SPATIAL INFORMATION INFERENCE IMPROVES ACTIVIATION ENERGY PREDICTIONS

Chua Wee Chong[1], Liu Wenkai[1], Liew Wen Xing Alvin[2], Chong Yihui[2], Chieu Hai Leong[2]
[1]NUS High School of Math and Science, 20 Clementi Ave 1, Singapore 129957
[2]DSO National Laboratories, 12 Science Park Dr, Singapore 118225

Abstract

Prediction of chemical properties is of interest due to their use in important tasks like screening candidates for novel drugs. In predicting, practical experiments are time-consuming and expensive, leading computational methods to be of recent interest. Recently, research to predict reaction properties given the reaction has shown promising results. Independently, research to infer spatial information using machine learning models has shown improvements in molecular predictions. Extracting ideas from both previous works, we aim to create a model to improve accuracy in reaction property predictions through inferring spatial information, using activation energy as our reaction property. We achieve this through a multi-step training process. First, we train two models, both with the objective to predict activation energy, one inputting only reaction information, termed the reaction model, and the other inputting only spatial information. Next, to implicitly encode spatial information into the reaction model, the models are contrastively trained to maximize their mutual information. Finally, to predict activation energy, the contrastively trained reaction model is trained on reaction inputs again. Our method produces statistically significant reductions in mean absolute error with a p-value of 0.002, reducing mean absolute error to 6.02, from the initial reaction model of 6.12, whilst maintaining similarly high speeds of prediction. Our work validates the possibility of spatial information inference with reaction models to improve activation energy predictions. Inferring spatial information would allow scalability of the model to reactions containing larger molecules and allows for use of higher quality conformers during training, without a trade-off in prediction speed.

Introduction

Prediction of reaction properties is of great importance for various processes, especially in organic synthesis planning for novel drugs or new materials. In predicting reaction properties, practical observation first comes to mind, however it would be time-consuming and expensive. As an alternative, computational methods have been of interest. One computational approach is to use

quantum simulations such as Density Functional Theory to predict chemical properties to a high level of accuracy. However, such methods are time consuming, taking up much more time with larger molecules. A second approach, which we are interested in, is data-driven machine learning methods. Such methods have been enabled by the emergence of large chemical datasets like [1].

Two past works are of our interest. In the first, cited in [2], Heid and Green propose a general architecture to take in reactions as SMILES strings, either balanced or unbalanced. They show that their method outperforms state-of-the-art models in accuracy in a range of reaction property predictions, such as activation energy, enthalpies, and rate constants.

In the second, cited in [3], Stärk *et al*. proposed a method to implicitly encode spatial information into molecular models to allow the molecular models to reason using spatial information, without having spatial information as input during test time. They show that their method does improve results over pure molecular models, but using full spatial information outperforms implicit information. However, their main advantage is in their speed and ability to predict molecular properties accurately in the absence of spatial information at test time.

For this paper, we focus on the prediction of activation energies when given only the reaction SMILES, which is a representation of the reaction. We would like to take inspiration from some ideas from both papers. Our work wishes to implicitly encode spatial information in encodings of reaction SMILES, inspired by but differing from Stärk's implicit encoding only onto molecular SMILES. The purpose is twofold: spatial information is highly useful in predicting chemical properties, we wish to implicitly encode it on top of reaction information to improve accuracy. At the same time, by implicitly encoding information, we improve the speed of the model at test time, as spatial information would not need to be generated during test time.

Methodology

To perform our training, we first need to generate spatial information for our reactions. Our training consists of three stages. We first created two models, one taking in reaction SMILES, the other taking in spatial information. Next, we perform contrastive learning between the two models. Finally, we train the reaction SMILES input model once again to obtain our final model.

Conformer Generation

To generate spatial information, we first generated conformers for the various molecules in the dataset using ETKDG [6]. We first split up the reaction into its component molecules, generating conformers for each molecule. We generated multiple conformers for each molecule, ensuring a range of candidates for the lowest energy conformer. Some conformers generated by ETKDG were similar, hence we pruned some of them off. Afterwards, we carried out energy minimization using MMFF94 [7]. We then chose the lowest energy conformer of the tuned conformers as the conformer for that molecule. These conformers were then put into a reaction representation, forming an analogue to the original reaction. Details for exact generation and pruning parameters can be found in the Appendix.

Unfortunately, MMFF94 failed to minimize energy for conformers of certain molecules, such as molecules containing boron. For this project, we chose to filter away the molecules for which conformers could not be generated for. This leaves us with a reduced dataset containing only reactions with all molecules having valid conformers. The rationale for filtering and results without the filtering of molecules is presented in the Appendix.

Reaction Model

To predict reaction properties from reaction SMILES strings, Heid and Green proposed a two-step method which we followed. First, the reaction SMILES is transformed into an encoding by the encoder. Then, the encoding is turned into the actual reaction property prediction by the feed-forward network. These two components form a model that takes in a reaction SMILES, and outputs the desired reaction property. Exact details can be found in [2].

For this project, we largely used the *chemprop* library on GitHub [8], directly modifying it to work for our purposes. For the reaction model, *chemprop*'s architecture contained Heid and Green's method to train a model using only reaction inputs, without any spatial information to predict activation energy. We term this model as our Reaction Model. Refer to Figure 1.
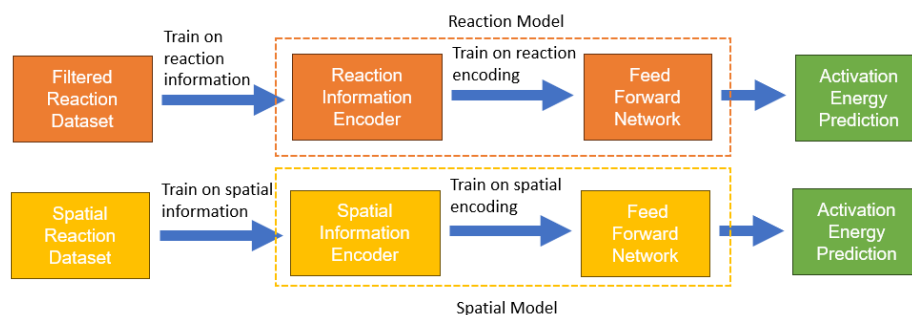
Figure 1: Training processes for the Reaction and Spatial Models

Spatial Model

From the previously generated conformers, we get the coordinates of each atom in space. However, we do not use coordinates directly as they would change with translation and rotation of the molecule. Hence, drawing inspiration from [3], we converted the coordinates generated into translation and rotation invariant distances. We first convert the coordinates of the reactants. From the molecular graph, we took the Euclidean distance between bonded atoms to find their bond length. We then created a distance matrix for all atoms in the reactants. For atoms within the same molecule, we took their Euclidian distance based on their coordinates. For atoms within different molecules, we set their distance to 100, to represent an infinitely large distance. The atom distance matrix and bond lengths are then mapped to higher dimensions, using the same method as [3], in hopes of improving results. We convert the products in a similar fashion to the reactants.

We then train the spatial model, which has a similar architecture to the reaction model, to predict activation energy. Refer to figure 1 for the architecture. We input the above generated atom and bond spatial information for both products and reactants to the spatial model. Additionally, we included the atomic number and bond type as inputs to the spatial model. Other molecular features were not present.

Contrastive Learning

From the previous two subsections, we have produced two models: one for predicting activation energy given a reaction equation: the reaction model, and another given mostly spatial information: the spatial model. Both models first turn the input into an encoding, we call this portion the encoder, then predict based on that encoding.

For contrastive learning, we wish for reaction and spatial encodings to have a high similarity if they correspond to the same reaction; a low similarity if they correspond to different reactions. This would accurately encode of spatial information implicitly into reaction encodings. It is important to note that we have deliberately trained the two models with mutually exclusive features. This ensures that the similarity of our encodings output by our reaction model after contrastive learning is only due to the similarity to spatial information.

To conduct our training, we use both trained encoders simultaneously. Initially, we use the trained weights of both encoders, adding a linear layer to change the size of the encoding. Refer to figure 2. To the molecule encoder, we feed in the reactions for which all molecules have conformers. To the spatial encoder, we feed in the conformers of the reactions. We calculate the loss between the two output lists of encodings. The loss function used is NTXloss [3], which has the formula:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \left[\log \frac{e^{sim(z_i^a, z_i^b)/\tau}}{\sum_{\substack{k=1 \\ k \neq i}}^{N} e^{sim(z_i^a, z_k^b)/\tau}} \right]$$

where $sim(z^a, z^b) = z^a \cdot z^b / (\|z^a\| \|z^b\|)$, the cosine similarity, and τ is a temperature parameter.

After training, we removed the last linear layer and ended up with two encoders: the augmented molecule encoder with implicit spatial information and the augmented spatial encoder with implicit molecule information. Of the two, we use only the augmented reaction encoder.

Final Training

Refer to figure 2. Here, we train an encoder, initializing its weights to the weights of the augmented reaction encoder. Then, we input these encodings to train a feed-forward network. To form a single model that predicts activation energy given a reaction SMILES, we put the trained encoder and the feed-forward network as a singular model. We term this model as our Final Model.
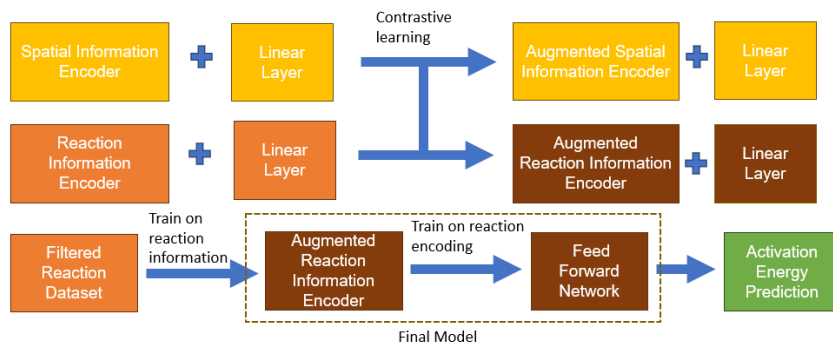
Figure 2: The contrastive learning process and the Final Training process

More details on training parameters for the methodology are within the Appendix.

Data Preparation

For this project, our data was sourced from [4]. Specifically, we used *fold_0* of the *ωB97X-D3* dataset to predict activation energies. The dataset contains a variety of reactions from unimolecular reactants to either unimolecular or multimolecular products, for a total of about 23000 reactions containing up to 7 heavy atoms in the reactant. The reaction data used Density Functional Theory to calculate activation energy to a high level of accuracy[5]. The train and validation data used the "*aam_train.csv*" file in the same folder. The test data used "*aam_test.csv*" in the same folder.

Results and Discussion

Through this project, we wish to answer 2 questions:

1. Does spatial information inference give greater accuracy on activation energy predictions?

We use a test set of 1682 reactions provided in the test file for a fold of *ωB97X-D3,* which has been filtered, ensuring that all reactions have molecules that can generate conformers. To test this hypothesis, we compare the reaction model and our final model. As an additional test, we also compare the spatial model. We compare them on two metrics, mean absolute error (MAE) and root-mean squared error (RMSE).

	Reaction Model	Spatial Model	Final Model
MAE	6.12	21.4	6.02
RMSE	9.56	26.9	9.46

Table 1: A comparison of the performances of the three trained models

To test if the improvement of the model is statistically significant, we perform a paired t-test between the absolute errors of reaction model and the final model at 1% significance level. The null hypothesis is that there is no reduction in the absolute error, the alternative hypothesis is the final model does reduce the absolute error, with 1681 degrees of freedom, giving us a p-value of 0.002. We reject the null hypothesis at 1% significance level, concluding that the final model shows statistically significant improvements over the reaction model.

We have shown with our model that the implicit inference of spatial information can improve activation energy predictions. However, discussed in the Appendix, using more training data, the reaction model outperformed our implicit inference training. Therefore, we recommend that spatial information inference be used best when there is a high availability of accurate data.

Of interest to note is how the spatial model is unable to accurately predict the activation energy. However, when the reaction model is made to contrastively learn from encodings from the spatial model, it performs better than without the contrastive learning. A possible reason is that when the conformer generated is converted into an encoding, it does not contain enough information to fully model the reaction. However, the encoding generated by the spatial model contains valuable information about properties of the reaction, not present within encodings of the reaction model. This would allow the contrastive learning to enhance the information within the reaction encoding, even as the spatial encoding itself is insufficient.

2. Would the model perform faster than explicit spatial information models?

For this project, we have used ETKDG to generate our conformers. To compare the runtimes between explicit and implicit, we will use our test set containing 1682 reactions. In explicit spatial information models, these conformers must be generated at test time for them to be available as input. To generate the conformers for the entire test set, we take over 3 minutes. In running our final model, taking only the reaction SMILES as input, it takes only 3 seconds for the entire test set. In other words, this is a 60-fold improvement in runtime compared to using explicit spatial information. It is worth noting that ETKDG is already one of the faster conformer generation tools available. Our final model improves runtime by not requiring spatial information as input, using only implicit spatial information to help with its predictions.

Future Work

For our project, conformers were generated rather quickly using the ETKDG algorithm. However, more accurate and time-consuming methods of generating conformers could be employed, without a trade-off in speed for our model. One such method of recent interest is Torsional Diffusion [9], shown to be more accurate, while still offering high speeds. Using more accurate conformers would allow the spatial model to model the actual spatial data more accurately, resulting in higher quality encodings, giving more accurate activation energy predictions.

Bond angle of the reactants could be a factor within the model to predict activation energy more accurately. For example, the reaction between cyclopropane and bromine has a lower activation energy compared to propane and bromine. This is as the removal of the ring strain in cyclopropane via reaction with bromine is highly favorable, resulting in a lower activation energy. Using bond angles would hence allow us to predict activation energy more accurately for reactions, especially those containing molecules with many small rings.

Activation energy predictions on datasets with larger molecules could also be explored. Explicit models find it difficult to scale up for larger molecules, since time taken for conformer generation increases greatly for more molecules. In contrast, we generate the conformers only during training, making it faster for the end user to predict using the model.

Conclusion

We have proposed an activation energy prediction model that accounts for spatial information implicitly. This model is 1. able to outperform state-of-the-art models with 2. lesser computation time as compared to models using explicit spatial information.

Acknowledgements

References

[1] Wu, Zhenqin, et al. "MoleculeNet: a benchmark for molecular machine learning." Chemical science 9.2 (2018): 513-530.

[2] Heid, Esther, and William H. Green. "Machine learning of reaction properties via learned representations of the condensed graph of reaction." *Journal of Chemical Information and Modeling* 62.9 (2021): 2101-2110.

[3] Stärk, Hannes, et al. "3d infomax improves gnns for molecular property prediction." *International Conference on Machine Learning*. PMLR, 2022.

[4] CSV Files of Datasets, Data Splits, Models, 2021.
https://github.com/hesther/reactiondatabase (accessed 12/2023)

[5] Grambow, Colin A., Lagnajit Pattanaik, and William H. Green. "Reactants, products, and transition states of elementary chemical reactions based on quantum chemistry." *Scientific data* 7.1 (2020): 137.

[6] Riniker, Sereina, and Gregory A. Landrum. "Better informed distance geometry: using what we know to improve conformation generation." *Journal of chemical information and modeling* 55.12 (2015): 2562-2574.

[7] Tosco, Paolo, Nikolaus Stiefl, and Gregory Landrum. "Bringing the MMFF force field to the RDKit: implementation and validation." *Journal of cheminformatics* 6 (2014): 1-4.

[8] Chemprop Software, 2021. https://github.com/chemprop/chemprop (accessed 12/2023)

[9] Jing, Bowen, et al. "Torsional diffusion for molecular conformer generation." *Advances in Neural Information Processing Systems* 35 (2022): 24240-24253.

[10] Conformer Generation Using RDKit, 2012.

www.rdkit.org/UGM/2012/Ebejer_20110926_RDKit_1stUGM.pdf. (accessed 12/2023)

Appendix

Parameter Details

Conformer Generation

To generate the conformers using ETKDG, we followed recommended parameters in [10]. We chose the number of conformers to generate for each molecule based on the formula below, featured in [10]:

$$n = \begin{cases} 50 & \text{if } n_{rot} \leq 7 \\ 200 & \text{if } n_{rot} \geq 8 \text{ and } n_{rot} \leq 12 \\ 300 & \text{otherwise} \end{cases}$$

where n_{rot} represents the number of rotatable bonds for the molecule that conformers is being generated for. As for the pruning threshold, we chose it to be:

$$threshold = \frac{n_{rot}}{6} + \frac{19}{60} - \frac{2}{10}$$

This is based on the linear regression of the graph of number of rotatable bond against the average minimum RMSD from crystallographic conformation found in [10].

Reaction Model

We train the reaction model on 100 epochs, using the *reaction* setting in *chemprop* [8], so that it can accept reactions. For the features used, we use the default features accepted by *chemprop*. For the reaction mode used, we used the setting to take the difference between the products and the reactants as input to feed into the model (*prod_diff*). The reasoning is that activation energy is largely affected by atoms and bonds that are targeted during the reaction and not as much by atoms and bonds that remain unchanged throughout the reaction. By using *prod_diff*, we give higher weightage to the atoms and bonds that are being changed, since we are taking their difference before and after the reaction. This would lead to the model being more able to deduce, as it does not have to infer which atoms and bonds are the most important when considering the activation energy. The loss function used is root mean square error. Other model parameters were set to default, using the parameters suggested in the original paper in [2].

Spatial Model

The parameters for the reaction and spatial model are the same, except for the input features. For the input features to the spatial model, we use the condensed graph of reaction representation for the distance matrix for atoms, atomic number, bond length and bond type.

Contrastive Learning

We train both models contrastively for 100 epochs. For the reaction mode used, we used the setting to take the difference between the products and the reactants as input to feed into the model (*prod_diff*) for a similar reason as above. The loss function used is NTXloss. In NTXloss, τ is a temperature parameter, we set $\tau = 0.1$, using the finding from [3] that and $\tau = 0.1$ worked best. Our linear layer output size is 256, to make the same as the encoding size in [3].

Final Model

We train this model similarly to the reaction model for 100 epochs. The only difference between training the reaction model and the final model is that the final model initializes the weights from the augmented reaction encoder from the contrastive learning. This is opposed to the reaction model, where the encoder is trained from scratch and initialized using random weights. The final model also uses the *reaction* setting, with *prod_diff* as the reaction mode, with a similar reasoning to the reaction and spatial models.

Additional Results

Training models without filtering

We use the same test set of 1682 reactions that was used for in *Results and Discussions.* We test four models. We test the reaction and final models, with and without filtering reaction SMILES for which conformers could not be generated for. Refer to figure 3. The difference in training without filtering reaction SMILES would be the input to the reaction and final models.
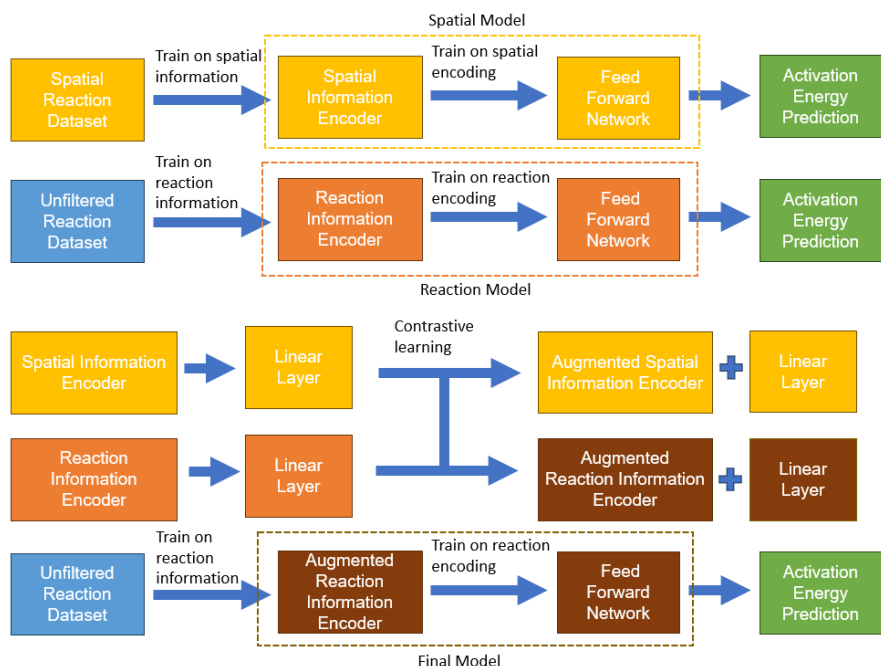
Figure 3: Diagram of the training process without filtering reaction SMILES for which conformers could not be generated for

We compare them on two metrics, mean absolute error (MAE) and root-mean squared error (RMSE).

	Reaction Model with filtering	Final Model with filtering	Reaction Model without filtering	Final Model without filtering
MAE	6.11	6.02	5.82	5.86
RMSE	9.56	9.46	9.16	9.16

Table 3: A comparison of the performances of the various models with and without filtering.

Our final model with filtering seems to perform worse than the reaction model without filtering. This could possibly be due to the greater availability of data from which the reaction model without filtering could learn from over the final model with filtering. Additionally, it also seems that the final model without filtering does worse than the reaction model without filtering. One possible hypothesis is that the contrastive learning step causes the encoder to "un-learn" the proper encoding of certain reactions that conformers could not be generated for. For example, reactions containing boron could be negatively impacted by the contrastive learning, as they

would not be accounted for during the contrastive step. As a result, their encodings would contain less essential information about them, leading to a worse outcome.