

A Unified View of Deep Learning for Reaction and Retrosynthesis Prediction: Current Status and Future Challenges

Ziqiao Meng^{1*}, Peilin Zhao^{2†}, Yang Yu² and Irwin King^{1†}

¹The Chinese University of Hong Kong

²Tencent AI Lab

{zqmeng, king}@cse.cuhk.edu.hk, masonzhao@tencent.com

Abstract

Reaction and retrosynthesis prediction are fundamental tasks in computational chemistry that have recently garnered attention from both the machine learning and drug discovery communities. Various deep learning approaches have been proposed to tackle these problems, and some have achieved initial success. In this survey, we conduct a comprehensive investigation of advanced deep learning-based models for reaction and retrosynthesis prediction. We summarize the design mechanisms, strengths, and weaknesses of state-of-the-art approaches. Then, we discuss the limitations of current solutions and open challenges in the problem itself. Finally, we present promising directions to facilitate future research. To our knowledge, this paper is the first comprehensive and systematic survey that seeks to provide a unified understanding of reaction and retrosynthesis prediction.

1 Introduction

Drug discovery is crucial to human healthcare, but the process is notoriously labor-intensive and costly. As Eroom's law suggests [Scannell *et al.*, 2012], the exploration of new drugs becomes increasingly slower and more expensive over time. Therefore, it is natural and significant to leverage machine learning techniques to accelerate the drug discovery process. In recent years, the use of deep learning approaches to enhance different stages of drug discovery has become prevalent due to the rise of deep learning. Among these stages, reaction prediction and retrosynthesis prediction are two fundamental steps that can benefit from deep learning tools.

In actual production environments, chemists often aim to design synthesis routes that can lead to target molecules through a series of chemical reactions. One common strategy is to decompose target molecules into simpler precursor structures that can be synthesized more easily, a process known as retrosynthetic analysis. Automating this planning

*This work is done when Ziqiao Meng worked as an intern in Tencent AI Lab.

†Corresponding authors: Peilin Zhao, and Irwin King.

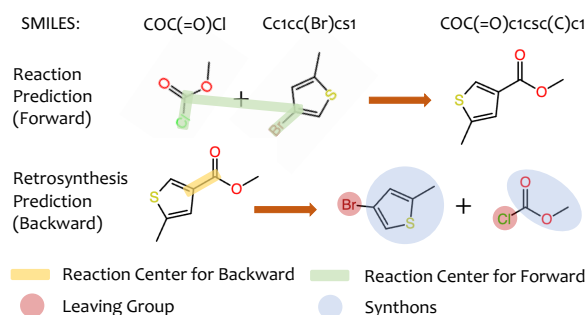


Figure 1: This figure illustrates the problem formulation of reaction and retrosynthesis prediction

process using deep learning is crucial for discovering and optimizing synthesis routes. Retrosynthetic planning involves two subtasks: multi-step retrosynthetic planning and single-step retrosynthesis prediction. In this survey, we focus on the latter, as multi-step planning is usually approached as a searching problem, which is fundamentally different from the conditional structured prediction problem posed by reaction prediction and single-step retrosynthesis prediction. For the remainder of this survey, we will refer to single-step retrosynthesis prediction as simply retrosynthesis prediction. Reaction prediction is another key task in organic synthesis analysis. A robust reaction prediction model can provide insight into the underlying mechanisms of biochemical reactions, as well as generate virtual reactions to expand the database for retrosynthetic planning. In summary, reaction and retrosynthesis prediction are interrelated and can enhance one another.

Several surveys exist on reaction and retrosynthesis prediction. Engkvist [Engkvist *et al.*, 2018] provides an overview of various computational approaches for reaction prediction, ranging from quantum computation to deep learning tools. However, this survey lacks detail for each approach and provides limited coverage of deep learning-based methods. On the other hand, [Dong *et al.*, 2021] reviews some deep learning-based retrosynthetic planning methods and datasets, but it falls behind the state-of-the-art research on retrosynthesis prediction. It is worth noting that DualTF [Sun *et al.*, 2021] provides a unified framework for retrosynthesis prediction from the perspective of energy-based models, but it does not offer a unified understanding of reaction and retrosynthe-

sis, nor does it provide specific discussions on the limitations and major challenges of each approach. Overall, the existing literature lacks a comprehensive and unified understanding of advanced reaction and retrosynthesis prediction models.

Compared to previously mentioned surveys and works, our survey unifies the formulation of reaction and retrosynthesis prediction for the first time. We systematically discuss the strengths and weaknesses of each approach from different perspectives for both problems. Additionally, our survey presents novel challenges and limitations that were not explicitly stated in previous works. Finally, based on the current status, we list several future directions for further enhancement, along with detailed analysis.

2 Preliminary and Problem Formulation

Reaction prediction and retrosynthesis prediction are dual tasks of each other. They are also known as forward reaction prediction and backward reaction prediction, respectively. To apply deep learning, both tasks are formulated as conditional generation tasks. In this section, we first introduce the problem formulation for both tasks. Then, we provide an introduction to the basic background knowledge required for understanding different methods.

Molecular Formulation. A chemical molecule \mathcal{M} can be represented in two major data formats: the *SMILES string* and the *molecular graph*. (1) For the SMILES format, a molecular structure \mathcal{M} is described as a sequence of characters m_i such that $\mathcal{M} := m_1 m_2 \dots m_L$, where L denotes the total length of the string. The sequence represents a spanning tree of the 2D molecular structure, and each character m_i denotes a structural element such as an atom element, chemical bond, branching notation, and so on. (2) A molecule can also be abstracted as an undirected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} = \{v_1, \dots, v_n\}$ denotes the set of n atoms and $\mathcal{E} = \{e_1, \dots, e_m\}$ denotes the set of m edges. Each node is associated with a feature vector $\mathbf{h}_i \in \mathbb{R}^d$ containing atomic information such as aromaticity and electric charge. Then, we have a feature matrix $\mathbf{H} \in \mathbb{R}^{n \times d}$ containing all-atom information. An adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n \times c}$ describes the topological structures of \mathcal{M} , where \mathbf{A}_{ijk} indicating the presence or absence of a chemical bond of type k between atom i and atom j . Multiple molecules can be easily represented by the above two formats. For SMILES format, multiple SMILES strings can be concatenated by a full stop “.” into one single SMILES sequence. For molecular graph, a set of molecules is regarded as a single disconnected graph with each molecule as an independent connected component.

Definition 1 (Reaction Prediction) Given a set of N reactant molecules $\{\mathcal{M}_i^R\}_{i=1}^N$, the target is to predict the set of M possible product molecules $\{\mathcal{M}_i^P\}_{i=1}^M$.

Definition 2 (Retrosynthesis Prediction) Given a set of M product molecules $\{\mathcal{M}_i^P\}_{i=1}^M$, the target is to predict a set of N reactant molecules $\{\mathcal{M}_i^R\}_{i=1}^N$ that can lead to $\{\mathcal{M}_i^P\}_{i=1}^M$.

Note that in real implementations, $M = 1$ because only the main product is recorded in public benchmark datasets. Unfortunately, this issue makes retrosynthesis prediction more difficult than reaction prediction because $M < N$ requires

retrosynthesis to attach newly appeared atoms, leading to a much larger combinatorial search space. In general, reaction prediction aims to model the conditional probability distribution $\mathcal{P}(\mathcal{M}^P | \{\mathcal{M}_i^R\}_{i=1}^N)$, while retrosynthesis prediction aims to model the distribution $\mathcal{P}(\{\mathcal{M}_i^R\}_{i=1}^N | \mathcal{M}^P)$.

Reaction Center & Reaction Template. In reaction prediction, a reaction center \mathcal{C} is a subset of atom pairs $\mathcal{C} = \{(v_i, v_j)\} \subseteq \mathcal{V} \times \mathcal{V}$ that change bond types when a chemical reaction occurs. In retrosynthesis prediction, a reaction center \mathcal{C} is defined as a subset of existing bonds $\mathcal{C} = \{e_i\} \subseteq \mathcal{E}$ that can be modified to obtain simpler structures. A reaction template pool \mathcal{T} is a set of reaction subgraph rules derived from a large chemical reaction database. A reaction template $T \in \mathcal{T}$ is an extracted subgraph pattern from the corresponding reaction center \mathcal{C} . Specifically, $T := t_1^R + t_2^R + \dots t_N^R \rightarrow t^P$, where t_i^R denotes the subgraph pattern inside the i th reactant, and t^P denotes the subgraph pattern inside the product.

Atom-mapping. Both reaction and retrosynthesis prediction follow the atom-mapping principle. This principle states that each atom in the reactants/products has exactly one corresponding atom in the products/reactants. This fundamental one-to-one mapping relation physically constrains reaction space and determines that chemical reactions are mainly about bond breaking and bond formation.

Synthon and Leaving Group. A target molecule can be decomposed into a set of synthons $\mathcal{S} = \{\mathcal{G}_i^S\}_{i=1}^N$, which are simpler precursor substructures that can constitute the target molecule with few additional bond connections. Note that the atom set \mathcal{V} covered by \mathcal{S} is exactly the same as the target molecule. Leaving group $\mathcal{L} = \{\mathcal{G}_i^L\}_{i=1}^N$ is a group of atoms or substructures in original reactants that do not appear in the target molecule after a reaction occurs. In short, synthon \mathcal{S} and leaving group \mathcal{L} can form reactants $\{\mathcal{G}_i^R\}_{i=1}^N$.

Evaluation Metrics. Both reaction and retrosynthesis prediction adopt top- k accuracies to evaluate the model performance. Top- k accuracy is the percentage of reactions that have the ground-truth product in the top- k predicted sets of molecules. As long as the top- k predicted products include the ground-truth main product, it would be counted as a correct prediction. Usually, the k is in range $\{1, 2, 3, 5, 10\}$.

3 Deep Learning on Reaction and Retrosynthesis Prediction

In this section, we discuss different approaches by classifying them into four categories which are template-based methods, sequence-based and graph-based autoregressive models, graph-based two-stage models, and graph-based non-autoregressive models. Specifically, we introduce their learning mechanism design decisions, weaknesses, and strengths.

3.1 Template-based Methods

Template-based (TB) methods are mainly leveraging a pool of reaction templates to deduce possible reaction centers. TB methods are mimicking how human experts conduct chemical reasoning. Assume we have a pool of reaction templates $\mathcal{T} =$

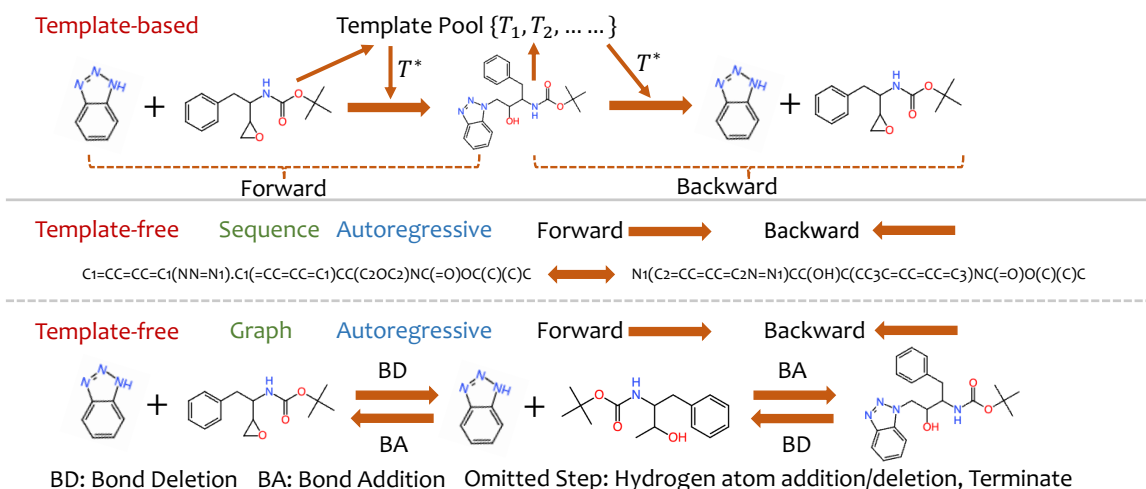


Figure 2: This figure illustrates template-based models, sequence-based autoregressive models, and graph-based autoregressive models.

$\{T_1, T_2, \dots\}$, then the inference process of TB methods aims at picking the best template as follows:

$$T^* = \arg \max_{T_i \in \mathcal{T}} \mathcal{P}(T = T_i | \{\mathcal{M}_i^R\}_{i=1}^N),$$

$$\hat{\mathcal{M}}^P = \arg \max_{\mathcal{M}_i^P \in \mathcal{M}^P} \mathcal{P}(\mathcal{M}^P = \mathcal{M}_i^P | T^*), \quad (1)$$

where \mathcal{M}^P is a pool of candidate product molecules. Eq. (1) describes the steps of choosing the best template T^* and deducing the possible product $\hat{\mathcal{M}}^P$ according to the given T^* . The critical part is matching molecules with templates based on some similarity measure. Specifically, $\mathcal{P}(T_i | \{\mathcal{M}_i^R\}_{i=1}^N)$ and $\mathcal{P}(\mathcal{M}_i^P | T^*)$ are evaluated in the following way:

$$\mathcal{P}(T_i | \{\mathcal{M}_i^R\}_{i=1}^N) = \frac{\exp(\text{sim}(T_i, \{\mathcal{M}_i^R\}_{i=1}^N))}{\sum_{T_j \in \mathcal{T}} \exp(\text{sim}(T_j, \{\mathcal{M}_i^R\}_{i=1}^N))},$$

$$\mathcal{P}(\mathcal{M}_i^P | T^*) = \frac{\exp(\text{sim}(T^*, \mathcal{M}_i^P))}{\sum_{\mathcal{M}_j^P \in \mathcal{M}^P} \exp(\text{sim}(T^*, \mathcal{M}_j^P))}, \quad (2)$$

where $\text{sim}(\cdot, \cdot)$ denotes a similarity measure function. It can be a simple inner product between molecular embedding and template embedding obtained from deep neural networks, or a more complicated similarity function evaluated by some sub-graph matching algorithms. TB retrosynthesis has the same learning mechanism but in a reverse direction:

$$T^* = \arg \max_{T_i \in \mathcal{T}} \mathcal{P}(T = T_i | \mathcal{M}^P),$$

$$\mathcal{P}(\mathcal{M}^R | T^*) = \prod_{i=1}^N \mathcal{P}(\mathcal{M}_i^R | t_i^R), \quad (3)$$

$$\hat{\mathcal{M}}_i^R = \arg \max_{\mathcal{M}_j^R \in \mathcal{M}^R} \mathcal{P}(\mathcal{M}^R = \mathcal{M}_j^R | t_i^R),$$

where \mathcal{M}^R denotes the pool of candidate reactant molecules, $T^* := t_1^R + t_2^R + \dots + t_N^R \rightarrow t^P$. The second and third equation in Eq. (3) indicates that each predicted reactant molecule $\hat{\mathcal{M}}_i^R$ matches with the subgraph template t_i^R with highest score.

Advantages. (1) TB methods are reliable since they are using extracted human knowledge, which can always provide good interpretations for predictions. (2) The training and inference process of TB methods are relatively simple, which is easy for domain experts to manipulate.

Disadvantages. (1) The performance of TB methods highly relies on the scale of template database. Therefore, the template database must be updated frequently, which is apparently very expensive. (2) TB methods have poor generalization to out-of-domain unseen reactions. (3) Templates are extracted local subgraph rules while ignoring global information over the reaction. Consequently, TB methods fail to capture global information interactions and easily make false predictions based on local rules. The illustration of TB methods is shown in Figure 2.

3.2 Sequence-based and Graph-based Autoregressive Models

Sequence-based autoregressive (SAR) models are widely adopted in both forward and backward prediction. It regards both problems as the neural machine translation problem. For reaction prediction, the input is SMILES strings of reactants $\mathcal{M}^R := m_1^R m_2^R \dots m_{L_1}^R$ with length L_1 and output is SMILES strings of products $\mathcal{M}^P := m_1^P m_2^P \dots m_{L_2}^P$ with length L_2 . The input source and output target are flipped for retrosynthesis prediction. Specifically, SAR models are estimating the following conditional probability distribution:

$$\mathcal{P}(\mathcal{M}^P) = \prod_{i=1}^{L_2} \mathcal{P}(m_i^P | m_{<i}^P, \mathcal{M}^R),$$

$$\mathcal{P}(\mathcal{M}^R) = \prod_{i=1}^{L_1} \mathcal{P}(m_i^R | m_{<i}^R, \mathcal{M}^P), \quad (4)$$

where $\mathcal{P}(m_i^P | m_{<i}^P, \mathcal{M}^R)$ and $\mathcal{P}(m_i^R | m_{<i}^R, \mathcal{M}^P)$ are approximated by the transformer [Vaswani *et al.*, 2017] model. Each generation step does a greedy search over token space and select the optimal token. To generate top- k candidates, we only

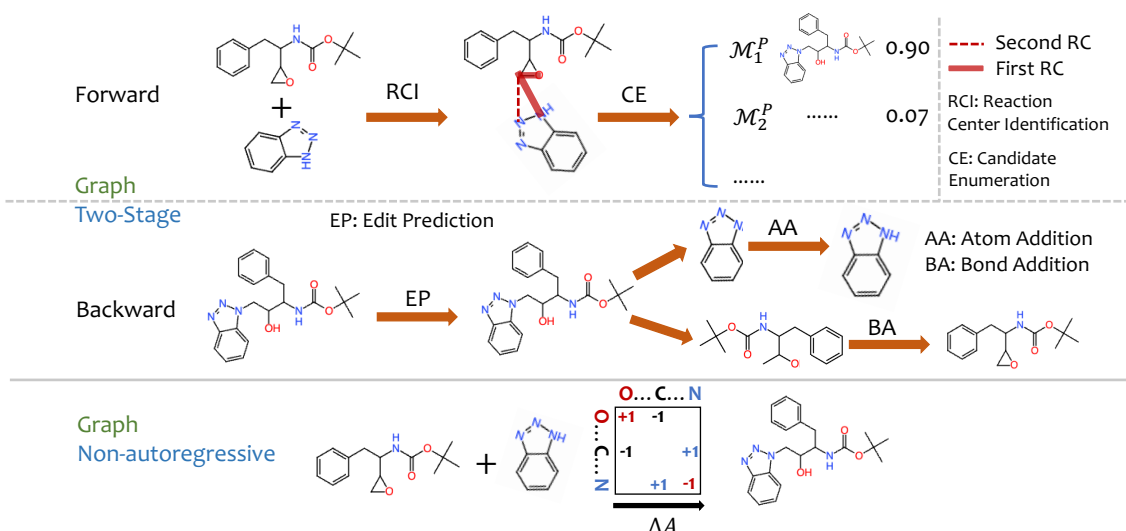


Figure 3: This figure illustrates graph-based two-stage models and graph-based non-autoregressive models.

need to conduct beam search over greedy results. Graph-based autoregressive (GAR) models have similar learning mechanisms but with different generation sequence definitions. GAR models first define an action space π including several edit actions like atom addition/deletion, bond addition/deletion, and termination. A sequence of these actions would transform reactants/products into their corresponding products/reactants. Thus, they are estimating the following probability distribution:

$$\mathcal{P}(\mathcal{G}^P) = \prod_{i=1}^{L_2} \mathcal{P}(\pi_i | \pi_{<i}, \mathcal{G}^{R,<i}), \quad (5)$$

$$\mathcal{P}(\mathcal{G}^R) = \prod_{i=1}^{L_1} \mathcal{P}(\pi_i | \pi_{<i}, \mathcal{G}^{P,<i}),$$

where $\mathcal{G}^{R,<i}$ and $\mathcal{G}^{P,<i}$ denote the states of edited reactants and products in previous $i - 1$ steps respectively. The sampling process is exactly the same as that of SAR models. Each step i selects the optimal action π_i^* and applies the optimal action to transform reactants $\mathcal{G}^{R,i-1}$ and products $\mathcal{G}^{P,i-1}$ to next state $\mathcal{G}^{R,i}$ and $\mathcal{G}^{P,i}$ respectively. The optimal action would be “terminate” when all generation steps are predicted to be over. Note that both Eq. (4) and Eq. (5) can be simplified by Markov assumptions.

Advantages. (1) Autoregressive models do not need atom-mapping information. (2) Autoregressive modeling has very natural sampling process, such as beam search. (3) Sequence-based modeling can directly utilize some mature techniques from natural language processing.

Disadvantages. (1) Autoregressive modeling can only generate predictions step-by-step, which is very inefficient. (2) Autoregressive modeling requires pre-defined generation orders. However, the molecular generation order is ambiguous. (3) Sequence-based modeling requires data augmentation techniques to improve performance. The illustration of both SAR models and GAR models are shown in Figure 2.

3.3 Graph-based Two-Stage Models

Reaction. For graph-based two-stage models in reaction prediction, they split reaction prediction into two stages which are reaction center identification stage and candidate ranking stage. For reaction center identification, it aims at selecting atom pairs assigned with high reactivity scores:

$$\hat{\mathcal{C}} = \text{top-}k(\{s(v_i, v_j | \{\mathcal{G}_i^R\}_{i=1}^N)\}_{ij}), \quad (6)$$

where $s(\cdot)$ denotes any function that outputs a score in range $(0, 1)$. Based on identified reaction centers $\hat{\mathcal{C}}$, a set of candidate products $\hat{\mathcal{G}}^P = \{\hat{\mathcal{G}}_1^P, \hat{\mathcal{G}}_2^P, \dots\}$ will be enumerated in a combinatorial way through hand-written rules or reaction templates. Then the second stage is learning to rank candidate products in $\hat{\mathcal{G}}^P$. To rank generated products, each pair $(\{\mathcal{G}_i^R\}_{i=1}^N, \hat{\mathcal{G}}_i^P)$ should have a score evaluated as follows:

$$s(\{\mathcal{G}_i^R\}_{i=1}^N, \hat{\mathcal{G}}_i^P) = \sigma(f(\{\mathcal{G}_i^R\}_{i=1}^N, \hat{\mathcal{G}}_i^P)), \quad (7)$$

where $f(\cdot, \cdot)$ can be complex neural networks and $\sigma(\cdot)$ denotes the sigmoid function.

Retrosynthesis. For graph-based two-stage models in retrosynthesis prediction, they split retrosynthesis into two stages which are edit prediction stage and synthon completion stage. For edit prediction, they select some existing edges with high predicted reactivity scores to decompose:

$$\hat{\mathcal{C}} = \text{top-}k(\{s(e_i^P)\}_{i=1}^m), \quad (8)$$

After obtaining predicted edit centers $\hat{\mathcal{C}}$, we decompose the target molecule by breaking the predicted edit centers, which will result in a set of synthons $\hat{\mathcal{S}} = \{\hat{\mathcal{G}}_i^S\}_{i=1}^N$, ($N = k + 1$). A basic assumption for this approach is that the number of synthons is the same as the number of reactants. For synthon completion stage, it models the following distribution:

$$\mathcal{P}(\mathcal{G}_i^R | \hat{\mathcal{G}}_i^S) = \mathcal{P}(\mathcal{G}_i^L | \hat{\mathcal{G}}_i^S). \quad (9)$$

The above Eq. (9) indicates that the synthon completion is equivalent to attaching the leaving group \mathcal{G}_i^L to the corresponding synthon \mathcal{G}_i^S . Leaving group attachment can be either formulated as an autoregressive conditional generation problem or a classification problem with a predefined dictionary. With predicted leaving group $\hat{\mathcal{G}}_i^L$, each predicted reactant $\hat{\mathcal{G}}_i^R$ can be recovered by attaching $\hat{\mathcal{G}}_i^L$ to the corresponding synthon $\hat{\mathcal{G}}_i^S$.

Advantages. (1) The inference process of graph-based two-stage models is very similar to the deduction process of chemists. (2) Graph-based two-stage models split a hard task into two simpler tasks that are easier to tackle.

Disadvantages. (1) Some two-stage models like WLDN [Jin *et al.*, 2017] require costly hand-crafted combinatorial enumerations. (2) The overall performance of two-stage models is restricted by the bottleneck of each stage. The illustration of the above methods are shown in Figure 3.

3.4 Graph-based Non-autoregressive Models

Only reaction prediction has a graph-based non-autoregressive model achieved by NERF [Bi *et al.*, 2021] and currently no non-autoregressive models for retrosynthesis. NERF reformulates the problem into electron redistribution modeling. Previous approaches adopt three dimensional adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n \times c}$ while NERF adopts two-dimensional adjacency matrix by converting one-hot type encoding to scalar values such that $\mathbf{A} \in \mathbb{R}^{n \times n}$. Formally, \mathbf{A}_{ij} is a scalar value in range $[0, 3]$, representing the number of shared electrons (number of bonds) between atom i and atom j . Note that aromatic bond is represented as $\mathbf{A}_{ij} = 1$ and atoms i, j will be marked as aromatic atoms. The key idea of NERF is predicting $\Delta \hat{\mathbf{A}}$ by combining self-attention mappings. NERF adopts the conditional variational autoencoder [Sohn *et al.*, 2015] (CVAE) architecture to approximate $\mathcal{P}(\mathcal{G}^P | \mathcal{G}^R)$ by introducing a latent variable z . Instead of directly maximizing the log-likelihood $\log \mathcal{P}(\mathcal{G}^P | \mathcal{G}^R)$, CVAE maximize its evidence-lower bound (ELBO):

$$\log \mathcal{P}(\mathcal{G}^P | \mathcal{G}^R) \geq \mathbb{E}_{q(z | \mathcal{G}^P, \mathcal{G}^R)} [\log \mathcal{P}(\mathcal{G}^P | \mathcal{G}^R, z)] - \text{KL}(q(z | \mathcal{G}^P, \mathcal{G}^R) || \mathcal{P}(z | \mathcal{G}^R)), \quad (10)$$

where $q(z | \mathcal{G}^P, \mathcal{G}^R)$ is the reaction encoder with reaction $(\mathcal{G}^R, \mathcal{G}^P)$ as input and low-dimensional representation \mathbf{h}^z as output, $\mathcal{P}(\mathcal{G}^P | \mathcal{G}^R, z)$ is product decoder with reactants \mathcal{G}^R and latent embedding \mathbf{h}^z as input, $\mathcal{P}(z | \mathcal{G}^R)$ denotes the prior distribution of latent variable z . KL term is minimizing the gap between $q(z | \mathcal{G}^P, \mathcal{G}^R)$ and $\mathcal{P}(z | \mathcal{G}^R)$. The backbone network architectures of $q(z | \mathcal{G}^P, \mathcal{G}^R)$ are combinations of GNNs and transformers. With this architecture, \mathcal{G}^R and \mathcal{G}^P will be projected to reactants embedding \mathbf{h}^R and products embedding \mathbf{h}^P respectively. A cross attention layer is used for mapping \mathbf{h}^R to latent \mathbf{h}^z with \mathbf{h}^P as teacher forcing during training. The conditional latent embedding is derived such that $\hat{\mathbf{h}}^z = \mathbf{h}^R + \mathbf{h}^z$. Then apply self-attention mechanism on $\hat{\mathbf{h}}^z$ to derive two electron redistribution matrix W^+ and W^- for bond increase and bond decrease respectively. Then $\Delta \hat{\mathbf{A}} = W^+ - W^-$ and $\hat{\mathbf{A}}^P = \mathbf{A}^R + \Delta \hat{\mathbf{A}}$.

Advantages. (1) Non-autoregressive models enable parallel sampling, which result in much faster sampling speed compared to autoregressive ones. (2) Non-autoregressive models have already achieved the best class of top-1 accuracy, which demonstrates that the non-autoregressive decoder is very powerful for reaction modeling. (3) Non-autoregressive models do not need pre-defined generation order.

Disadvantages. (1) The uncertainty modeling of non-autoregressive models is very tricky. The top-k sampling process is not as natural as beam search in autoregressive modeling. (2) Non-autoregressive models rely on the atom-mapping information while this also requires additional algorithms for alignment. The illustration of NERF learning mechanism is shown in Figure 3.

4 Limitations and Challenges

In this section, we discuss some important limitations and challenges existing in current solutions.

4.1 Side Products

Reaction. Side products are missing in public USPTO benchmark datasets, which results in incomplete supervision signals. Particularly for non-autoregressive modeling, missing side products makes electron redistribution matrix violate the conservation rule and thereby the reaction space is not fully constrained. How to complete and infer this missing information is a fundamental challenge for reaction prediction.

Retrosynthesis. Retrosynthesis directly use reaction data in USPTO and thereby all single-step analysis only contains one single outcome. Missing side products in the outcome side may not affect the general process of retrosynthesis prediction. However, side products still provide important information about leaving groups.

4.2 Limitations in Dataset

Reaction. The USPTO-479K dataset has two major issues. First, reaction types are very imbalanced. Indicated by Bi [Bi *et al.*, 2021], reactions with linear topology dominate reaction types while few reactions with cyclic topology exist in dataset. How to learn transferable knowledge from rare reactions is an important lesson for reaction modeling. Second, in real applications, same set of reactants can result in different products under different physical conditions, which is called multi-modality in reaction prediction. Multi-modality can provide rich information for conditional generative models to generate valid and diverse candidate products. However, most reactions in UPSTO-MIT are one-to-one mapping, which means that same set of reactants can only result in a unique major product.

Retrosynthesis. The USPTO-50K dataset has two major limitations. First, the size of USPTO-50K is not large enough, which only contains 50k backward reactions. Considering many recent approaches only have slight numerical differences in top-k accuracies, the current small-scale dataset is not adequate for testing model capability. Second, the current dataset will bias edit predictions and leaving group selections. Most backward reactions only have one single edit

	Template Usage	Approaches	End-to-End	Generation	Graph/Sequence	Top-1	Top-5
Reaction Prediction (Forward)	Template-based	NN-reaction [Wei <i>et al.</i> , 2016]	×	NA	Sequence	NR	NR
		NeuralSym [Segler and Waller, 2017]	×	NA	Sequence	NR	NR
		Symbolic [Qian <i>et al.</i> , 2020]	×	NA	Sequence	90.4	95.0
		LocalTransform [Chen and Jung, 2022]	×	NA	Graph	90.8	96.3
	Template-free	WLDN [Jin <i>et al.</i> , 2017]	×	NA	Graph	79.6	89.2
		GTPN [Do <i>et al.</i> , 2019]	✓	A	Graph	83.2	86.5
		MEGAN [Sacha <i>et al.</i> , 2020]	✓	A	Graph	89.3	95.6
		ELECTRO [Bradshaw <i>et al.</i> , 2019]	✓	A	Graph	77.8	94.7
		Motif-Reaction [Zhao <i>et al.</i> , 2022]	✓	A	Sequence	91.0	95.7
		MT-base [Schwaller <i>et al.</i> , 2019]	✓	A	Sequence	88.8	94.4
		MT [Schwaller <i>et al.</i> , 2019]	✓	A	Sequence	90.4	95.3
		Graph2SMILES [Tu and Coley, 2021]	✓	A	Sequence	90.3	94.8
		Chemformer [Irwin <i>et al.</i> , 2022]	✓	A	Sequence	91.3	93.7
		Stranformer [Lee <i>et al.</i> , 2019]	×	A	Graph	NR	NR
		ReactionT5 [Lu and Zhang, 2022]	✓	A	Sequence	88.9	95.2
		Aug-Transformer [Tetko <i>et al.</i> , 2020]	✓	A	Sequence	90.6	96.1
NERF [Bi <i>et al.</i> , 2021]	✓	NA	Graph	90.7	93.7		
ReactionSink [Meng <i>et al.</i> , 2023]	✓	NA	Graph	91.3	94.0		
Retrosynthesis Prediction (Backward)	Template-based	RetroSim [Coley <i>et al.</i> , 2017]	×	NA	Sequence	52.9/37.3	81.2/63.3
		RetroComposer [Yan <i>et al.</i> , 2022]	×	A	Graph	65.9/54.5	89.5/83.2
		NeuralSym [Segler and Waller, 2017]	×	NA	Graph	55.3/44.4	81.4/72.4
		GLN [Dai <i>et al.</i> , 2020]	×	NA	Graph	64.2/52.5	85.2/75.6
		LocalRetro [Chen and Jung, 2021]	×	A	Graph	63.9/53.4	92.4/85.9
		DualTB [Sun <i>et al.</i> , 2021]	×	NA	Sequence	67.7/55.2	88.9/80.5
	Template-free	MEGAN [Sacha <i>et al.</i> , 2020]	✓	A	Graph	60.7/48.1	87.5/78.4
		AutoSynRoute [Lin <i>et al.</i> , 2020]	✓	A	Sequence	54.6/43.1	80.2/71.8
		SCROP [Zheng <i>et al.</i> , 2020]	✓	A	Sequence	59.0/43.7	78.1/65.2
		LV-Transformer [Chen <i>et al.</i> , 2019]	✓	A	Sequence	NR/40.5	NR/72.8
		DualTF [Sun <i>et al.</i> , 2021]	✓	A	Sequence	65.7/53.6	84.7/74.6
		GET [Mao <i>et al.</i> , 2021]	✓	A	Graph	57.4/44.9	74.8/62.4
		Retroprime [Wang <i>et al.</i> , 2021]	✓	NA	Sequence	64.8/51.4	85.0/74.0
		GTA [Seo <i>et al.</i> , 2021]	×	A	Graph	NR/51.1	NR/74.8
		G2Gs [Shi <i>et al.</i> , 2020]	×	A	Graph	61.0/48.9	86.0/72.5
		RetroXPert [Yan <i>et al.</i> , 2020]	×	A	Graph	62.1/50.4	78.5/62.3
		Retroformer [Wan <i>et al.</i> , 2022]	✓	A	Sequence	64.0/53.2	86.7/76.6
		Tied-transformer [Kim <i>et al.</i> , 2021]	✓	A	Sequence	NR/47.1	NR/73.1
		RetroLSTM [Liu <i>et al.</i> , 2017]	✓	A	Sequence	NR/37.4	NR/57.0
		G2GT [Lin <i>et al.</i> , 2022]	✓	A	Graph	NR/54.1	NR/74.5
		Stranformer [Lee <i>et al.</i> , 2019]	✓	A	Graph	NR/43.8	NR/NR
		GraphRetro [Somnath <i>et al.</i> , 2021]	×	A	Graph	63.9/53.7	85.2/72.2
		MARS [Liu <i>et al.</i> , 2022]	✓	A	Sequence	66.2/54.6	90.2/83.3

Table 1: A table of recent approaches on forward and backward prediction. “End-to-end” indicates whether the model inference process is in an end-to-end manner. “Generation” indicates whether the predicted product is generated in an autoregressive or non-autoregressive manner. “Graph/Sequence” denotes the molecular representation format used in modeling. “NA” and “A” denote non-autoregressive generation and autoregressive generation respectively. We report reaction prediction results on USPTO-MIT dataset and retrosynthesis prediction results on USPTO-50K dataset. For retrosynthesis prediction model, we report both results with reaction class known and unknown. “NR” means not reported. Note that we do not put “Semi-Template-based” as an independent category.

while very few have multiple edits, which results in poor prediction accuracy in multiple-edit cases. Additionally, leaving group distribution is very imbalanced, which makes graph-based models tend to select a few frequently occurring atoms.

4.3 Limitations in Evaluation

Reaction. [Kovacs *et al.*, 2021] points out that the random split of USPTO-MIT is not sufficient enough for evaluations since most of the reaction types in the testing set are already covered in the training set. Therefore, harder dataset split like scaffold splits and time splits should be taken into account for future evaluations of reaction predictions. Scaffold split is testing whether the model can generalize well under out-of-distribution settings, in which the training data distribution is very different from testing data distribution. Time split is splitting reactions in the order of discovery time. This split aims at testing whether the model can truly discover new chemical reactions or not.

Retrosynthesis. Current evaluation metrics are not sufficient to evaluate retrosynthesis predictions since they require exact matching between predicted reactants and ground-truth

reactants. This will lead to low diversity of generations. Although some predicted reactants are distinct from given ground-truth reactants, they are acceptable solutions for synthesis choices. Since the single-step retrosynthesis model is finally served for multi-step retrosynthetic planning purposes, low diversity for each step would make the whole planning monotonous and inflexible.

4.4 Challenges in Non-autoregressive Modeling

Reaction. Non-autoregressive models [Bi *et al.*, 2021; Meng *et al.*, 2023] achieve the state-of-the-art top-1 accuracy among template-free approaches, which demonstrates its high potential in reaction modeling. However, the top-k accuracies of these models are not comparable to that of autoregressive ones. These experimental results indicate that NERF has a powerful encoder-decoder architecture while its uncertainty modeling is not successful. They adopt the CVAE framework to inject uncertainty modeling. However, training CVAE will easily encounter the posterior collapse issue. In fact, the KL divergence loss term gets close to zero with few iterations during training. Also, imposing uncertainty on latent can lead to the uncontrollable generation of the follow-

ing electron redistribution prediction, which will lower the validity of predicted products. To sum up, uncertainty estimation is an important challenge for non-autoregressive models. Non-autoregressive modeling with correct uncertainty estimation will revolutionize reaction prediction research.

Retrosynthesis. Non-autoregressive retrosynthesis model is very tricky since the total number of atoms in reactants must be pre-determined before attaching leaving groups. However, the distribution over the total number of reactant atoms is biased by the benchmark dataset. Therefore, to generate reactants in a non-autoregressive manner, enough blank entries should be reserved for predicted leaving groups. Unfortunately, if there are too many blank entries in the 3D adjacency matrix and atom feature matrix, the task would become a very tough imbalanced classification task since only a few blank entries will be filled up with predicted atoms and bonds.

4.5 Challenges in Generalization

There are two common challenges in generalization for reaction and retrosynthesis prediction. The first challenge is out-of-distribution prediction. Out-of-distribution mainly evaluates whether the model trained on main reaction types can generalize to rare reaction types. This imbalance issue naturally exists in chemical reactions since we can expect that some reactions easily occur while some reactions rarely occur. Therefore, it is important to design stable reaction and retrosynthesis prediction models under different distribution splits. The second challenge comes from the low-level representation quality. Both problems require low-level molecular and reaction representation learning techniques. Therefore, it is significant to derive powerful molecular and reaction representations with strong generalization to related tasks.

5 Conclusion and Future Directions

From the above discussions, we know that current approaches still have some major shortcomings although they have achieved promising results. Therefore, we list several future directions for further refinement of current solutions.

5.1 3D Molecular Information

A chemical molecule is inherently a point cloud consisting of a set of 3D points with Cartesian coordinates. However, 3D molecular information has not been used in the literature while sequence and graph representation of molecules have been widely explored. 3D position vectors provide important complementary distance information for each pair of atoms. The relative pairwise distance can be very different in 3D Euclidean geometry compared to 2D molecular graphs. For example, atom v_1 and atom v_2 may be distant from each other in a non-euclidean molecular graph while they might be close to each other in 3D Euclidean space. This is particularly useful for reaction center ranking. Therefore, effectively incorporating 3D molecular information into modeling can facilitate the more accurate reaction and retrosynthesis predictions.

5.2 Diverse Benchmark Dataset and New Evaluation Metrics

For reaction prediction, the size of the current USPTO-479K dataset is large enough. However, its modality and diversity

are still not sufficient. New benchmark datasets should include more reaction types and more complex reactions. Also, different dataset splits, such as scaffold split and time split, should be included for cross-validation. For retrosynthesis prediction, the USPTO-50K dataset is at small-scale. A new benchmark dataset should be a large-scale dataset containing at least 100K samples. Furthermore, future benchmark datasets should contain more target molecules with multiple edits. In addition, a new evaluation metric for retrosynthesis is necessary and urgent. FusionRetro [Liu *et al.*, 2023] attempts to evaluate single-step retrosynthesis models in the context of multi-step planning. More diverse evaluation metrics can be designed in the future.

5.3 Non-autoregressive Modeling

Just as mentioned in the previous section, the uncertainty estimation of non-autoregressive models is still inaccurate. Thus, it is crucial to explore more effective uncertainty modeling solutions for non-autoregressive models. We believe it will be a ground-breaking moment if the top-k accuracies of non-autoregressive models can be comparable to that of autoregressive ones. Non-autoregressive retrosynthesis prediction has not even been explored in current works. Due to the small size of leaving group in dataset, the significance of inference speed have been neglected in current research. We expect that the inference efficiency will be emphasized when analyzing complicated synthesis routes.

5.4 Self-supervised Learning

As previously discussed in limitations, reaction and retrosynthesis prediction are suffering from some issues brought by the dataset itself. Leveraging self-supervised learning (SSL) strategies on unlabeled datasets to overcome these limitations is a naturally motivated direction to explore. Although various SSL strategies have been presented for general molecular representation learning, currently seldom specific SSL strategies are designed for reaction and retrosynthesis prediction. PMSR [Jiang *et al.*, 2021] explores SSL strategies for retrosynthesis but it fails to achieve a strong performance. Therefore, exploring more powerful SSL strategies for both problems is a feasible direction.

5.5 Multi-task Learning

Combining both tasks with other related tasks is a promising direction. DualTF [Sun *et al.*, 2021] gives a first attempt by presenting a dual model to show that reaction and retrosynthesis prediction model can reinforce each other. FusionRetro [Liu *et al.*, 2023] and GNN-Retro [Han *et al.*, 2022] leverages reaction contexts and molecular contexts respectively to improve retrosynthetic analysis. Lu and Zhang [Lu and Zhang, 2022] show that multi-task learning can further improve reaction prediction accuracy. Intuitively, reaction classification, reaction yield prediction, and many other tasks have close connections with reaction and retrosynthesis predictions. How to leverage these tasks together is crucial for further improvements on both problems.

Acknowledgments

The work described here was partially supported by grants from the National Key Research and Development Program of China (No. 2018AAA0100204) and from the Research Grants Council of the Hong Kong Special Administrative Region, China (CUHK 14222922, RGC GRF, No. 2151185).

References

- [Bi *et al.*, 2021] Hangrui Bi, Hengyi Wang, Chence Shi, Connor W. Coley, Jian Tang, and Hongyu Guo. Non-autoregressive electron redistribution modeling for reaction prediction. In *ICML 2021*, volume 139 of *Proceedings of Machine Learning Research*, pages 904–913. PMLR, 2021.
- [Bradshaw *et al.*, 2019] John Bradshaw, Matt J. Kusner, Brooks Paige, Marwin H. S. Segler, and José Miguel Hernández-Lobato. A generative model for electron paths. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.
- [Chen and Jung, 2021] Shuan Chen and Yousung Jung. Deep retrosynthetic reaction prediction using local reactivity and global attention. *JACS Au*, 1(10):1612–1620, 2021.
- [Chen and Jung, 2022] Shuan Chen and Yousung Jung. A generalized-template-based graph neural network for accurate organic reactivity prediction. *Nature Machine Intelligence*, 4:1–9, 09 2022.
- [Chen *et al.*, 2019] Benson Chen, Tianxiao Shen, Tommi S. Jaakkola, and Regina Barzilay. Learning to make generalizable and diverse predictions for retrosynthesis. *CoRR*, abs/1910.09688, 2019.
- [Coley *et al.*, 2017] Connor W. Coley, Luke Rogers, William H. Green, and Klavs F. Jensen. Computer-assisted retrosynthesis based on molecular similarity. *ACS Central Science*, 3(12):1237–1245, 2017.
- [Dai *et al.*, 2020] Hanjun Dai, Chengtao Li, Connor W. Coley, Bo Dai, and Le Song. Retrosynthesis prediction with conditional graph logic network. *CoRR*, abs/2001.01408, 2020.
- [Do *et al.*, 2019] Kien Do, Truyen Tran, and Svetha Venkatesh. Graph transformation policy network for chemical reaction prediction. In *KDD 2019*, pages 750–760. ACM, 2019.
- [Dong *et al.*, 2021] Jingxin Dong, Mingyi Zhao, Yuansheng Liu, Yansen Su, and Xiangxiang Zeng. Deep learning in retrosynthesis planning: datasets, models and tools. *Briefings in Bioinformatics*, 23(1), 2021.
- [Engkvist *et al.*, 2018] Ola Engkvist, Per-Ola Norrby, Nidhal Selmi, Yu hong Lam, Zhengwei Peng, Edward C. Sherer, Willi Amberg, Thomas Erhard, and Lynette A. Smyth. Computational prediction of chemical reactions: current status and outlook. *Drug Discovery Today*, 23(6):1203–1218, 2018.
- [Han *et al.*, 2022] Peng Han, Peilin Zhao, Chan Lu, Junzhou Huang, Jiaxiang Wu, Shuo Shang, Bin Yao, and Xiangliang Zhang. Gnn-retro: Retrosynthetic planning with graph neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(4), 2022.
- [Irwin *et al.*, 2022] Ross Irwin, Spyridon Dimitriadis, Jiazhen He, and Esben Jannik Bjerrum. Chemformer: a pre-trained transformer for computational chemistry. *Machine Learning: Science and Technology*, 3(1):015022, jan 2022.
- [Jiang *et al.*, 2021] Yinjie Jiang, Ying Wei, Fei Wu, Zhengxing Huang, Kun Kuang, and Zhihua Wang. Learning chemical rules of retrosynthesis with pre-training. In *AAAI 2021*, pages 531–539. AAAI Press, 2021.
- [Jin *et al.*, 2017] Wengong Jin, Connor W. Coley, Regina Barzilay, and Tommi S. Jaakkola. Predicting organic reaction outcomes with weisfeiler-lehman network. In *NIPS 2017*, pages 2607–2616, 2017.
- [Kim *et al.*, 2021] Eunji Kim, Dongseon Lee, Youngchun Kwon, Min Sik Park, and Youn-Suk Choi. Valid, plausible, and diverse retrosynthesis using tied two-way transformers with latent variables. *Journal of Chemical Information and Modeling*, 61(1):123–133, 2021.
- [Kovacs *et al.*, 2021] David Kovacs, William McCorkindale, and Alpha Lee. Quantitative interpretation explains machine learning models for chemical reaction prediction and uncovers bias. *Nature Communications*, 12, 03 2021.
- [Lee *et al.*, 2019] Alpha Albert Lee, Qingyi Yang, Vishnu Sresht, Peter Bolgar, Xinjun Hou, Jacquelyn L. Klug-McLeod, and Christopher R. Butler. Molecular transformer unifies reaction prediction and retrosynthesis across pharma chemical space. *Chemical communications*, 2019.
- [Lin *et al.*, 2020] Kangjie Lin, Youjun Xu, Jianfeng Pei, and Luhua Lai. Automatic retrosynthetic route planning using template-free models. *Chemical Science*, 11, 03 2020.
- [Lin *et al.*, 2022] Zaiyun Lin, Shiqiu Yin, Lei Shi, Wenbiao Zhou, and Yingsheng J. Zhang. G2gt: Retrosynthesis prediction with graph to graph attention neural network and self-training. *ArXiv*, abs/2204.08608, 2022.
- [Liu *et al.*, 2017] Bowen Liu, Bharath Ramsundar, Prasad Kawthekar, Jade Shi, Joseph Gomes, Quang Luu Nguyen, Stephen Ho, Jack Sloane, Paul Wender, and Vijay Pande. Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Central Science*, 3(10):1103–1113, 2017.
- [Liu *et al.*, 2022] Jiahua Liu, Chaochao Yan, Yang Yu, Chan Lu, Junzhou Huang, Le Ou-Yang, and Peilin Zhao. MARS: A motif-based autoregressive model for retrosynthesis prediction. *CoRR*, abs/2209.13178, 2022.
- [Liu *et al.*, 2023] Songtao Liu, Zhengkai Tu, Minkai Xu, Zuobai Zhang, Lu Lin, Rex Ying, Jian Tang, Peilin Zhao, and Dinghao Wu. Fusionretro: Molecule representation fusion via in-context reactions for retrosynthetic planning. In *ICML 2023*, Proceedings of Machine Learning Research. PMLR, 2023.

- [Lu and Zhang, 2022] Jieyu Lu and Yingkai Zhang. Unified deep learning model for multitask reaction predictions with explanation. *J. Chem. Inf. Model.*, 62(6):1376–1387, 2022.
- [Mao et al., 2021] Kelong Mao, Xi Xiao, Tingyang Xu, Yu Rong, Junzhou Huang, and Peilin Zhao. Molecular graph enhanced transformer for retrosynthesis prediction. *Neurocomputing*, 457:193–202, 2021.
- [Meng et al., 2023] Ziqiao Meng, Peilin Zhao, Yang Yu, and Irwin King. Doubly stochastic graph-based non-autoregressive reaction prediction. In *IJCAI 2023*. ijcai.org, 2023.
- [Qian et al., 2020] Wesley Wei Qian, Nathan T. Russell, Claire L. W. Simons, Yunan Luo, Martin D. Burke, and Jian Peng. Integrating deep neural networks and symbolic inference for organic reactivity prediction. *ChemRxiv*, 2020.
- [Sacha et al., 2020] Mikolaj Sacha, Mikolaj Blaz, Piotr Byrski, Pawel Wlodarczyk-Pruszyński, and Stanislaw Jastrzebski. Molecule edit graph attention network: Modeling chemical reactions as sequences of graph edits. *CoRR*, abs/2006.15426, 2020.
- [Scannell et al., 2012] Jack Scannell, Alex Blanckley, Helen Boldon, and Brian Warrington. Diagnosing the decline in pharmaceutical r&d efficiency. *Nature reviews. Drug discovery*, 11:191–200, 03 2012.
- [Schwaller et al., 2019] Philippe Schwaller, Teodoro Laino, Théophile Gaudin, Peter Bolgar, Christopher A. Hunter, Costas Bekas, and Alpha A. Lee. Molecular transformer: A model for uncertainty-calibrated chemical reaction prediction. *ACS Central Science*, 5(9), 2019.
- [Segler and Waller, 2017] Marwin H. S. Segler and Mark P. Waller. Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chemistry – A European Journal*, 23(25):5966–5971, 2017.
- [Seo et al., 2021] Seung-Woo Seo, You Young Song, June Yong Yang, Seohui Bae, Hankook Lee, Jinwoo Shin, Sung Ju Hwang, and Eunho Yang. GTA: graph truncated attention for retrosynthesis. In *AAAI 2021*, pages 531–539. AAAI Press, 2021.
- [Shi et al., 2020] Chence Shi, Minkai Xu, Hongyu Guo, Ming Zhang, and Jian Tang. A graph to graphs framework for retrosynthesis prediction. In *ICML 2020*, volume 119 of *Proceedings of Machine Learning Research*, pages 8818–8827. PMLR, 2020.
- [Sohn et al., 2015] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 3483–3491, 2015.
- [Somnath et al., 2021] Vignesh Ram Somnath, Charlotte Bunne, Connor W. Coley, Andreas Krause, and Regina Barzilay. Learning graph models for retrosynthesis prediction. In *NeurIPS 2021*, 2021.
- [Sun et al., 2021] Ruoxi Sun, Hanjun Dai, Li Li, Steven Kearnes, and Bo Dai. Towards understanding retrosynthesis by energy-based models. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 10186–10194, 2021.
- [Tetko et al., 2020] Igor Tetko, Pavel Karpov, Ruud Deursen, and Guillaume Godin. State-of-the-art augmented nlp transformer models for direct and single-step retrosynthesis. *Nature Communications*, 11, 11 2020.
- [Tu and Coley, 2021] Zhengkai Tu and Connor W. Coley. Permutation invariant graph-to-sequence model for template-free retrosynthesis and reaction prediction. *CoRR*, abs/2110.09681, 2021.
- [Vaswani et al., 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS 2017*, pages 5998–6008, 2017.
- [Wan et al., 2022] Yue Wan, Chang-Yu Hsieh, Ben Liao, and Shengyu Zhang. Retroformer: Pushing the limits of end-to-end retrosynthesis transformer. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 22475–22490. PMLR, 2022.
- [Wang et al., 2021] Xiaorui Wang, Yuquan Li, Jiezhong Qiu, Guangyong Chen, Huanxiang Liu, Benben Liao, Chang-Yu Hsieh, and Xiaojun Yao. Retroprime: A diverse, plausible and transformer-based method for single-step retrosynthesis predictions. *Chemical Engineering Journal*, 420:129845, 2021.
- [Wei et al., 2016] Jennifer N. Wei, David Duvenaud, and Alán Aspuru-Guzik. Neural networks for the prediction of organic chemistry reactions. *ACS Central Science*, 2(10):725–732, 2016.
- [Yan et al., 2020] Chaochao Yan, Qianggang Ding, Peilin Zhao, Shuangjia Zheng, Jinyu Yang, Yang Yu, and Junzhou Huang. Retroxpert: Decompose retrosynthesis prediction like a chemist. In *NeurIPS 2020*, 2020.
- [Yan et al., 2022] Chaochao Yan, Peilin Zhao, Chan Lu, Yang Yu, and Junzhou Huang. RetroComposer: Composing templates for template-based retrosynthesis prediction. *Biomolecules*, 12(9):1325, sep 2022.
- [Zhao et al., 2022] Ming Zhao, Lei Fang, Li Tan, Yves Lepage, and Jian-Guang Lou. Leveraging reaction-aware substructures for retrosynthesis and reaction prediction. *CoRR*, abs/2204.05919, 2022.
- [Zheng et al., 2020] Shuangjia Zheng, Jiahua Rao, Zhongyue Zhang, Jun Xu, and Yuedong Yang. Predicting retrosynthetic reactions using self-corrected transformer neural networks. *Journal of Chemical Information and Modeling*, 60(1):47–55, 2020.