Minchan Kim

Ph.D. Student

Address Seoul, Republic of Korea

Phone 010-2560-5374

E-mail zhdzhd5373@gmail.com

GitHub https://github.com/goldtan

Research interests

Network analysis, Graph neural networks, Natural language processing

Education

2024-03 - current	**Ph.D. in Data Science** Seoul National University	Seoul, Republic of Korea
2022-09 - 2024-02	**M.S. in Data Science** Seoul National University of Science & Technology GPA: 4.45/4.5	Seoul, Republic of Korea
2020-03 - 2022-08	**B.S. in Computer Science and Engineering** Seoul National University of Science & Technology	Seoul, Republic of Korea
2017-03 - 2022-08	**B.B.A. in Business Administration** Seoul National University of Science & Technology GPA: 4.25/4.5 Rank: 1/30	Seoul, Republic of Korea

Publications

1. **Kim, M.** and Lee, H. (in preparation). Treasure hunting in the talent ocean: Automating talent acquisition for on-demand developers from GitHub, Submitted at *Expert Systems With Applications.*

Conference presentations

2. Kim, K., Choi, J., **Kim, M.**, and Lee, H. (2023). Electric Vehicle Technology Prediction Based on Graph Neural Network, 2023 Spring Conference of the Korean Institute of Industrial Engineers (KIIE).
1. **Kim, M.** and Lee, H. (2023). Automatic Recommendation Framework for Developer Headhunting Candidates Based on GitHub Data, 2023 Spring Conference of the Korean

Institute of Industrial Engineers (KIIE).

Patents

Domestic

1. Developer head-hunting candidate automation system
   Application Number 0060095, 2023

Research Projects

2024-01 - current **Developing a personalized knowledge state analysis and question recommendation system for learners with Knowledge Tracing**
Onuii
• Improves the existing SAINT model by employing graph-based embedding techniques to represent relationships between concepts and integrating a memory structure to handle longer sequences.
• Researches ways to more effectively visualize the service representing the learner's real-time knowledge state.

2023-09 - current **Development of machine learning-based information processing analysis methods and systems for B2B business environment analysis**
LG Electronics
• Proposed utilizing SentiWordNet, VADER, BERT, and T5 models for sentiment analysis and introduced strategies to employ explainable AI tools like LIME and SHAP to assign keyword scores, enhancing interpretability and insight extraction.
• Evaluated frequency-based YAKE!, graph-based TextRank, and embedding-based KeyBERT and KeyphraseTransformer (T5) for keyword extraction, assessing their performance to identify the most effective model.
• Constructed a network based on co-occurrence frequency and performed link prediction by reflecting the relevance to internal documents on the edges.

2022-07 - 2023-05 **Development of Machine Learning-based Graph Learning and Link Prediction Methodology for Text-based Technology Trend Information Analysis**
Hyundai Motor Company
• Proposed various methods, including Wikipedia redirection and Noun Phrase chunking, to refine keyword extraction and enhance the precision and relevance of retrieved data.
• Conducted experiments to improve link prediction accuracy by utilizing various node embedding algorithms, including GCN, GraphSAGE, Attri2vec, and Node2vec.
• Enhanced validation accuracy by approximately 0.15 by integrating machine learning model ensemble with embeddings generated via GCN.

2022-03 - 2024-02 **Nowcasting of Artificial Intelligence with Open Source Project Data**
Ministry of Science and ICT
• Explored efficient methods for collecting data from GitHub and proposed various

strategies using models like BERT and CodeBERT to leverage READMEs, code, commits, forks, and stars for enhanced data utilization and analysis.

• Performed research using GitHub and library data to facilitate companies' securing the talent they seek and enhance recruitment strategies for optimal outcomes.

2021-09 – 2021-12 **Analysis of the university's industry-academia research capabilities big data and identification of key areas**

Seoul National University of Science and Technology

• Utilized Doc2Vec and LDA for document embedding, applied k-means and hierarchical clustering for data segmentation, and employed PCA and t-SNE, among other dimensionality reduction techniques, to visualize the clusters in 2D.

• Utilized network visualization tools like Gephi and Tableau to visualize keywords, effectively enhancing data interpretation and insights.

Skills

Programming Language & Tools

Python (Numpy, Pandas, Scikit-learn, Pytorch)	★★★★★
SQL (MySQL, NoSQL(MongoDB))	★★★★
R	★★★★

Visualization Tools

Gephi (Graph visualization tool)	★★★★★
Tableau	★★★

Awards

2022 • Grand Prize, Idea Hackathon for Customs Clearance Reading Using AI, National IT Industry Promotion Agency (NIPA), Data Science and Business Potential Education Research Center, Data Analysis-based Electronic Manufacturing Specialist Training Center

Grants

2022 - 2023 • Electronics Manufacturing Specialist Training Program, Korea Institute for Advancement of Technology (KIAT)

• BK21 Scholarship, National Research Foundation of Korea (NRF)

Other Experiences

Individual project

2022 - current **Supporting victims of cyberbullying and reducing violence on social media**

TLAB

• Utilize various APIs from platforms like Twitter and YouTube to gather data and employ Python visualization libraries, including matplotlib, seaborn, and pyvis, to visualize the collected data intuitively.

2022 **Platform services for solopreneurs**

• Implemented an RNN-based text generation model to automatically create hashtags or descriptions by taking product names as inputs.

• Collaborated with backend engineers, utilizing the Flask web framework in Python to develop and optimize web applications, improving functionality and user experience.

2022 **Improving access to medicines for the visually impaired**

• Utilized yolov5 and state-of-the-art text recognition models from clovaai to implement OCR technology, enabling text extraction from medicine labels and facilitating the transfer of this information to the backend for additional processing.

2021 **Cryptocurrency price prediction using sentiment analysis**

• Conducted sentiment analysis using the lexicon-based VADER and embedding-based BERT models and engaged in feature engineering to transform the results into valuable features for model enhancement and predictive analysis.

Volunteer

2021 - current • Rescue organization for stray cats in redevelopment areas

2020 • Seoultech Overseas Volunteer in Vietnam

• Entrepreneurial Education Corps for middle and high school students

References

Prof. Hakyeon Lee

Professor of Industrial Engineering & Data Science at Seoul National University of Science & Technology

Ph.D. in Industrial Engineering, Seoul National University

E-mail: hylee@seoultech.ac.kr