HANKYU JANG

Applied Scientist @ Amazon

jhkmath@gmail.com | (+1) 319-512-6129 | Seattle, WA | 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

CERTIFICATIONS

Practical Data Science on the AWS Cloud Specialization (3 courses) | Coursera
08 2023 | Credential

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

BEST PAPER AWARDS

2nd Best@ASONAM22

Best Paper@ASONAM19

EXPERIENCE

Applied Scientist | Amazon.com Services, Inc.
12 2023 – current | Seattle, WA, USA

Machine Learning Intern | Pivot Bio
05 2023 – 08 2023 | Berkeley, CA, USA

Skills: AutoML, ExplainableAI | Tree Boosting Algorithms, Permutation Importance

- Discovered key features that impact the performance of the product

Applied Scientist Intern | Amazon.com Services, Inc.
05 2022 – 08 2022 | Seattle, WA, USA

Skills: Clustering, Community Detection | Graph Neural Networks

- Implemented a fraud community detection pipeline via retail order embeddings

Machine Learning and Data Science Intern | American Family Insurance
05 2021 – 08 2021 | Madison, WI, USA

Skills: Multi-class Classification, Data Validation | Graph Attention Networks, BERT

- Provided an ML solution to detect suspicious claim data entries

Graduate Research and Teaching Assistant | University of Iowa
08 2018 – 05 2023 | Iowa City, IA, USA

Skills: Collaboration, Leadership, Research, Teaching

- Developed computational methods (algorithms, data mining, machine learning) to model, make inferences about and predict various aspects of healthcare-associated infections
- Collaborated in an interdisciplinary group with specialists in medicine and statistics
- Advised students on a graduate-level course: Computational Epidemiology
- 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)

AWARDS

Data Analysis Winner at Indiana Medicaid Data Challenge

- Discovered imbalance in capacity and demand of mental health treatment
- Our solution is publicly available on the Indiana State government webpage

Presentation
Solution
Tableau visualization

SCHOLARSHIPS AND FELLOWSHIPS

Dissertation Fellowship
Post-Comp Fellowship
Top 1% - HGU 2015
Top 10% - HGU 2014-2015

DEEP LEARNING

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

MACHINE LEARNING

CATBoost | LightGBM
XGBoost | CART | KNN
Random Forest | K-means
Logistic | Linear Regression
PCA | NMF | t-SNE
LIME | SHAP

MACHINE LEARNING RESEARCH PROJECTS

Predictive modeling of an onset of an infection @ UIowa

Skills: Classification, Network Embedding, Continual Learning | Neural Networks, BERT

Designed ML models and algorithmic approaches to predict the onset of an infection

- *Continual*: Learn continually adaptive patient embedding via *DECEnt* on electronic health records and BERT on clinical notes | CIKM 23
- *DECEnt*: Learn patient embedding via co-evolving neural networks and autoencoders to capture the medical history | ASONAM 22 | | award
- *2-stage model*: Construct new features that captures the “exposure" to detected asymptomatic infections | epiDAMIK@KDD 20
- *Optimization*: Detect asymptomatics via graph mining | KAIS 22 | ICDM 21 |

Fraud detection of e-commerce retail orders @ Amazon

Skills: Clustering, Community Detection | Graph Neural Networks

- Implemented a semi-supervised fraud community detection pipeline
- Detected dozens of fraud communities with high fraud ratio
- Parallelized the pipeline by using 48 CPUs and 4 GPUs for fast, scalable inference

Suspicous data entry detection @ AmFam

Skills: Multi-class classification | Graph Attention Networks, BERT

- Developed a machine learning system that detects incorrectly classified claims
- Encoded claim description (free text) using sentence-BERT
- Applied Graph Attention to learn structural information among features for each claim
- Achieved 75% accuracy on classifying 13K claims into over 200 classes
- Distributed the solution to a partnering company for use in practice

Discover key features that derive product performance @ PivotBio

Skills: Classification, Regression, AutoML, Explainable AI | CatBoost, Permutation Importance

- Designed a pipeline for identifying key features that affect the product performance
- Prepared 672 datasets by engineering data from 13 different sources
- Applied variance inflation factor (VIF) to handle collinearity
- Trained 115 ML models and found key features via permutation importance

DATA SCIENCE PROJECTS

Image Captioning | | | Poster

Skills: Transfer Learning | CNN, ResNet50, VGG19, LSTM

- Encoded Flickr8k using ResNet50, then decoded the embeddings using LSTM to generate captions. The encoder-decoder scheme was implemented from scratch using Keras

Dog Breed Classification |

Skills: Transfer Learning | CNN

- Achieved 79% accuracy for classifying 8K dog images into 133 categories

TOOLS

AWS Deep Learning AMI
AWS EC2, S3, SageMaker
Python | MySQL | SQLite
Bash | PowerShell Script
Jupyter Notebook | Docker
TensorFlow Extended (TFX)

PACKAGES

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

POSTER AND DATA PUBLICATIONS

Mobility Data
Kaggle 20

Sensor Data
ICHE 20 | Poster

IMDB Movie Reviews Sentiment Classification |

Skills: Term Frequency, Multi-hot Encoding | Neural Networks

- Achieved 86% accuracy of predicting (+) review of 50K IMDB reviews

Daily Bike Rental Ridership Prediction |

Skills: Regression | Neural Networks

- Predicted hourly bike rental counts for 10 consecutive days

Kaggle Competition: Iceberg Classifier Challenge | |

Skills: Image Classification, Dimensionality Reduction | CNN, KNN, Random Forests, SVM, PCA

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

Single Cell Classification | |

Skills: Multi-class classification, Dimensionality Reduction | KNN, Random Forest, SVM, PCA

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

PUBLICATIONS

Hankyu Jang and S. Pemmaraju, *"Identifying Central Nodes in the Spread of Healthcare Associated Infections," in submission*

A. Choudhuri, ***Hankyu Jang*** et al., *"Continually-Adaptive Representation Learning Framework for Time-Sensitive Healthcare Applications"* | CIKM 23

Hankyu Jang et al., *"Detecting Sources of Healthcare Associated Infections"* | AAAI 23 | | Poster

Hankyu Jang et al., *"Dynamic Healthcare Embeddings for Improving Patient Care"* | ASONAM 22 | | Best Student Paper Runner Up

Hankyu Jang et al., *"Risk-aware Temporal Cascade Reconstruction to Detect Asymptomatic Cases"* | KAIS 22 | ICDM 21 |

Hankyu Jang et al., *"COVID-19 modeling and non-pharmaceutical interventions in an outpatient dialysis unit"*| PLoS CompBio 21 | | Kaggle data publication

Hankyu Jang et al., *"A Data-driven Approach to Identifying Asymptomatic C. diff Cases"* | epiDAMIK@KDD 20

S. Lee, ***Hankyu Jang*** et al., *"Link Predictions in an Online Health Community for Smoking Cessation"* | MLG@KDD 20 | DataScience@INFORMS19

Hankyu Jang et al., *"Evaluating Architectural Changes to Alter Pathogen Dynamics in a Dialysis Unit"* | ASONAM 19 | **Best Paper Award**