Jian Wang

https://www.linkedin.com/in/jianwang92/

Email: jianwang.data@gmail.com
Mobile: (734) 548-7758

Experience

DoorDash, Senior Machine Learning Engineer Sep 2024 – present

- **Attribute Extraction**: Training LLM to extract product attributes from unstructured data

GoPuff, Senior Data Scientist, E-commerce Search and Discovery Jan 2022 – Jul 2024

- **Product Ranking**: Developed machine learning models (XGBoost) for personalized product ranking in browsing and search. Initial launch and iterations led to $600k savings and 1.03% conversion gain
- **Search Retrieval**: Implemented transformers large language models (LLMs) to retrieve products based on query/ product embedding, intent classification and entity tagging. Improved search conversion by 33 bps
- **Carousel Layout Ranking**: Built machine learning models (XGBoost, deep learning) for personalized carousel layout ranking in browsing. Lifted 27 bps in conversion and 20 bps in margin in A/B tests
- **Query Understanding**: Created BERT transformer model for query to taxonomy mapping and query entity tagging on human and LLM annotated data. Powering 5 core services and lifting 0.90% in revenue
- **Out-of-Stock Substitution**: Incorporated LLM finetuning to build product similarity embeddings and power shopping-time and order-time product out-of-stock substitution. Reduced order cancellation by 1.19%
- **Multiobjective Optimization**: Introduced linear layers after machine learning rankers to explicitly trade off conversion, margin, and exploration. Tuned linear layer using simulations. Lifted margin by 31 bps
- **Data ETL Pipelines**: Engineered data ETL pipelines to process terabytes of big data everyday (SQL, Pyspark, Pandas) and power a dozen production services. Feature iterations lifted 0.75% in conversion
- **Wide and Deep Neural Network**: Pioneered state-of-the-art deep learning models with feature engineering to replace tree-based models. Improves conversion by 0.90% and margin by 0.47% in simulations
- **Personalized Recommender Systems**: Leveraged collaborative filtering and two-tower models to generate user and product embeddings and product recommendations. Improved 24 bps in conversion
- **Chat-To-Order**: Led team of three to implement SMS-ordering backend system in hackathon with natural language processing, ChatGPT, and internal APIs. Won highest prize among all 15 teams

LivePerson, Data Scientist II, Conversational AI Nov 2020 – Jan 2022

- **Multilingual Natural Language Understanding (NLU)**: Built transformer intent classification and entity tagging models in 8 languages. Increased intent detection performance by a vast 8.1%
- **Goal-Oriented Chatbot**: Improved Chatbot for contact centers with NLU output and templated question-answering paradigm. Improved chatbot intent resolution accuracy by 5.2% for 4 verticals
- **Few-Shot Learning**: Created class-leading entity tagging solution to predict with as few as 5 training examples per type (contrastive learning). Decreased customer training effort by 50.1%

Amazon, Software Development Engineer, Machine Learning Apr 2020 – Nov 2020

- **Machine Learning Platform**: Developed privacy-centric internal machine learning platform for model training and model deployment (Java, AWS serverless)

LivePerson, Data Scientist, Conversational AI Nov 2018 – Apr 2020

- **Text Classification**: Built deep learning package for intent detection and text classification for customer service platform (PyTorch, BERT finetuning). Fast <5 ms inference speed for a range of model architectures
- **Anomaly Detection**: Improved anomaly detection algorithm to identify out-of-topic content. Ourperformed previous solution by 2% accuracy

Skills

Programming: Python, Java; **Platforms**: Databricks, Snowflake, AWS, Microsoft Azure, Google Cloud
Machine Learning: PyTorch, TensorFlow, XGBoost, Huggingface, Scikit-Learn; **Big Data**: SQL, Pyspark

Awards

Gold Medal, Chinese Physics Olympiad (2011): 51 winners in China among thousands of competitors

Publications and Patents

Domain adaptation of AI NLP encoders with knowledge distillation. Kristen Howell, **Jian Wang**, Matthew Dunn, Joseph Bradley. *United States Patent US-11568141-B2, 2023.*

Domain-Specific Knowledge Distillation Yields Smaller and Better Models for Conversational Commerce. Kristen Howell, **Jian Wang**, Akshay Hazare, Joseph Bradley, Chris Brew, Xi Chen, Matthew Dunn, Beth Ann Hockey, Andrew Maurer, Dominic Widdows. *e-Commerce and NLP (ECNLP), 2022.*

Think Visually: Question Answering through Virtual Imagery. Ankit Goyal, **Jian Wang**, Jia Deng. *Association for Computational Linguistics (ACL), 2018.*

Premise Selection for Theorem Proving by Deep Graph Embedding. Mingzhe Wang, Yihe Tang, **Jian Wang** Jia Deng. *Neural Information Processing Systems (NeurIPS), 2017.*

Education

University of Michigan	Ann Arbor, MI
Master of Science in Computer Science	Sep 2015 – Aug 2018
Peking University	Beijing, China
Bachelor of Science in Physics	Sep 2011 – Jun 2015