

JialuLiu

E-mail: remenberl@gmail.com

WWW: jialu.info

Research Interests

Utilizing weak human supervision and self-guidance to enhance the comprehension of textual and multi-modal data, for the sake of aligning large language models.

Professional Experience

2023.12 - **Google Research, NYC** Senior Staff Research Engineer
Teach Gemini models to solve problems by integrating with search engine.

2016.5 - 2023.12 **News Understanding at Google Research, NYC** Senior Staff Research Engineer, Manager
Develop SOTA research technologies and drive innovations across Google where news matters.

- Generating factual and genuine multi-doc summary for clustered news stories (headline, briefing, timeline).
- Drafting news for journalists (long text).
- Embedding and classifying news articles across languages and modalities for tasks like story clustering, content type categorization, topic grouping, etc.
- Modeling evolving news story via temporal reasoning.
- Driving research on new paradigm of alignment and distillation of LLMs.

2011.9 - 2016.5 **Data Mining Research Group at UIUC** Research Assistant
Thesis: Constructing and Modeling Text-Rich Information Networks: A Phrase Mining-based Approach
Advisor: Jiawei Han

- Proposed an embedding method for text-rich information networks. (ICDM16)
- Modelled document representations via latent keyphrase inference. (WWW16)
- Designed a phrase mining framework from massive text corpora. (SIGMOD15)
- Proposed a generative clustering algorithm for information networks. (SDM15)
- Built an entity disambiguation system for information networks. (KDDCup13)
- Proposed a spectral clustering algorithm for information networks. (IJCAI13)
- Proposed a multi-view clustering algorithm for information networks. (SDM13)

Education

2011 - 2016 **Ph.D.** University of Illinois at Urbana-Champaign
Majoring in Computer Science, Specialization in Data Mining

2007 - 2011 **B.Eng.** Zhejiang University
Majoring in Computer Science

Publications

conference/proceedings

Tianqi Liu, Yao Zhao, Rishabh Joshi, Misha Khalman, Mohammad Saleh, Peter J. Liu and **Jialu Liu**, “Statistical Rejection Sampling Improves Preference Optimization”, arXiv preprint arXiv:2309.06657, 2023.

Zhen Qin, Rolf Jagerman, Kai Hui, Honglei Zhuang, Junru Wu, Jiaming Shen, Tianqi Liu, **Jialu Liu**, Donald Metzler, Xuanhui Wang and Michael Bendersky, “Large Language Models are Effective Text Rankers with Pairwise Ranking Prompting”, arXiv preprint arXiv:2306.17563, 2023.

Rongzhi Zhang, Jiaming Shen, Tianqi Liu, **Jialu Liu**, Michael Bendersky, Marc Najork and Chao Zhang, “Do Not Blindly Imitate the Teacher: Using Perturbed Loss for Knowledge Distillation”, arXiv preprint arXiv:2305.05010, 2023.

Jiaming Shen, **Jialu Liu**, Dan Finnie, Negar Rahmati, Michael Bendersky and Marc Najork, “‘Why is this misleading?’: Detecting News Headline Hallucinations with Explanations”, Proc. of the the Web Conf. 2023 (WebConf’23), Austin, TX, April 2023.

Jiaxin Huang, Tianqi Liu, **Jialu Liu**, Adam D. Lelkes, Cong Yu and Jiawei Han, “All Birds with One Stone: Multi-task Text Classification for Efficient Inference with One Forward Pass”, arXiv preprint arXiv:2205.10744, 2022.

Liyuan Liu, **Jialu Liu** and Jiawei Han, “Multi-head or Single-head? An Empirical Comparison for Transformer Training”, arXiv preprint arXiv:2106.09650, 2021.

Jialu Liu, Tianqi Liu and Cong Yu, “NewsEmbed: Modeling News through Pre-trained Document Representations”, Proc. of 2021 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’21), Singapore, August 2021.

Jiaming Shen, **Jialu Liu**, Tianqi Liu, Cong Yu and Jiawei Han, “Training ELECTRA Augmented with Multi-word Selection”, Findings of the Association for Computational Linguistics: ACL 2021, Bangkok, Thailand, August 2021.

Xiaotao Gu, Yuning Mao, Jiawei Han, **Jialu Liu**, You Wu, Cong Yu, Daniel Finnie, Hongkun Yu, Jiaqi Zhai and Nicholas Zukoiski, “Generating Representative Headlines for News Stories”, Proc. of the the Web Conf. 2020 (WWW’20), Taipei, Taiwan, April 2020.

Silu Huang, **Jialu Liu**, Flip Korn, Xuezhi Wang, You Wu, and Cong Yu, “Contextual Fact Ranking and Its Applications in Table Synthesis and Compression”, Proc. of 2019 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’19), Anchorage, AK, August 2019.

Jingbo Shang, Meng Qu, **Jialu Liu**, Lance M. Kaplan, Jiawei Han and Jian Peng, “Meta-Path Guided Embedding for Similarity Search in Large-Scale Heterogeneous Information Networks”, arXiv preprint arXiv:1610.09769, 2017.

Fernando Chirigati, **Jialu Liu**, Flip Korn, You Wu, Cong Yu and Hao Zhang, “Knowledge Exploration using Tables on the Web”, Proc. 2017 Int. Conf. on Very Large Data Bases (VLDB’17), 10(3), pp.193-204, Munich, Germany, Aug 2017.

Huan Gui, **Jialu Liu**, Fangbo Tao, Meng Jiang, Brandon Norick and Jiawei Han, “Large-Scale Embedding Learning in Heterogeneous Event Data”, Proc. of 2016 Int. Conf. on Data Mining (ICDM’16), pp.907-912, Barcelona, Spain, Dec 2016.

Jialu Liu, Xiang Ren, Jingbo Shang, Taylor Cassidy, Clare Voss and Jiawei Han, “Representing Documents via Latent Keyphrase Inference”, Proc. of the 25th Int. Conf. on World Wide Web (WWW’16), pp.1057-1067, Montreal, Canada, April 2016.

Jialu Liu, Jingbo Shang, Chi Wang, Xiang Ren and Jiawei Han, “Mining Quality Phrases from Massive Text Corpora”, Proc. of 2015 ACM SIGMOD Int. Conf. on Management of Data (SIGMOD’15), pp.1729-1744, Melbourne, Australia, May 2015.

Jialu Liu, Chi Wang, Jing Gao, Quanquan Gu, Charu Aggarwal, Lance Kaplan and Jiawei Han, “GIN: A Clustering Model for Capturing Dual Heterogeneity in Networked Data”, Proc. of 2015 SIAM Int. Conf. on Data Mining (SDM’15), pp.388-396, Vancouver, Canada, Apr. 2015.

Jialu Liu, Charu Aggarwal and Jiawei Han, “On Integrating Network and Community Discovery”, Proc. of 2015 Int. Conf. on Web Search and Data Mining (WSDM’15), pp.117-126, Shanghai, China, Feb. 2015.

Xiang Ren, **Jialu Liu**, Xiao Yu, Urvashi Khandelwal, Quanquan Gu, Lidan Wang and Jiawei Han, “ClusCite: Effective Citation Recommendation by Information Network-Based Clustering”, Proc. of 2014 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’14), pp.821-830, New York, NY, August 2014.

Chi Wang, Marina Danilevsky, **Jialu Liu**, Nihit Desai, Heng Ji and Jiawei Han, “Constructing Topical Hierarchies in Heterogeneous Information Networks”, Proc. of 2013 IEEE Int. Conf. on Data Mining (ICDM’13), pp.767-776, Dallas, TX, December 2013.

Quanquan Gu, Charu Aggarwal, **Jialu Liu**, and Jiawei Han, “Selective Sampling on Graphs for Classification”, Proc. of 2013 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD’13), pp.131-139, Chicago, IL, August 2013.

Jialu Liu, Chi Wang, Marina Danilevsky and Jiawei Han, “Large-Scale Spectral Clustering on Graphs”, Proc. of 2013 Int. Joint Conf. on Artificial Intelligence (IJCAI’13), pp.1486-1492, Beijing, China, August 2013.

Jialu Liu, Chi Wang, Jing Gao and Jiawei Han, “Multi-View Clustering via Joint Nonnegative Matrix Factorization”, Proc. of 2013 SIAM Data Mining Conf., pp.252-260, (SDM’13), Austin, TX, May 2013.

Chi Wang, Hongning Wang, **Jialu Liu**, Ming Ji, Lu Su, Yuguo Chen and Jiawei Han, “On the Detectability of Node Grouping in Networks”, Proc. of 2013 SIAM Data Mining Conf., (SDM’13), pp.713-721, Austin, TX, May 2013.

Jialu Liu, Deng Cai and Xiaofei He, “Gaussian Mixture Model with Local Consistency”, Proc. of the

Twenty-Fourth AAAI Conf. on Artificial Intelligence (AAAI'10), pp.512-517, Atlanta, GA, July 2010.

book and chapter

Jialu Liu, Jingbo Shang and Jiawei Han, "Phrase Mining from Massive Text and Its Applications", (Book) in Synthesis Lectures on Data Mining and Knowledge Discovery, Morgan & Claypool Publishers, March 2017.

Jialu Liu and Jiawei Han, "Spectral Clustering", (Book Chapter) in Charu C. Aggarwal and Chandan K. Reddy (eds.) Data Clustering: Algorithms and Applications., CRC Press, 2013, pp.177-200.

journal

Jingbo Shang, **Jialu Liu**, Meng Jiang, Xiang Ren, Clare Voss and Jiawei Han, "Automated Phrase Mining from Massive Text Corpora", IEEE Transactions on Knowledge and Data Engineering (TKDE), 2018.

Huan Gui, **Jialu Liu**, Fangbo Tao, Meng Jiang, Brandon Norick, Lance Kaplan and Jiawei Han, "Embedding Learning with Events in Heterogeneous Information Networks", IEEE Transactions on Knowledge and Data Engineering (TKDE), 2017.

Chi Wang, **Jialu Liu**, Nihit Desai, Marina Danilevsky and Jiawei Han, "Constructing Topical Hierarchies in Heterogeneous Information Networks", Knowledge and Information Systems (KAIS), 44(3), pp.529-558, Springer London, 2014.

workshop paper

Jialu Liu, Chi Wang, Jeffrey Yufei Liu, Kin Hou Lei and Jiawei Han, "Ranking-Based Name Matching for Author Disambiguation in Bibliographic Data", Proc. of KDD Cup 2013 Workshop (KDDCUP'13), Chicago, IL, August 2013.

Jialu Liu and Jiawei Han, "HINMF: A Matrix Factorization Method for Clustering in Heterogeneous Information Networks", Proc. of 2013 Int. Joint Conf. on Artificial Intelligence Workshop (IJCAI-HINA'13), Beijing, China, August 2013.

demo paper

Fangbo Tao, George Brova, Jiawei Han, Heng Ji, Chi Wang, Brandon Norick, Ahmed El-Kishky, **Jialu Liu**, Xiang Ren and Yizhou Sun, "NewsNetExplorer: Automatic Construction and Exploration of News Information Networks", Proc. of 2014 ACM SIGMOD Int. Conf. on Management of Data (SIGMOD'14), pp.1091-1094, Snowbird, UT, June 2014.

Fangbo Tao, Kin Hou Lei, Jiawei Han, ChengXiang Zhai, Xiao Cheng, Marina Danilevsky, Nihit Desai, Bolin Ding, Jing Ge, Heng Ji, Rucha Kanade, Anne Kao, Qi Li, Yanen Li, Cindy Xide Lin, **Jialu Liu**, Nikunj Oza, Ashok Srivastava, Rod Tjoelker, Chi Wang, Duo Zhang, and Bo Zhao, "EventCube: Multi-Dimensional Search and Mining of Structured and Text Data", Proc. of 2013 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'13), pp.1494-1497, Chicago, IL, August 2013.