
- **Chenwei Zhang**, Yong Hu, Felix T. S. Chan, Rehan Sadiq and Yong Deng. A New Method to Determine Basic Probability Assignment Using Core Samples. Knowledge-Based Systems (2014): 140-149.

Google Scholar: https://scholar.google.com/citations?user=u_bIiBQAAAAJ&hl=en
(Citation:904, H-index:16 as of 11/22/2021)

SELECTED HONORS
AND AWARDS

- SIGKDD Student Travel Award 2018
- Outstanding Reviewer, Neurocomputing 2018
- SIGIR Student Travel Grant 2016
- UIC Student Presenter Award 2016
- Outstanding Undergraduate Award, China Computer Federation 2013
- National Scholarship 2011, 2012, 2013
- Foxconn Scholarship 2012
- First Prize, 2012 Contemporary Undergraduate Mathematical Contest in Modeling 2012

PROFESSIONAL
ACTIVITIES

- Program Committee
 - The ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, *KDD 2022*.
 - The 15th ACM International Conference on Web Search and Data Mining, *WSDM 2022*.
 - The Annual Meeting of the Association for Computational Linguistics, *ACL 2019-2021*.
 - The Annual Conference of the North American Chapter of the Association for Computational Linguistics, *NAACL 2020-2021*.
 - The Annual Conference on Empirical Methods in Natural Language Processing, *EMNLP 2020-2021*.
 - The annual conference of the Asia-Pacific chapter of the Association for Computational Linguistics, *AAACL 2020*.
 - The ACM International Conference on Information and Knowledge Management, *CIKM 2019-2020*.
 - AAAI Conference on Artificial Intelligence, *AAAI 2019-2022*.
 - The 13th ACM Recommender Systems Conference, *RecSys 2018-2020*.
 - The 19th China National Conference on Computational Linguistics *CCL 2020*.
- Organizing Committee
 - KR2ML Workshop at NeurIPS2020
- Reviewer
 - ACM Transactions on Knowledge Discovery from Data, *TKDD 2015, 2016, 2020*.
 - IEEE Transactions on Knowledge and Data Engineering, *TKDE 2017, 2018, 2020*.

- The International Journal on Very Large Data Bases, *VLDB 2019-2020*.
 - ACM Transactions on Information Systems, *TOIS 2021*.
 - Elsevier Neurocomputing, *2017-2020*.
 - IEEE Transactions on Big Data, *TBD 2017*.
 - IEEE Transactions on Signal Processing, *TSP 2018*.
 - IEEE Access, *2019*.
 - IEEE Journal of Biomedical and Health Informatics, *2019*.
 - Journal of Computer Science and Technology, *2020*.
 - ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, *KDD 2019*.
 - ACM Conference on Recommender Systems, *RecSys 2018-2020*.
 - The AAAI Conference on Artificial Intelligence, *AAAI 2019-2020*.
 - The 14th International AAAI Conference on Web and Social Media, *ICWSM 2020*.
- External Reviewer
- IEEE International Conference on Data Mining, *ICDM 2017*.
 - Pacific-Asia Conference on Knowledge Discovery and Data Mining, *PAKDD 2017*.
 - International Conference on Big Data Computing and Communication, *BigCom 2017*.
 - IEEE International Conference On Machine Learning and Applications, *IMCLA 2017*.
 - International Conference on Advances in Social Network Analysis and Mining, *ASONAM 2016*.
 - The 30th IEEE International Conference on Tools with Artificial Intelligence, *ICTAI 2018*.
 - The First IEEE International Conference on Artificial Intelligence and Knowledge Engineering, *AIKE 2018*.

COURSES

Spring 2017: Research Methods in Computer Science (CS 590)
 Fall 2016: Advanced Data Mining (CS 584), Advanced Topics in Machine Learning (CS 594)
 Spring 2016: Artificial Intelligence II (CS 511)
 Fall 2015: Artificial Intelligence I (CS 411), Advanced Data Mining (CS 584)
 Spring 2015: Intro to Machine Learning (CS 412), Database Systems (CS 480)
 Fall 2014: Computer Algorithms I (CS 401), Data Mining and Text Mining (CS 583)

TEACHING EXPERIENCE

Graduate Teaching Assistant

Department of Computer Science, University of Illinois at Chicago

- | | |
|--|-------------|
| • CS480 Database Systems | Spring 2017 |
| • CS141 Programming Design II | Spring 2016 |
| • CS401 Introduction to Algorithms | Fall 2015 |
| • CS251 Data Structures | Spring 2015 |
| • CS341 Programming Language Design and Implementation | Fall 2014 |
| • CS201 Data Structures and Discrete Mathematics | Fall 2014 |

PERSONAL LIFE

I was born and raised in Hangzhou, China. I am a partially-corrected lefty: holds pens/scissors using my right hand and throw/bat/cut mostly lefty. I had spent a significant amount of spare time during my childhood drawing and painting. I am an avid hiking enthusiast and love traveling. I like visual arts in general and appreciate different kind of architectures. I am a fan of different styles of music.

REFERENCES

Xin Luna Dong, Ph.D.

Sr. Principle Scientist
 500 9th Ave N, Seattle, WA 98109
 Email: lunadong@amazon.com
 Homepage: <http://lunadong.com/>

Philip S. Yu, Ph.D.

UIC Distinguished Professor and Wexler Chair in Information Technology
Department of Computer Science, University of Illinois at Chicago
851 S. Morgan St., Rm 1138 SEO, Chicago, IL 60607
Email: psyu@cs.uic.edu
Homepage: <https://www.cs.uic.edu/PSYu/>

Wei Fan, Ph.D.

Executive Director, Tencent Medical AI Lab, Palo Alto, CA 94301
Email: davidwfan@tencent.com

Yong Deng, Ph.D.

Professor, School of Computer and Information Science, Southwest University
No. 2 Tiansheng Road, Chongqing, China 400715
Email: ydeng@swu.edu.cn