John Boaz Lee

jtlee@wpi.edu | johnboaz.github.io

Senior machine learning engineer and researcher working primarily on deep learning,
sequence and graph modeling, representation learning, and data mining.

Education

08/2015—12/2019   *Ph.D., Worcester Polytechnic Institute*
Computer Science, GPA 4.0/4.0
- *Thesis*: Deep Learning on Graph-structured Data
- *Committee*: Xiangnan Kong (advisor), Elke Rundensteiner, Yanhua Li, Ryan Rossi

06/2010—06/2012   *M.Sc., University of the Philippines – Diliman*
Computer Science, GPA 4.0/4.0
- *Thesis*: A Link Prediction Algorithm for Heterogeneous Bibliographic Information Networks
- *Committee*: Henry Adorna (advisor), Prospero Naval Jr., Jaderick Pabico, Rommel Feria

06/2004—06/2008   *B.Sc., Silliman University*
Computer Science, GPA 3.1/4.0
- *Thesis*: Dumaguete e-Traveler – A Knowledge-based Decision Support System
- *Advisor*: Dave Marcial

Experience

01/2020—Present   *Facebook*
Senior Machine Learning Engineer, Menlo Park
- Working primarily on graph learning for ads retrieval and ranking. I have experience with a wide range of graph modeling techniques. These include GNNs, self-supervised node representation learning algorithms like node2vec, and graph traversal methods like Personalized PageRank — including low-latency variants suitable for real-time use. In total, over the years, I have contributed to efforts that led to >$1B in incremental revenue.
- Led the effort to deploy user intent prediction models across 3 surfaces (jobs, marketplace, shops) for 13 separate use cases. Our models were found to be 24% better than existing methods. One notable adoption was for promotions targeting ($2.4M/year in savings).
- Launched ~50 ML solutions for a wide variety of problems (related ads sourcing, multimodal topic prediction, scraping detection). Our anti-scraping models were instrumental in the takedown of 2 malicious third-party apps with >100M downloads.
- Core code contributor to an internal PyTorch framework for sequence modeling.
- Some of my major responsibilities include designing and developing advanced ML models, mentoring junior engineers, building data pipelines, and performing ad hoc data analysis.

06/2019—08/2019   *Facebook*
Machine Learning Researcher, Menlo Park (Internship)
- Deployed a recurrent pipeline which generated person-level graph embeddings for >1B users — leveraging their interactions from across FB's family of apps.
- Demonstrated usefulness of approach by showing consistent AUROC gains across 9 tasks.

05/2018—08/2018   *Adobe*
Machine Learning Researcher, San Jose (Internship)
- Worked on a deep learning model for entity alignment which was deployed internally. This resulted in two research publications and a patent (approved in 2023).

05/2017—08/2017   *Xerox PARC*
Machine Learning Researcher, Palo Alto (Internship)
- Led a project on graph classification and collaborated on a second one on dynamic network representation learning. Published two research papers from these projects.

08/2015—06/2019   *Worcester Polytechnic Institute*
Research Assistant, Worcester
- Researched and developed deep learning techniques for various types of relational data. Results were published at top data mining venues like TKDE, TKDD, KDD, and CIKM.
- Assisted professors in the following courses: Data Mining, Computer Networks, Software Security, Temporal Logic, and Theory of Computation (teaching assistant during initial two years at WPI).

10/2012—10/2012   *Nara Institute of Science and Technology*
Research Assistant, Nara (Internship)
- Proposed and built a system for automatic software patch reviewer recommendation.
- Finished a draft research publication within a month, the paper was later accepted for publication and won a "Best Presentation Paper" award.

06/2012—07/2015   *Ateneo de Manila University*
Instructor, Metro Manila
- Taught the following college-level courses: Data Structures and Algorithms, Java Programming, C# Programming, Design Patterns, and Linux Systems Programming.
- Coached, in a minor capacity, competitive programming team (13/53 in ICPC regionals).

06/2010—06/2012   *University of the Philippines – Diliman*
Instructor, Metro Manila
- Handled the following courses: C Programming, Software Engineering, Data Structures and Algorithms. Taught primarily to an audience without a computing background.

04/2008—04/2009   *Interprise Solutions*
Software Engineer, Cebu
- Designed and developed core tools used in the application development framework of an ERP system. Promoted after six months at the company.

04/2006—06/2006   *Neri & Hu Design and Research Office*
Software Engineer, Shanghai (Internship)
- Designed and implemented an in-house Inventory Management System from scratch.

Publications

Citation statistics available on Google Scholar
Published at top data mining venues including: KDD (3), SDM (3), TKDD (2), TKDE (1), and CIKM (1)
Patents: US 11,544,535 B2

TKDE'22   Role-based Graph Embeddings
N. K. Ahmed, R. A. Rossi, **J. B. Lee**, T. L. Willke, R. Zhou, X. Kong, H. Eldardiry
IEEE Transactions on Knowledge and Data Engineering. 2022.

KDD'21   Energy-efficient Models for High-dimensional Spike Train Classification using Sparse Spiking Neural Network
H. Yin, **J. B. Lee**, X. Kong, T. Hartvigsen, S. Xie
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2021.

KDD'20   Recurrent Networks for Guided Multi-Attention Classification
X. Dai, X. Kong, T. Guo, **J. B. Lee**, X. Liu, C. Moore
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2020.

TKDD'20   On Proximity and Structural Role-based Embeddings in Networks: Misconceptions, Techniques, and Applications
R. A. Rossi, D. Jin, S. Kim, N. K. Ahmed, D. Koutra, **J. B. Lee**
ACM Transactions on Knowledge Discovery from Data. 2020.

SDM'20a   Deep Parametric Model for Discovering Group-cohesive Functional Brain Regions
J. B. Lee, X. Kong, C. M. Moore, N. K. Ahmed
SIAM International Conference on Data Mining. 2020.

SDM'20b   Dual-Attention Recurrent Networks for Affine Registration of Neuroimaging Data
X. Dai, X. Kong, X. Liu, **J. B. Lee**, C. Moore
SIAM International Conference on Data Mining. 2020.

TKDD'19   Attention Models in Graphs: A Survey
J. B. Lee, R. A. Rossi, S. Kim, N. K. Ahmed, E. Koh
ACM Transactions on Knowledge Discovery from Data. 2019.

CIKM'19   Graph Convolutional Networks with Motif-based Attention
J. B. Lee, R. A. Rossi, X. Kong, S. Kim, E. Koh, A. Rao
ACM International Conference on Information and Knowledge Management. 2019.

ANS'19   Learning Compact Graph Representations via an Encoder-Decoder Network
J. B. Lee, X. Kong
Applied Network Science. 2019.

BigData'18   Dynamic Network Embeddings: From Random Walks to Temporal Random Walks
G. Nguyen, **J. B. Lee**, R. Rossi, N. Ahmed, E. Koh, S. Kim
IEEE International Conference on Big Data. 2018.

KDD'18   Graph Classification using Structural Attention
J. B. Lee, R. A. Rossi, X. Kong
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. *2018.*

BigNet @ WWW'18   Continuous-Time Dynamic Network Embedding
G. Nguyen, **J. B. Lee**, R. Rossi, N. Ahmed, E. Koh, S. Kim
International Workshop on Learning Representations for Big Networks @ *The Web Conference. 2018.*

SDM'17   Identifying Non-linear Contrasting Networks from Time-Series Data: Application to Brain Network Analysis
J. B. Lee, X. Kong, Y. Bao, C. Moore
SIAM International Conference on Data Mining. 2017.

IWESEP @ APSEC'13   Patch Reviewer Recommendation in OSS Projects
J. B. Lee, A. Ihara, A. Monden, K. Matsumoto
International Workshop on Empirical Software Engineering in Practice @ Asia-Pacific Software Engineering Conference. 2013. **(Best Presentation Award)**

JoC'13   Understanding the Behavior of Filipino Twitter Users during Disaster
J. B. Lee, M. Ybañez, M. M. de Leon, M. R. E. Estuar
Journal on Computing. 2013.

CGAT'13   Characterizing Behavior and Features of Participants and Observers during Disaster on Twitter
J. B. Lee, M. Ybañez
International Conference on Computer Games Multimedia and Allied Technologies. 2013.

ASONAM'12 Link Prediction in a Modified Heterogeneous Bibliographic Network
J. B. Lee, H. Adorna
International Conference on Advances in Social Networks Analysis and Mining. 2012.

JUCS'12 Uncovering the Social Dynamics of Online Elections
J. B. Lee, G. Cabunducan, R. Castillo, F. G. Cabarle, J. A. Malinao
Journal of Universal Computer Science. 2012.

JIP'12 A Quantitative Analysis-based Algorithm for Optimal Data Signature Construction of Traffic Data Sets
J. A. Malinao, R. A. B. Juayong, E. R. F. Oquendo, R. M. U. Tadlas, **J. B. Lee**, J. B. Clemente, M. S. Gaabucayan-Napalang, J. R. F. Regidor, H. N. Adorna
Journal of Information Processing. 2012.

ASONAM'11 Voting Behavior Analysis in the Election of Wikipedia Admins
G. Cabunducan, R. Castillo, **J. B. Lee**
International Conference on Advances in Social Networks Analysis and Mining. 2011. **(Best Poster Paper)**

Professional Service

Program Committee **WWW**: The Web Conference (**2022**)

CIKM: International Conference on Information and Knowledge Management (**2020**)

BigData: IEEE International Conference on Big Data (**2018**, **2020**)

Journal Reviewer **TKDD**: ACM Transactions on Knowledge Discovery from Data (**2017**, **2021**)

TNNLS: IEEE Transactions on Neural Networks and Learning Systems (**2021**)

VLDB: International Journal on Very Large Data Bases (**2020**)

TKDE: IEEE Transactions on Knowledge and Data Engineering (**2017**)

JUCS: International Journal of Universal Computer Science (**2012**)

External Reviewer **KDD**: Conference on Knowledge Discovery and Data Mining (**2016 — 2019**)

ICLR: International Conference on Learning Representations (**2018 — 2019**)

IJCAI: International Joint Conference on Artificial Intelligence (**2017 — 2019**)

ICDM: IEEE International Conference on Data Mining (**2019**)

SDM: SIAM International Conference on Data Mining (**2017 — 2019**)

NeurIPS: Conference on Neural Information Processing Systems (**2016**, **2018**)

ASONAM: Conference on Adv. in Social Networks Analysis & Mining (**2017 — 2018**)

ICDE: IEEE International Conference on Data Engineering (**2017**)

AISTAT: International Conference on Artificial Intelligence and Statistics (**2017**)

CIKM: International Conference on Information and Knowledge Management (**2016**)

AAAI: AAAI Conference on Artificial Intelligence (**2016**)

Other Application Reviewer, *Facebook PhD Fellowship – Applied Statistics* (**2021**)

Application Reviewer, *Facebook Request for Proposal – Applied Statistics* (**2020**)

Awards

2018 ***Graduate Research Innovation Exchange Awards Honorable Mention***, Worcester Polytechnic Institute CS Department

2013 ***Best Presentation Award***, IWESEP: International Workshop on Empirical Software Engineering in Practice (awarded to a single paper at workshop)

2012 ***Honors Graduate***, Phi Kappa Phi Honor Society (awarded to top 10% of graduating class)

2011 ***Best Poster Paper***, ASONAM: International Conference on Advances in Social Networks Analysis and Mining (awarded to a single poster paper)

Relevant Skills

Programming	Python, Java, C#, C, SQL (Presto, Spark), LaTeX
Frameworks	PyTorch, TensorFlow, Keras, Airflow, scikit-learn, NumPy, SciPy, pandas, matplotlib
Tools	Jupyter Notebook, VS Code, GNU Screen, Vim, virtualenv, Git, Mercurial
Languages	Cebuano (native), English (professional), Tagalog (professional), Mandarin (intermediate), Hokkien (intermediate)

References

A full list of references is available upon request

Updated November 3, 2024 ∎