Yupeng Gu

Contact Information

3551 Boelter Hall,
University of California, Los Angeles,
Los Angeles, CA, 90095

Homepage: http://www.yupenggu.me/
Mobile: 1(617)-982-8994
E-mail: ypgu@cs.ucla.edu

Education

University of California, Los Angeles (UCLA) — Los Angeles, CA
Ph.D in Computer Science (cont.) — *Aug. 2016 - present*

Northeastern University — Boston, MA
Ph.D in Computer Science — *Aug. 2013 - Aug. 2016*
Core Courses: Special Topics in Data Mining · Information Retrieval
Machine Learning · Advanced Algorithm

University of Science and Technology of China — Hefei, China
Bachelor of Science in Applied Mathematics (with honors), — *Aug. 2009 - June 2013*
School of Gifted Young

Research Experience

Computer Science Department, UCLA — Los Angeles, CA
Research Assistant — *Aug. 2016 - present*
Advisor: Prof. Yizhou Sun

- Research on ranking-based embedding and proximity-based embedding in information networks: Network embedding methods seek to find a lower dimensional vector representation of nodes while preserving proximity of the network. However, some links (especially those to highly-connected nodes) are the result of the popularity of the latter instead of homophily. Here we explicitly model the popularity factor, detach it from the proximity-based embedding, and propose a principled approach to explain the generation of a link. The embedding learned from our approach is shown to outperform several strong baselines on large-scale real-world datasets.

College of Computer and Information Science, Northeastern University — Boston, MA
Research Assistant — *Aug. 2013 - Aug. 2016*
Advisor: Prof. Yizhou Sun

- Research on latent feature propagation in dynamic networks: I proposed a novel Bayesian approach on continuous latent feature estimation for dynamic networks. Social influence was taken into account explicitly and an efficient training algorithm was proposed to solve the optimization problem.
- Research on political ideology of Twitter users: I collected information for politics-related users and U.S. Congressmen in the past few years on Twitter, including their tweets and friends/followers etc., and formulated a unified graph generation model to handle heterogeneous types of links on Twitter.
- Research on legislative voting network for U.S. Congress: I analyzed the dataset of voting records in U.S. congress from 1990 to 2013, a heterogeneous network with politicians, bills, terms and links between them. A novel model was proposed to learn the topic distributions for bills from both voting records and bill text, as well as congressmen's political ideology from the heterogeneous network.

Pandora Media Inc. — Oakland, CA
Scientist Intern — *June 2017 - Sep. 2017*
Mentor: Dr. Kristiana Schneck and Dr. Chaitanya Chemudugunta

- Tag prediction for ingestion songs: genre and subgenre of a song is crucial in understanding its content and all kinds of recommendation tasks. Unfortunately they are usually manually input by music providers and are often inconsistent or even missing. I implemented a matrix factorization based network embedding method to build feature vectors for artists, incorporated them with features extracted from the audio of a song, and built a machine learning classifier on top of these features. Accuracy for genre and subgenre prediction was enhanced by 25% and 37% over the baseline method respectively.

LinkedIn Corporation Sunnyvale, CA
Software Engineer Intern *May 2015 - Aug. 2015*
Mentor: Dr. Bo Zhao and Dr. David Hardtke

- Optimizing content representation for job recommendation: I proposed a multi-layer logistic regression model to learn global term weights for content-based recommendation, which is efficient in handling large-scale dataset. The AUC of recommendation evaluation was improved by over 17% on real data from LinkedIn job recommendation system. This work was published in the research track of 25th International World Wide Web conference (WWW'16).

Microsoft Research Asia Beijing, China
Research Intern *July 2012 - May 2013*
Mentor: Dr. Jingdong Wang

- Research on image similarity search and hashing: I assisted in implementing a demonstration in similar image searching with scale-invariant feature transform. A Windows 8 application was also developed using color sketching technology, where I accounted for the implementation of the algorithm and parts of the user interface.

Publications

- **Yupeng Gu**, Yizhou Sun, Yanen Li, Yang Yang, *Social Rank Regulated Large-scale Network Embedding*, Proceedings of the 27th International Conference on World Wide Web (WWW'18), Lyon, France, Apr. 2018 (to appear).
- Rui Dong, Yizhou Sun, Lu Wang, **Yupeng Gu**, Yuan Zhong, *Weakly-Guided User Stance Prediction via Joint Modeling of Content and Social Interaction*, Proceedings of the 26th ACM International Conference on Information and Knowledge Management (CIKM'17), Singapore, Nov, 2017.
- **Yupeng Gu**, Yizhou Sun, Jianxi Gao, *The Co-Evolution Model for Social Network Evolving and Opinion Migration*, Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17), Halifax, NS, Canada, Aug. 2017.
- **Yupeng Gu**, Ting Chen, Yizhou Sun, Bingyu Wang, *Ideology Detection for Twitter Users with Heterogeneous Types of Links*, Proceedings of 2017 International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction (SBP'17), Washington D.C., USA, July, 2017.
- **Yupeng Gu**, Bo Zhao, David Hardtke, Yizhou Sun, *Learning Global Term Weights for Content-based Recommender Systems*, Proceedings of the 25th International Conference on World Wide Web (WWW'16), Montreal, Canada, Apr. 2016.
- **Yupeng Gu**, Yizhou Sun, Ning Jiang, Bingyu Wang, Ting Chen, *Topic-Factorized Ideal Point Estimation Model for Legislative Voting Network*, Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14), New York, NY, Aug. 2014.

Patents

- Bo Zhao, **Yupeng Gu**, David Hardtke, *Term weight optimization for content-based recommender systems*, US 20170177708 A1.

Professional Experience

- Program Committee: NLPCC (2016, 2018), SoCal NLP (2018).
- External Reviewer: ICDM (2017).

Computer Knowledge

- Programming Languages: Java, Python (scipy/keras), Scala, C/C++, Unix shell scripting (bash/awk), SQL
- Operating Systems: Linux (Ubuntu, Arch Linux)
- Distributed Processing Systems: Hive, Hadoop, Spark (mllib).