

# 基于多层注意力和消息传递网络的药物相互作用预测方法

饶晓洁<sup>1</sup> 张通<sup>1,2</sup> 孟献兵<sup>1</sup> 陈俊龙<sup>1,2</sup>

**摘要** 药物相互作用 (Drug-drug interaction, DDI) 是指不同药物存在抑制或促进等作用. 现有 DDI 预测方法往往直接利用药物分子特征表示预测 DDI, 而忽略药物分子中不同原子对 DDI 的影响. 为此, 提出基于多层次注意力机制和消息传递神经网络的 DDI 预测方法. 该方法将 DDI 建模为通过学习基于序列表示的药物分子特征实现 DDI 预测的链接预测问题. 首先, 建立基于注意力机制和消息传递神经网络的原子特征网络, 结合提出的基于分子质心的位置编码, 学习不同原子及其相关联化学键的特征, 构建基于图结构的药物分子特征表示; 然后, 设计基于注意力机制的分子特征网络, 并通过监督和对比损失学习, 实现 DDI 预测; 最后, 通过实验证明该方法的有效性和优越性.

**关键词** 药物相互作用预测, 多层次注意力机制, 消息传递神经网络, 位置编码

**引用格式** 饶晓洁, 张通, 孟献兵, 陈俊龙. 基于多层注意力和消息传递网络的药物相互作用预测方法. 自动化学报, 2023, 49(12): 2507-2519

**DOI** 10.16383/j.aas.c220371

## Drug-drug Interaction Prediction Method Based on Multi-level Attention Mechanism and Message Passing Neural Network

RAO Xiao-Jie<sup>1</sup> ZHANG Tong<sup>1,2</sup> MENG Xian-Bing<sup>1</sup> CHEN C. L. Philip<sup>1,2</sup>

**Abstract** Drug-drug interaction (DDI) denotes the presence of inhibitory or promoting effects between different drugs. The existing DDI prediction methods often directly use drug molecular feature representation, while ignoring the different effects of different atoms within drug molecule on DDI. To solve this problem, a DDI prediction method is proposed based on multi-level attention mechanism and message passing neural network. This method models the task as a link prediction problem of predicting DDI by extracting the drug molecular features from their sequence representations. First, the atomic feature network is developed based on attention mechanism and message passing neural network. Through integration with the proposed positional encoding based on molecular centroid, the proposed network can learn from different atoms and the correlated chemical bonds to construct drug molecular graph features. Second, attention mechanism-based molecular feature network is designed, and the DDI prediction can then be realized by using supervision and contrastive loss learning. Finally, experiments demonstrate the effectiveness and superiority of the proposed method.

**Key words** Drug-drug interaction prediction, multi-level attention mechanism, message passing neural network, positional encoding

**Citation** Rao Xiao-Jie, Zhang Tong, Meng Xian-Bing, Chen C. L. Philip. Drug-drug interaction prediction method based on multi-level attention mechanism and message passing neural network. *Acta Automatica Sinica*, 2023, 49(12): 2507-2519

### 药物相互作用 (Drug-drug interaction, DDI)

收稿日期 2022-05-06 录用日期 2022-07-25

Manuscript received May 6, 2022; accepted July 25, 2022

国家重点研发计划 (2019YFB1703600), 国家自然科学基金 (62006081), 广东省自然科学基金面上项目 (2022A1515011317), 中国博士后科学基金 (2020M672630) 资助

Supported by National Key Research and Development Program of China (2019YFB1703600), National Natural Science Foundation of China (62006081), Natural Science Foundation of Guangdong Province (2022A1515011317), and China Postdoctoral Science Foundation Project (2020M672630)

本文责任编辑 何晖光

Recommended by Associate Editor HE Hui-Guang

1. 华南理工大学计算机科学与工程学院 广州 510006 2. 琶洲实验室 广州 510355

1. School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006 2. Pazhou Laboratory, Guangzhou 510335

针对的是两种或两种以上药物进行混合时, 某些药物的性能受到其他药物的影响, 而发生协同或拮抗作用<sup>[1-2]</sup>. DDI 引发的药物副作用可能会降低药物疗效, 诱发不良反应, 甚至影响到患者的身体健康. 由于传统基于实验的方法进行 DDI 预测存在成本高、检测周期长等问题<sup>[3-4]</sup>, 近年来, 基于计算机辅助的计算方法日益得到广泛应用<sup>[2, 5]</sup>. 通过传统的机器学习和深度学习等计算机辅助计算方法从已有药物相互作用数据中学习建立模型并实现 DDI 预测, 可以大大提高 DDI 预测任务的效率<sup>[3]</sup>. 因此, 研究基于计算机辅助的 DDI 预测方法具有重要的理论和应用价值.

DDI 预测任务主要包括对药物分子式进行量化并提取其特征, 以及选择合适的模型预测 DDI<sup>[6-7]</sup>. 不同于一般的结构化数据, 药物分子具有确定的分子结构及其生物化学性质. 因此, 解决 DDI 预测问题的关键是学习药物分子结构及其特征信息, 并建立预测模型. 目前, 已有大量 DDI 预测任务相关研究成果, 其中涉及的方法可以概括为基于传统机器学习的方法和基于深度学习的方法<sup>[1, 8-9]</sup>.

在现有 DDI 预测方法中, 利用药物的生化特征信息, 包括药物靶点、酶、药物转运体及蛋白质等相关信息构建模型并预测 DDI, 是一种有效的研究思路<sup>[2, 10-11]</sup>. 这类方法通过充分利用药物相关联的生化特征信息可以提高 DDI 预测的精度, 不过也存在一些局限性<sup>[3]</sup>, 比如此类生化特征信息的获取成本较高<sup>[6]</sup>. 如何在没有这些生化特征信息的前提下提高 DDI 预测的精度, 是值得深入研究的问题. 此外, 由于 DDI 预测任务涉及多个药物分子, 且不同药物分子内部又具有不同的原子信息, 因此有必要深入挖掘药物分子内不同原子和不同药物分子之间的相关关系, 并研究通过融合原子和分子等不同层次的特征信息, 提高 DDI 预测准确率.

针对上述问题, 本文提出一种基于多层次注意力机制和消息传递神经网络的药物相互作用预测方法. 为了充分挖掘药物的分子结构信息, 本文从药物分子内不同原子和不同药物分子之间两个层面研究 DDI 预测方法. 通过基于注意力机制的消息传递神经网络学习药物分子内包含不同原子和化学键的图结构特征, 并结合基于多头注意力机制的神经网络提取不同药物分子间相关关系的特征信息, 实现从药物分子内和分子间两个不同层次进行药物分子特征提取, 从而完成 DDI 预测任务.

本文主要贡献是同时从原子和分子层面设计基于注意力机制的特征学习策略, 提出基于多层次注意力机制和消息传递神经网络的 DDI 预测方法. 主要创新点包括两个方面: 一方面, 通过考虑不同原子及其相关联化学键的不同相互作用信息, 设计基于分子质心的位置编码策略, 提出基于注意力机制和消息传递神经网络的原子特征网络; 另一方面, 通过考虑分子之间的不同相互作用关系, 提出基于注意力机制和密集连接结构的分子特征网络. 具体来说, 相比直接在分子层面学习不同药物分子相互作用关系的方法<sup>[10, 12]</sup>, 本文方法选择同时从原子和分子层面学习预测 DDI. 考虑到不同药物分子的相互作用关系本质上取决于其内部原子和化学键的相互作用, 这里首先学习药物分子内不同原子、化学键间的相互作用关系, 并通过设计基于分子质心的

位置编码策略, 辅助学习药物分子的图结构特征表示. 这种方法有助于模型学习得到丰富的药物分子表示, 从而更有利于挖掘出不同药物分子的潜在相互作用关系. 虽然已有研究工作也同时从原子和分子层面研究 DDI 预测方法, 不过这些方法在原子层面只考虑了原子节点特征, 并没有充分利用不同原子相关联的化学键特征<sup>[13]</sup>, 无法在原子和分子层面同时提取不同原子及其相关联化学键的不同注意力信息, 因此其预测能力也有限. 大量对比实验和消融实验验证了本文方法的有效性, 以及相比现有方法的优越性.

## 1 相关工作

基于计算机辅助计算的 DDI 预测方法可以概括为基于传统机器学习的方法和基于深度学习的方法. 基于传统机器学习的方法主要分为 3 类, 即基于传统分类器的方法和基于回归的方法和基于矩阵分解的方法<sup>[8, 14-15]</sup>. 在基于传统分类器的方法和基于回归的方法中, 通常会使用相似性度量方法 (基于内积或基于余弦值的相似度度量) 度量两种药物之间的相似度, 并通过不同的分类或回归算法, 预测得出不同药物相互作用的概率<sup>[8, 14]</sup>. 在基于矩阵分解的方法中, DDI 预测任务可以建模为矩阵补全任务: 将一个含有缺失值的矩阵恢复为一个完全的矩阵, 目的是对未观察到的相互作用进行预测<sup>[15]</sup>. 基于传统机器学习的方法虽然能有效解决 DDI 预测任务, 但因其提取深层特征的能力有限, 也存在一定的局限性, 比如在不平衡数据以及大规模数据集上存在表现能力不足的问题<sup>[6]</sup>. 此外, 这类方法往往直接利用药物分子的特征信息, 而忽略药物分子内部的原子特征信息, 这在一定程度上也会影响 DDI 预测的效果.

不同于基于传统机器学习的方法, 基于深度学习的 DDI 预测方法能够提取更深层次的特征, 在实际应用中, 往往可以更好地预测出潜在的 DDI. 如, Ryu 等<sup>[1]</sup>提出了一种基于深度学习的 DDI 预测模型, 通过学习不同药物的结构相似度信息, 实现 DDI 预测. Deng 等<sup>[10]</sup>通过计算药物结构、基因本体和目标基因这三种相似度信息, 结合深度神经网络, 实现药物分子特征提取, 并用于 DDI 预测. Lee 等<sup>[6]</sup>利用每种药物不同的相似度信息训练模型, 使用自动编码器和深度前馈网络实现 DDI 预测. 这些方法借助深度学习深层特征提取能力<sup>[1, 6, 10, 12]</sup>, 虽然也能解决 DDI 预测问题, 但忽略了药物基于图结构的数据本质, 实际应用中效果往往有限.

近年来, 基于图结构的深度学习方法相继提出并成功应用于 DDI 预测<sup>[5, 9]</sup>. 这类方法通过将不同

药物分子作为节点、相互作用关系作为边, 构建药物分子的图网络, 从而实现 DDI 预测<sup>[7, 16]</sup>. 如, Liu 等<sup>[9]</sup> 使用多模态深度自编码器, 将每个药物数据源视为一个药物特征网络, 在每个网络中利用图结构的邻接矩阵做图嵌入, 从多个药物特征网络中学习药物的统一表示, 并在此基础上构建模型, 实现 DDI 预测. Lin 等<sup>[16]</sup> 设计出一种基于知识图谱的图卷积神经网络, 通过学习基于不同药物分子的知识图谱, 获取药物潜在的相互作用关系. Karim 等<sup>[7]</sup> 提出一种 DDI 预测模型, 通过知识图谱学习药物的重要特征, 并通过集成卷积神经网络和长短期记忆递归神经网络进行学习, 得到不同药物的相互作用关系. 这类方法虽然考虑了药物分子的图结构信息, 但也存在一定的局限性. 例如, 上述方法从药物分子层面进行特征提取, 而忽略了药物分子内原子层面的特征学习. 此外, 这些方法并没有考虑通过区分不同药物分子或药物原子的重要性来预测 DDI. 针对上述问题, 本文提出了基于药物分子内和药物分子间的多层次注意力机制和特征提取方法.

## 2 本文方法

本文研究的 DDI 预测任务是预测给定药物分子集合中任意 2 个药物分子的相互作用关系. 这里采用简化分子线性输入规范 (Simplified molecular input line entry specification, SMILES) 表示每个药物分子, 并将 DDI 任务建模为通过提取 SMILES 文本序列表示的药物分子特征预测得出两种药物是否存在相互作用的链接预测问题. 理论上来说, 不同药物是否存在相互作用关系取决于药物的分子图结构信息及其相关生化性质. 因此, 对于基于 SMILES 文本序列的 DDI 预测方法来说, 从序列中准确提取药物分子图结构信息对提高 DDI 预测方法的精度具有十分重要的作用. 为此, 本文基于 Transformer 的注意力机制<sup>[17]</sup> 和消息传递神经网络<sup>[18]</sup>, 设计基于多层次注意力机制和消息传递神经网络的 DDI 预测方法, 旨在实现分子图结构特征提取及 DDI 预测.

本文方法整体框架如图 1 所示. 首先, 从 SMILES 文本序列中提取药物分子的图结构信息, 包括原子和化学键相关联的特征<sup>[19]</sup>; 然后, 在原子特征层面, 利用消息传递神经网络, 并通过融合 Transformer 的注意力机制以及本文提出的基于分子质心的位置编码方法, 实现药物分子内不同原子和化学键特征的学习更新; 最后, 在分子特征层面, 进一步利用注意力机制, 并通过监督学习和对比学习, 挖掘出不同药物分子的潜在相互作用关系, 从而实现任意两个药物分子  $i$  和  $j$  的相互作用预测.

### 2.1 基于注意力机制的消息传递原子特征网络

药物分子内部由不同原子及原子之间相关联的化学键组成. 对于具有图结构的药物分子来说, SMILES 这种一维线性的序列表示无法直接反映不同原子和化学键在药物分子图结构中的相对位置信息. 为解决这个问题, 并提取药物分子内不同原子和化学键的特征信息, 本文建立基于注意力机制的消息传递原子特征网络, 学习得到基于图结构的药物分子特征表示.

原子特征网络如图 2 所示. 首先, 使用药物原子和化学键信息进行节点和边的特征嵌入, 同时设计基于分子质心的位置编码方法编码具有图结构的原子和化学键特征信息; 然后, 利用结合图结构交互式注意力机制的消息传递神经网络, 学习不同节点和边的特征信息; 最后, 通过不断更新迭代模型, 得到药物分子的特征表示. 下面具体从两个方面进行介绍.

#### 2.1.1 基于分子质心的位置编码

由于药物分子是由不同原子和化学键组成的图结构数据, 因此在 Transformer 注意力机制中处理时序数据的位置编码方法不一定适合药物分子数据. 此外, 由于表示药物分子的 SMILES 是一维线性化的序列, 因此仅仅利用药物分子 SMILES 序列中字符的输入先后顺序作为位置编码的依据<sup>[18]</sup> 是不合理的. 为了有效提取药物分子中不同原子的位置信息, 本文提出一种基于分子质心的位置编码方法, 通过计算原子与分子质心的距离来表示原子之间的相对位置, 得到基于该距离的排序结果并用于设计位置编码.

给定具有  $n$  个原子的药物分子, 该分子可表示为有向图  $G = (X, E)$ , 其中  $X, E$  是通过 RDKit<sup>[19]</sup> 化学信息库分别得到的原子和化学键的初始特征.  $X$  由药物分子内  $n$  个原子组成,  $x_i \in X \in \mathbf{R}^{n \times f_n}$  表示原子节点  $i$  的特征嵌入信息, 包括通过 RDKit 获得原子的杂化方式、形式电荷、连接数等信息.  $E$  代表原子之间相关联的化学键, 包括化学键的类型、是否为芳香键、是否成环等特征信息.  $e_{uv} \in E \in \mathbf{R}^{n \times n \times f_e}$  表示从原子节点  $i$  到节点  $j$  的边特征嵌入信息.  $f_n, f_e$  分别表示节点和边的特征维度. 拓扑连接矩阵  $ADJ \in \mathbf{R}^{n \times n}$  由两个原子之间的最短路径组成. 基于分子质心的位置编码方法, 具体如下:

首先, 通过计算分子中  $n$  个原子二维坐标  $v_i (i = 1, 2, 3, \dots, n)$  的均值  $v$ , 得到分子质心的坐标  $s_0$ ; 其次, 计算每个原子和分子质心之间的欧氏距离  $d_i$ , 再按照  $n$  个原子与分子质心  $s_0$  之间的距离值  $d_i$  由近到远进行排序, 得到距离分子质心由近到远的原

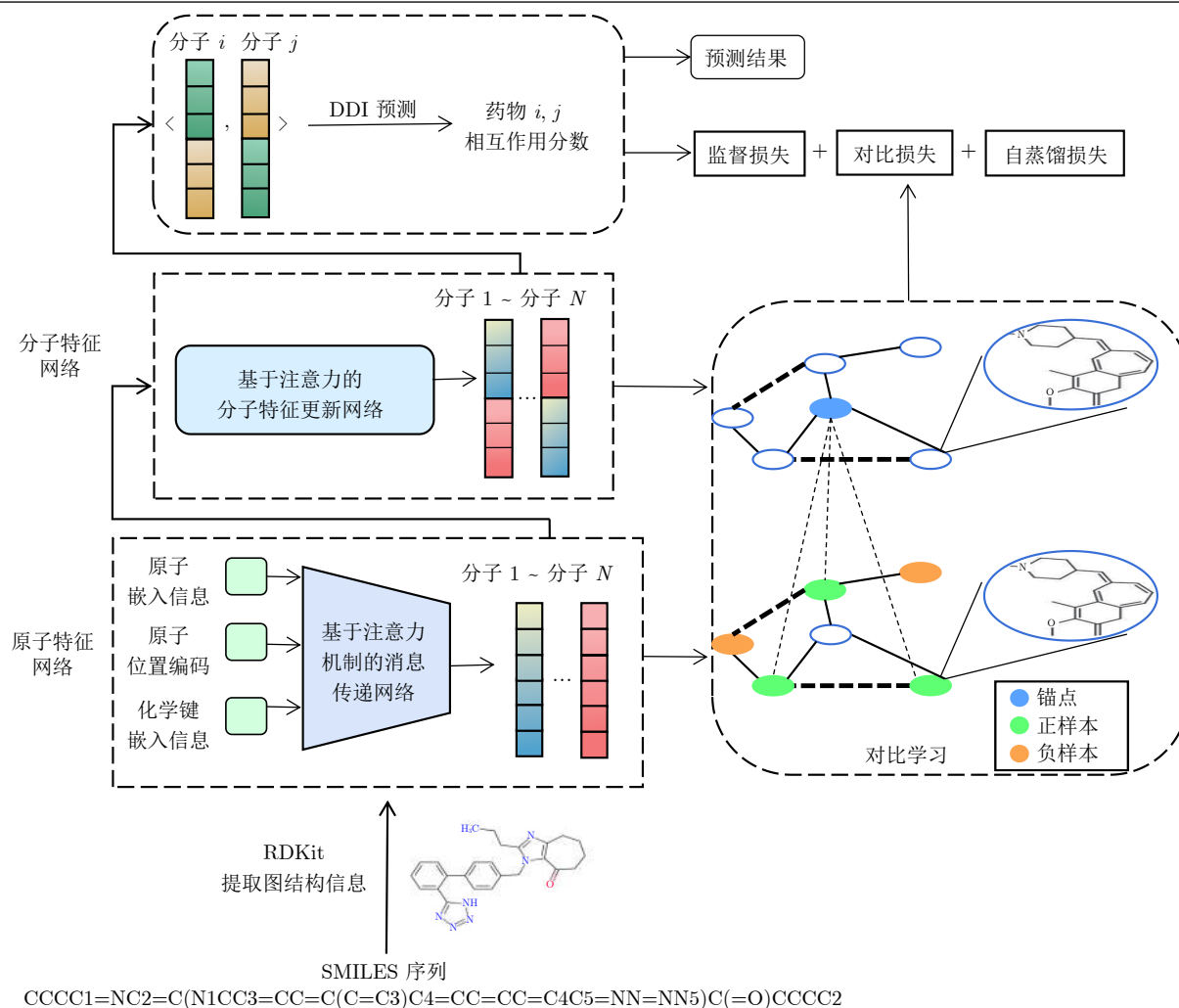


图 1 模型框架图

Fig.1 Framework of the proposed model

子索引;最后,将该原子索引顺序作为  $n$  个原子的位置编码,经过词嵌入得到  $n$  个原子的位置编码  $pos_i$ . 基于分子质心位置编码的原子特征  $h(x_i)$  和边特征  $h(e_{uv})$  表示方法分别为

$$h(x_i) = x_i + pos_i \quad (1)$$

$$h(e_{uv}) = e_{uv} + h(x_u) \quad (2)$$

这种位置编码方法通过利用分子图结构的空  
间信息描述不同原子的相对位置关系,在一定程度  
上可以改善分子 SMILES 序列表示方法仅以  
SMILES 序列中字符输入先后顺序作为位置编码的  
不足,为原子的特征表示学习提供更多的信息.

### 2.1.2 基于注意力机制的消息传递神经网络

针对传统消息传递神经网络不能有效区分不同  
节点和边的不同作用信息的问题<sup>[13]</sup>,本文借鉴基于  
图交互式的消息传递神经网络的思想<sup>[18]</sup>,将药物分子  
中不同原子及其相关联的化学键(边)表示为区

分入边和出边的有向图,利用基于 Transformer 注  
意力机制的消息传递神经网络,结合本文提出的基  
于分子质心的位置编码策略,计算药物分子中不同  
原子和化学键之间相互作用的注意力分数,并进行  
基于图结构的消息传递,不断学习更新节点和边  
的特征,从而得到药物分子的特征表示.通过将药  
物分子表示为有向无环图,节点和边的信息只会沿  
着确定的方向传递,不会出现因无向图中的环路造  
成节点和边信息的循环更新问题,从而提高节点和  
边信息的学习更新效率.此外,由式(1)和式(2)可  
知,通过区分入边和出边,不同节点及其相关联  
的边将具有不同的信息.此时,通过基于注意力机  
制的图结构信息传递和更新学习,更容易学习得  
到不同原子及其相关联边的不同作用信息,从而  
提取到更有效的药物分子特征表示.

在基本消息传递神经网络中,通常利用邻接矩  
阵进行消息的聚合及更新<sup>[20]</sup>.该方式存在只考虑节

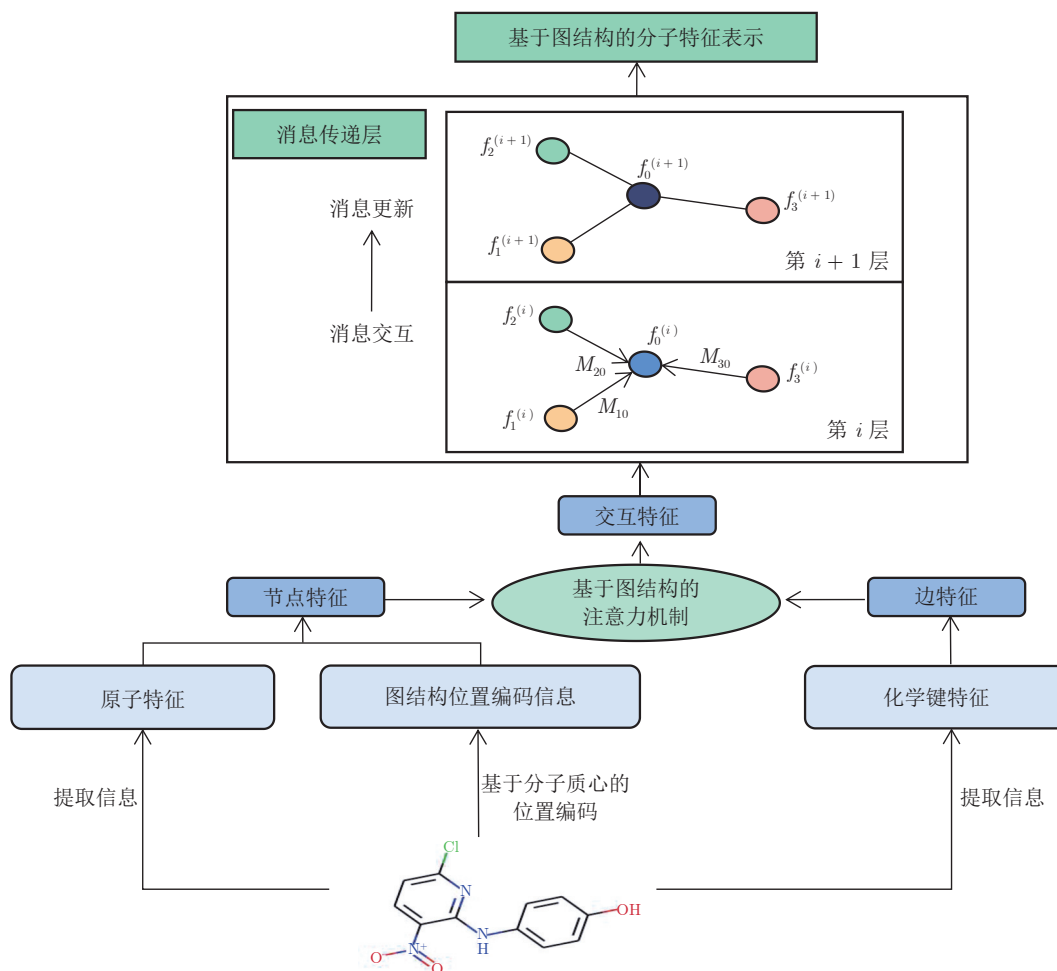


图 2 基于注意力机制的消息传递原子特征网络

Fig. 2 Framework of the message passing atomic feature network base on attention mechanism

点特征而忽略边信息的缺点. 同时, 在进行消息传递时无法自适应调节节点之间信息传递权重. 为了解决上述问题, 本文引入了注意力机制. 其中, 计算  $Q$ ,  $K$ ,  $V$  的方式为

$$\begin{cases} [Q, V] = h(X)[W^Q, W^V] \\ K = h(E)W^K \end{cases} \quad (3)$$

其中,  $Q, V \in \mathbf{R}^{n \times d}$ ,  $K \in \mathbf{R}^{n \times n \times d}$ .  $W^Q, W^V \in \mathbf{R}^{f_n \times d}$ ,  $W^K \in \mathbf{R}^{f_e \times d}$  为可学习的权重参数.

为了进一步挖掘图结构信息, 本文将每个分子视为一个有向图, 并将边分为入边和出边两个类型, 利用式 (3) 求得的  $Q, K, V$  矩阵计算相应的消息传递分数矩阵  $M_i$  和  $M_o$ , 即

$$\begin{cases} M_i = \text{einsum}(Q_i, K_i) \\ M_o = \text{einsum}(Q_o, K_o) \end{cases} \quad (4)$$

其中,  $M_i, M_o \in \mathbf{R}^{n \times n}$ , einsum 为爱因斯坦求和约定,  $Q_i, K_i, Q_o, K_o$  分别表示入边、出边相应的矩

阵. 最终的消息传递分数矩阵  $M$  为

$$M = \text{softmax}(M_o + M_i - \text{diag}\{M_o\}) \quad (5)$$

其中,  $M \in \mathbf{R}^{n \times n}$ , softmax 将数值向量归一化为概率分布,  $\text{diag}\{M_o\}$  表示只保留矩阵  $M_o$  的对角线元素.

考虑到在消息传递过程中, 不同传递深度的消息携带的信息不同, 为了模拟随着消息传递层数加深而导致的信息量减少的现象, 本文引入消息衰减机制<sup>[18]</sup>, 并假设距离越远的两个原子之间交互分数衰减得越快. 带有衰减机制的消息传递矩阵  $M$  计算式为

$$M(u, v) = M(u, v)e^{-\gamma \text{ADJ}(u, v)} \quad (6)$$

其中,  $\gamma$  为表示衰减程度的参数,  $\text{ADJ}(u, v)$  表示原子  $u$  和  $v$  之间的最短路径.

经过单层消息传递后, 得到的节点和边特征的更新式为

$$\begin{cases} h(X) = \text{matmul}(M, V) \\ h(E) = M \odot K \end{cases} \quad (7)$$

其中,  $\text{matmul}$  表示矩阵乘法,  $\odot$  表示两个矩阵对应元素相乘. 最后, 本文使用平均池化的方式生成原子特征网络输出的分子表示, 即

$$Fa_i = \frac{1}{K_i} \sum_{k=1}^{K_i} h(X^m)_k \quad (8)$$

其中,  $Fa$  是所有分子特征组成的矩阵,  $Fa_i$  是第  $i$  个分子的特征表示,  $K_i$  表示第  $i$  个分子的最大原子数量,  $h(X^m)_k$  表示经过上述网络更新后的第  $m$  个分子中第  $k$  个原子的特征.

## 2.2 基于多头注意力机制的分子特征网络

原子特征网络只考虑了单个分子内部的结构信息. 如果直接基于原子特征网络输出的  $Fa$  进行 DDI 预测, 将会丢失不同药物分子之间的交互信息, 从而影响最终的 DDI 预测结果精度. 为了解决这个问题, 本文将原子特征网络输出的分子表示作为分子特征网络的输入, 通过设计多头注意力模块学习不同药物分子间的交互信息, 并利用学习得到的药物分子关系更新每个药物分子的向量表示, 最终基于不同药物分子的特征学习得到 DDI 预测的结果.

通过对原子特征网络输出的药物分子特征表示  $Fa$  进行线性变换, 得到  $Q'$ ,  $K'$ ,  $V'$  矩阵, 并计算多头注意力, 即

$$\text{head} = \text{softmax} \left( \frac{Q'K'^T}{\sqrt{d_k}} \right) V' \quad (9)$$

$$\text{MultiHead} = \text{concat}(\text{head}_1, \text{head}_2, \text{head}_3) \quad (10)$$

其中,  $Q' \in \mathbf{R}^{n' \times d'}$ ,  $K' \in \mathbf{R}^{n' \times d'}$ ,  $V' \in \mathbf{R}^{n' \times d'}$ ,  $n'$  表示分子数量,  $\text{concat}$  表示拼接操作.

为防止网络层数过深导致的梯度消失, 且实现不同层次特征的融合, 本文在多层感知机 (Multi layer perceptron, MLP) 网络中线性层之间使用密集连接结构, 计算最终的输出: 药物分子特征表示  $Fm$ . 具体为

$$x_l = \text{layer}^l \left( \sum_{k=0}^{l-1} x_k \right) \quad (11)$$

$$Fm = \text{DenseMLP}_N(x_0) = \text{layer}^N \left( \sum_{k=0}^{N-1} x_k \right) \quad (12)$$

其中,  $x_0$  是多头注意力层的输出,  $x_l$  表示每个密集连接层的计算式,  $\text{layer}^l$  表示第  $l$  个线性层,  $\text{DenseMLP}_N$  表示一个具有  $N$  层密集连接的 MLP 网络.

$$S_{ij} = E_i \odot E_j \quad (13)$$

其中,  $E = Fm$ .

给定药物分子对  $i$  和  $j$ , 通过式 (13) 计算得到  $S_{ij}$ , 再将其经过一个 MLP 和 sigmoid 函数, 得到最终的链接预测结果  $p_{ij}$ .

本文模型的复杂度主要包括原子特征网络和分子特征网络以及对比学习这 3 部分的计算. 在原子特征和分子特征网络中主要计算量是式 (8) 和 (12), 其相应的计算复杂度都是  $O(NBf + Hd)$ , 其中,  $N$  指数据集中的药物分子数量,  $B$  是药物分子中的化学键数量,  $H$  表示能发生相互作用的药物对数量,  $f$  和  $d$  分别表示输入的特征维数和药物嵌入的特征维数. 对比学习部分的计算复杂度是  $O(Nk)$ , 其中  $k$  是正样本数量. 因此, 模型复杂度为  $O(N(Bf + k) + Hd)$ .

## 2.3 模型的训练优化

为了提高模型的泛化性能, 本文在传统二元交叉熵损失函数 (Binary cross entropy, BCE)  $L_{\text{label}}$  的基础上, 引入 2 种无监督损失函数, 包括基于自蒸馏的正则化约束  $L_{\text{un}}$  和基于对比学习的无监督损失  $L_{\text{c}}$ . 模型整体的损失函数  $L$  为

$$L = L_{\text{label}} + \alpha L_{\text{c}} + \beta L_{\text{un}} \quad (14)$$

其中,  $\alpha$  和  $\beta$  是对应损失的权重系数.

$L_{\text{label}}$  使用二元交叉熵损失函数衡量模型的误差损失, 即

$$L_{\text{label}} = \sum_{(i,j) \in Tr} (\text{BCE}(r_{ij}, y_{ij}) + \text{BCE}(p_{ij}, y_{ij})) \quad (15)$$

其中,  $Tr$  表示训练集的样本对集合,  $y_{ij}$  表示样本对  $(i, j)$  的真实标签,  $r_{ij}$  和  $p_{ij}$  分别表示原子特征网络和分子特征网络输出的预测结果.

对于基于自蒸馏的正则化约束<sup>[21]</sup>, 利用分子特征网络的输出对原子特征网络输出进行蒸馏学习, 进一步提高原子特征网络输出特征的质量. 这里通过 KL 散度 (Kullback-Leibler divergence, KL) 表示  $L_{\text{un}}$ , 即

$$L_{\text{un}} = \sum_{(i,j) \in D \setminus Tr} \text{KL}(p_{ij} \parallel r_{ij}) \quad (16)$$

其中,  $\text{KL}(p_{ij} \parallel r_{ij})$  表示两个概率分布之间的 KL 散度, 用来衡量两个分布之间的分布差异.  $D \setminus Tr$  表示除训练集之外的样本对集合.

基于对比学习的无监督损失函数设计思想如下. 对于每个药物分子, 选取其在分子特征网络的输出特征作为锚点, 并将其一阶邻居和非一阶邻居在原子特征网络的输出特征分别作为正样本和负样本. 通过对比损失学习, 使得锚点与其正样本相接近, 与其负样本区分开来. 具体为

$$(\hat{\omega}, \hat{\phi}, \hat{\varphi}) = \arg \max_{\omega, \phi, \varphi} \sum_{i \in D} \frac{1}{\tilde{C}(i)} \times \sum_{j \in C(i) \cap \{i\}} \hat{I}_{\omega, \phi, \varphi} (Fa_j^{\phi}; Fm_j^{\varphi}) \quad (17)$$

其中,  $C(i)$  和  $\tilde{C}(i)$  分别表示节点  $i$  的一阶邻居集合和非一阶邻居集合.  $\phi, \varphi$  分别定义了原子特征网络和分子特征网络的参数,  $\hat{I}_{\omega, \phi, \varphi}$  定义了互信息估计器.

由于无法直接优化互信息, 本文使用 JS 散度 (Jensen-shannon divergence, JSD) 优化互信息の下界<sup>[22]</sup>. 通过最小化对比损失函数  $L_c$ , 使得互信息最大化, 对比损失函数的计算式为

$$L_c = \frac{1}{|Tr|} \sum_{i \in Tr} \left( \frac{1}{|C(i)|} \sum_{j \in C(i)} \text{JSD}(Fm_i, Fa_j) - \frac{1}{|\tilde{C}(i)|} \sum_{k \in \tilde{C}(i)} \text{JSD}(Fm_i, Fa_k) \right) \quad (18)$$

模型训练过程的伪代码如算法 1 所示.

### 算法 1. 端到端的药物相互作用预测模型

**输入.** DDI 网络  $D = (G_i, G_j, y)$ .  $G_i, G_j \in G$  表示一组药物对,  $y_{ij}$  为表示两个药物分子是否发生相互作用的标签.

**输出.** 优化后的模型及参数.

**步骤 1.** 构建原子特征网络: 在每个药物分子中, 由式 (1) 和式 (2) 分别得到节点和边特征  $h(X)$  和  $h(E)$ ;

**步骤 2.** 使用基于注意力机制的消息传递神经网络计算节点和边之间的交互信息, 由式 (3) ~ (6) 得到交互分数矩阵  $M$ ;

**步骤 3.** 由式 (6) 和式 (7) 更新每个分子中的所有节点特征  $h(X)$  和边特征  $h(E)$ , 并由式 (8) 得到原子特征网络生成的分子特征矩阵  $Fa$ ;

**步骤 4.** 构建分子特征网络: 以每个分子为节点, 由式 (9) 和式 (10) 得到不同分子间的注意力分数;

**步骤 5.** 由式 (11) 和式 (12) 得到分子特征网络输出的分子特征矩阵  $Fm$ , 由式 (13) 得到 DDI 预测结果  $p$ ;

**步骤 6.** 由式 (14) ~ (18) 得到总损失函数, 通过梯度下降更新模型参数.

## 3 实验结果及分析

为了验证本文方法的有效性和优越性, 我们选择两个常用的 DDI 数据集, 即 ZhangDDI<sup>[11]</sup> 和 ChCh-Miner<sup>[23]</sup>, 进行对比和消融实验分析. ZhangDDI 包含 548 种药物和 48 548 组药物相互作用关系数据, ChCh-Miner 包含 1 514 种药物和 48 514 组药物相互作用关系数据. 评价指标包括 ROC (Receiver operating characteristic curve) 下面积

(Area under ROC, AUROC)、PRC (Precision-recall curve) 下面积 (Area under PRC, AUPRC) 和 F1 分数 (F1-score, F1).

### 3.1 对比方法介绍及实验设置

本文选择 13 种具有代表性的 DDI 预测方法作为对比方法, 分为基于传统机器学习的 DDI 预测方法和基于图结构的深度学习 DDI 预测方法.

基于传统机器学习的 DDI 预测方法包括以下 6 种方法: 基于子结构相似性的 DDI 预测方法 NN (Nearest neighbor)<sup>[24]</sup>; 基于标签传播的 DDI 预测方法, 这里包括 3 个基于不同相似性的方法 (LP-Sub (Label propagation substructure)、LP-SE (Label propagation side effect)、LP-OSE (Label propagation off-label side effect))<sup>[25]</sup>; 基于混合集成模型的 DDI 预测方法 MF-Ens (Multi-feature ensemble)<sup>[11]</sup>; 基于结构相似性轮廓的 DDI 预测方法 SSP-MLP (Structural similarity profile and multi-layer perceptron)<sup>[1]</sup>.

基于图结构的深度学习 DDI 预测方法又分为两类, 即基于分子特征网络的 DDI 预测方法、基于原子特征和分子特征网络的 DDI 预测方法. 第一类方法包括以下 4 种方法: 基于图卷积网络的 DDI 预测方法 GCN (Graph convolutional network)<sup>[26]</sup>、基于图同构网络的 DDI 预测方法 GIN (Graph isomorphism network)<sup>[27]</sup>、基于图自动编码器的 DDI 预测方法 Att-auto (Attentive graph autoencoder)<sup>[12]</sup>、基于图注意力网络的 DDI 预测方法 GAT (Graph attention network)<sup>[28]</sup>. 第二类方法包括以下 3 种方法: 基于层次图表示学习的 DDI 预测方法 SEAL-CI (Semi-supervised hierarchical graph classification)<sup>[29]</sup>、基于分子指纹和图卷积网络的 DDI 预测方法 NFP-GCN (Molecular fingerprint graph convolutional network)<sup>[30]</sup>、基于键感知消息传递神经网络和图卷积网络的 DDI 预测方法 MIRACLE (Multi-view graph contrastive representation learning)<sup>[13]</sup>.

对于 ZhangDDI 和 ChCh-Miner 数据集, 本文参照文献 [13] 的数据划分方式, 所有数据样本按照 4:1 的比例分为训练集和测试集, 并在训练集中随机选择 1/4 的样本作为验证集. 当连续训练 10 轮且模型在验证集上的最佳精度没有改变时, 模型停止训练. 所有实验结果都是通过 5 次独立实验进行统计分析得到. 在原子特征网络中, 原子特征维度设置为 115, 化学键特征维度设置为 13. 在分子特征网络中, 注意力头数设置为 3. 目标函数中的系数  $\alpha$  和  $\beta$  分别设置为 1 和 0.8, 实验基于 Pytorch 1.6.0.

### 3.2 对比实验分析

在 ZhangDDI 和 ChCh-Miner 两个数据集上的实验结果分别如表 1 和表 2 所示. 由表 1 可知, 相比于 12 种对比算法, 本文在所有指标上都取得最好结果. 与基于键感知消息传递神经网络和图卷积网络的 DDI 预测方法<sup>[13]</sup> 相比, 本文方法虽然在 AUPRC 指标上取得次优结果, 但是在 AUROC 和 F1 指标上表现更好、更鲁棒. 由表 2 的实验结果可知, 当药物种类数显著增加时, 本文方法超过了所有对比方法, 且优势更加明显.

表 1 ZhangDDI 数据集上的对比实验结果  
Table 1 Comparison experimental results on ZhangDDI dataset

模型	AUROC	AUPRC	F1
NN <sup>[24]</sup>	67.81±0.25	52.61±0.27	49.84±0.43
LP-Sub <sup>[25]</sup>	93.39±0.13	89.15±0.13	79.61±0.16
LP-SE <sup>[25]</sup>	93.48±0.25	89.61±0.19	79.83±0.61
LP-OSE <sup>[25]</sup>	93.50±0.24	90.31±0.82	80.41±0.51
MF-Ens <sup>[11]</sup>	95.20±0.14	92.51±0.15	85.41±0.16
SSP-MLP <sup>[1]</sup>	92.51±0.15	88.51±0.66	80.69±0.81
GCN <sup>[26]</sup>	91.91±0.62	88.73±0.84	81.61±0.39
GIN <sup>[27]</sup>	81.45±0.26	77.16±0.16	64.15±0.16
Att-auto <sup>[12]</sup>	92.84±0.61	90.21±0.19	70.96±0.39
GAT <sup>[28]</sup>	91.49±0.29	90.69±0.10	80.93±0.25
SEAL-CI <sup>[29]</sup>	92.93±0.19	92.82±0.17	84.74±0.17
NFP-GCN <sup>[30]</sup>	93.22±0.09	93.07±0.46	85.29±0.38
MIRACLE <sup>[13]</sup>	98.95±0.15	<b>98.17±0.06</b>	93.20±0.27
本文方法	<b>99.14±0.01</b>	97.97±0.02	<b>93.79±0.28</b>

表 2 ChCh-Miner 数据集上的对比实验结果  
Table 2 Comparison experimental results on ChCh-Miner dataset

模型	AUROC	AUPRC	F1
GCN <sup>[26]</sup>	82.84±0.61	84.27±0.66	70.54±0.87
GIN <sup>[27]</sup>	70.32±0.87	72.41±0.63	65.54±0.97
GAT <sup>[28]</sup>	85.84±0.23	88.14±0.25	76.51±0.38
SEAL-CI <sup>[29]</sup>	90.93±0.19	89.38±0.39	84.74±0.48
NFP-GCN <sup>[30]</sup>	92.12±0.09	93.07±0.69	85.41±0.18
MIRACLE <sup>[13]</sup>	96.15±0.29	95.57±0.19	92.26±0.09
本文方法	<b>98.45±0.31</b>	<b>99.79±0.04</b>	<b>96.51±0.84</b>

1) 与 6 种基于传统机器学习 DDI 预测方法相比, 本文方法在 ZhangDDI 数据集上取得的 3 项指标结果至少提高 4%, 5%, 8%. 这是由于基于相似性的 DDI 预测方法是通过传统机器学习计算多种药物特征的相似度从而预测出 DDI 结果, 而药物分子往往具有复杂的结构特性, 不能简单地由一种或几种特征刻画, 且传统机器学习提取深层特征的能

力有限, 因此这类方法效果并不好. 不同于此类方法, 本文方法没有选择具体的药物分子特征, 而是通过深度学习方法同时从药物原子和分子层面学习药物分子的深层特征.

2) 与 4 种基于分子特征网络的 DDI 预测方法相比, 本文方法在 ZhangDDI 数据集上取得的 3 项指标结果至少提高 6%, 7%, 12%, 而在 ChCh-Miner 数据集上, 相应结果至少提高 13%, 12%, 20%. 基于分子特征网络的 DDI 预测方法直接从药物分子层面学习不同药物分子的相互作用关系, 忽略了分子内部的结构特性, 算法性能受制于药物分子特征表示的好坏. 而本文方法首先从原子层面学习得到每个药物分子的特征, 然后结合监督学习和对比学习, 不断优化得到的药物分子特征, 并基于这些分子特征学习得到不同药物分子的相互作用关系, 在理论上更具优势. 实验结果也证明了本文方法的优越性.

3) 与 3 种基于原子特征和分子特征网络的 DDI 预测方法相比, 本文方法除 1 个指标取得次优结果外, 都能取得最好的结果. 虽然这些方法都能从原子和分子层面提取药物分子特征, 但它们都缺乏明确的机制学习原子和化学键之间、分子之间的不同注意力信息. 例如, 基于键感知消息传递神经网络和图卷积网络的 DDI 预测方法 MIRACLE<sup>[13]</sup>, 虽然在 AUPRC 指标上具有良好的竞争力, 但其整体性能不如本文方法. 这是因为 MIRACLE 只考虑不同原子间的消息传递, 并没有考虑边的特征信息. 而本文方法可同时考虑不同原子及其相关边之间的消息传递, 且可学习不同原子间的注意力信息. 本文方法可以在原子和分子层面同时进行具有不同作用的注意力学习, 因此, 本文方法在综合性能表现上更优越.

### 3.3 消融实验分析

#### 3.3.1 多层注意力机制的消融实验

为验证本文提出的多层次注意力网络的有效性, 我们在两个数据集上针对基于注意力机制的原子特征网络和分子特征网络分别进行消融实验.

关于原子特征网络和分子特征网络的消融实验结果分别见表 3 和表 4. 实验结果表明, 无论是原子特征网络, 还是分子特征网络, 删除其注意力机制后, 模型性能都会显著下降. 如果缺乏基于注意力机制的原子特征网络, 那么在最终的分子表示中将会丢失分子内部原子和边的特征信息, 而这会直接影响药物分子的特征质量; 同样, 如果没有基于注意力机制的分子特征网络, 那么将会丢失分子间的相互作用信息, 导致模型只会根据两个独立的分子

表 3 原子特征网络的消融实验结果

Table 3 Ablation experimental results on atomic feature network

数据集	算法	AUROC	AUPRC	F1
ZhangDDI	无注意力的原子网络	98.70±0.20	96.89±0.50	90.46±1.18
	本文方法	<b>99.14±0.01</b>	<b>97.97±0.02</b>	<b>93.79±0.28</b>
ChCh-Miner	无注意力的原子网络	95.90±0.99	99.18±0.15	96.23±0.34
	本文方法	<b>98.45±0.31</b>	<b>99.79±0.04</b>	<b>96.51±0.84</b>

表 4 分子特征网络的消融实验结果

Table 4 Ablation experimental results on molecular feature network

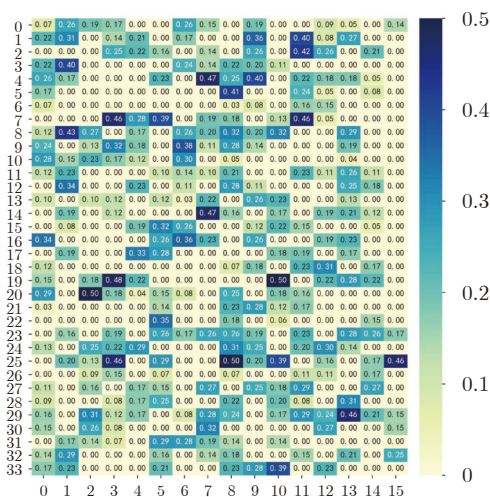
数据集	算法	AUROC	AUPRC	F1
ZhangDDI	无注意力的分子网络	98.82±0.27	97.18±0.68	91.60±1.84
	本文方法	<b>99.14±0.01</b>	<b>97.97±0.02</b>	<b>93.79±0.28</b>
ChCh-Miner	无注意力的分子网络	95.78±1.29	99.19±0.38	95.19±1.45
	本文方法	<b>98.45±0.31</b>	<b>99.79±0.04</b>	<b>96.51±0.84</b>

特征来进行 DDI 预测. 根据上述分析, 本文提出的基于多层次注意力机制的原子特征和分子特征网络确实有助于提高药物分子的特征质量和模型性能.

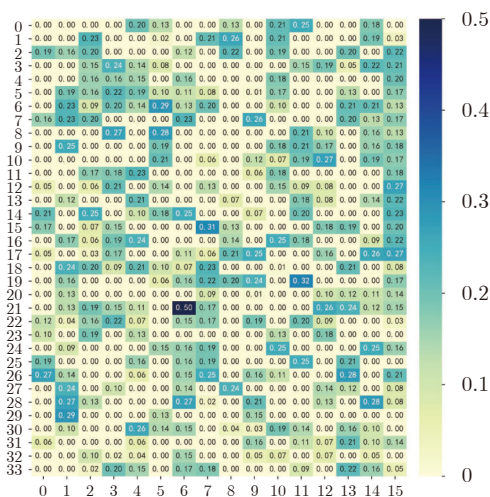
为进一步验证本文提出注意力机制的有效性, 我们将分子特征网络经过注意力机制计算得到的分子之间相互作用的注意力分数进行可视化. 作为示例说明, 这里展示在 ZhangDDI 数据集上抽取的一个分子 (记为 A) 和另外 542 个分子之间的注意力分数, 其中与 A 发生相互作用和不发生相互作用的分子各占一半数量. 图 3 是分子 A 与其他 542 个分子之间的注意力分数经过归一化后的可视化结果, 其中, 图 3(a) 是与 A 发生相互作用的药物分子的注意力分数可视化, 图 3(b) 是不与 A 发生相互作用的药物分子的注意力分数可视化. 经过计算可得, 与 A 发生相互作用的药物分子的注意力分数之和占注意力分数总和的 56.87%, 平均注意力分数是 0.21; 而不与 A 发生相互作用的药物分子的注意力分数之和则占注意力分数总和的 43.13%, 平均注意力分数是 0.16. 即当药物分子之间存在相互作用时, 其注意力分数大于没有相互作用时的注意力分数. 这在一定程度上说明通过本文提出的多层次注意力机制计算得到两个药物分子之间的注意力分数越大, 则两个药物分子发生相互作用的可能性也越大.

### 3.3.2 位置编码的消融实验

为了验证本文提出的基于分子质心的位置编码方法的有效性, 本节将研究在有无位置编码和传统位置编码条件下模型性能的差别. 表 5 是在两个数



(a) 与 A 发生相互作用的药物分子的注意力分数  
(a) Attention scores of drug molecules that interact with A



(b) 不与 A 发生相互作用的药物分子的注意力分数  
(b) Attention scores of drug molecules that do not interact with A

图 3 药物分子之间注意力分数的可视化

Fig. 3 Visualization of attention scores between drug molecules

据集上, 本文方法在有无基于分子质心的位置编码和传统位置编码条件下的实验结果. 可以看到, 本文提出的位置编码方法可以显著提高模型的性能, 且使得模型具有更稳定的性能表现.

如果删去本文提出的位置编码, 在原子特征网络学习过程中, 将有可能丢失原子节点在药物分子图结构中的相对位置信息; 而采用传统位置编码, 将使模型仅依赖于各原子节点在 SMILES 序列表示中出现的先后顺序, 进行原子特征学习. 正如前面提到, SMILES 序列是分子的一维线性化表示, 因此仅依赖 SMILES 序列中原子的先后次序进行原子特征学习, 无法充分学到分子的图结构特征信息.

表 5 位置编码对模型性能影响的对比结果

Table 5 Comparison results of the impact of positional encoding on model performance

数据集	算法	AUROC	AUPRC	F1
ZhangDDI	无位置编码	98.91±0.26	97.46±0.60	91.38±2.11
	传统位置编码	99.02±0.21	97.68±0.53	92.70±1.35
	本文方法	<b>99.14±0.01</b>	<b>97.97±0.02</b>	<b>93.79±0.28</b>
ChCh-Miner	无位置编码	95.62±2.79	99.11±0.63	96.12±0.58
	传统位置编码	97.54±0.46	99.66±0.06	94.73±0.54
	本文方法	<b>98.45±0.31</b>	<b>99.79±0.04</b>	<b>96.51±0.84</b>

此外, 本文提出的位置编码也有助于提高模型的收敛速度. 图 4 是本文方法在有无位置编码条件下模型性能的收敛曲线, 可以清楚地看到, 本文提出的位置编码可以显著提高模型的收敛速度.

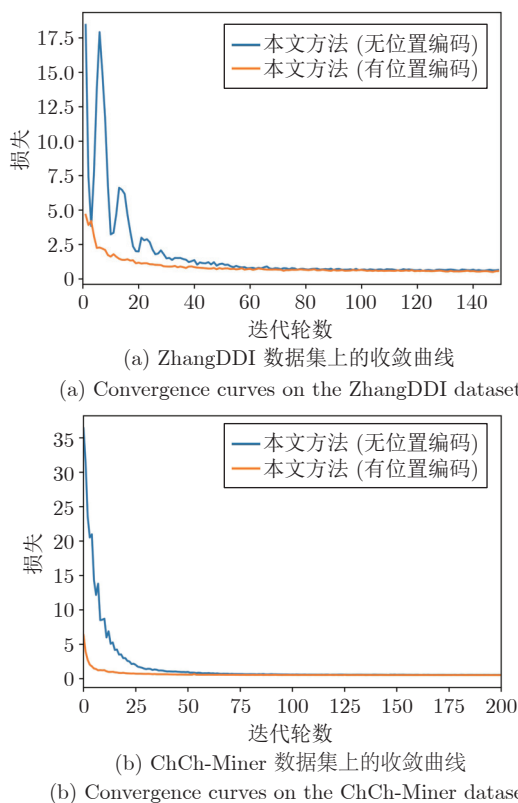


图 4 位置编码对模型收敛性能的影响

Fig. 4 The effect of positional encoding on model convergence performance

与无位置编码的模型相比, 本文方法可以在更少的迭代轮数条件下取得更快更好的模型性能. 综合表 5 和图 4 可知, 实验结果从侧面进一步证明: 基于分子质心的位置编码可以显著提高药物分子中不同原子的编码效率, 进而提高模型的收敛速度; 同时, 通过分子质心引入药物分子的空间结构信息, 有助于模型提取更丰富的药物分子结构特征, 从而

进一步提升模型的 DDI 预测精度.

综合上述实验结果可知, 本文提出的多层次注意力机制和基于分子质心的位置编码方法都是有效且不可或缺, 有助于提高药物分子中不同原子的编码效率和不同药物分子相互作用预测的精度.

### 3.3.3 损失函数的消融实验

为验证本文引入的自蒸馏约束项和对比学习损失项的有效性, 我们在两个数据集上分别针对 2 项损失函数进行消融实验, 并进一步将对比学习损失项替换为基于互信息的噪声对比估计模型 (Mutual information noise contrastive estimation, infoNCE)<sup>[22]</sup>, 检验不同对比损失函数对模型性能的影响. 此外, 我们还改变正负样本采样方式, 检验采样方式对本文对比学习损失函数的影响. 对于每个药物分子, 选取其在分子特征网络的输出特征作为锚点. 我们这里选择 2 种正负样本的采样方式, 其中一种是将锚点的一阶邻居和非一阶邻居在原子特征网络的输出特征分别作为正样本和负样本, 即本文实验采用的方法; 另外一种是将锚点的二阶邻居在原子特征网络的输出特征作为正样本, 其他节点作为负样本, 进行对比学习.

表 6 是不同损失函数对模型性能影响的对比实验结果. 可以看到, 无论对于 ZhangDDI 还是 ChCh-Miner 数据集, 在没有基于自蒸馏的正则化约束项或基于对比学习的损失项时, 模型性能都有一定程度的降低, 即这 2 项损失函数对提高模型的性能都不可或缺. 对于 2 种正负样本采样方式来说, 实验结果表明本文选取的采样方式更好. 这个结果也表明正负样本采样方式对模型性能具有一定的影响. 对于不同对比学习损失函数来说, 在 ChCh-Miner 数据集上, 本文方法与 infoNCE 相比, 取得次优结果; 但是在 ZhangDDI 数据集上, 本文方法取得最优结果, 且相对更稳定. 这是因为 infoNCE 通过自归一化重要性采样来优化互信息的下界, 需要相对较多的负样本; 而本文采用的 JSD 方法则对负样本数相对不敏感, 性能也相对更稳定. 由此可见, 不同对比学习损失函数对模型性能具有一定的影响, 本文采用的对比学习损失函数具有一定优势.

### 3.4 参数敏感性分析

考虑到损失函数会直接影响模型的性能, 本节选择对损失函数中的参数  $\alpha$  和  $\beta$  进行敏感性分析. 在上述实验中,  $\alpha$  和  $\beta$  分别取值 1 和 0.8. 为分析 2 个参数的敏感性,  $\alpha$  取值范围为 {0.2, 0.4, 0.6, 0.8, 1},  $\beta$  取值范围为 {0.2, 0.4, 0.6, 0.8, 1}. 当分析  $\alpha$  的敏感性时,  $\beta$  固定, 取值为 0.8; 当分析  $\beta$  的敏感性时,  $\alpha$  固定, 取值为 1. 图 5 和图 6 分别是 2 个参

表 6 损失函数对模型性能影响的对比结果  
Table 6 Comparison results of the impact of loss function on model performance

数据集	算法	AUROC	AUPRC	F1
ZhangDDI	无自蒸馏约束项	98.71±0.01	96.87±0.02	89.53±0.54
	无对比学习损失项	94.19±0.06	79.62±0.31	73.59±0.39
	infoNCE 对比损失项	99.10±0.05	97.91±0.09	92.87±0.60
	不同采样方式的对比损失项	99.13±0.02	97.97±0.04	93.15±0.61
	本文方法	<b>99.14±0.01</b>	<b>97.97±0.02</b>	<b>93.79±0.28</b>
ChCh-Miner	无自蒸馏约束项	97.55±3.24	99.48±0.87	96.34±1.01
	无对比学习损失项	58.70±5.00	90.30±1.06	94.89±0.57
	infoNCE 对比损失项	<b>98.59±0.20</b>	<b>99.80±0.03</b>	<b>97.31±0.20</b>
	不同采样方式的对比损失项	98.38±0.01	99.78±0.00	95.67±0.09
	本文方法	98.45±0.31	99.79±0.04	96.51±0.84

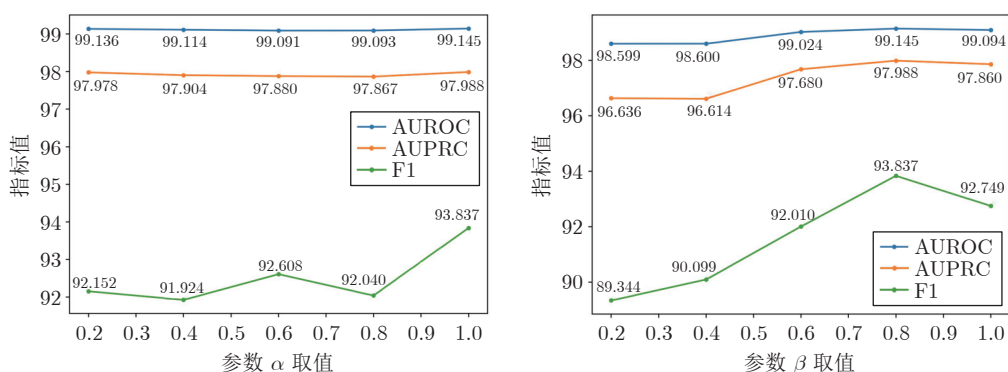


图 5 在 ZhangDDI 数据集上不同  $\alpha$  和  $\beta$  取值对模型性能的影响

Fig. 5 The effects of different  $\alpha$  and  $\beta$  on model performance on ZhangDDI dataset

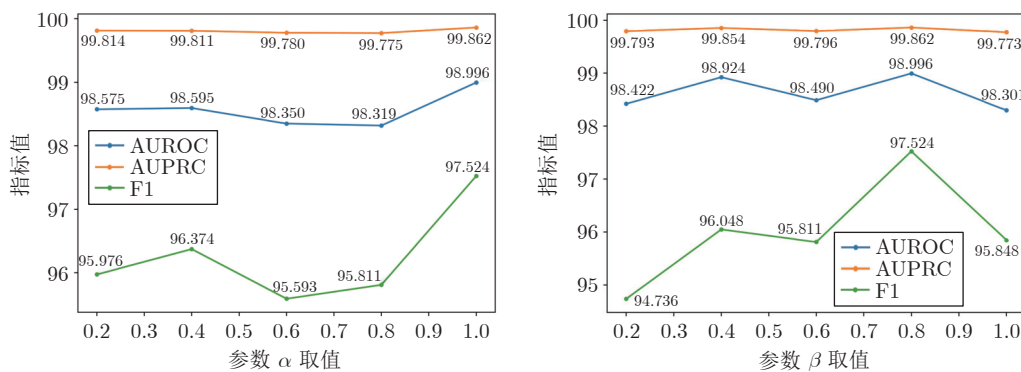


图 6 在 ChCh-Miner 数据集上不同  $\alpha$  和  $\beta$  取值对模型性能的影响

Fig. 6 The effects of different  $\alpha$  and  $\beta$  on model performance on ChCh-Miner dataset

数在两个数据集上的实验结果. 可以看到, 在两个数据集上, 不同参数  $\alpha$  和  $\beta$  的取值对于本文方法在指标 AUROC 和 AUPRC 上的结果影响相对较小, 而在 F1 指标上的结果影响相对较大. 综合  $\alpha$  和  $\beta$  对模型性能的敏感性分析可知, 在  $\alpha$  和  $\beta$  分别取值 1 和 0.8 时, 本文方法可以取得最好的实验结果.

## 4 总结与展望

针对药物相互作用预测的应用研究需求和不同药物分子及其内部不同原子对 DDI 预测结果具有不同作用等问题, 本文提出一种基于多层次注意力机制和消息传递神经网络的药物相互作用预测方

法。通过设计基于注意力机制的原子特征网络和分子特征网络,从两个不同层次分别学习分子内不同原子和化学键以及不同分子间的特征信息,并结合本文提出的基于分子质心的位置编码,提高药物分子编码的效率,从而提高 DDI 预测结果的准确性。通过大量对比实验和消融实验验证了本文方法的有效性和优越性。

本文提出的方法虽然可以从原子和分子层面提取药物分子特征信息,但是利用的仅仅只是包含药物分子信息的 SMILES 序列,并没有充分利用其他的药物相关信息。下一步的研究工作可以考虑如何充分利用药物分子结构式之外的信息,如同时利用药物分子结构式和包含药物相互作用关系的文本等多种信息,进一步提高模型预测潜在药物相互作用关系的能力。

## References

- 1 Ryu J Y, Kim H U, Lee S Y. Deep learning improves prediction of drug-drug and drug-food interactions. *The National Academy of Sciences (NAS)*, 2018, **115**(18): E4304–E4311
- 2 Sun M Y, Zhao S D, Gilvary C, Elemento O, Zhou J Y, Wang F. Graph convolutional networks for computational drug development and discovery. *Briefings in Bioinformatics*, 2020, **21**(3): 919–935
- 3 Qiu Y, Zhang Y, Deng Y F, Liu S C, Zhang W. A comprehensive review of computational methods for drug-drug interaction detection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2022, **19**(4): 1968–1985
- 4 Abbas K, Abbasi A, Dong S, Niu L, Yu L H, Chen B L, et al. Application of network link prediction in drug discovery. *BMC Bioinformatics*, 2021, **22**(1): 1–21
- 5 Hou Mei-Hao. Prediction of Adverse Drug Interactions Based on Graph Neural Network [Master thesis], Shandong University, China, 2020.  
(侯美好. 基于图神经网络的药物不良相互作用预测 [硕士学位论文], 山东大学, 中国, 2020.)
- 6 Lee G, Park C, Ahn J. Novel deep learning model for more accurate prediction of drug-drug interaction effects. *BMC Bioinformatics*, 2019, **20**(1): 1–8
- 7 Karim M R, Cochez M, Jares J B, Uddin M, Beyan O, Decker S. Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-lstm network. In: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. New York, USA: ACM, 2019. 113–123
- 8 Yan C, Duan G H, Zhang Y Y, Wu F X, Pan Y, Wang J X. IDNDDI: An integrated drug similarity network method for predicting drug-drug interactions. In: Proceedings of the 15th International Symposium on Bioinformatics Research and Applications. Cham, Switzerland: Springer, 2019. 89–99
- 9 Liu S C, Huang Z Y, Qiu Y, Chen Y P P, Zhang W. Structural network embedding using multi-modal deep auto-encoders for predicting drug-drug interactions. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. San Diego, USA: IEEE, 2019. 445–450
- 10 Deng Y F, Xu X R, Qiu Y, Xia J B, Zhang W, Liu S C. A multimodal deep learning framework for predicting drug-drug interaction events. *Bioinformatics*, 2020, **36**(15): 4316–4322
- 11 Zhang W, Chen Y L, Liu F, Luo F, Tian G, Li X H. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinformatics*, 2017, **18**(1): 1–12
- 12 Ma T F, Xiao C, Zhou J Y, Wang F. Drug similarity integration through attentive multi-view graph auto-encoders. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: Morgan Kaufmann, 2018. 3477–3483
- 13 Wang Y H, Min Y S, Chen X, Wu J. Multi-view graph contrastive representation learning for drug-drug interaction prediction. In: Proceedings of the Web Conference (WWW'21). New York, USA: ACM, 2021. 2921–2933
- 14 Yan C, Duan G H, Zhang Y Y, Wu F X, Pan Y, Wang J X. Predicting drug-drug interactions based on integrated similarity and semi-supervised learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2022, **19**(1): 168–179
- 15 Jin B, Yang H Y, Xiao C, Zhang P, Wei X P, Wang F. Multi-task dyadic prediction and its application in prediction of adverse drug-drug interaction. In: Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI, 2017. 367–373
- 16 Lin X, Quan Z, Wang Z J, Ma T F, Zeng X X. KGNN: Knowledge graph neural network for drug-drug interaction prediction. In: Proceedings of the 29th International Joint Conference on Artificial Intelligence. Yokohama, Japan: Morgan Kaufmann, 2020. 2739–2745
- 17 Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 30th Annual Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates, 2017. 5998–6008
- 18 Chen J W, Zheng S J, Song Y, Rao J H, Yang Y D. Learning attributed graph representations with communicative message passing transformer. In: Proceedings of the 30th International Joint Conference on Artificial Intelligence. Montreal, Canada: Morgan Kaufmann, 2021. 2242–2248
- 19 Landrum G. Rdkit documentation [Online], available: <https://www.rdkit.org/>, July 29, 2022
- 20 Gilmer J, Schoenholz S S, Riley P F, Vinyals O, Dahl G E. Neural message passing for quantum chemistry. In: Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: ACM, 2017. 1263–1272
- 21 Zhang L F, Song J B, Gao A, Chen J W, Bao C L, Ma K S. Be your own teacher: Improve the performance of convolutional neural networks via self-distillation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019. 3713–3722
- 22 Hjelm R D, Fedorov A, Lavoie-Marchildon S, Grewal K, Bachman P, Trischler A, et al. Learning deep representations by mutual information estimation and maximization. In: Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA: OpenReview.net, 2019.
- 23 Zitnik M, Sosis R, Leskovec J. BioSNAP Datasets: Stanford biomedical network dataset collection [Online], available: <http://snap.stanford.edu/biodata>, July 30, 2022
- 24 Vilar S, Harpaz R, Uriarte E, Santana L, Rabadan R, Friedman C. Drug-drug interaction through molecular structure similarity analysis. *Journal of the American Medical Informatics Association*, 2012, **19**(6): 1066–1074
- 25 Zhang P, Wang F, Hu J Y, Sorrentino R. Label propagation prediction of drug-drug interactions based on clinical side effects. *Scientific Reports*, 2015, **5**(1): 1–10

- 26 Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. In: Proceedings of the 5th International Conference on Learning Representations. Toulon, France: OpenReview.net, 2017.
- 27 Xu K Y L, Hu W H, Leskovec J, Jegelka S. How powerful are graph neural networks? In: Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA: OpenReview.net, 2019.
- 28 Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. In: Proceedings of the 6th International Conference on Learning Representations. Montréal, Canada: OpenReview.net, 2018.
- 29 Li J, Rong Y, Cheng H, Meng H L, Huang W B, Huang J Z. Semi-supervised graph classification: A hierarchical graph perspective. In: Proceedings of the World Wide Web Conference. San Francisco, USA: ACM, 2019. 972–982
- 30 Duvenaud D, Maclaurin D, Iparraguirre J, Bombarell R, Hirzel T, Aspuru-Guzik A, et al. Convolutional networks on graphs for learning molecular fingerprints. In: Proceedings of the 28th Annual Conference on Neural Information Processing Systems. Montreal, Canada: MIT, 2015.



**饶晓洁** 华南理工大学计算机科学与工程学院硕士研究生. 主要研究方向为机器学习与人工智能在生物医药领域的应用.

E-mail: rxj19971214@163.com

**(RAO Xiao-Jie** Master student at the School of Computer Science and

Engineering, South China University of Technology. Her research interest covers application of machine learning and artificial intelligence in the field of bio-medicine.)

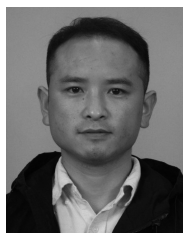


**张通** 华南理工大学计算机科学与工程学院教授. 2016 年获得澳门大学软件工程专业博士学位. 主要研究方向为情感计算, 进化计算, 神经网络和其他机器学习技术及其应用.

E-mail: tony@scut.edu.cn

**(ZHANG Tong** Professor at the

School of Computer Science and Engineering, South China University of Technology. He received his Ph.D. degree in software engineering from University of Macau in 2016. His research interest covers affective computing, evolutionary computation, neural network, and other machine learning techniques and their applications.)



**孟献兵** 华南理工大学计算机科学与工程学院助理研究员. 2019 年获中南大学博士学位. 主要研究方向为计算智能和人工智能算法及其应用. 本文通信作者.

E-mail: axbmeng@gmail.com

**(MENG Xian-Bing** Assistant re-

search fellow at the School of Computer Science and Engineering, South China University of Technology. He received his Ph.D. degree from Central South University in 2019. His research interest covers computational intelligence and artificial intelligence algorithms and their applications. Corresponding author of this paper.)



**陈俊龙** 华南理工大学计算机科学与工程学院教授. 主要研究方向为控制论, 智能系统和计算智能

E-mail: philipchen@scut.edu.cn

**(CHEN C. L. Philip** Professor at the School of Computer Science and Engineering, South China Univer-

sity of Technology. His research interest covers cybernetics, intelligent systems, and computational intelligence.)