

Zheng Gao

| | | |
|---------------------|---|---|
| CONTACT INFORMATION | School of Informatics, Computing, and Engineering Indiana University Bloomington IN, 47408 | (412)638-3401 gao27@indiana.edu 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 | <ul style="list-style-type: none">• Languages: Python, Java, SQL, Shell, \LaTeX• Tools: TensorFlow, Spark, Maven, Lucene, MySQL, MongoDB, Neo4j | |
| EDUCATION | Indiana University Bloomington, USA | 08/2015 - 05/2020 (expected) |
| | Ph.D. in Information Science, Advisor: Xiaozhong Liu <ul style="list-style-type: none">• Minor in Computer Science, Advisor: Xiaofeng Wang | |
| | University of Pittsburgh, USA | 08/2013 - 05/2015 |
| | M.S. in Information Science, Advisor: Daqing He | |
| INDUSTRY EXPERIENCE | Shanghai International Studies University, China | 08/2009 - 05/2013 |
| | B.S. in Information Management and System | |
| | Data Scientist Intern , Alexa AI, Amazon | 06/2019 - 09/2019 |
| | <ul style="list-style-type: none">• 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. | |
| TEACHING EXPERIENCE | Research Intern , DAMO Academy / AI Lab, Alibaba | 02/2018 - 03/2019 |
| | <ul style="list-style-type: none">• 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.• All three aforementioned projects have been registered as Chinese Patents. | |
| | Indiana University Bloomington | |
| | Instructor / Lecturer <ul style="list-style-type: none">• ILS-Z534: Search (Spring/Fall 2018, Spring/Fall 2019) | |
| PUBLICATION | Associate Instructor <ul style="list-style-type: none">• ILS-Z534: Search (Fall 2015, Fall 2016, Fall 2017)• INFO-I590: Topics in Informatics (Summer 2016, Spring/Summer 2017) | |
| | | |
| | | |
| | | |
| | <p>[1] Zheng Gao, Chun Guo, Xiaozhong Liu. Efficient Personalized Community Detection via Genetic Evolution. <i>The Genetic and Evolutionary Computation Conference (GECCO)</i>, 2019.</p> <p>[2] 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. <i>BMC Bioinformatics</i>, 2019. (impact factor = 2.511).</p> <p>[3] Yongzhen Wang, Xiaozhong Liu, Zheng Gao. Neural Related Work Summarization with a Joint Context-driven Attention Mechanism. <i>Conference on Empirical Methods in Natural Language Processing (EMNLP)</i>, 2018.</p> | |

- [4] 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.
- [5] **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.
- [6] 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.
- [7] 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.
- [8] **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.
- [9] 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.
- [10] **Zheng Gao**, John Wolohan, Fast NLP-based Pattern Matching in Real Time Tweet Recommendation. *Text REtrieval Conference (TREC)*, 2017.
- [11] **Zheng Gao**, Rui Bi. University of Pittsburgh at TREC 2014 Microblog Track. *Text REtrieval Conference (TREC)*, 2014.
- [12] **Zheng Gao**, Xiaozhong Liu. Personalized Community Detection in Scholarly Network. *International Conference on Information*, 2017.
- [13] 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.
- [14] Nan Li, Naren Suri, **Zheng Gao**, Tian Xia, Xiaozhong Liu, Katy Borner. Enter a Job, Get Course Recommendations. *International Conference on Information*, 2017.
- [15] Chenwei Zhang, **Zheng Gao**, Xiaozhong Liu. How Others Affect Your Twitter #hashtag Adoption? Examination of Community-based and Context-based Information Diffusion in Twitter. *International Conference on Information*, 2015.
- [16] **Zheng Gao**, Patrick C. Shih. Communities of Support: Social Support Discussion in a HIV Online Forum. *International Symposium of Chinese CHI*, 2019.
- [17] Satoshi Tsutsui, **Zheng Gao**, Yuzhuo Wang, Guilin Meng, Ying Ding. A Case Study on Viziometrics: Whats the Role of Western Blots in Alzheimers Disease Literature. *International Conference on Information*, 2018.

SERVICE

Conference Reviewer

- The Web Conference (WWW 2018, 2019, 2020)
- ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018)

Journal Reviewer

- Journal of the Association for Information Science and Technology (JASIST 2018)
- BMC Bioinformatics (2019)
- Social Network Analysis and Mining (SNAM 2018, 2019)
- Medical Science Monitor (2019)

Administrative Service

- Chair of Doctoral Student Association (DSA) at Department of Information and Library Science, Indiana University Bloomington (2016 - 2018)

AWARD

- Tung-li Yuan Memorial Fellowship, Indiana University Bloomington (2015 - 2018)
- Clayton A. Shepherd Scholarship, Indiana University Bloomington (2018 - 2019)
- NetSci Student Award (2017)
- IUB GPSG Student Award (2017)