Mingda Li

201-702-3208 | mingdali@pinterest.com | linkedin.com/in/mingdali456

Education

New Jersey Institute of Technology Newark, NJ
Ph.D. in Computer Science, GPA 3.875 / 4.0 *Sept. 2018 – Aug. 2020*

Work Experiences

Machine Learning Engineer Aug. 2020 – Present
Pinterest, Inc. *San Francisco, CA*

The role is to leverage cutting edge technologies in machine learning to understand the user's intent on a Pin Closeup and recommend other content related to an initial Pin and the user context.

Machine Learning Software Engineer Intern May 2019 – Aug. 2019
Facebook, Inc. *Seattle, WA*

Goal: To improve the user experience of news feed by understanding the content of the posts.

- Extend the post classification pipelines and workflows using **Hive SQL** to support Spanish.
- Train new post classifiers by using additional features and a new workflow **(AUCs are above 97%)**.
- Deliver high quality post classification models, an inspect tool, and a **GraphQL API** using **Hack**.
- Improve the performance of post classification models by label filtering and transfer learning **(98% to 99%)**.

Technical Skills

Languages: Python, C/C++, Java, Scala, PHP/Hack, Hive SQL, Presto, GraphQL, HTML/CSS
Frameworks/Libraries: Hadoop, Spark, Storm, TensorFlow
Developer Tools: Git/Mercurial, Google Cloud/Amazon EC2/Microsoft Azure, VS Code, PyCharm, IntelliJ

Research Experience

Analyzing and Assisting Patient Decision-Making in Online Health Community May 2018 – Aug. 2020

- Proposed a decision-making thread classifier with a higher AUC than existing methods **(0.915 vs 0.582, 0.695)**.
- Analyzed the influence that OHC users received during their decision-making processes by developing a framework and deep learning techniques to identify influence relationships among posts.
- Proposed a novel thread recommender system that models user interest in topic and concept dimensions leveraging **Convolutional Neural Network (CNN)** and **Latent Dirichlet allocation (LDA)** with **TensorFlow**.

Efficient Top-k Path Search in Large Knowledge Bases Mar. 2017 – Jun. 2017

- Proposed an algorithm to process join operations in parallel using **Scala** and **Spark** built on **Hadoop YARN**.
- Evaluated the method on **Amazon EC2** and demonstrated that it was **five times faster** than existing work.

Constructing Target-Aware Results for Keyword Search on Knowledge Graphs Sept. 2016 – Dec. 2016

- Developed an index building function to import INEX IMDB data into Oracle Berkeley Database.
- Designed a ranking function which improved the mean average generalized precision from **3% to 43%**.

Analysis Methods of the Disease Feature of Omics Datasets Sept. 2012 – Jun. 2015

- Built a model leveraging different omics sources to predict diseases with the accuracy more than **95%**.
- Proposed an analysis approach to predict disease development with a higher precision **(70% vs 46%, 53%)**.

Publications

- **Mingda Li**, Weiting Gao, and Yi Chen. A Topic and Concept Integrated Model for Thread Recommendation in Online Health Communities. The 29th ACM international Conference on Information and Knowledge Management **(CIKM)**, 2020.
- **Mingda Li**, Jinhe Shi, and Yi Chen. Analyzing Patient Decision Making in Online Health Communities. The 7th IEEE International Conference on Healthcare Informatics **(ICHI)**, 2019.
- Yi Shan, **Mingda Li**, and Yi Chen. Constructing target-aware results for keyword search on knowledge graphs. Data & Knowledge Engineering **(DKE)**, 2017.
- **Mingda Li**, Haoran Zheng. A mid-level fusion method for omics dataset. Beijing Biomedical Engineering, 2016.
- Yingtao Zhang, Jianhua Huang, **Mingda Li**, et al. Novel R-wave Detection Algorithm of DCG Signal. Journal of Tianjin University Science and Technology, 2014, 47(1).