[AICP 2022] Final Report

❒ Team Information

Research Topic	EN	Discovering eco-friendly hydrate inhibitors using molecular dynamics and deep learning approaches		
	KR	분자동역학 및 딥러닝을 이용한 친환경 하이드레이트 저해제 탐색		
Team Name		**시너지**		
Professor		서용원	Dept.	도시환경공학과
TA		Student No.	Name	Dept.
Team Members		Student No.	Name	Dept.
	1	20185009	고우진	도시환경공학과
	2	20181353	김예진	산업공학과
	3	20181360	남은슬	경영학과
	4			
	5			
	6			

❒ Abstract

Keywords	Molecular dynamics (MD), graph neural network (GNN), eco-friendly hydrate inhibitors, flow assurance

This study shows that the combination of molecular dynamics (MD) and deep learning (DL) can be used to search eco-friendly hydrate inhibitors. The MD simulation results provided detailed information about the molecules and the hydrate structures, and numerous maximum pulling force data. Then, the graph neural network (GNN) approaches were used to identify the most effective inhibitors according to their molecular graph data. We believe that the combination of these methods can help to find the best possible hydrate inhibitors and predict the pulling force faster and more accurately.

이번 연구는 분자동역학 (MD)과 딥러닝 (DL)의 조합을 활용할 시, 친환경 수화물 억제제를 탐색할 수 있음을 보여주고 있습니다. MD 시뮬레이션 결과는 분자 및 하이드레이트 구조에 대한 자세한 정보와 수많은 최대 인장력 데이터를 제공했습니다. 그런 다음 분자 그래프 데이터에 따라 가장 효과적인 억제제를 식별하기 위해 그래프 신경망(GNN) 접근법을 사용했습니다. 우리는 이러한 방법의 조합이 가능한 최상의 수화물 억제제를 찾는 데 도움이 될 수 있고 인장력을 더 빠르고 정확하게 예측하는 데 도움이 될 수 있다고 생각합니다.

Research Contents

1. Introduction

(1) An introduction to gas hydrates

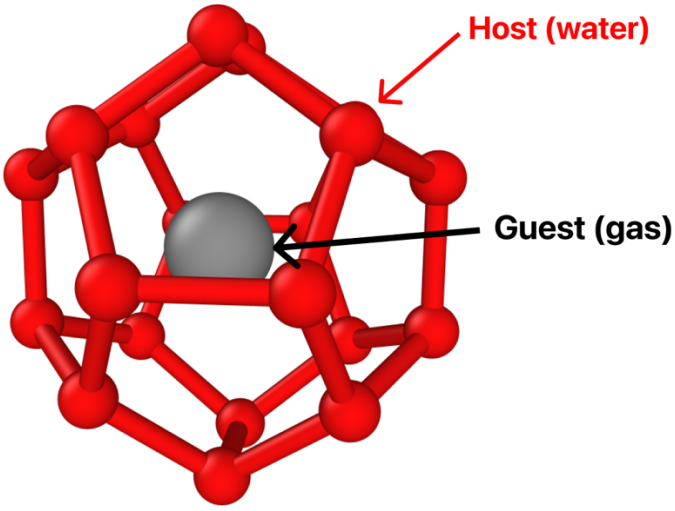


Figure 1. Gas hydrates and their components

Gas hydrates are crystalline solids that form when water molecules surround and trap small gas molecules, such as methane or carbon dioxide [1]. The structure of gas hydrates is similar to that of ice, with water molecules forming a lattice around the gas molecules. Gas hydrates are found in marine sediments and permafrost, and their formation can be used to trap and store methane or carbon dioxide. Recent advances in gas hydrate research have focused on their potential as a renewable energy source. In particular, gas hydrates could be used to store methane, which

could then be used as a fuel. Additionally, gas hydrates could be used to store carbon dioxide, which would help to mitigate climate change.

It should also be noted that the gas hydrate can also be formed during the oil and gas transmission through the pipelines since most deep-sea conditions satisfy the favorable thermodynamic condition where hydrates can be readily formed [2]. In this case, the gas hydrate must be removed in order for the pipeline to function properly and prevent the system from malfunctioning. These trials and efforts are known as flow assurance. There are several methods to prevent pipeline blockage known these days, but utilizing gas hydrate inhibitors could be a good way to reduce the risk of hydrate formation in the first place.

Figure 2. (left) the pipeline blockage due to the gas hydrate formation, (right) a potential disaster could be happened due to the pipeilne blockage (Deep water horizon oil spill)

(2) Types of gas hydrate inhibitors and main research context

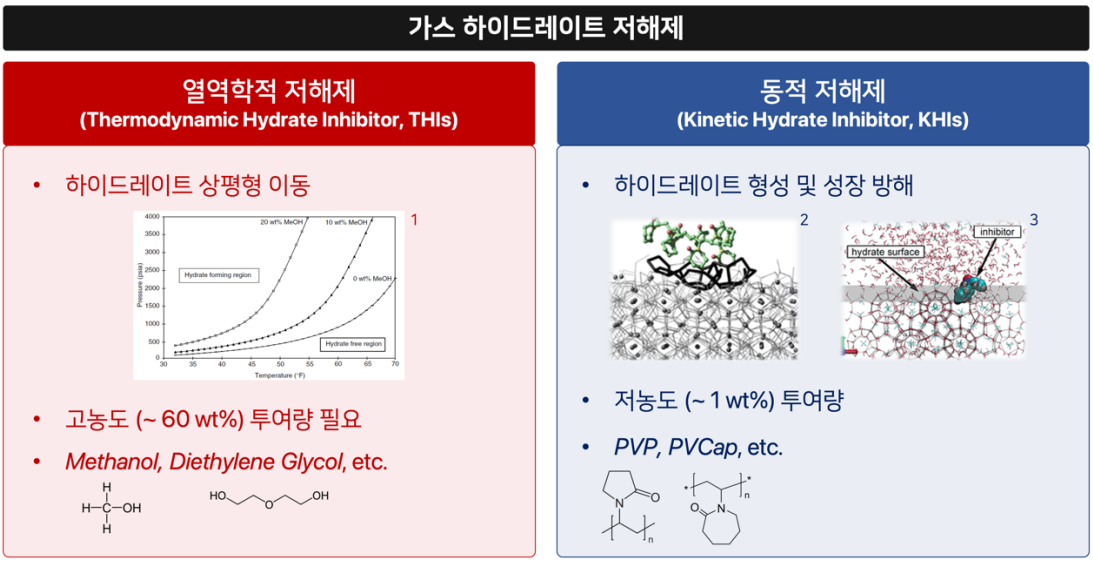


Figure 3. Types of gas hydrate inhibitors

There are two types of gas hydrate inhibitors: thermodynamic hydrate inhibitors (THIs) and kinetic hydrate inhibitors (KHIs). THIs work by lowering the thermodynamic equilibrium between water and gas molecules, while KHIs work by inhibiting the nucleation and growth of hydrate crystals. The former types require much higher concentrations to be effective (up to 60 wt%), while the latter can be effective at lower concentrations (0.1-1 wt%). Well-known examples of THIs are glycols, MEG and DEG, and KHIs mostly include water-soluble polymers, amino acids, and so on.

The inhibition of gas hydrate formation has been an active area of research in recent years, as the potential for gas hydrates to cause environmental and economic damage has become more apparent. There are two main research contexts in this area: discovering environmentally-friendly inhibitors and identifying the main mechanism(s) by which these inhibitors work.

In terms of eco-friendly inhibitors, research has focused on finding alternatives to traditional THIs, which are low-dosage inhibitors like KHIs. But conventional KHIs like high-molecular-weight polymers now should be avoided since they are viscous and may not be readily degraded in the marine environment. Recently, a promising avenue of research is the use of molecular dynamics (MD) simulations to screen for potential inhibitors. This approach has the advantage of being relatively fast and cheap and can be used to screen a large number of molecules. To understand the mechanism(s) by which inhibitors work, recent research has focused on the role of water molecules in the hydrate nucleation process. It has been suggested that water molecules play a key role in the formation of hydrate crystals and that by inhibiting the interaction of water molecules with gas molecules, it may be possible to prevent hydrate formation.

(3) Facilitating the discovery of effective and environmental-friendly hydrate inhibitors by introducing artificial intelligence (AI)

In recent years, there were a bunch of trials to apply machine learning (ML) or AI to the field of computational chemistry or material science with the aim of accelerating the process of new materials discovery. Expecting protein structures or binding affinity of protein-ligand substances are currently well-studied and representative research fields where AI applications have been successful. However, in the field of gas hydrate, there are not many successful examples of AI applications yet. The

reason for this could be that the gas hydrate system is more complex than other systems, and the gas hydrate system consists of hydrogen bond networks of water molecules. In addition, the hydrate inhibitor is not only adsorbed on the surface of the hydrate but also interacts with the water molecules to form a "cage" to prevent the hydrate from growing, which makes it difficult to find an effective hydrate inhibitor with only AI-based methods. Note that is almost impossible to predict the hydrate inhibition performance by seeing only the structure of the candidate chemical. Therefore, in this study, MD simulations were performed to obtain the interaction (pulling force) between the candidate chemicals and water molecules, and this interaction energy was used as a descriptor to train the machine learning model. In addition, we used a deep learning approach to improve the prediction accuracy of the model. The convolutional layer for Graph structure data like HOGNNs and GCNs are used to conduct the molecular regression problem. By doing so, it is available to speed up the screening speed of candidate inhibitors with MD approaches, which were computationally expensive. Furthermore, it is expected that the introduction of AI to the screening of gas hydrate inhibitors will contribute to the development of new environmental-friendly hydrate inhibitors.

2. Research Purpose

The problem that our research seeks to address is the lack of effective and eco-friendly hydrate inhibitors. Current hydrate inhibitors are often toxic and/or difficult to dispose of, which creates environmental and safety concerns. However, evaluating and comparing different kinds of molecules in terms of their hydrate-inhibiting properties is a complex task. Therefore, in this work, we simplified our interested molecules by considering only tripeptides as our interested chemicals. The tripeptide is a molecule consisting of three amino acids, and it is a nature-driven substance. Since the number of essential amino acids in nature is 20, there are maximum 8,000 of different kinds of tripeptides that can be formed. We expect that tripeptides would be biodegradable since amino acids have been investigated by several researchers, and tripeptides are nature-driven substances. More complex molecules would investigated in further studies, but those are beyond the scope of this work.

3. Research Methods

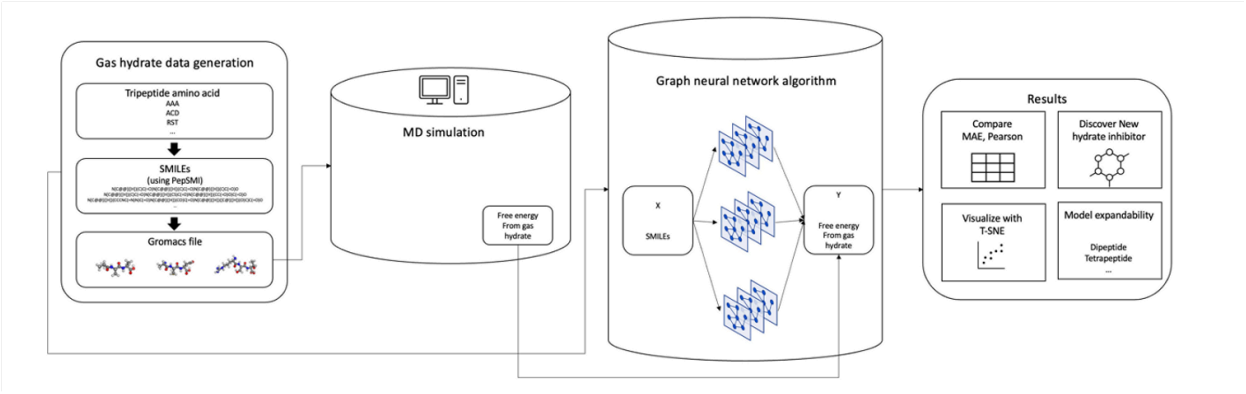


Figure 4. The overall illustration of experimental workflow

(1) SMILES to 3D structure

The overall experimental procedures that we have proceeded are described in the figure 4. Initially, we generated randomized sequences of arbitrary amino acids (note that the we only treated the essential 20 amino acids); thus, several 3 capital letters consisted of abbreviated alphabet symbols of essential amino acids were generated and enrolled in the list. Next, the PepSMI plugin developed by NovoPro was utilized to convert generated capital letters to SMILES strings. After than, all SMILE strings saved in the list were transmitted to the LigParGen server to generate 3D structures (`.pdb` file format) and force field parameters (`.itp` file format) of each tripeptide. Since more than 2,000 of tripeptide information should be downloaded, our team also designed web-scraping code that can simultaneously send our SMILES information to the server, and download `.pdb` and `.gro` files to our local computer. After successfully downloading numerous force field data from the server, the massive molecular dynamics (MD) simulations could be proceeded in the aid of Unist Supercomputing Center (USC).

(2) MD simulation details

To calculate the maximum pulling force between the sI hydrate structure and several tripeptide molecules, we performed all-atom molecular dynamics (MD) simulations with GROMACS 2021.3 package [3]. The TIP4P-ice model was used to describe the water molecules, and the OPLS-AA force field was utilized to describe the CH_4 and tripeptide molecules [4, 5]. The Lorentz-Berthelot mixing rules were employed to calculate the cross-interactions between different species. Short-range interactions

were truncated at 1.0 nm, and long-range electrostatic interactions were calculated with a Fourier spacing of 0.12 nm. The V-rescale thermostat and Parrinello-Rahman barostat were used to control the temperature and pressure of each system. The motion equations were integrated using the leapfrog algorithm with a time step of 2 fs. The LINCS algorithm was used as the standard constraint algorithm for all bonds in the system.

The CH_4 hydrate slab consisted of 414 water molecules and 72 CH_4 molecules were located on the left side of periodic boundary conditions. Adjacent to the CH_4 hydrate slab, A single tripeptide molecule was attached directly to the surface. A different kind of tripeptide was attached to each system, and this kind of work could be easily done by using the PACKMOL source code [6]. When the initial configuration was determined, energy minimization was performed by using the steepest descent algorithm. After the minimization procedure was completed, consecutive 100 ps of NVT and NPT relaxation processes were deployed. The target tripeptide was pulled with the force of 800 $kJ/mol^{-1}nm^{-1}$ at the rate of 0.0075 nm per 1 ps. The total pulling period was 200 ps; therefore, the tripeptide was pulled up to 1.5 nm during the simulation. When the tripeptide molecule was pulled, the force was recorded according to the pulling coordinate. By retrieving the maximum pulling force, it was available to obtain every maximum pulling force depending on the chemical type we were interested in.

(3) GNN (Graph Neural Networks)

Deep learning has had a lot of success recently, which has led to more research in areas like pattern recognition and data mining. Tasks like object detection, machine translation, and speech recognition, that were previously difficult to do, are now possible with deep learning paradigms like CNN, RNN, or autoencoders. Deep Learning is particularly good at finding hidden patterns in Euclidean data (images, text, videos). But what about applications where data is generated from non-Euclidean domains, represented as graphs with complex relationships and interdependencies between objects? That's where Graph Neural Networks (GNN) come in.

Graph Neural Networks (GNNs) are a class of neural networks that operate on graph-structured data, using neural networks to learn from such data. They are used to learn features of the graph, which are a natural way to represent many real-world phenomena, such as social networks, molecular structures, and dependencies

between objects. And it can be used for a variety of tasks such as node classification, link prediction, and molecular regression.

The molecular dataset is well-suited for graph neural network because molecules can be represented as graphs, and it has a lot of features that can be learned by the neural network. For example, the molecular dataset has a lot of information about the atoms and their bonds. This information can be used to learn the chemical properties of the molecules. In addition, the molecular dataset has a lot of 3D information. This 3D information can be used to learn the shape of the molecules.

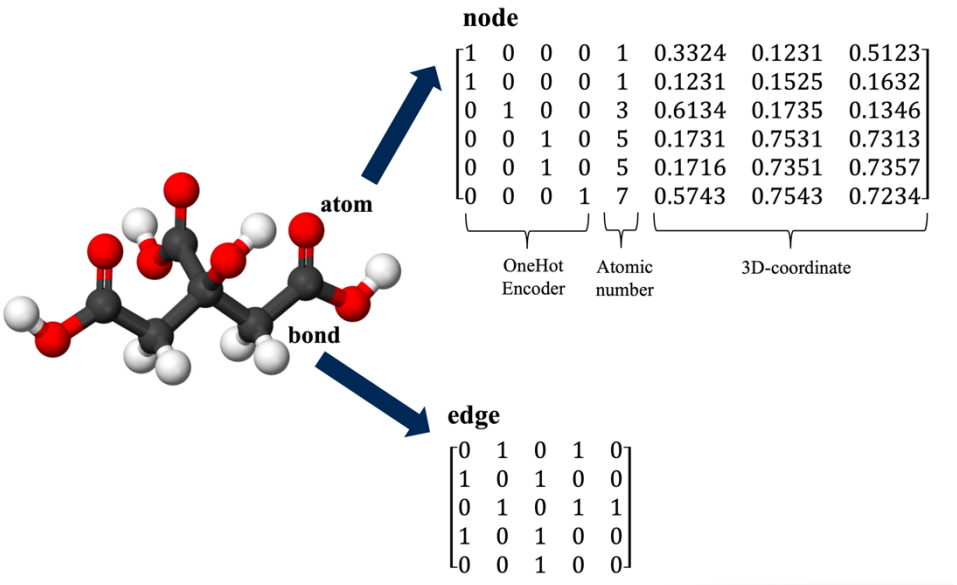


Figure 5. An example of our molecular data in graph form [7]

The above figure is an example of our molecular data in graph form. In molecules, atoms mean nodes and bonds between atoms mean edges. Information containing what each atom means is stored in the form of a node feature. The node feature may contain information such as a one-hot encoder vector, atomic number, and 3D coordinates of atoms that distinguish each atom. The edge contains information about which nodes are connected. So, it means which atoms are bonded together.

(4) Molecular Regression Problem

What we are going to solve is a molecular regression problem, the task is to predict a property of a molecule, such as its pulling force, based on its structure. Molecular regression is a statistical technique used to predict the values of dependent variables based on the values of independent variables. The coefficients of the independent

variables are estimated by minimizing the sum of squared residuals. The dataset for this problem consists of 2,872 molecules, each represented by Molecular Graph $G = (V, E)$. All the molecular graphs are passed by Data Loader and split into batch size. Each batch data passed Graph Convolution layer and aggregated the neighbor node information and updated itself. After the updating, all the node feature embeddings passed global pooling layer. And then, they passed fully-connected network and the y-value was predicted. The dependent variable is the graph structure of the molecule, and the independent variables are the pulling forces of the each molecule. The goal was to build a model that can predict the pulling force of a newly presented molecule, based on its chemical properties.

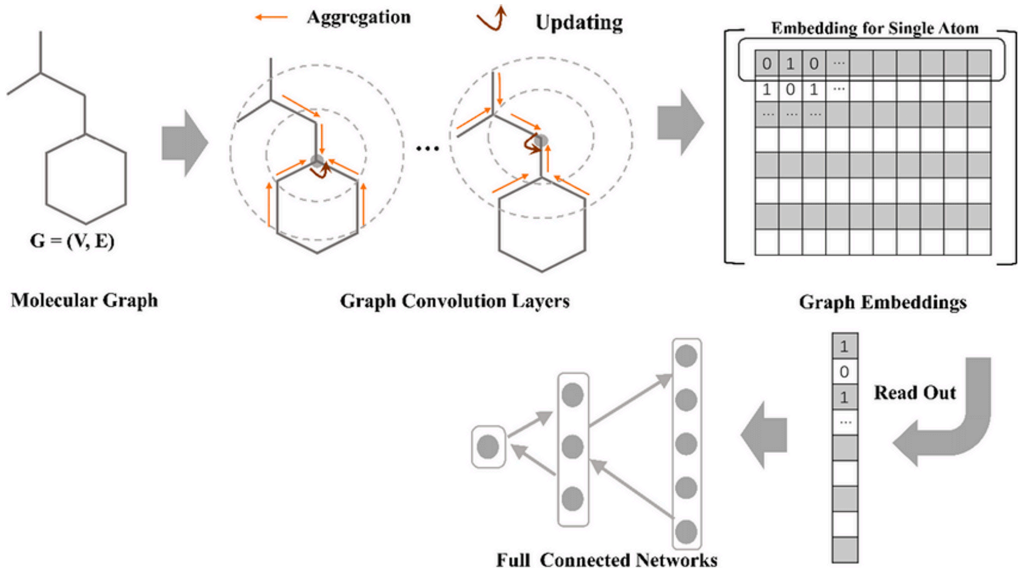


Figure 6. The brief overview of the molecular regression problem workflow

The first step of the molecular regression was to split the dataset into a training set and a test set. The training set was used to build the model and the test set was used to evaluate the performance of the model. We splitted the training set and test set as a ratio of 8:2. The next step was to choose a graph neural network algorithm and train the model on the training set. The performance of the model was then evaluated on the test set. The final step was to use the model to predict the pulling force of a new molecule. The results of the molecular regression problem showed that the model could accurately predict the pulling force of new molecules.

(5) Graph Convolution Layer

The graph convolution layer is a type of neural network layer that is used to perform

convolution on graphs. The convolution operation on a graph is like the convolution operation that is performed on images. Just as in image convolution, the graph convolution operation takes in a graph as input and produces an output graph. The output graph is computed by applying a filter to the input graph. The filter is a matrix that is used to weight the edges of the input graph. In this experiment, we used two Graph Convolution Layer. The first is HOGNNs from “Weisfeiler and Leman Go Neural: Higher-order Graph Neural Networks” paper [8]. The paper introduced a new neural network architecture for learning on graphs, called higher-order graph neural networks (HOGNNs). HOGNNs are like standard graph neural networks (GNNs), but they can operate on graphs with higher-order structure (i.e., edges between edges). This enables HOGNNs to capture more intricate relationships between nodes in a graph. The authors evaluated the performance of HOGNNs on several benchmark datasets and showed that they outperformed standard GNNs. Mathematically it is defined as $x_i' = W_1 x_i + W_2 \sum_{j \in N(i)} e_{ji} x_j$ where e_{ji} denotes the edge weight from source node j to target node i.

Next is GCNConv from Semi-Supervised Classification with Graph Convolutional Networks [9]. This paper proposed a semi-supervised classification method using graph convolutional networks (GCNs). GCNs are a type of neural network that can operate on graph-structured data and have been shown to be effective for various tasks such as node classification and link prediction. The paper demonstrated that GCNs can be used for semi-supervised classification and achieves state-of-the-art performance on several benchmark datasets. Mathematically it is defined as $h_i^{l+1} = \sigma\left(b^{(l)} + \sum_{j \in N(i)} \frac{1}{c_{ji}} h_j^{(l)} W^l\right)$ where $N(i)$ is the set of neighbors of nodes i, $c_{\{ji\}}$ is the product of square root of node degress, and σ is an activation function.

(6) Tool

All the experiments were performed with the Pytorch geometric. PyG (PyTorch Geometric) is a library built upon PyTorch to easily write and train Graph Neural Networks (GNNs) for a wide range of applications related to structured data. It consisted of various methods for deep learning on graphs and other irregular structures, also known as geometric deep learning, from a variety of published papers.

4. Research Schedule

Month	Major Activities	Note
April	Understanding Gas Hydrate	
May	Understand and implement machine learning, deep learning	
June	Molecular dynamics (MD) simulations & learning PyTorch Geometric	
July	Molecular dynamics (MD) projector	
August	Graph data pre-processing	
September	Hyper parameter tunning and implement GNN model	
October	Write the final report and poster	
November	AICP Festival	

5. Research Results

(1) Accuracy

	MAE	**MSE**
GCNs	0.7031	0.8466
HOGNNs	0.7073	0.8554

To compare the models, we used two accuracy indicator which are mean absolute error (MAE) and mean squared error(MSE). The result was as follow: MAE of GraphConv was 0.7031, MSE of GraphGonv was 0.8466, MAE of GCNConv was 0.7073 and MSE of GCNGonv was 0.8554. The result suggested that the GraphConv model was slightly better than the GCNConv model in terms of MAE and MSE.

There are several reasons why the MAE and MSE of GCN were better than HOGNNs. Firstly, GCN uses a more sophisticated technique called "graph convolution" to update node representations, while HOGNNs uses a simple matrix multiplication. Graph convolution is more effective at capturing relationships between nodes, which leads to better node representations and hence better predictions. Secondly, GCN was specifically designed to handle graph data, while HOGNNs was a general-purpose neural network that can be

applied to any kind of data. This design difference was led that GCN could take advantage of the structure of graph data, which was led to better predictions.

(2) Plot the edge weights

Now to explain the predictions, we looked at the attribution methods according to case study. We calculated the gradient of the output with respect to the edge weights W_{ei} . Edge weights are initially one for all edges. For the saliency method, we used the absolute value of the gradient as the attribution value for each edge $\text{Attribution}_{e_i} = \left| \frac{\partial F(x)}{\partial W_{e_i}} \right|$. Where x is the input and $F(x)$ is the output of the GNN model on input x.

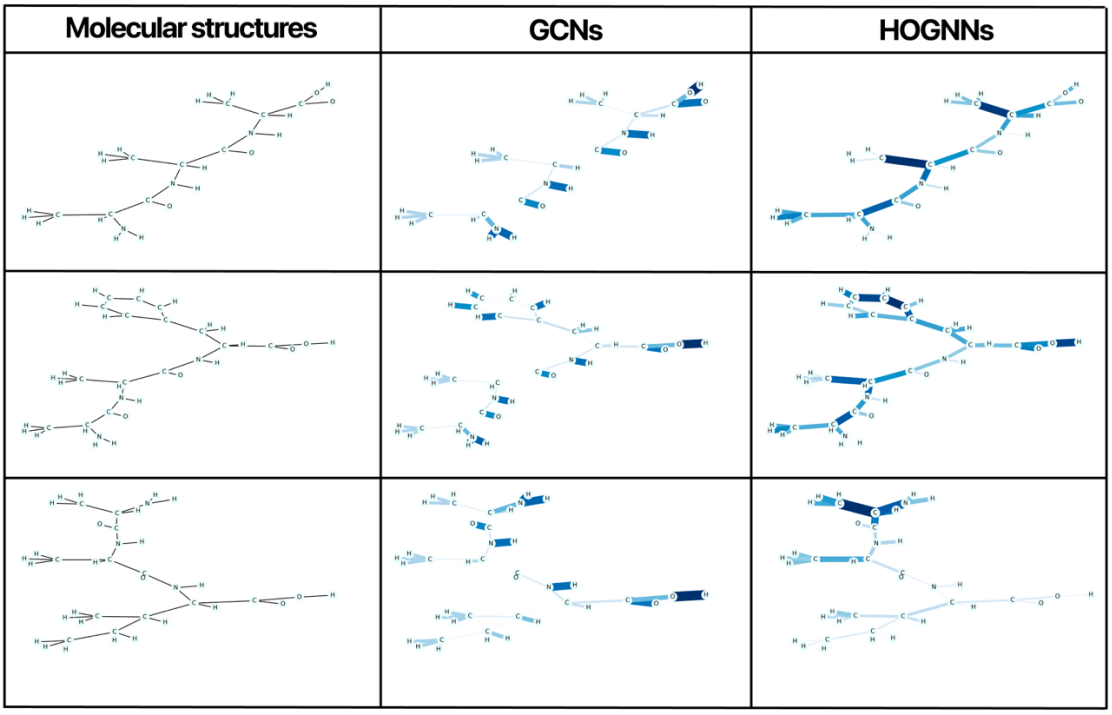


Figure 7. A case study through three molecular structures.

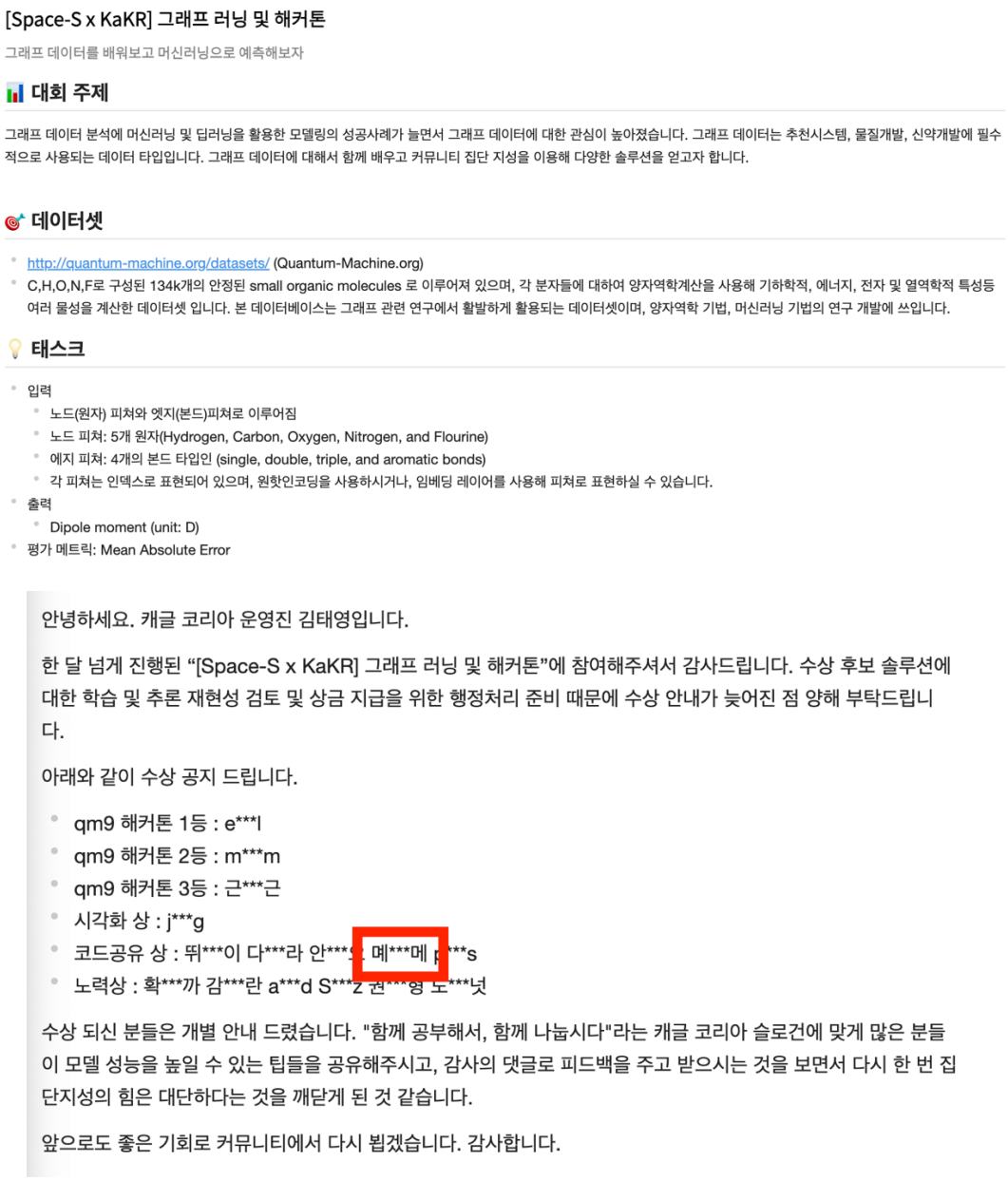
We conducted a case study through three molecular structures. The figure 7 shows that in the case of GCNs, the edges of the part in contact with the hydrate molecular were highly important. On the other hand, in the case of HOGNNs, edge importance was not well learned except for some edge edges and long-tailed edges. The difference between these two models influenced accuracies such as MAE and MSE. In our case, the edge parts in which

functional groups encounter the molecule were more important since the main task was to predict the pulling force between the hydrate slab and interested molecule.

6. Research Achievements

- [Space-S x KaKR] 그래프 러닝 및 해커톤

■ 코드공유상 수상



[Space-S x KaKR] 그래프 러닝 및 해커톤

그래프 데이터를 배워보고 머신러닝으로 예측해보자

📊 대회 주제

그래프 데이터 분석에 머신러닝 및 딥러닝을 활용한 모델링의 성공사례가 늘면서 그래프 데이터에 대한 관심이 높아졌습니다. 그래프 데이터는 추천시스템, 물질개발, 신약개발에 필수적으로 사용되는 데이터 타입입니다. 그래프 데이터에 대해서 함께 배우고 커뮤니티 집단 지성을 이용해 다양한 솔루션을 얻고자 합니다.

🎯 데이터셋

- http://quantum-machine.org/datasets/ (Quantum-Machine.org)
- C,H,O,N,F로 구성된 134k개의 안정된 small organic molecules 로 이루어져 있으며, 각 분자들에 대하여 양자역학계산을 사용해 기하학적, 에너지, 전자 및 열역학적 특성등 여러 물성을 계산한 데이터셋 입니다. 본 데이터베이스는 그래프 관련 연구에서 활발하게 활용되는 데이터셋이며, 양자역학 기법, 머신러닝 기법의 연구 개발에 쓰입니다.

💡 태스크

- 입력
  - 노드(원자) 피쳐와 엣지(본드)피쳐로 이루어짐
  - 노드 피쳐: 5개 원자(Hydrogen, Carbon, Oxygen, Nitrogen, and Flourine)
  - 에지 피쳐: 4개의 본드 타입인 (single, double, triple, and aromatic bonds)
  - 각 피쳐는 인덱스로 표현되어 있으며, 원핫인코딩을 사용하시거나, 임베딩 레이어를 사용해 피쳐로 표현하실 수 있습니다.
- 출력
  - Dipole moment (unit: D)
- 평가 메트릭: Mean Absolute Error

안녕하세요. 캐글 코리아 운영진 김태영입니다.

한 달 넘게 진행된 “[Space-S x KaKR] 그래프 러닝 및 해커톤”에 참여해주셔서 감사드립니다. 수상 후보 솔루션에 대한 학습 및 추론 재현성 검토 및 상금 지급을 위한 행정처리 준비 때문에 수상 안내가 늦어진 점 양해 부탁드립니다.

아래와 같이 수상 공지 드립니다.

- qm9 해커톤 1등 : e***l
- qm9 해커톤 2등 : m***m
- qm9 해커톤 3등 : 근***근
- 시각화 상 : j***g
- 코드공유 상 : 뛰***이 다***라 안***요 메***메 p***s
- 노력상 : 확***까 감***란 a***d S***z 권***형 도***넛

수상 되신 분들은 개별 안내 드렸습니다. "함께 공부해서, 함께 나눕시다"라는 캐글 코리아 슬로건에 맞게 많은 분들이 모델 성능을 높일 수 있는 팁들을 공유해주시고, 감사의 댓글로 피드백을 주고 받으시는 것을 보면서 다시 한 번 집단지성의 힘은 대단하다는 것을 깨닫게 된 것 같습니다.

앞으로도 좋은 기회로 커뮤니티에서 다시 뵙겠습니다. 감사합니다.

7. Expected Contribution / Future Plans

In this project, we both used MD and DL approaches to discover eco-friendly hydrate inhibitors. The hydrate formation in the oil and gas pipelines is a major problem, and conventional hydrate inhibitors are known as toxic and expensive. Our goal was to find new, environmentally friendly hydrate inhibitors that are both effective and affordable. To do so, we conducted numerous inhibitor-pulling scenarios to predict the pull force under the assumption that the stronger the pull force between hydrate and a molecule, the better the inhibition.

MD simulation was a powerful tool that can enabled us to simulate the molecules and hydrate structures. We used this tool to get the information of molecular and pulling force. After the simulation, we demonstrated Graph Neural Network to solve the molecular regression problem. We think that DL is a powerful artificial intelligence technique that can help us identifying the most effective inhibitors. By combining MD and DL approaches, we hope to find the best possible hydrate inhibitors and predict the pulling force faster and more accurately. According to the simulation, the best molecular was Asp-Trp-Asp tripeptide. It showed the largest pulling force among other tripeptides.

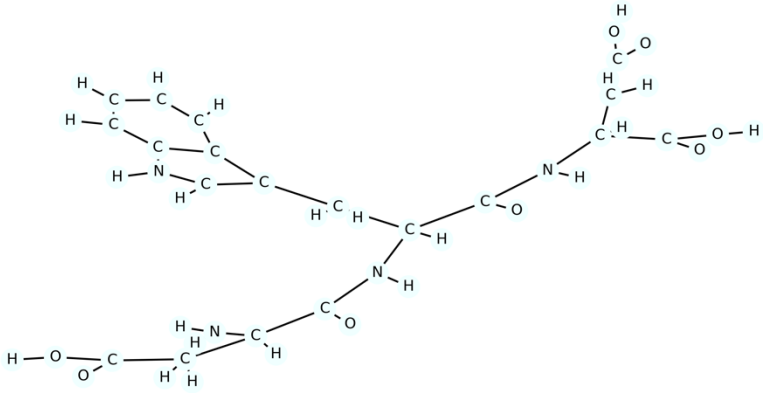


Figure 8. The Asp-Trp-Asp tripeptide which showed the largest pulling force between the hydrate inhibitor and the slab

The long-term goal of this research is to develop more environmentally friendly hydrate inhibitors using MD and DL approaches. In the short term, we focused on finding more accurate models for predicting pulling force and on understanding the mechanisms of MD simulation and PyTorch Geometric. In further research,

we hope to discover new hydrate inhibitor molecules based on generated model that are more effective and less toxic and new model that have more expandability to longer oligopeptide data for scalability.

8. Member's Role

	Name	Role	Details
1	고우진	Team Leader	Gas hydrate and MD simulation - Responsible for MD Simulation Research - Problem-solving activities in the research topic - Feasibility study and progression of research topics - Identification of differentiation from the existing gas hydrate problem approach
2	김예진		Machine learning and deep learning - Overall research subject design and research progress - Understand and implement machine learning, deep learning - Responsible for team member research - Check the results of individual studies
3	남은슬		External activities and administrative - the overall internal and external activities of the project - the publication of papers and the administration of academic participation - Administrative processing related to the purchase of necessary goods

9. References

[1] Sloan Jr, E. D., & Koh, C. A. (2007). *Clathrate hydrates of natural gases*. CRC press.

[2] Sloan, E. D. (2010). *Natural gas hydrates in flow assurance*. Gulf Professional Publishing.

[3] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1*, 19-25.

[4] Abascal, J. L. F., Sanz, E., García Fernández, R., & Vega, C. (2005). A potential model for the study of ices and amorphous water: TIP4P/Ice. *The Journal of chemical physics*, *122*(23), 234511.

[5] Robertson, M. J., Tirado-Rives, J., & Jorgensen, W. L. (2015). Improved peptide and protein torsional energetics with the OPLS-AA force field. *Journal of chemical theory and computation*, *11*(7), 3499-3509.

[6] Martínez, L., Andrade, R., Birgin, E. G., & Martínez, J. M. (2009). PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry*, *30*(13), 2157-2164.

[7] Jiang, D., Wu, Z., Hsieh, C. Y., Chen, G., Liao, B., Wang, Z., ... & Hou, T. (2021). Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. *Journal of cheminformatics*, *13*(1), 1-23.

[8] Morris, C., Ritzert, M., Fey, M., Hamilton, W. L., Lenssen, J. E., Rattan, G., & Grohe, M. (2019, July). Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 4602-4609).

[9] Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

10. Participants' Comments

(Professor)

This is a very well-conceived and executed project. The use of molecular dynamics and deep learning approaches is very innovative and promises to be a fruitful research direction. I am very impressed with the work that has been done so far and I look forward to seeing the results of future work in this area.

(Team members)

- (고우진) The project proceeded well during this year and the team did a great job. We had made great progress in discovering eco-friendly hydrate inhibitors using molecular dynamics and deep learning approaches. It was the first time I had run so many simulations at once, so the automation of the series was a challenge for me. I am deeply grateful to the ai innovation center for supporting my team so much.
- (김예진) Thank you so much for creating an environment where we can research with

generous support for a year. As I continued to study deep learning, I was able to build a lot of knowledge about it and set the direction of my research. It was a great honor to lead this research with the professor who gave generous advice and the best team members.

- (남은슬) I was very excited about the project, and it was amazing how the team used molecular dynamics and deep learning approaches to discover eco-friendly hydrate inhibitors. If there is a chance in the future, I would like to conduct a study like this again. I would like to say thank you to the ai Innovation Center for giving our team their generous supports!

We hereby submit the final report for the AICP in 2022 as stated above.

Date: 11 / 27 / 2022

Professor: 서용원 (Signature)

Student leader: 고우진 (Signature)