Mingda Li

201-702-3208 | mingdali@pinterest.com | mingdali456.github.io

Education

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

Work Experience

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

Goal: To improve the engagement and relevance of the related pins feed and closeup stream.

- Built an end-to-end pipeline of relevance modeling improving relevance metrics by **2x%**.
- Created new labels to the multi-head ranking model powering **3x%** on product metrics and **5%** on engagement metrics.
- Proposed a two-tower model for candidate retrieval boosting engagement metrics by **5%**.
- Developed a real-time user signal increasing engagement metrics by **1-2%**.
- Migrated multiple Hadoop pipelines to Spark with both **Scala and Pyspark**.
- Enabled data logging of a stream feed and established a real-time flow for **content type distribution control**.

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, TensorFlow, Pytorch, Keras
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)**.
- Developed a deep learning system to identify the influence relationships user received in the community.
- 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%**.

Publications

- **Mingda Li**, Jinhe Shi, and Yi Chen. *Identifying Influences in Patient Decision-Making Processes in Online Health Communities.* Journal of Medical Internet Research **(JMIR)**, 2022.
- **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.