JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2019, 59, 5026−5033

pubs.acs.org/jcim

Prediction and Interpretable Visualization of Retrosynthetic Reactions Using Graph Convolutional Networks

Shoichi Ishida,[†] Kei Terayama,[‡,§,∥] Ryosuke Kojima,[∥] Kiyosei Takasu,[†] and Yasushi Okuno*[,∥,§,⊥]

[†]Graduate School of Pharmaceutical Sciences, Kyoto University, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan

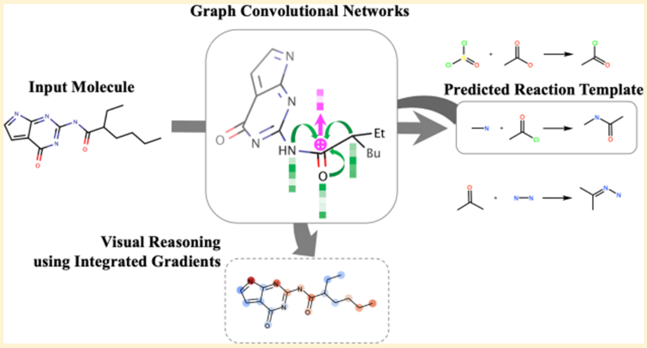
[‡]RIKEN Center for Advanced Intelligence Project, Chuo-ku, Tokyo 103-0027, Japan

[§]Medical Sciences Innovation Hub Program, RIKEN Cluster for Science, Technology and Innovation Hub, Tsurumi-ku, Kanagawa 230-0045, Japan

[∥]Graduate School of Medicine, Kyoto University, Shogoin-kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan

[⊥]Foundation for Biomedical Research and Innovation at Kobe, Center for Cluster Development and Coordination, Chuo-ku, Kobe, Hyogo 650-0047, Japan

Supporting Information


Graph Convolutional Networks
Input Molecule
Predicted Reaction Template
Visual Reasoning
using Integrated Gradients
N
O
HN
Et
Bu

ABSTRACT: Recently, many research groups have been addressing data-driven approaches for (retro)synthetic reaction prediction and retrosynthetic analysis. Although the performances of the data-driven approach have progressed because of recent advances of machine learning and deep learning techniques, problems such as improving capability of reaction prediction and the black-box problem of neural networks persist for practical use by chemists. To spread data-driven approaches to chemists, we focused on two challenges: improvement of retrosynthetic reaction prediction and interpretability of the prediction. In this paper, we propose an interpretable prediction framework using graph convolutional networks (GCN) for retrosynthetic reaction prediction and integrated gradients (IG) for visualization of contributions to the prediction to address these challenges. As a result, from the viewpoint of balanced accuracies, our model showed better performances than the approach using an extended-connectivity fingerprint. Furthermore, IG-based visualization of the GCN prediction successfully highlighted reaction-related atoms.

INTRODUCTION

Designing chemical synthesis is a crucial area of synthetic organic chemistry. Corey formalized the concept of retrosynthetic analysis, which is a common approach to plan a synthetic route.[1,2] Many research groups have tried to develop various computer-aided synthesis planning systems,[3−10] and they are divided into two major approaches: rule-based and data-driven strategies.[11] The rule-based approach is adopted in a successful retrosynthetic tool, Synthia, which can flexibly support chemists in constructing retrosynthetic routes.[12] Besides, chemists have discovered new routes by Synthia, and these routes have shown great results, such as offering yield improvements. On the other hand, researchers have recently focused on the data-driven approaches because of the advancement of deep learning (DL) and search algorithms, and these approaches have shown excellent results.[8,9] DL has also produced remarkable achievements in various fields such as medical,[13,14] translation system,[15] and agriculture.[16]

A one-step retrosynthetic reaction prediction is an essential step for implementing retrosynthetic analysis in data-driven approaches. The retrosynthetic reaction prediction in data-driven approaches corresponds to the rules in rule-based approaches, and we need to repeat the retrosynthetic reaction prediction for retrosynthesis. Hence, the quality of the retrosynthetic reaction prediction strongly influences the quality of a synthetic route obtained by data-driven approaches because the prediction errors accumulate because of the multiple predictions. For addressing the retrosynthetic reaction prediction, DL-based approaches have been investigated and have achieved excellent performances.[17−24]

However, there is a gap between the process of retrosynthetic reaction prediction by DL and the basic chemist's knowledge of the reaction mechanism. When chemists design a proposed synthetic route, they consider not only local structures of a molecule but its whole structure because chemical reactions could be affected by atoms and functional groups that appear to be irrelevant to the reactive center of the molecule.[2] Therefore, because of the lack of the whole chemical structural information, molecular fingerprints like extended-connectivity fingerprint (ECFP)[25,26] are considered inadequate as an input feature for the reaction prediction.

Received: July 3, 2019
Published: November 26, 2019

DOI: 10.1021/acs.jcim.9b00538

However, for handling a molecular structure in machine learning (ML) or DL, ECFP is still used because it is handy and effective.[17,18,27]

Besides, a black-box problem often arises[28] while DL methods have achieved higher prediction accuracy than conventional ML methods. The black-box problem causes difficulty in the interpretability of the prediction reason. Hence, the black-box problem would make the prediction by DL less acceptable to chemists, whereas it is easy to explain why the reaction is selected by the rule-based approach. Despite these problems, the data-driven approaches have been showing comparable performance to the rule-based approaches.[8,10] To make the data-driven approaches more accessible to chemists, it is essential to solve the above problem. Therefore, this study aims to address the above two issues: (1) improving performance of the retrosynthetic reaction prediction and (2) developing an interpretable visualization system to resolve the black-box problem.

In this paper, we propose a new framework for a retrosynthetic reaction prediction framework based on graph convolutional networks (GCN)[29] with integrated gradients (IG)[30] for visualization to address the above two problems. It is reported that a GCN is a state-of-the-art method that treats a molecule as a graph structure in various tasks[31,32] A GCN can be a solution to the problem derived from the fingerprints because a GCN uses the whole molecular structure information for the prediction. For the black-box problem of DL prediction, visualization methods of feature contributions for each sample are applicable.[30,33−35] IG is an architecture-free visualization method; that is, this method can be applied to any differentiable neural networks, including GCN, and does not affect the neural network performance. Although IG has been proposed and evaluated mainly in the application of image recognition,[30] no study has yet been conducted to analyze chemical properties using IG quantitatively. In addition, in the studies for retrosynthetic reaction prediction,[17,18,27] comparison of the performances of GCN and ECFP for the retrosynthetic reaction prediction, that is, which molecular representation is suited for the reaction prediction, is not reported.

In this study, we demonstrate the effectiveness of our framework by combining GCN and IG using United States patent dataset[36] that has been employed in many studies for the data-driven approaches.[18,27,37] Following the previous studies,[17,18] we extracted reaction templates from the patent dataset. We trained the GCN model and ECFP model, which predict the reaction template from a given molecule. As a result, the GCN model showed better performance of retrosynthetic reaction prediction toward retrosynthetic analysis compared to the ECFP model. We also demonstrated that IG-based visualization of the GCN prediction successfully highlighted reaction-related atoms. Furthermore, the visualization of GCN prediction showed the contribution of the reaction-related atoms to the prediction quantitatively. Our implementations are available on GitHub at https://github.com/clinfo/kGCN and https://github.com/clinfo/extract_reaction_template, and we also provide KNIME node extension of our framework. Our method will contribute to retrosynthetic analysis based on the data-driven approaches.

METHODS

Data Set. We define a reaction template as a reactive center and the first neighboring atoms and bonds in a reaction. We refer to a product in a reaction template as a reaction center. To create the dataset of the reaction templates, we employed the set of 1 808 937 reactions of the United States patents published between 1976 and September 2016, prepared by Lowe.[36] The reaction set contains many duplicate reactions, and solvents and chemical agents are inconsistently registered in the reaction set. Therefore, we only used reactants and products in the reactions for simplicity. As a procedure, we removed the duplicates (resulted in 1 105 130 reactions), reduced all reactions to reactants and products, and kept only reactions that have a product. In detail, we filtered the reactions by two steps. First, we remove agents, such as solvent, reagents, and catalysts, defined by SMILES syntax. Then, we removed salts in the reactions because the salts influenced the condition of keeping only reactions that have a product. We defined the salts with reference to ChemAxon's default salts (Supporting Information, Table S1). After the refinement of the reaction set, a total of 1 072 175 reactions remained, and we extracted reaction templates from the reactions using Automapper in ChemAxon API.[38] According to the previous studies,[8,18] we used unique 1752 templates occurring at least 50 times in the 1 072 175 extracted reaction templates and 371 003 molecules that have a corresponding correct template in the unique templates as the input dataset.

Graph Representation of Molecules for GCN. A molecule is formalized as a tuple $\mathcal{M} \equiv (V, E, F)$, where V is a set of nodes. A node represents an atom in a molecule. A node has features $f_i \in F(i \in V)$, and F is a set of feature vectors representing properties of an atom. We employed the features used in DeepChem[39] (Table 1). E is a set of edges. An edge $e \in E$ represents a bond between atoms, that is, $e \in V \times V \times T$, where T is a set of bond types. We used an adjacency matrix $\mathbf{A}^{(t)}$ defined as follows

$$(\mathbf{A}^{(t)})_{i,j} = \begin{cases} 1 & (v_i, v_j, t) \in E \\ 0 & (v_i, v_j, t) \notin E \end{cases}$$

where $(\cdot)_{i,j}$ represents the j-th element of the i-th row. Using this matrix, a molecule is represented by $\mathcal{M}' = (\mathbf{A}, F)$ where $\mathbf{A} = \{\mathbf{A}^{(t)} | t \in T\}$. In this paper, we used RDKit[40] to create an adjacency matrix and a feature matrix and employed $\mathcal{M}'$ as input for GCN.

Representation of Molecules for ECFP. ECFP is a circular topological fingerprint for molecular characterization[25] and is commonly used in a wide variety of research studies. In this paper, we set the maximum diameter to four and prepared different bit-length ECFPs which have 2048, 4096, and 8192 bits. We used the ChemAxon API for calculating the ECFP.

RETROSYNTHETIC REACTION PREDICTION

We aimed to predict the reaction templates from a molecule (product) in a reaction correctly. We build two models: a model using graph representation of molecules (GCN model) as input and a model with ECFP (ECFP model) as input for comparison (see Figure 1). For the prediction performance evaluation, we used fivefold cross-validation. In detail, the data set was split into three sets in each fold: 65% of the dataset for train data, 15% for validation data, and 20% for test data.

GCN AND ECFP MODELS

GCN Model. We define a graph convolution layer, a graph fully connected layer, and a graph gather layer with the below

Table 1. List of Atom Features

RDKit atom class method	possible value	dimension
GetSymbol()	C,N,O,S,F,Si,P,Cl,Br,Mg,Na,Ca,Fe,As,Al,I,B,V,K,Tl,Yb,Sb, Sn,Ag,Pd,Co,Se,Ti,Zn,H,Li,Ge,Cu,Au,Ni,Cd,In,Mn,Zr,Cr,Pt,Hg,Pb,unknown	44
GetDegree()	0,1,2,3,4,5,6,7,8,9,10	11
GetImplicitValence()	0,1,2,3,4,5,6	7
GetFormalCharge()	formal charge value	1
GetNumRadicalElectrons()	number of radical electrons	1
GetHybridization()	SP,SP2,SP3,SP3D,SP3D2	5
GetIsAromatic()	0,1	1
GetTotalNumHs()	0,1,2,3,4	5

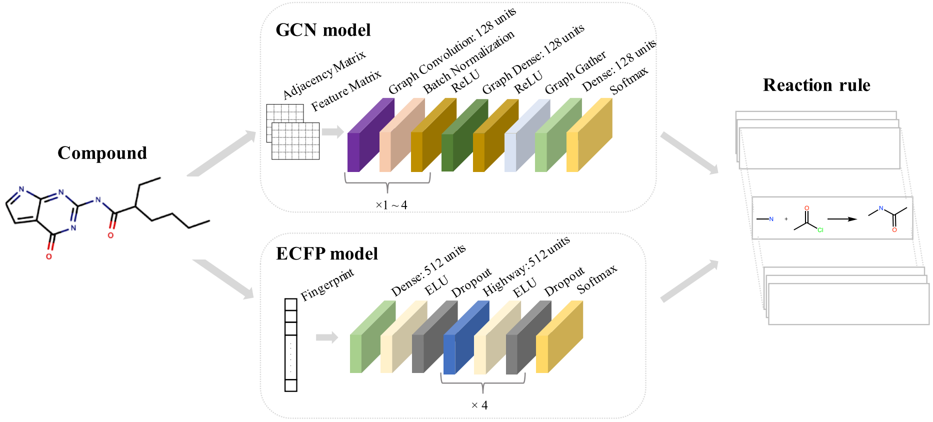


Figure 1. Summary of the GCN and the ECFP models.

descriptions. Our implementation of the layers followed Kipf's model.[29]

Graph Convolution Layer. The graph convolution is calculated from the input $\mathbf{X}^l$ as follows

$$\mathbf{X}^{l+1} = \sigma\left(\sum_t \tilde{\mathbf{A}}^{(t)}\mathbf{X}^{(l)}\mathbf{W}_t^{(l)}\right)$$

where $\mathbf{X}^l$ is an $N \times D^{(l)}$ matrix, $\mathbf{W}_t^{(l)}$ is a parameter matrix $(D^{(l)} \times D^{(l+1)})$ for a bond type t, σ is an activation function, and $\tilde{\mathbf{A}}^{(t)}$ is a normalized adjacency matrix $(N \times N)$.

Graph Dense Layer. $\mathbf{X}^l$ is an input for the graph dense layer. $\mathbf{X}^{l+1}$ is calculated as follows

$$\mathbf{X}^{l+1} = \mathbf{X}^{(l)}\mathbf{W}^{(l)}$$

where $\mathbf{X}^l$ is an $N \times D^{(l)}$ matrix, $\mathbf{W}^{(l)}$ is a parameter matrix $(D^{(l)} \times D^{(l+1)})$.

Graph Gather Layer. This layer converts a graph into a vector,[31] that is, the input $\mathbf{X}^l$ is an $N \times D^{(l)}$ matrix and $\mathbf{X}^l$

$$(\mathbf{X}^{(l+1)})_j = \sum_j (\mathbf{X}^{(l)})_{ij}$$

where $(\cdot)_i$ represents an i-th element of a vector. This operation converts a matrix into a vector.

The GCN model is a neural network consisting of three graph convolutional layers with the batch normalization[41] and the ReLU activation, a graph dense layer with the ReLU activation, a graph gather layer and dense layer with softmax activation. Each layer has 128 units. We set hyperparameters as epochs = 100, batch size = 128, and learning rate = 0.0001, and used early stopping with the patience of three. To implement this model, TensorFlow[42] was used.

ECFP Model. The ECFP model is a neural network according to previous research studies[8,17] and consists of a dense layer with ELU activation and five highway network layers with ELU activation. The dense layer has 512 units with a dropout ratio of 0.3. The highway network layers have 512 units with a dropout ratio of 0.1. We set hyperparameters as epochs = 1,000, batch size = 128, and learning rate = 0.001, and used early stopping with the patience of three. To implement this model, Keras[43] was employed.

VISUALIZATION

To confirm which features of the molecules influenced the prediction result, we developed a visualization system using IG.[30] After learning the model for the retrosynthetic reaction prediction, we can visualize the attributes of the prediction result on the molecular structure. We also quantitatively evaluate the IGs of 10 000 molecules, which were correctly predicted.

Integrated Gradients. We define IGs $\boldsymbol{I}$ as follows

$$\boldsymbol{I}(x_t) = \frac{\Delta x_t}{M}\sum_{k=1}^{M} \nabla_x S_l\left(\frac{k}{M}(\Delta x_t + x^0)\right)$$

where x_t is an atom of a molecule, x^0 is a reference atom whose feature matrix is a zero matrix, M is the number of a division of the reference atom, and $S_l\left(\frac{k}{M}(\Delta x_t + x^0)\right)$ is a score value of the softmax layer's l-th neuron for the predicted reaction template. We defined the atom importance as a sum of IGs of atom features. Our implementation followed Sundararajan's model.[30] We can calculate IGs for each reaction template individually.

DOI: 10.1021/acs.jcim.9b00538

Quantitative Evaluation of Visualization by IGs. To quantify the visualization results of IGs, we calculated the average of IGs of atom features in a reaction center. If some reaction centers exist in a molecule, we chose the reaction center that has the highest average of IGs among them. We compared the average of IGs in each reaction center to the IGs of all the atoms in the molecules by a histogram. To make comparison easier, we standardized the IGs in each of the molecules and plotted Gaussian kernel density estimates of each histogram.

RESULT

Retrosynthetic Reaction Prediction. The GCN model showed better performance than the ECFP model in the balanced accuracy, as shown in Figure 2. In top-*n* balanced accuracy, we regard the prediction that contains the correct reaction template among the top-*n* reaction templates with the softmax probability as a correct prediction. The best GCN model was a three-convolutional-layer model, and its top-one balanced accuracy was 0.249, the top-10 balanced accuracy was 0.510, and top-30 balanced accuracy was 0.662. The best ECFP model was the 2048 dimension, and its top-one balanced accuracy was 0.217, top-10 balanced accuracy was 0.473, and top-30 balanced accuracy was 0.642. Table 2 shows the detailed results of the retrosynthetic reaction prediction of the GCN and the ECFP models.

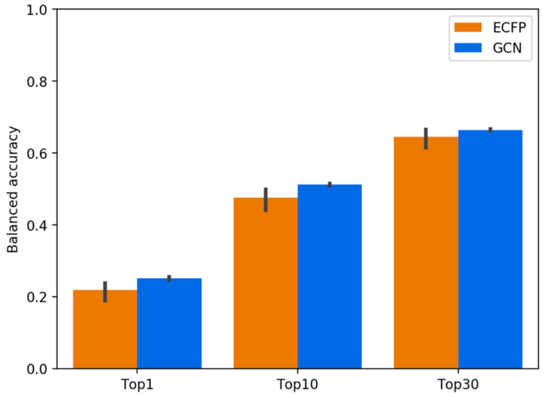


Figure 2. Comparison of the balanced accuracies between the GCN (blue) and the ECFP (orange) models.

To elucidate the difference between the GCN and ECFP models, we show the detailed prediction results in Figure 3. We compared the accuracy of each reaction template in the top-10 balanced accuracy of the best GCN and ECFP models. Figure 3 represents the difference between the accuracies of the GCN model and that of the ECFP model. We also show the top-one and top-30 cases in Supporting Information Figure S1. Figure 3a shows that the GCN model predicted more precisely than the ECFP model around an accuracy ranging from 0.7 to 1.0. Moreover, compared to the ECFP model, the GCN model reduced the number of templates that have a lower accuracy. Figure 3b shows the scatter plot of the accuracies between the GCN model and the ECFP model. To clarify the effects of the number of molecules included in the reaction template to the prediction result, we also added the color information of a logarithmic number of the molecules. We can see in this figure that both the GCN model and the ECFP model tended to accurately predict the templates with a large number of molecules per the reaction template and predict poorly on the templates with a small number of molecules per the reaction template. In the dataset we used, the frequency of occurrence per reaction template was quite different, as shown in Supporting Information Figure S2. If a dataset is biased, conventional ML methods tend to learn classes containing many train data mainly. We can see in Figure 3b that the ECFP model showed this tendency and predicted more accurately the reaction templates containing many train data than the GCN model. Besides, Figure 3b shows that the GCN model predicted more precisely the reaction templates that were not predicted accurately by the ECFP model. To clearly show the difference between the GCN and ECFP model predictions, we show distributions of the top-10 accuracies of top-100/bottom-100 reaction templates (ranked by the frequency of occurrence per reaction template) between GCN and ECFP in Supporting Information Figure S3. Additionally, to clarify which reaction templates the GCN and ECFP models can predict with high accuracy, and which reaction templates the models cannot, we selected top reaction templates and bottom reaction templates and counted duplicate product structure in the reaction template. We defined the top reaction templates as the 129 templates predicted with more than 90% accuracy and the bottom reaction templates as the 125 templates predicted with less than 10% accuracy. Figure 3c shows that the top reaction templates have various unique product structures in the templates, that is, the prediction tasks were easy, and the bottom reaction templates have many duplicate product structures in the template. Examples of the top and bottom reaction templates are in Figure 3d. We can see that the reaction templates with the same product structure have significant negative effects on both GCN and ECFP model performances.

VISUALIZATION

We visualized the contributions of atom features in a molecule to a retrosynthetic reaction prediction. We show typical examples in Figure 4 and examples of top-100 and bottom-100 prediction probabilities in the Supporting Information. We selected examples on the condition that the GCN model correctly predicted the reaction template in Figure 4a,b. Figure 4a shows examples in which the reaction center and the atom contributions do match, and Figure 4b shows examples in which the reaction center and the atom contributions do not match. Figure 4c shows examples of the incorrect prediction,

Table 2. Top-*n* Balanced Accuracy of the GCN and the ECFP Models

	ECFP			GCN			
descriptor	2048 dim	4096 dim	8192 dim	1 conv layer	2 conv layers	3 conv layers	4 conv layers
top 1 balanced accuracy	0.217	0.205	0.192	0.192	0.237	0.249	0.128
top 10 balanced accuracy	0.473	0.464	0.451	0.421	0.489	0.510	0.347
top 30 balanced accuracy	0.642	0.634	0.623	0.569	0.641	0.662	0.496

DOI: 10.1021/acs.jcim.9b00538
J. Chem. Inf. Model. 2019, 59, 5026−5033

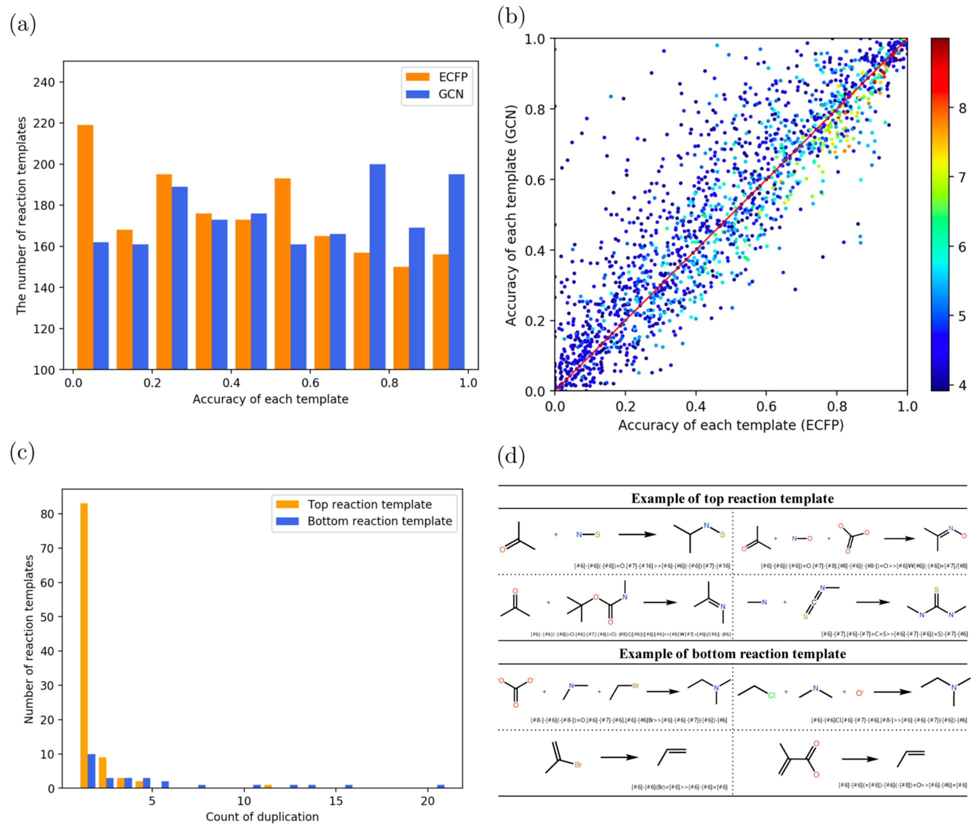


Figure 3. Distribution of the top-10 accuracies of each template between GCN and ECFP. (a) Histogram of the accuracies of each template between GCN (blue) and ECFP (orange). (b) Scatter plot of the accuracies with the color bar on the right. The color bar represents a logarithmic number of molecules in the template. (c) Histogram of the count of duplicate reaction templates between the top reaction template (orange) and the bottom reaction template (blue). (d) Example of top/bottom reaction templates. The reaction templates in a SMARTS under the drawn reaction templates.

so the correct reaction template and the predicted reaction template are shown in the reaction template column. The red color means positive contributions for the prediction, and the blue color means negative contributions for the prediction. The light-green part corresponds to the substructure of the product in the correct reaction template and the light-purple part in the incorrectly predicted reaction template. For defining the light-green or light-purple part of atoms, we performed substructure matching for the compound (e.g., right column in Figure 4a) using a reaction center which was defined in the Data Set section (e.g., left column in Figure 4a). Then, we colored the matching part in the compounds.

Quantitative Evaluation of Visualization by IGs. To quantitatively evaluate the visualization performance, we calculated the average of IGs of atom features in a reaction center. Figure 5 shows the histogram of standardized IGs of each atom in the molecules (orange) and the histogram of standardized averages of IGs in the reaction centers (blue). Here, the standardized average of IGs in a reaction center means the average of standardized IGs of atoms in the reaction center. The orange and blue lines show Gaussian kernel density estimations of the standardized IGs of each atom in the molecules and the standardized averages of IGs in the reaction centers, respectively. The average IGs of the reaction center is 3.71, and that of all the atoms is 0.0 because the IGs in each molecule were standardized. The distribution of the standardized averages in the reaction center was shifted positively. This result suggests that our system successfully recognizes the reaction center. The reason why both distributions did not separate completely is that not all atoms in the reaction center have a positive IG (see Figure 4).

■ DISCUSSION

From the result shown in Figures 2 and 3, the GCN model showed a higher prediction performance than the ECFP model in the retrosynthetic reaction prediction. In prediction tasks of molecular properties, many previous studies have shown that graph-based approaches overcome conventional ML approaches.[32] Besides, graph-based approaches have been performing well in multitask learning for hundreds of classes.[32] In this study, compared to conventional neural network methods with ECFP, the graph-based approach was also effective in the retrosynthetic reaction prediction and showed good performance in almost 1800 class predictions. Moreover, although the dataset was biased, the GCN model tends to predict a wide range of reaction templates correctly. That may be caused by the property that the graph-based methods are generally hard to overfit a dataset.[44] In general, this tendency is important in the retrosynthetic analysis because important reactions do not always appear frequently in the reaction templates.

We showed that the proposed system using IGs successfully recognized the reaction center for a retrosynthetic reaction prediction, as shown in Figures 4 and 5. Although data-driven retrosynthetic analysis has not shown sufficient reasons for a retrosynthetic reaction prediction, our system using IG is

DOI: 10.1021/acs.jcim.9b00538
J. Chem. Inf. Model. 2019, 59, 5026−5033

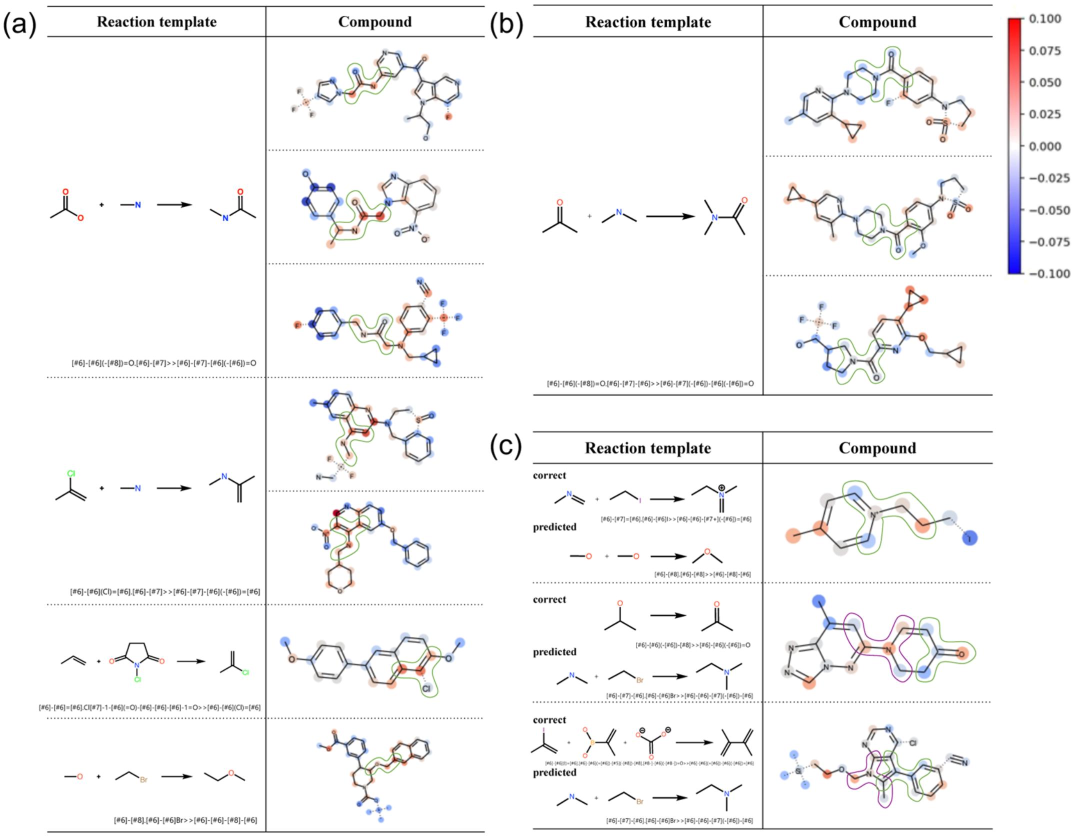

(a)
Reaction template
Compound
[#6]-[#6](-[#8])=O.[#6]-[#7]>>[#6]-[#7]-[#6](-[#6])=O
[#6]-[#6](Cl)=[#6].[#6]-[#7]>>[#6]-[#7]-[#6](-[#6])=[#6]
[#6]-[#6]=[#6].Cl[#7]-1-[#6](=O)-[#6]-[#6]-[#6]-1=O>>[#6]-[#6](Cl)=[#6]
[#6]-[#8].[#6]-[#6]Br>>[#6]-[#6]-[#8]-[#6]
(b)
Reaction template
Compound
[#6]-[#6](-[#8])=O.[#6]-[#7]-[#6]>>[#6]-[#7](-[#6])-[#6](-[#6])=O
0.100
0.075
0.050
0.025
0.000
−0.025
−0.050
−0.075
−0.100
(c)
Reaction template
Compound
correct
[#6]-[#7]=[#6].[#6]-[#6]I>>[#6]-[#6]-[#7+](-[#6])=[#6]
predicted
[#6]-[#8].[#6]-[#8]>>[#6]-[#8]-[#6]
correct
[#6]-[#6](-[#6])-[#8]>>[#6]-[#6](-[#6])=O
predicted
[#6]-[#7]-[#6].[#6]-[#6]Br>>[#6]-[#6]-[#7](-[#6])-[#6]
correct
predicted
[#6]-[#7]-[#6].[#6]-[#6]Br>>[#6]-[#6]-[#7](-[#6])-[#6]

Figure 4. Visualization of contributions of atom features in a molecule to retrosynthetic reaction prediction. The light-green colored part of atoms in a molecule corresponds to the reaction center in the reaction template. The color bar represents the value of IGs. The reaction templates in a SMARTS format under the drawn reaction templates. (a) Examples of the correct predictions that show the atom contributions and the reaction center do match. (b) Examples of the correct predictions that show the atom contributions and the reaction center do not match. (c) Examples of incorrect predictions. The light-green and light-purple colored parts of atoms represent the correct and predicted reaction centers, respectively.

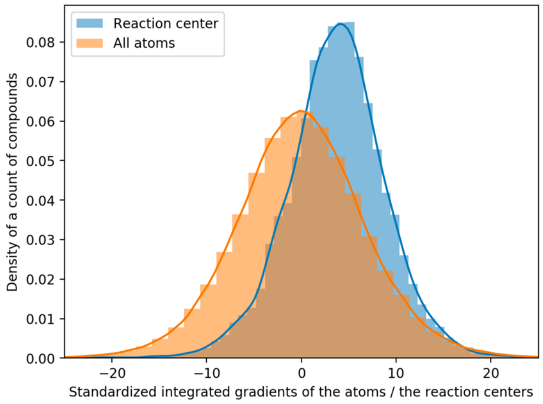


Figure 5. Histogram of standardized averages of IGs in reaction centers (blue) and the standardized IGs of all the atoms (orange) in the molecules.

considered to be a basic method to estimate the reason for each step of the proposed synthetic route. Even if the contributions of the prediction using ECFP can be visualized, we can only visualize the contributions by the substructure unit. When the contributions are visualized by the substructure unit, we cannot discuss the influences of neighboring atoms on the reaction center because the substructures in the fingerprint are not related to each other. Conversely, we can visualize the contributions by an atom unit using our model, considering the influences of the neighboring atoms on the reaction center. We believe that this improvement of interpretability would be essential to make data-driven approaches more accessible to chemists.

Figure 4 suggests that the common substructure which exists in molecules in the same reaction template contributed positively to the retrosynthetic reaction prediction. If various molecules are included in the same reaction template, as shown in Figure 4a, IGs are considered to be able to reflect the common reaction center for visualization. However, if similar molecules are included in the same reaction template, the GCN model tends to predict a reaction template by recognizing a characteristic substructure (e.g., cyclopropyl group in Figure 4b) other than the reaction center, as shown in Figure 4b. Therefore, using larger-scale chemical reaction databases such as Reaxys and SciFinder is considered as one solution to the above problem. Larger databases would ensure the diversity of molecules in the same reaction template and the GCN model is considered to predict the correct class with recognizing a common reaction center.

To confirm the performance of the GCN model for natural products, we performed the retrosynthetic reaction prediction to four natural products with different structural complexities: benzylpenicillin, erythromycin A, morphine, and prostaglandin E1 (Supporting Information Figure S4). The prediction for benzylpenicillin is thought to be reasonable. However, the other predicted results are considered to be unreasonable. The reason why it cannot be predicted is that the model could not

DOI: 10.1021/acs.jcim.9b00538

learn important features well for natural products because the USPTO reaction dataset has a few natural products.

For future work, we will focus on improving our model performance by the following three points. The first one is oversampling the lower reaction templates. The second one is setting a postfiltering parameter with IGs to rerank the predicted reaction templates. The last one is developing a molecular representation that considers precise local charge and chemical structural information, such as rotamers, topoisomers, and steric hindrance. These methods are expected to improve the top-*n* balanced accuracy, and the improved model will be more suitable for the chemist-friendly retrosynthetic analysis. We also plan to compare the improved GCN model with other advanced DL approaches, including transformer models.[23,24] We are going to address the prediction tasks of reaction conditions, yield, and multistep routes using our system.

CONCLUSIONS

We succeeded in the development of the GCN-based interpretable retrosynthetic reaction prediction system using IG. The prediction performance of our GCN-based model was compared with that of the traditional ECFP model. From the results, the prediction accuracy of the GCN model was higher than that of the ECFP model, and the GCN prediction was less influenced by the dataset bias. Additionally, the visualization of the GCN prediction using IG successfully showed the atom's contributions to a retrosynthetic reaction prediction. Through the visualization of the contributions, we can estimate the reasons for the retrosynthetic reaction prediction, which are expected to help chemists' understanding of a retrosynthetic reaction prediction based on a data-driven approach. Our model is expected to be a cornerstone for constructing a high-quality model for a retrosynthetic reaction prediction and important for searching retrosynthetic routes.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.9b00538.

Supporting information (PDF)

Examples of the visualization in top-100 and bottom-100 prediction probabilities (ZIP)

AUTHOR INFORMATION

Corresponding Author

*E-mail: okuno.yasushi.4c@kyoto-u.ac.jp.

ORCID

Shoichi Ishida: 0000-0002-5638-3579

Kei Terayama: 0000-0003-3914-248X

Kiyosei Takasu: 0000-0002-1798-7919

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Nobuo Cho for helpful discussions from a viewpoint of synthetic organic chemistry. This paper is based on a part of results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This article is also supported by MEXT, as "Priority Issue 1 on Post-K computer" (Building Innovative Drug Discovery Infrastructure through Functional Control of Biomolecular Systems).

REFERENCES

(1) Robinson, R. LXIII.—A Synthesis of Tropinone. *J. Chem. Soc. Trans.* **1917**, *111*, 762−768.

(2) Corey, E. J.; Cheng, X. M. *The Logic of Chemical Synthesis*; Wiley: New York, 1995.

(3) Corey, E. J.; Cramer, R. D.; Howe, W. J. Computer-assisted Synthetic Analysis for Complex Molecules. Methods and Procedures for Machine Generation of Synthetic Intermediates. *J. Am. Chem. Soc.* **1972**, *94*, 440−459.

(4) Wipke, W. T.; Ouchi, G. I.; Krishnan, S. Simulation and Evaluation of Chemical Synthesis-SECS: An Application of Artificial Intelligence Techniques. *Artif. Intell.* **1978**, *11*, 173−193.

(5) Funatsu, K.; Sasaki, S.-I. Computer-assisted Organic Synthesis Design and Reaction Prediction System, "AIPHOS". *Tetrahedron Comput. Methodol.* **1988**, *1*, 27−37.

(6) Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem., Int. Ed.* **2016**, *55*, 5904−5937.

(7) Judson, P.; Hirst, J.; Lim, C.; Jordan, K. D.; Thiel, W. Knowledge-Based Expert Systems in Chemistry. *Theoretical and Computational Chemistry Series*; The Royal Society of Chemistry: Cambridge, 2009.

(8) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature* **2018**, *555*, 604−610.

(9) Coley, C. W.; Green, W. H.; Jensen, K. F. Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **2018**, *51*, 1281−1289.

(10) Schreck, J. S.; Coley, C. W.; Bishop, K. J. M. Learning Retrosynthetic Planning through Simulated Experience. *ACS Cent. Sci.* **2019**, *5*, 970−981.

(11) Feng, F.; Lai, L.; Pei, J. Computational Chemical Synthesis Analysis and Pathway Design. *Front. Chem.* **2018**, *6*, 199.

(12) Klucznik, T.; Mikulak-Klucznik, B.; McCormack, M. P.; Lima, H.; Szymkuć, S.; Bhowmick, M.; Molga, K.; Zhou, Y.; Rickershauser, L.; Gajewska, E. P.; Toutchkine, A.; Dittwald, P.; Startek, M. P.; Kirkovits, G. J.; Roszak, R.; Adamski, A.; Sieredzińska, B.; Mrksich, M.; Trice, S. L. J.; Grzybowski, B. A. Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. *Chem* **2018**, *4*, 522−532.

(13) Litjens, G.; Kooi, T.; Bejnordi, B. E.; Setio, A. A. A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J. A. W. M.; van Ginneken, B.; Sánchez, C. I. A Survey on Deep Learning in Medical Image Analysis. *Med. Image Anal.* **2017**, *42*, 60−88.

(14) Ching, T.; Himmelstein, D. S.; Beaulieu-Jones, B. K.; Kalinin, A. A.; Do, B. T.; Way, G. P.; Ferrero, E.; Agapow, P.-M.; Zietz, M.; Hoffman, M. M.; Xie, W.; Rosen, G. L.; Lengerich, B. J.; Israeli, J.; Lanchantin, J.; Woloszynek, S.; Carpenter, A. E.; Shrikumar, A.; Xu, J.; Cofer, E. M.; Lavender, C. A.; Turaga, S. C.; Alexandari, A. M.; Lu, Z.; Harris, D. J.; DeCaprio, D.; Qi, Y.; Kundaje, A.; Peng, Y.; Wiley, L. K.; Segler, M. H. S.; Boca, S. M.; Swamidass, S. J.; Huang, A.; Gitter, A.; Greene, C. S. Opportunities and Obstacles for Deep Learning in Biology and Medicine. *J. R. Soc. Interface* **2018**, *15*, 20170387.

(15) Wu, Y.; Schuster, M.; Chen, Z.; Le, Q. V.; Norouzi, M.; Macherey, W.; Krikun, M.; Cao, Y.; Gao, Q.; Dean, J. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. **2016**, arXiv:abs/1609.08144. Computing Research Repository (CoRR).

(16) Kamilaris, A.; Prenafeta-Boldú, F. X. Deep Learning in Agriculture: A Survey. *Comput. Electron. Agric.* **2018**, *147*, 70−90.

(17) Segler, M. H. S.; Waller, M. P. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chem.—Eur. J.* **2017**, *23*, 5966−5971.

(18) Coley, C. W.; Barzilay, R.; Jaakkola, T. S.; Green, W. H.; Jensen, K. F. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **2017**, *3*, 434−443.

(19) Coley, C. W.; Rogers, L.; Green, W. H.; Jensen, K. F. Computer-Assisted Retrosynthesis Based on Molecular Similarity. *ACS Cent. Sci.* **2017**, *3*, 1237−1245.

(20) Liu, B.; Ramsundar, B.; Kawthekar, P.; Shi, J.; Gomes, J.; Luu Nguyen, Q.; Ho, S.; Sloane, J.; Wender, P.; Pande, V. Retrosynthetic Reaction Prediction using Neural Sequence-to-Sequence Models. *ACS Cent. Sci.* **2017**, *3*, 1103−1113.

(21) Coley, C. W.; Jin, W.; Rogers, L.; Jamison, T. F.; Jaakkola, T. S.; Green, W. H.; Barzilay, R.; Jensen, K. F. A Graph-convolutional Neural Network Model for the Prediction of Chemical Reactivity. *Chem. Sci.* **2018**, *10*, 370−377.

(22) Lin, K.; Xu, Y.; Pei, J.; Lai, L. Automatic Retrosynthetic Pathway Planning using Template-free Models. **2019**, arXiv:abs/1906.02308.

(23) Karpov, P.; Godin, G.; Tetko, I. A Transformer Model for Retrosynthesis. **2019**, ChemRxiv, 10.26434/chemrxiv.8058464.v1.

(24) Zheng, S.; Rao, J.; Zhang, Z.; Xu, J.; Yang, Y. Predicting Retrosynthetic Reaction using Self-Corrected Transformer Neural Networks. **2019**, arXiv:abs/1907.01356.

(25) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.

(26) Hu, Y.; Lounkine, E.; Bajorath, J. Improving the Search Performance of Extended Connectivity Fingerprints through Activity-Oriented Feature Filtering and Application of a Bit-Density-Dependent Similarity Function. *ChemMedChem* **2009**, *4*, 540−548.

(27) Baylon, J. L.; Cilfone, N. A.; Gulcher, J. R.; Chittenden, T. W. Enhancing Retrosynthetic Reaction Prediction with Deep Learning using Multiscale Reaction Classification. *J. Chem. Inf. Model.* **2019**, *59*, 673−688.

(28) Ribeiro, M. T.; Singh, S.; Guestrin, C. Why Should I Trust You?. *Explaining the Predictions of Any Classifier*; Knowledge Discovery and Data Mining (KDD), 2016; pp 1135−1144.

(29) Kipf, T. N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *International Conference on Learning Representations (ICLR)*, 2017.

(30) Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic Attribution for Deep Networks. *International Conference on Machine Learning (ICML)*, 2017.

(31) Altae-Tran, H.; Ramsundar, B.; Pappu, A. S.; Pande, V. Low Data Drug Discovery with One-Shot Learning. *ACS Cent. Sci.* **2017**, *3*, 283−293.

(32) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: A Benchmark for Molecular Machine Learning. *Chem. Sci.* **2018**, *9*, 513−530.

(33) Zeiler, M. D.; Fergus, R. Visualizing and Understanding Convolutional Networks. *European Conference on Computer Vision (ECCV)*, 2014; pp 818−833.

(34) Simonyan, K.; Vedaldi, A.; Zisserman, A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. **2013**, arXiv:abs/1312.6034. Computing Research Repository (CoRR).

(35) Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *IEEE International Conference on Computer Vision (ICCV)*, 2017; pp 618−626.

(36) Lowe, D. Chemical Reactions from US Patents (1976−Sep 2016), 2017, https://doi.org/10.6084/m9.figshare.5104873.v1, accessed Aug 21, 2019.

(37) Avramova, S.; Kochev, N.; Angelov, P. RetroTransformDB: A Dataset of Generic Transforms for Retrosynthetic Analysis. *Data* **2018**, *3*, 14.

(38) ChemAxon. 2018, http://www.chemaxon.com (accessed Aug 21, 2019).

(39) Ramsundar, B.; Eastman, P.; Walters, P.; Pande, V.; Leswing, K.; Wu, Z. *Deep Learning for the Life Sciences*; O'Reilly Media: Sebastopol, 2019.

(40) Landrum, G. RDKit: Open-source Cheminformatics. 2018, http://www.rdkit.org (accessed Aug 21, 2019).

(41) Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *International Conference on Machine Learning (ICML)*, 2015; pp 448−456.

(42) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Zheng, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015, https://www.tensorflow.org/ (accessed Aug 21, 2019).

(43) Chollet, F. Keras, https://keras.io, 2015.

(44) Ishiguro, K.; Maeda, S. i.; Koyama, M. Graph Warp Module: an Auxiliary Module for Boosting the Power of Graph Neural Networks. **2019**, arXiv:abs/1902.01020.