JIAN WANG

Email: jianwang.data@gmail.com

Skills

Programming Languages: Python, Java, C/C++, SQL, PHP

Frameworks: PyTorch, TensorFlow, Keras, XGBoost, Spark

Platforms: Amazon Web Services (AWS), Google Cloud Platform, Microsoft Azure, Databricks, Snowflake

Experience

Senior Data Scientist, GoPuff Jan 2022 – present

- **Query Understanding**: Applied transformers deep learning model to resolve query-indicated product categories. Extended search results with relevant products. Boosted average order volume by 0.8%
- **Product ranker**: Designed XGBoost model for personalized product ranking and recommendation on top of Spark-processed features. Increased product recall by 15%

Data Scientist II, LivePerson Nov 2020 – Jan 2022

- **Multilingual Text Classification**: Built deep learning systems that understand human intents in eight languages. Outperformed Watson and DialogFlow by 5% on conversational data
- **Entity Recognition**: Applied transformer-based model to resolve general entities in conversations. Achieved 88% f1 score on more than 20 entities. Outperforming Spacy by 13% on conversational data
- **Few-Shot Learning**: Designed contrastive-learning methods to solve entity tagging tasks with only five labels per entity type. Outperform previous state-of-the-art by 2% on these data scarce tasks

Software Development Engineer, Amazon Apr 2020 – Nov 2020

- **Machine Learning Workflow**: Developed data-secure applications to orchestrate computing resources, for automated machine learning model training and releasing
- **Serverless Application**: Built applications purely on AWS serverless services. Applied AWS Lambda for computing, with REST API as input point, SQS as connection, and S3, DynamoDB as data storage

Natural Language Understanding (NLU) Scientist, LivePerson Nov 2018 – Apr 2020

- **Text Classification**: Created deep learning NLU library for intent detection and text classification. Achieved fast inference speed by searching among model architectures. Increased accuracy by 5%
- **Anomaly Detection**: Improved anomaly detection algorithm to identify out-of-topic content. Replaced former algorithm in intent classification and increased accuracy by 2%
- **Model Training Automation**: Automated model training workflow and model hyperparameter searching processes. Achieved 4x speedup in training by optimizing hardware usage

Vision and Learning Lab Research Assistant, University of Michigan Summers 2017 & 2018

- **Question Answering**: Designed spatial-aware deep learning architecture for question answering. Created datasets for spatial-relation understanding. Improved 6% accuracy than former state-of-the-art model
- **Math Theorem Proving**: Proposed siamese neural network to assist math theorem proving. Outperformed former state-of-the-art model by 7% accuracy on premise selection dataset

Awards

Gold Medal, Chinese Physics Olympiad (2011): Excelled in both physics theory and experiment and won the competition. Only 51 winners in China among thousands of competitors

Gold Medal, Peking University Math Modeling Competition (2013): Modeled and simulated basketball shooting. Achieved top 9% among 82 teams

Publications

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 T. 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**, and Jia Deng.. *Association for Computational Linguistics (ACL), 2018.*

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

Education

University of Michigan Ann Arbor, MI
Master of Science in Computer Science GPA: 3.87 Sep 2015 – Aug 2018

Peking University Beijing, China
Bachelor of Science in Physics GPA: 3.73 Sep 2011 – Jun 2015