Nan Li, Ph.D.

Relevance & Personalization
Airbnb Inc.

✉ *nan.li2@airbnb.com*
✉ *lilxine0416@gmail.com*
https://whoisnanli.github.io

Interests

Machine learning and its various applications, such as search & recommendation, personalization, NLP, and predictive modeling

---

Education

2008.9.5-
2013.12.13
Ph.D. in Computer Science, *University of California Santa Barbara (UCSB)*
◇ Research areas: Applied machine learning, data mining, and graph mining
◇ Thesis title: Uncovering Interesting Attributed Anomalies in Large Graphs
◇ GPA: **3.99 / 4.0**

2005.9.1-
2008.7.1
M.S. in Computer Science, *Peking University*
◇ GPA: Overall, **88.8 / 100**, Major, **90.0 / 100** (Rank: **1 out of 36**)

2001.9.1-
2005.6.30
B.S. in Computer Science, *Wuhan University*
◇ GPA: Overall, **90.0 / 100**, Major, **92.5 / 100** (Rank: **1 out of 452**)

Professional Experiences

2019.12-Now
Relevance & Personalization, Airbnb, *San Francisco*, CA
Machine Learning Engineer
Overview: Improving homes search ranking relevance using deep learning

● Designed and built customized cross neural modules to capture higher-order feature interactions in our first-pass ranking model; launched it with a 0.56% gain in global bookings

● Created a new sequence-of-IDs feature type to encode user engagement histories for better deep personalization; launched it with a 0.68% gain in global bookings

● Various feature engineering projects launched with a combined 1.2% gain in global bookings

● Other work includes building foundational ranking infra and pipelines such as the incremental learning framework, sparse entity representation learning, etc.

2019.02-
2019.11
Messenger Business Ranking, Facebook, *Menlo Park, CA*
Research Scientist
Overview: Person-to-business (P2B) in-thread conversation understanding and automation using NLP and statistical modeling

● Built intent detection models by fine-tuning pre-trained FB BERT models; launched the models for the EN and ES P2B threads with a combined 0.97% gain in global appointments booked

● Causal inference on P2B experimental results using generalized linear models to counter page-side biases

2017.04- 2019.02	**News Feed Integrity, Facebook**, *Menlo Park, CA* *Research Scientist* Overview: Optimizing human labeling workflows using machine learning ● Launched model-assisted sampling to increase the fraction of positive examples for human labeling and to tighten the estimated prevalence CI ● Launched a novel labeling workflow using a generative model to estimate aggregated labels with uncertainty and labeler reliability, allowing skipping expensive disagreement escalations, ergo significantly reducing the labeling cost
2015.02- 2017.04	**Core Methods @ Core Data Science, Facebook**, *Menlo Park, CA* *Data Scientist (Machine Learning & Algorithms)* Overview: Ranking, embedding models, text classification, spatial clustering, etc.
2014.05- 2015.02	**Applied Machine Learning, Apple**, *Cupertino, CA* *Data Scientist* Overview: Recommender systems, predictive modeling and analytics for Apple Online Store
2013.08- 2014.05	**Data Products & Research, oDesk**, *Redwood City, CA* *Data Scientist* Overview: Building statistical and machine learning models to solve various data problems in a large online work marketplace
2012.12-2013.3	**Microsoft Research**, *Cambridge, UK* *Research Intern* Project: User skill ranking and competition outcome prediction
2012.6-2012.9	**Bing Indexing and Knowledge Team, Microsoft**, *Bellevue, WA* *Research and Software Development Intern* Project: Full-document entity extraction and disambiguation
2010.6-2010.9	**Customer Insight & Data Analytics Team, IBM Research**, *Yorktown Heights, NY* *Research Intern* Project: Customer lifetime value maximization using action proxy-driven reinforcement learning

Selected Publications

Conference Publications

- **Nan Li**, Huan Sun, Kyle Chipman, Jemin George, and Xifeng Yan, "A Probabilistic Approach to Uncovering Attributed Graph Anomalies", *Proc. of the 2014 SIAM International Conference on Data Mining (SDM'14)*, Philadelphia, Pennsylvania, April 2014.

- **Nan Li**, Ziyu Guan, Lijie Ren, Jian Wu, Jiawei Han, and Xifeng Yan. gIceberg: Towards Iceberg Analysis in Large Graphs. *Proc. of the 2013 IEEE International Conference on Data Engineering (ICDE'13)*, pp. 1021-1032, Brisbane, Australia, April 2013.

- **Nan Li**, Xifeng Yan, Zhen Wen, and Arijit Khan. Density index and proximity search in large graphs. *Proc. of the 2012 ACM International Conference on Information and Knowledge Management (CIKM'12)*, pp. 235-244, Maui, HI, USA, October 2012.

- Arijit Khan, **Nan Li**, Xifeng Yan, Ziyu Guan, Supriyo Chakraborty, and Shu Tao. Neighborhood based fast graph search in large networks. *Proc. of the 2011 International Conference on Management of Data (SIGMOD'11)*, pp. 901-912, Athens, Greece, June 2011.

- **Nan Li** and Naoki Abe. Temporal cross-sell optimization using action proxy-driven reinforcement learning. *Proc. of the ICDM 2011 Workshop on Optimization Based Methods for Emerging Data Mining Problems (ICDMW'11)*, pp. 259-266, Vancouver, Canada, December 2011.

- Charu Aggarwal and **Nan Li**. On node classification in dynamic content-based networks. *Proc. of the 2011 SIAM International Conference on Data Mining (SDM'11)*, pp. 355-366, Phoenix, AZ, USA, April 2011.

- **Nan Li**, Yinghui Yang, and Xifeng Yan. Cross-selling optimization for customized promotion. *Proc. of the 2010 SIAM International Conference on Data Mining (SDM'10)*, pp. 918-929, Columbus, Ohio, USA, April 2010.

Archive

- Tobias G. Tiecke, Xianming Liu, Amy Zhang, Andreas Gros, **Nan Li**, Gregory Yetman, Talip Kilic, Siobhan Murray, Brian Blankespoor, Espen B. Prydz, Hai-Anh H. Dang, "Mapping the world population one building at a time", *CoRR abs/1712.05839 (2017)*.

Journal Publications

- Charu Aggarwal and **Nan Li**. On supervised mining of dynamic content-based networks. *Statistical Analysis and Data Mining*, 5(1):16–34, 2012.

- **Nan Li** and Desheng Dash Wu. Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, 48(2):354–368, 2010.

- **Nan Li**, Xun Liang, Xinli Li, Chao Wang, and Desheng Dash Wu. Network environment and financial risk using machine learning and sentiment analysis. *Human and Ecological Risk Assessment*, 15(2):227–252, 2009.

Academic Experiences & Awards

2009.1-2013.9 — **Research Assistant at Department of Computer Science, UCSB**,
Advisor: *Prof. Xifeng Yan*

Topics: Data mining, applied machine learning, graph mining, anomaly detection, social network analysis, business analytics and optimization

• *Project area: Graph anomaly detection using machine learning*
a) gAnomaly: a regularized mixture model for anomaly detection in graphs.

• *Project area: Creating algorithms to efficiently and effectively index and query large-scale graphs*
a) gDensity: label-based proximity search via density indexing;
b) gIceberg: graph iceberg search via local aggregate scoring.

Speaker/Attendee — **Workshops**
• 2012 Grace Hopper Celebration of Women in Computing, Baltimore, MD, Oct 3-6, 2012.
• Invited speaker at 2009 Google Workshop for Women Engineers, Mountain View, CA, Jan 22-25, 2009.
• 2009 Grad Cohort Program, San Mateo, CA, Mar 27-28, 2009.

Reviewer — **Journals**
IEEE Transactions on Neural Networks, Journal of Neurocomputing.
Conferences
VLDB'14, KDD'13, WWW'13, ICDM'12, SIGMOD'11, SDM'11, KDD'10, SDM'10, SIGMOD'10, ICDM'10, ICDM'09, ICDE'09, ISNN'07.

2012 — 2012 CIKM Student Travel Grant

2012 — 2012 Grace Hopper Scholarship

2011 — 2011 SDM Conference Travel Award

2010 — 2010 SDM Conference Travel Award

2008-2009 — UCSB Department of Computer Science Merit Fellowship

UCSB Department of Computer Science Teaching Assistantship