HANKYU JANG

PhD Candidate | Applied Scientist Intern @ Amazon 22' | Machine Learning Intern @ AmFam 21'

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

in 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

Graph Mining

Network Embedding

Community Detection

Node Classification

Link Prediction

Deep Learning

Machine Learning

Social Network Analysis

Submodular Optimization

Data Preprocessing

Parallel Computing

Bash Scripting

Database | Data Mining

Classification | Clustering

Recommender System

Predictive Modeling

Data Visualization

Model Development

Reinforcement Learning

EXPERIENCE

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 Assistant | University of Iowa
📅 01 2019 – 12 2022   📍 Iowa City, IA, USA

- Developed approximation algorithms for submodular function optimization | AAAI 23
- Designed co-evolving neural networks to learn dynamic embedding of patients for predictive modeling | Achieved 48% gain | 🏆 ASONAM 22 📄
- Proposed data mining method for missing case detection on large graphs with 1.5M edges | Achieved 360% gain | KAIS 22 📄 | 🏆 ICDM 21 📄
- Developed disease simulators | PLoS CompBio 21 📄 | 🏆 ASONAM 19 📄

Graduate Teaching Assistant | University of Iowa
📅 08 2018 – 12 2018, 01 2023 – 05 2023   📍 Iowa City, IA, USA

- Advised 11 groups of students (31 total) on their research projects at a graduate-level Computational Epidemiology course
- Guided undergraduates with math in Discrete Structures course

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)

MACHINE LEARNING ALGORITHMS

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

DEEP LEARNING ALGORITHMS

TGN · GNN · GCN · GAT · CNN · RNN · LSTM · ANN · Encoder-Decoder · Autoencoder

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

AWARDS

Data Analysis Winner | 2017 Indiana Medicaid Data Challenge

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

Best Papers

- Best Student Paper Award (Runner up) | ASONAM 2022 | 🔗
- One of the Best Ranked Papers | ICDM 2021
- Best Paper Award | ASONAM 2019

Scholarships and Fellowships

- Ballard and Seashore Dissertation Fellowship | University of Iowa
- Post-Comprehensive Research Fellowship | University of Iowa | 🔗
- AAAI Student Scholarship - Travel Funds | AAAI 2023

PUBLICATIONS

Hankyu Jang, A. Fu, J. Cui, M. Kamruzzaman, B.A. Prakash, A. Vullikanti, B. Adhikari, S.V. Pemmaraju, *"Detecting Sources of Healthcare Associated Infections,"* AAAI 2023 | 🜂 | Poster 🔗

Hankyu Jang, S. Lee, H. Hasan, P. M. Polgreen, S. V. Pemmaraju, B. Adhikari, *"Dynamic Healthcare Embeddings for Improving Patient Care,"* ASONAM 2022 | 🜂 | 📄 | 🏆 **Best Student Paper Award (Runner Up)**

Hankyu Jang, S. Pai, B. Adhikari, S. V. Pemmaraju, *(*Extended version from ICDM 2021) "Risk-aware Temporal Cascade Reconstruction to Detect Asymptomatic Cases,"* KAIS 2022 | 🜂 | 📄

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 | 🜂 | 📄 | Data 🔗 (published in Kaggle)

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 | 📄

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 | 📄

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

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 | 📄

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**

PACKAGES

Deep Graph Library (DGL)

PyTorch | Tensorflow

Keras | Scikit-Learn

Numpy | Pandas | Scipy

Matplotlib | Seaborn

Hugging Face | NLTK

igraph | NetworkX

POSTER AND DATA PUBLICATIONS

Healthcare Personnel Movement Data
Kaggle 2020
Data 🔗

Sensor Data - Inform Mathematical Models
ICHE 2020
📄 | Poster 🔗

MACHINE LEARNING CERTIFICATIONS

Machine Learning Specialization (3 courses) | Coursera

📅 10 2022 | Credential 🔗

1. Supervised Machine Learning: Regression and Classification | 🔗
2. Advanced Learning Algorithms | 🔗
3. Unsupervised Learning, Recommenders, Reinforcement Learning | 🔗

PyTorch (2 courses) | edX

📅 5 2022

1. Deep Learning with Python and PyTorch | 🔗
2. PyTorch Basics for Machine Learning | 🔗

Deep Learning Specialization (5 courses) | Coursera

📅 4 2022 | Credential 🔗

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

Data Analysis with R | Udacity

📅 3 2015 | 🔗

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

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 | 🔗 | 📄

- 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 | 🔗 | 📄

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

Single Cell Classification | 🔗 | 📄

- 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