Sumin Lee

sumin.lee@lgresearch.ai Research Scientist linkedin

Deep Learning Research Scientist with an extensive background in computer science and chemistry. Skilled in machine learning, deep learning, and programming. Loves the challenge of applying deep learning to various domains by collaborating with colleagues and strives for best results.

Research Interest

Deep Learning, Representation Learning, Deep Reinforcement Learning

Education

M.S in Computer Science and Engineering, Korea University, Seoul, South Korea **Sep. 2020 — Oct. 2022**
Data Mining & Information Systems Lab
Advisor: Prof. Jaewoo Kang

Bachelor of Science in Environmental Science and Ecological Engineering, Korea University **Mar. 2015 — Sep. 2020**
Bachelor of Engineering in Computer Science and Engineering

Sejong Science High School, Seoul, South Korea **Mar. 2013 — Feb. 2015**

Experience

LG AI Research, Materials Intelligence Lab, *Research Scientist* **Sep. 2022 — present**
- Research and development of Deep Learning algorithms for predicting and designing material properties
- Developing platform to accelerate material development research

LG AI Research, Materials Informatics Lab, *Internship* **Oct. 2021 — Feb. 2022**
- TADF molecule property optimization using reinforcement learning for OLED material discovery
- Model performance achieved beyond target performance with increased diversity in a 2000 times wider chemical space

Big Data Academy, Korea Data Agency (한국데이터산업진흥원 청년데이터캠퍼스) **Jun. 2019 — Sep. 2019**
- Completed the course on big data-based intelligent information system development.
- Project leader for *Movie beat*, a Movie search system based on analysis of famous quotes. Analyzed the change of popularity of each line throughout the movie running time by extracting 12 thousand famous quotes from over 1.1 thousand movies.
- Code repository : github.com/iamchosenlee/moviebeat

Foreign Exchange Student Program **Aug. 2017 — Dec. 2017**
Degree program in Biology, University of Helsinki Helsinki, Finland

Publication

Geometrically Aligned Transfer Encoder for Inductive Transfer in Regression Tasks **2024 ICLR**
- Sung Moon Ko, Sumin Lee, Dae-Woong Jeong, Woohyung Lim, Sehui Han
- Proposed a novel transfer learning technique, the Geometrically Aligned Transfer Encoder, by extending transfer learning to regression tasks through a differential geometry-based approach
- Interpretation of latent vectors on a Riemannian curved manifold facilitates knowledge transfer between tasks, ensuring stable behavior in both latent space and extrapolation regions for diverse molecular graph datasets

ArkDTA: Attention Regularization guided by non-Covalent Interactions for Explainable Drug-Target Binding Affinity Prediction **2023 Bioinformatics**
- Mogan Gim, Junseok Choe, Seungheun Baek, Jueon Park, Chaeeun Lee, Minjae Ju, Sumin Lee, Jaewoo Kang
- Proposed a protein-ligand binding affinity prediction model utilizing attention regularization guided by NCIs
- Significant enhancements in model explainability achieved, along with insights into the model's internal understanding of chemical systems through qualitative analysis of attention maps

Improved Binding Affinity Prediction Using Non-Covalent Interactions and Graph Integration **2022 IEEE BigComp**
- Junseok Choe, Keonwoo Kim, Minjae Ju, Sumin Lee and J. Kang
- Proposed a binding complex representation for binding affinity prediction by building a bipartite drug-target interaction graph
- Utilizes non-covalent residue-residue interactions in protein graphs and formulates an extended form of drug-target link prediction involving non-covalent atom-residue interactions

Bond Property Optimization Using Reinforcement Learning **2022 Thesis**

- Sumin Lee, J. Kang
- Proposed a model to optimize Bond Dissociation Energy(BDE) which maximizes the minimum BDE among all bonds
- he model takes account of other bonds by using a considerate Q-value while learning to optimize a single bond and shows effectiveness in both toy environment and molecule environment

Projects

Universal Molecular Design (*LG AI Research*) **Oct. 2022 — present**

- Developing technology capable of predicting properties from molecular structures, designing desired molecules, and even forecasting synthesis outcomes between novel types of molecules

Inter-regional flow prediction analysis for geofencing advertising (*Samsung Research*, *Industry Project*) **Oct. 2021 — Oct. 2022**

- Crowd flow prediction model and S/W development through user movement pattern modeling in virtual area.
- Train a graph neural network to predict crowd flow with data in graph form to consider inter-regional influence and topology.

Molecule Optimization using Reinforcement Learning in Synthetic Chemical Space (*personal*) **Jun. 2021 — Oct. 2021**

- Investigate a synthetically constrained molecule generator and optimized the binding affinity score with REINFORCE algorithm.
- Investigate chemical reaction based generation model to increase synthesis feasibility and practical applicability.
- Code repository: github.com/iamchosenlee/reinvent-scaffold-decorator

In silico cancer drug discovery pipeline development **Oct. 2020 — Oct. 2022**

- Propose an *in silico* anticancer drug discovery model that mimics the drug development process.
- Investigate generative models based on Variational Autoencoder, Recurrent Neural Networks, and Genetic Algorithm and optimize with Bayesian Optimization and Reinforcement Learning.
- Final optimal molecules are obtained by screening with toxicity prediction models based on message passing neural networks.
- Code repository : github.com/iamchosenlee/MolDQN-pytorch

Binding Affinity Prediction **Jun. 2020 — Jan. 2022**

- A Multimodal deep learning with a semi-supervised learning setting for drug-target interaction prediction.
- Propose a graph convolutional neural network model using multi-dimensional edge features extracted from protein structures to overcome the sparse binding interaction information.

ADMET Prediction **Jan. 2020 — Oct. 2020**

- Achieved state-of-the-art accuracy in MoleculeNet dataset, a benchmark for molecular property prediction, by using a pre-trained Graph Neural Networks(GNN) with Graph Isomorphism Network(GIN).

Programming Skills

Language	Python, SQL
Framework	PyTorch , Tensorflow

Language

Korean	Native
English	Fluent (TOEIC 980)

Awards & Honors

2017	KU Real Scholarship
2015, 2016	Academic Achievement Award