Zheng Gao

CONTACT INFORMATION

Applied Scientist
Amazon Alexa AI
Seattle, WA

(412)638-3401
woshigaozheng@gmail.com
https://zhenggao.io

RESEARCH INTEREST

My research interests are primarily in the area of Graph Mining and Natural Language Processing (NLP). Particularly, I am applying deep learning techniques on their interdisciplinary field therein to solve Community Detection, Information Retrieval and Recommendation related tasks.

SKILL

- **Languages**: Python, Java, SQL, Shell, LaTeX
- **Tools**: TensorFlow, MXNet, PyTorch, Spark, Maven, Lucene, MySQL, MongoDB, Neo4j

EDUCATION

Indiana University Bloomington, United States — 08/2015 - 06/2020
Ph.D. in Information Science, Advisor: Xiaozhong Liu
- Minor in Computer Science, Advisor: Xiaofeng Wang

University of Pittsburgh, United States — 08/2013 - 05/2015
M.S. in Information Science

Shanghai International Studies University, China — 08/2009 - 05/2013
B.M. in Information Management and System

EXPERIENCE

Applied Scientist II, Alexa AI, Amazon — 06/2020 - now
- Lead a large scale voice application ranking project to customer utterances for improving Alexa device natural understanding capability, which offers voice application enablement suggestion to skill developers to better serve customer requests in unpopular regions.
- Contributed session-based features to Alexa core Natural Language Understanding (NLU) pipeline for customer utterance interpretation, including domain & intent classification and slot detection.

Data Scientist Intern, Alexa AI, Amazon — 06/2019 - 09/2019
- Applied deep language models (i.e. Bert, ELMo) and state-of-art clustering methods to extract influential text patterns from user requests, which entirely replaced the existing human manual interpretation on annotated datasets.
- Built up an automatic pipeline by Spark and Shell scripts to enable training models on multiple data resources (i.e. Amazon S3 and Redshift) under Alexa restricted environment.

Research Intern, DAMO Academy / AI Lab, Alibaba — 02/2018 - 03/2019
- Generated product review summary from user consecutive behaviors by leveraging dynamic matrix factorization, deep reinforcement learning (Policy Gradient) and sequence to sequence model (Neural Machine Translation) with Attention techniques.
- Proposed an end-to-end pairwise ranking model with transfer learning techniques to detect communities in targeted sparse graphs.
- Detected multilevel anomalies from high dimensional dynamic use logs via Adversarial Autoencoder and Attention-based hierarchical representation learning.

SELECTED PUBLICATIONS

[1] Xiyao Ma, **Zheng Gao**, Qian Hu, Mohamed AbdelHady, HCL: Hybrid Contrastive Learning for Graph-based Recommendation. *International Joint Conference on Neural Networks (IJCNN)*, 2022.

[2] Xiyao Ma, Qian Hu, **Zheng Gao**, Mohamed AbdelHady, Contrastive co-training for diversified recommendation. *International Joint Conference on Neural Networks (IJCNN)*, 2022.

[3] Xiyao Ma, **Zheng Gao**, Qian Hu, Mohamed AbdelHady, Contrastive knowledge graph attention network for request-based recipe recommendation. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

[4] **Zheng Gao**, Chun Guo, Shutian Ma, Xiaozhong Liu. Improving Community Detection Performance in Heterogeneous Music Network by Learning Edge-type Usefulness Distribution. *International Conference on Information*, 2022.

[5] **Zheng Gao**, Radhika Arava, Qian Hu, Xibin Gao, Thahir Mohamed, Wei Xiao, Mohamed Abdel-Hady. Paraphrase Label Alignment for Voice Application Retrieval in Spoken Language Understanding. *Interspeech*, 2021.

[6] **Zheng Gao**, Mohamed AbdelHady, Radhika Arava, Xibin Gao, Qian Hu, Wei Xiao, and Thahir Mohamed. X-SHOT: Learning to Rank Voice Applications via Cross-Locale Shard-based Co-Training. *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2021.

[7] Xibin Gao, Radhika Arava, Qian Hu, Thahir Mohamed, Wei Xiao, **Zheng Gao** and Mohamed AbdelHady. Graphire: Novel Intent Discovery with Pretraining on Prior Knowledge using Contrastive Learning. *KDD Workshop on Pretraining: Algorithms, Architectures, and Applications*, 2021.

[8] Wei Xiao, Qian Hu, Thahir Mohamed, **Zheng Gao**, Xibin Gao, Radhika Arava, Mohamed Abdel-Hady. Two-stage Voice Application Recommender System for Unhandled Utterances in Intelligent Personal Assistant. *KDD 2nd International Workshop: Industrial Recommendation Systems*, 2021.

[9] Qian Hu, Thahir Mohamed, Wei Xiao, **Zheng Gao**, Xibin Gao, Radhika Arava, Xiyao Ma, Mohamed AbdelHady. Collaborative Data Relabeling for Robust and Diverse Voice Apps Recommendation in Intelligent Personal Assistants. *EMNLP Third Workshop on NLP for Conversational AI*, 2021.

[10] **Zheng Gao**, Hongsong Li, Zhuoren Jiang, Xiaozhong Liu. Detecting User Community in Sparse Domain via Cross-Graph Pairwise Learning. *ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 2020.

[11] **Zheng Gao**, Lujun Zhao, Heng Huang, Hongsong Li, Changlong Sun, Luo Si, Xiaozhong Liu. Behavior based Dynamic Summarization on Product Aspects via Reinforcement Neighbour Selection. *European Conference on Artificial Intelligence (ECAI)*, 2020.

[12] Zhuoren Jiang, **Zheng Gao**, Jinjiong Lan, Hongxia Yang, Yao Lu and Xiaozhong Liu. Task-Oriented Genetic Activation for Large-Scale Complex Heterogeneous Graph Embedding. *The Web Conference (WWW)*, 2020.

[13] **Zheng Gao**, Chun Guo, Xiaozhong Liu. Efficient Personalized Community Detection via Genetic Evolution. *The Genetic and Evolutionary Computation Conference (GECCO)*, 2019.

[14] **Zheng Gao**, Gang Fu, Chunping Ouyang, Satoshi Tsutsui, Xiaozhong Liu, Jeremy Yang, Christopher Gessner, Brian Foote, David Wild, Ying Ding, Qi Yu. edge2vec: Representation Learning Using Edge Semantics for Biomedical Knowledge Discovery. *BMC Bioinformatics*, 2019. (impact factor = 3.24).

[15] Yongzhen Wang, Xiaozhong Liu, **Zheng Gao**. Neural Related Work Summarization with a Joint Context-driven Attention Mechanism. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.

[16] Zizhe Gao, **Zheng Gao**, Heng Huang, Zhuoren Jiang, Yuliang Yan. An End-to-end Model of Predicting Diverse Ranking On Heterogeneous Feeds. *eCOM Workshop at ACM SIGIR Conference on Research and Development in Information Retrieval (eCom-SIGIR)*, 2018.

[17] **Zheng Gao**, Lin Guo, Chi Ma, Xiao Ma, Kai Sun, Hang Xiang, Xiaoqiang Zhu, Hongsong Li, Xiaozhong Liu. AMAD: Adversarial Multiscale Anomaly Detection on High-Dimensional and Time-Evolving Categorical Data. *Deep Learning Practice for High-Dimensional Sparse Data Workshop at ACM SIGKDD Conference on Knowledge Discovery and Data Mining (DLP-KDD)*, 2019.

[18] Zhuoren Jiang, Liangcai Gao, Ke Yuan, **Zheng Gao**, Zhi Tang, Xiaozhong Liu. Mathematics Content Understanding for Cyberlearning via Formula Evolution Map. *ACM International Conference on Information and Knowledge Management (CIKM)*, 2018.

[19] Xiaozhong Liu, Xing Yu, **Zheng Gao**, Tian Xia, Johan Bollen. Comparing Community-based Information Adoption and Diffusion across Different Microblogging Sites. *ACM Conference on Hypertext and Social Media*, 2016.

[20] **Zheng Gao**, Vincent Malic, Shutian Ma, Patrick Shih. How to Make a Successful Movie: Factor Analysis from both Financial and Critical Perspectives. *International Conference on Information*, 2019.

[21] Yongzhen Wang, Yan Lin, **Zheng Gao**, Yan Chen. A Two-stage Iterative Approach to Improve Crowdsourcing-based Relevance Assessment. *Arabian Journal for Science and Engineering*, 2019.

[22] **Zheng Gao**, John Wolohan, Fast NLP-based Pattern Matching in Real Time Tweet Recommendation. *Text REtrieval Conference (TREC)*, 2017.

[23] **Zheng Gao**, Rui Bi. University of Pittsburgh at TREC 2014 Microblog Track. *Text REtrieval Conference (TREC)*, 2014.

[24] **Zheng Gao**, Xiaozhong Liu. Personalized Community Detection in Scholarly Network. *International Conference on Information*, 2017.

[25] Tian Xia, Xing Yu, **Zheng Gao**, Yijun Gu, Xiaozhong Liu. Internal/External Information Access and Information Diffusion in Social Media. *International Conference on Information*, 2017.

Service

Conference Reviewer & PC Member

- iConference (2023)
- ACM International Conference on Web Search and Data Mining (WSDM 2023)
- AAAI Conference on Artificial Intelligence (AAAI 2022,2023)
- Workshop on Information Extraction from Scientific Publications (WIESP-AACL 2022)
- The Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)
- China Conference on Knowledge Graph and Semantic Computing (CCKS 2022)
- Workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE 2022)
- IEEE International Conference on Multimedia and Expo (ICME 2022)
- The Web Conference (WWW 2019, 2020)
- ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018, 2022)
- IEEE International Conference on Big Data (BigData 2020, 2022)
- Joint Conference on Digital Libraries (JCDL 2021, 2022)
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining Workshop (DLP-KDD 2020,2021; IWKG-KDD 2020)
- Workshop on Scholarly Document Processing (SDP-NAACL 2021, SDP-COLING 2022)
- International Conference on Information Systems (ICIS 2021)
- China Conference on Information Retrieval (CCIR 2021)

Journal Reviewer

- Data Intelligence (2022)
- The Social Science Journal (2022)
- Journal of Informetrics (JOI 2021)
- Computers in Industry (2021)
- Journal of the Association for Information Science and Technology (JASIST 2019, 2021)
- PeerJ Computer Science (2020)
- PLoS ONE (2020, 2021)
- BMC Bioinformatics (2019, 2020, 2022)
- Social Network Analysis and Mining (SNAM 2019, 2020, 2021)
- Medical Science Monitor (2019)
- ACM Transactions on Computing for Healthcare (2020)

Funding Reviewer

- Amazon Research Awards (ARA 2022)

Award

- Tung-li Yuan Memorial Fellowship, Indiana University Bloomington
- Clayton A. Shepherd Scholarship, Indiana University Bloomington