HANKYU JANG

PhD Candidate | Former Data Science Intern

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

hankyujang | HankyuJang | hankyujang.github.io

SKILLS

Machine Learning
Submodular Optimization
Deep Learning | Database
Data Mining | Graph Mining
Classification | Clustering
Recommender System
Network Embedding
Social Network Analysis
Predictive Modeling
Reinforcement Learning
Data Visualization
Data Preprocessing
Model Development

MACHINE LEARNING ALGORITHMS

Random Forest | XGBoost
AdaBoost | PCA | t-SNE
Decision Tree | Naive Bayes
Support Vector Machine
Losigtic Regression
K Nearest Neighbors
K-means Clustering
Linear Regression

DEEP LEARNING ALGORITHMS

GNN | GCN | GAT
CNN | RNN | LSTM
ANN | Encoder-Decoder
Autoencoder

EXPERIENCE

Applied Scientist Intern | Amazon.com Services, Inc.

05 2022 – 08 2022 | Seattle, WA, USA

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 GAT on claims data to detect then correct suspicious entries
- Transformed unstructured text into vectors using Sentence-BERT and tf-idf

Graduate Research Assistant | University of Iowa

01 2019 – Current | Iowa City, IA, USA

- Developed auto-encoding heterogeneous co-evolving dynamic neural networks that learn patient representation for predictive modeling | Achieved 48% gain
- Proposed data mining method for missing case detection on large graphs with 1.5M edges | Achieved 360% gain | IEEE ICDM 21
- Developed disease simulators | PLoS CompBio 21 | IEEE/ACM ASONAM 19

EDUCATION

Ph.D. in Computer Science | University of Iowa | GPA: 3.94

08 2018 – 05 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 | 2017 Indiana Medicaid Data Challenge

10 2017 | Sponsors: FSSA, Indiana Chapter of HIMSS, Regenstrief Institute, and KSM Consulting

- Discovered imbalance in capacity and demand of mental health treatment in the Indiana state | Published Solution | Tableau Visualization | Presentation

Best Paper Awards | IEEE/ACM ASONAM 2019

08 2019

Post-Comprehensive Research Fellowship | University of Iowa

02 2021 - 06 2021

NATURAL LANGUAGE PROCESSING

BERT | Sentence BERT | Transformer | Word2Vec | Word Embedding | Sentence Embedding | tf-idf | bag-of-words | sentiment analysis

TOOLS

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

PACKAGES

PyTorch | Tensorflow | Keras | Scikit-Learn | Numpy | Pandas | Scipy | Matplotlib | Seaborn | Hugging Face | NLTK | igraph | NetworkX | Deep Graph Library (DGL)

PROFESSIONAL SERVICE

Journal Reviewer | SNAM

11 2019 – Current

Program Committee Member | epiDAMIK Workshop at KDD 21'

08 2021

PUBLICATIONS

Hankyu Jang, S. Pai, B. Adhikari, S. V. Pemmaraju, *"Risk-aware Temporal Cascade Reconstruction to Detect Asymptomatic Cases,"* ICDM 2021 | | **One of the Best Ranked Papers**

Hankyu Jang, P. M. Polgreen, A. M. Segre, S. V. Pemmaraju, *"COVID-19 modeling and non-pharmaceutical interventions in an outpatient dialysis unit,"* PLoS CompBio 2021 | | Paper | Data (published in Kaggle)

D.M.H. Hasan, A. Rohwer, ***Hankyu Jang***, T. Herman, P. M. Polgreen, D. K. Sewell, B. Adhikari, S. V. Pemmaraju, *"Modeling and Evaluation of Clustering Patient Care into Bubbles,"* ICHI 2021 | Paper

Hankyu Jang, P. M. Polgreen, A. M. Segre, D. K. Sewell, S. V. Pemmaraju, *"A Data-driven Approach to Identifying Asymptomatic C. diff Cases,"* epiDAMIK 2020 | Paper

S. Lee, ***Hankyu Jang***, K. Zhao, M. Amato and A. Graham, *"Link Predictions in an Online Health Community for Smoking Cessation,"* MLG 2020 | Paper

S. Lee, ***Hankyu Jang***, K. Zhao, M. Amato and A. Graham, *"Multi-Relational Link Prediction for an Online Health Community,"* INFORMS Workshop on Data Science 2019 | Paper

Hankyu Jang, S. Justice, P. M. Polgreen, A. M. Segre, D. K. Sewell, and S. V. Pemmaraju, *"Evaluating Architectural Changes to Alter Pathogen Dynamics in a Dialysis Unit,"* ASONAM 2019 | **Best Paper Award Paper**

MACHINE LEARNING CERTIFICATIONS

Deep Learning with Python and PyTorch | edX

5 2022 | Show Credential

PyTorch Basics for Machine Learning | edX

5 2022 | Show Credential

Deep Learning Specialization (Completed 5 courses) | Coursera

4 2022 | Show Credential

1. Neural Networks and Deep Learning | Show Credential
2. Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization | Show Credential
3. Structuring Machine Learning Projects | Show Credential
4. Convolutional Neural Networks | Show Credential
5. Sequence Models | Show Credential

Data Analysis with R | Udacity

3 2015 | Show Credential

POSTER AND DATA PUBLICATIONS

Healthcare Personnel Movement Data
Kaggle 2020
Data

Sensor Data - Inform Mathematical Models
ICHE 2020
Abstract | Poster

DATA SCIENCE PROJECTS

Image Captioning | | Pdf | Poster

- Implemented encoder-decoder framework that generates image captions
- Applied transfer learning using ResNet50 to encode images
- Used LSTM to decode image embedding to generate text

Dog Breed Classification |

- Achieved 79% accuracy for classifying 8K dog images into 133 categories
- Used transfer learning to get 315% performance gain over CNN

IMDB Movie Reviews Sentiment Classification |

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

Daily Bike Rental Ridership Prediction |

- Accurately predicted hourly bike rental counts for 10 days using MLP for regression

Kaggle Competition: Iceberg Classifier Challenge | | Pdf

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

Identification and Localization of Ambulance Siren | | Pdf

- Proposed a framework to detect ambulance siren in noisy audio signals
- Reduced data dimension using NMF, then trained SVM for detection

Single Cell Classification | | Pdf

- Achieved 96% accuracy on 3K brain cell classification into 9 categories using SVM
- Reduced data dimension from 5K to 50 using PCA without loss of model accuracy