HANKYU JANG

PhD Candidate | Applied Scientist Intern @ Amazon 22' | ML Intern @ Pivot Bio 23', AmFam 21'

jhkmath@gmail.com | hankyu-jang@uiowa.edu | (+1) 319-512-6129 | Iowa City, IA (willing to relocate)

hankyujang | HankyuJang | hankyujang.github.io

PROFESSIONAL SERVICE

PC Member | AAAI

08 2022 - Current

PC Member | epiDAMIK @ KDD

08 2021 – Current

Journal Reviewer | SNAM

11 2019 – Current

SKILLS

Predictive Modeling
Deep Learning
Machine Learning
Database
Data Mining
Classification
Clustering
Data Preprocessing
Parallel Computing
Social Network Analysis
Graph Mining
Network Embedding
Submodular Optimization

MACHINE LEARNING ALGORITHMS

Random Forest
XGBoost
Losigtic Regression
K Nearest Neighbors
K-means Clustering
Linear Regression
PCA
t-SNE

EXPERIENCE

Machine Learning Intern | Pivot Bio

05 2023 – 08 2023 | Berkeley, CA, USA

- Discovered key features that affect the product performance via ML modeling
- Implemented an ML modeling pipeline that trains 115 models on 672 datasets
- Engineered data from 13 different sources to capture a holistic view of each field

Applied Scientist Intern | Amazon.com Services, Inc.

05 2022 – 08 2022 | Seattle, WA, USA

- Implemented fraud community detection pipeline that scales to raw data in 1.1 TB
- Detected 100% fraud community from heavily imbalanced 271 MM purchase orders
- Detected dozens of fraud communities with high fraud ratio ($>$ 30%)
- Achieved high quality results via graph embedding and local community detection
- Parallelized the pipeline by using 48 CPUs and 4 GPUs for fast inference

Machine Learning and Data Science Intern | American Family Insurance

05 2021 – 08 2021 | Madison, WI, USA

- Achieved 75% accuracy on classifying 13K claims into over 200 classes
- Applied Graph Attention Networks on claims data to detect suspicious entries
- Learned embedding of unstructured text data using Sentence-BERT and tf-idf

Graduate Research and Teaching Assistant | University of Iowa

08 2018 – 05 2023 | Iowa City, IA, USA

- Published data-driven, machine learning and algorithmic solutions in top conferences
- Advised students on their research projects at a graduate-level course
- Managed a paper reading group to adapt track novel ML techniques (AlgoEpi)

EDUCATION

Ph.D. in Computer Science | University of Iowa | GPA: 3.93

08 2018 – 12 2023 | Iowa City, IA, USA

M.S. in Data Science | Indiana University | GPA: 3.80

08 2016 – 05 2018 | Bloomington, IN, USA

B.S. in Computer Science & Management | Handong Global University

03 2009 – 06 2016 | Pohang, Korea (GPA: 3.94 | Cum Laude)

DEEP LEARNING ALGORITHMS

TGN | GNN | GCN | GAT | CNN | RNN | LSTM | ANN | Autoencoder | BERT | Transformer

TOOLS

AWS Deep Learning AMI | AWS EC2, Athena, S3 | Python | MySQL | SQLite | Jupyter Notebook | Docker | Google Colab | Rstudio | Tableau | Terraform | Bash Script | PowerShell Script

PACKAGES

PyTorch | Tensorflow | Keras | Scikit-Learn | Numpy | Pandas | Scipy | Matplotlib | Seaborn | Hugging Face | NLTK | Deep Graph Library | Autogluon

POSTER AND DATA PUBLICATIONS

Mobility Data
Kaggle 20

Sensor Data
ICHE 20 | Poster

AWARDS

Data Analysis Winner | 2017 Indiana Medicaid Data Challenge

- Discovered imbalance in capacity and demand of mental health treatment
- Published ML solution to Indiana state | Solution | Visualization | ppt

Scholarships and Fellowships

- Ballard and Seashore Dissertation Fellowship | University of Iowa
- Post-Comprehensive Research Fellowship | University of Iowa |
- Top 1% in Spring 2015, Merit Scholarhip (2014 - 2015) | Handong Global University

PUBLICATIONS

Infection source detection | AAAI 23 | | Poster

Patient embedding | CIKM 23 | ASONAM 22 | |

Missing infections | KAIS 22 | ICDM 21 | epiDAMIK 20 |

Disease modeling | PLoS CompBio 21 | ICHI 21 | ASONAM 19 | | | Kaggle data publication

Link prediction | MLG20@KDD | DS19@INFORMS

MACHINE LEARNING CERTIFICATIONS

Machine Learning Specialization (3 courses) | Coursera

10 2022 | Credential

Deep Learning Specialization (5 courses) | Coursera

4 2022 | Credential

PyTorch (2 courses) | edX

5 2022 | Credential

DATA SCIENCE PROJECTS

Image Captioning | | | Poster

- Applied transfer learning to encode 8K images from Flickr8k using ResNet50
- Used LSTM to decode embeddings to generate captions

Kaggle Competition: Iceberg Classifier Challenge | |

- Achieved 90% accuracy using CNN, classifying satellite images into iceberg or ship
- Evaluated KNN, Random Forests, and SVM on PCA dimension reduced data

Single Cell Classification | |

- Achieved 96% accuracy on 3K brain cell classification into 9 categories using SVM