

基于药物互作网络的协同与拮抗预测研究

刘文斌*^{①②} 陈杰^① 方刚^② 石晓龙^② 许鹏^{②③}

^①(温州大学计算机与人工智能学院 温州 325035)

^②(广州大学计算科技研究院 广州 510006)

^③(黔南民族师范学院计算机与信息学院 都匀 558000)

摘要: 药物的协同与拮抗关系预测, 有助于药物的使用安全及组合用药的发展。该文从药物互作网络(DDINet)出发, 基于网络拓扑结构构造分类特征, 提出一种预测药物协同和拮抗关系的方法。从特征选择结果可知, 根据药物与其公共邻居节点关系构造的特征表现出了明显的正负样本分布差距, 能有效地反映出药物的协同或拮抗关系。在使用不同特征分类器的分类结果中, 最优AUC和分类精度值分别达到了0.9687和0.9187。而在协同与拮抗关系预测结果中, 其预测精度值达到了0.45和0.75以上。这说明基于网络拓扑结构的方法能有效对药物协同和拮抗关系进行分类和预测。与传统基于药物功能、结构、靶基因等相似性特征的方法相比, 该方法计算简单高效, 将会有效促进组合用药的发展。

关键词: 药物相互作用预测; 网络拓扑结构; 药物协同; 药物拮抗

中图分类号: TP301

文献标识码: A

文章编号: 1009-5896(2020)06-1420-08

DOI: 10.11999/JEIT190867

Prediction of Drug Synergy and Antagonism Based on Drug-Drug Interaction Network

LIU Wenbin^{①②} CHEN Jie^① FANG Gang^② SHI Xiaolong^② XU Peng^{②③}

^①(College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325035, China)

^②(Institute of Computing Science and Technology, Guangzhou University, Guangzhou 510006, China)

^③(School of Computer Science and Information Technology, Qiannan Normal University for Nationalities, Duyun 558000, China)

Abstract: Accurately predicting the synergistic and antagonistic relationship of drugs is helpful to the safety of drug use and the development of drug combination. A method for predicting drug synergy and antagonistic is proposed, which based on the Drug-Drug Interaction Network (DDINet) and its topological structure. From the result of feature selection, it can be seen that the feature constructed based on the interaction between the drug and its common neighbor node shows an obvious difference in the distribution of positive and negative samples, which can effectively reflect the drug synergy or antagonism. In the classification results using different feature classifiers, the optimal Area Under the Curve (AUC) and classification accuracy value reached 0.9687 and 0.9187 respectively. In the prediction results of synergy and antagonism, the prediction accuracy also reached above 0.45 and 0.75. This shows that the method based on network topology can effectively classify and predict the synergistic and antagonistic effects of drugs. Compared with the traditional methods based on similarity features of drug function, structure, target gene, etc, this method is simple and efficient to calculate, and can effectively promote the development of combination drugs.

Key words: Prediction Drug-Drug Interactions(DDIs); Network topology; Synergy; Antagonism

收稿日期: 2019-11-01; 改回日期: 2020-01-15; 网络出版: 2020-02-18

*通信作者: 刘文斌 wbliu6910@126.com

基金项目: 国家重点研发计划(2019YFA0706402), 国家自然科学基金(61572367, 61573017, 61972107, 61972109)

Foundation Items: The National Key R&D Program of China (2019YFA0706402), The National Natural Science Foundation of China (61572367, 61573017, 61972107, 61972109)

1 引言

药物相互作用(Drug-Drug Interactions, DDIs)是指患者在一定时间内先后服用两种或两种以上药物后所发生的药效变化^[1]。在临床实践中,一般将DDIs分为协同(Synergism)作用和拮抗(Antagonism)作用^[2]。协同作用可以增强疗效或减轻药物不良反应,而拮抗作用可导致疗效降低或副作用增强,还可能发生一些异常反应,干扰治疗,加重病情^[3,4]。组合药物治疗能有效提高药物疗效,改善患者身体素质,如今已经被大范围采用^[5],但是随着批准药物数量的增加,药物之间潜在的相互作用变多,这在一定程度上增加了组合用药的风险。据统计,超过30%的药物副作用是由DDIs导致^[6]。因此,发现和预测DDIs不仅可以在临床实践中提高用药安全,也可以促进组合用药治疗在临床上的进一步发展。

传统的DDIs发现依赖于临床实验,但是这种方式需要花费大量的时间和成本,难以用于大规模发现DDIs。随着DDIs数据的积累、药物相关信息的完善以及大数据技术的发展,目前很多研究都是基于药物与蛋白质互作机制及其结构信息计算相似性特征^[7-10],并结合统计学^[11]、机器学习^[12]和深度学习^[13]的方法来预测DDIs,其主要思想是“分子结构相似或靶蛋白相似的药物更有可能具有相似的药理性质”。文献^[9]提出了基于药物分子结构相似性的方法来预测DDIs,之后又针对DrugBank数据库的DDIs数据,提出了一种基于药物相互作用指纹图谱(drug Interaction Profile Fingerprints, IPFs)方法^[10],进一步提高了DDIs的预测精度。Luo等人^[14]提出了一种基于化合物-蛋白质相互作用(Chemical-Protein Interaction, CPI)的方法来预测单个药物与其他药物的DDIs。Zhang等人^[15]从美国食品药品监督管理局(Food and Drug Administration, FDA)和有机小分子生物活性数据库(PubChem)中分别提取出药物不良反应信息和化学结构信息,结合标签传播框架来预测DDIs导致的副作用。Liu等人^[16]根据药物的化学互作、药物靶点互作和药物靶点在KEGG通路的富集信息,结合随机森林算法(Random Forest, RF)实现对DDIs的预测。除此之外,还有一些研究是从药物代谢动力学和药效学原理来对DDIs进行预测^[17,18]。Gottlieb等人^[18]针对一种常见代谢酶(Cytochrome P450, CYP)提出了一种“药物相互作用推断(INferring Drug Interactions, INDI)”的方法,这种方法可以同时推断与CYP相关的药物代谢动力学DDIs和与CYP无关的药效学DDIs。这些方法的优势在于它们都是基于现有的DDIs数据来对潜在DDIs进行预测,这缩短了发现

周期和减少研究成本,有利于药物的使用安全和管理。

现有的DDIs预测方法是从药物功能、结构、靶向关系等出发计算相似性特征进行DDIs关系预测。虽然这些方法能预测出部分潜在DDIs,但是忽略了药物交互网络自身具有的分布特征。本文从网络拓扑结构出发,提出一种基于药物交互网络(Drug-Drug Interaction Network, DDINet)拓扑结构的DDIs预测方法。把在DrugBank中提取出的DDIs数据,按照DDIs本身的协同或拮抗关系分为正负样本,将DDIs预测问题转换为一个二分类问题。为了更好地实现DDIs的分类和预测,本文基于DDIs关系构建DDINet,根据拓扑结构信息构造特征并选出重要特征,应用支持向量机(Support Vector Machine, SVM)模型实现对DDIs的预测。在结果中,AUC和分类精度的最优值分别为0.9687和0.9187,而对正负样本的最优平均预测精度为0.5007和0.7619。这表明基于网络拓扑结构构造特征的方法能有效实现对现有DDIs的分类和预测,也进一步说明了本文方法的合理性和可行性。

2 研究方法

2.1 实验数据集

ATC(Anatomical Therapeutic Chemical)是药物的解剖学、治疗学及化学分类学的简称,由7位字母和数字组成,共分为5个级别。其中第一级由字母表示,代表了药物作用的解剖学主族。本文从DrugBank(<https://www.drugbank.ca>)数据库^[19]中提取了所有经过FDA认证批准的小分子药物ATC编码,选择ATC分类系统中解剖学主族为“L”(抗肿瘤药和免疫机能调节药)的药物,得到267种小分子药物,并获取这些药物相关的所有DDIs数据信息。

本文将有利于药效提升、副作用减轻等协同关系的DDIs设为正样本(Positive);将加快药物代谢导致药效降低、增加副作用等拮抗关系的DDIs设为负样本(Negative),以此构建药物交互网络(Drug-Drug Interaction Network, DDINet)。网络中的每一个节点代表一种药物,每一条边代表一种药物相互作用关系。因此,DDINet中的每条边均为一个样本,其中协同关系的边为正样本,拮抗关系的边为负样本。为了增加文章的可读性,本文在下文中的分类和预测部分,统一将协同和拮抗关系的边称为正负样本,而不再描述样本的具体含义。为了更好地分类和预测DDIs,本文过滤掉了DDINet中两个节点所共有1阶邻居节点数小于20的边(样本),最终得到一个顶点数为247,总边数为19403

条的DDINet, 其中协同关系边数(正样本数)为7031, 拮抗关系边数(负样本数)为12372。

2.2 构造特征

2.2.1 基于邻居节点构造特征

假设 D_i 和 D_j 是DDINet中的两个节点(药物), 则他们1阶邻居节点之间的主要拓扑关系总共有以下两种可能: (1) 存在公共邻居节点, 这些节点内部存在相互关系(如图1(a)所示); (2) 存在直接相互关系的非公共邻居节点(如图1(b)所示)。

本文根据 D_i 和 D_j 的1阶邻居节点拓扑关系构造了5个基础特征。期望这些特征能在一定程度上反映药物间的相互关系, 并用于后期DDIs分类和预测。特征的具体构造规则如下:

x_1 : D_i 和 D_j 的非公共邻居节点之间是协同关系(Positive)的边数;

x_2 : D_i 和 D_j 的非公共邻居节点之间是拮抗关系(Negative)的边数;

x_3 : D_i 和 D_j 与他们的公共邻居节点同时为协同关系(Positive)的边数;

x_4 : D_i 和 D_j 与他们的公共邻居节点同时为拮抗关系(Negative)的边数;

x_5 : D_i 和 D_j 与他们的公共邻居节点正好是相反关系(Positive & Negative)的边数。

假设药物 D_i 和 D_j 的1阶邻居节点拓扑结构如图1所示, 则根据特征定义可知 $x_1 = 4$, $x_2 = 4$, $x_3 = 3$, $x_4 = 1$, $x_5 = 2$ 。

5个基础特征中, x_1 和 x_2 是为了研究药物邻居节点之间的相互作用对药物相互作用的反映程度, 而 x_3 , x_4 和 x_5 则是研究药物本身与邻居节点的关系对药物相互作用的反映程度。

2.2.2 基于邻居节点拓扑结构进一步构造特征

x_1, x_2, \dots, x_5 是直接基于网络拓扑结构构造的基础特征, 这种特征可能容易受到其他因素的干

扰, 并影响模型的性能和预测精度。因此, 为了提升特征的分类性能, 增强特征的抗干扰能力, 本文基于 $x_1 \sim x_5$ 这5个基本特征, 再结合其他拓扑结构关系重新构造了一系列特征 y_1, y_2, \dots, y_5 , 公式为

$$y_1 = \ln(x_1/x_2) \quad (1)$$

$$y_2 = \ln(x_3/x_4) \quad (2)$$

$$y_3 = x_3/(x_3 + x_4 + x_5) \quad (3)$$

$$y_4 = x_4/(x_3 + x_4 + x_5) \quad (4)$$

$$y_5 = n_{ab}/(n_a + n_b - n_{ab}) \quad (5)$$

其中 n_a, n_b, n_{ab} 分别表示 D_i 的1阶邻居数、 D_j 的1阶邻居数和 D_i, D_j 的公共1阶邻居数。

2.2.3 结合公共邻居节点子网络构造特征

从邻居节点构造特征是最常使用的一种方法, 然而这种方法对网络拓扑结构的挖掘不够深入, 难以进一步提高特征的分类性能, 存在一定的局限性。因此, 本文将 D_i 和 D_j 的公共邻居节点构成的子网络从DDINet中分离出来, 并进一步进行分析。本文首先计算出每两个药物公共节点构成的子网络密度 d , 以及子网络中协同关系网络密度 d_s 和拮抗关系网络密度 d_a 。它们之间的关系可表示为 $d = d_s + d_a$ 。但是, 原始DDINet网络中协同和拮抗关系边数的不平衡, 可能导致 d_s 值始终比 d_a 值小。为了减少由样本不平衡导致的误差, 本文将 d_s 和 d_a 按照以下式(6)和式(7)进行标准化处理

$$d'_s = d_s/d_{s_max} \quad (6)$$

$$d'_a = d_a/d_{a_max} \quad (7)$$

其中, d'_s 和 d'_a 表示标准化结果, d_{s_max} 和 d_{a_max} 分别表示 d_s 和 d_a 中的最大值。最后本文基于 d'_s 和 d'_a 构造了另一个特征 z_1

$$z_1 = d'_s/d'_a \quad (8)$$

可以看出, z_1 与先前构造的特征不同, 它是基于子网络的密度值进行构造。基于密度构造的特征不容易受到节点数的干扰, 相反它也不能反映出药物公共节点数的多少, 因此, 构造 z_1 这个特征, 更多的是想从网络节点之外的信息去辅助探讨和分析节点之间的关系。

2.3 模型评估

本文使用支持向量机(Support Vector Machine, SVM)^[20-22]作为分类和预测模型, 模型中的分类核函数使用高斯核(Radial Basis Function, RBF), 参数 C 和参数 γ 使用默认值, 模型的度量标准使用AUC(Area Under the Curve)^[23]和分类精度(precision)^[24]。使用十折交叉验证方法对模型分类

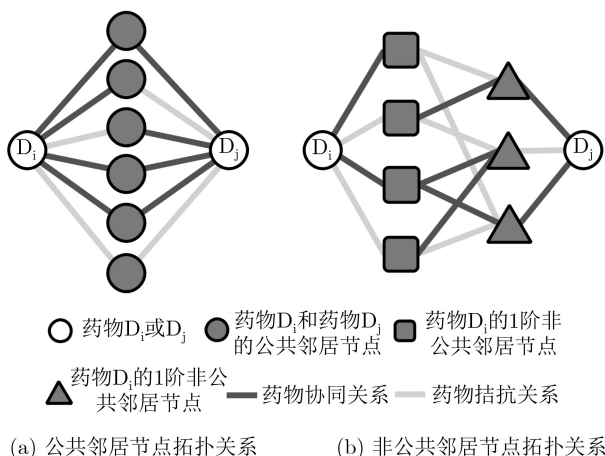


图1 药物 D_i 和 D_j 的1阶邻居节点拓扑关系示意

和预测性能进行验证^[25]。在每一次模型构建过程中，将DDINet中已知关系的边分为训练集(train set)和测试集(test set)，其中训练集占90%，测试集占10%，并且训练集和测试集中的正负样本比例与原始数据集中保持一致。此外，每个样本的特征值都是基于训练集构成的DDINet计算得到。因此，在每次训练中，所有样本的特征值都会随着训练集的变化而变化。

为了验证方法对潜在DDIs的预测性能，本文定义了一个预测精度指标 p ，表示在前 L 个预测结果中，正确预测样本的概率^[26]。在本文中，将测试集中的样本(DDIs)和未知关系样本(未知DDIs)作为探针集(probe set)样本，使用训练好的模型对其预测，将预测结果按照从大到小排序，取出排名前 L 个样本，如果有 l 个样本(测试集中)被成功预测，则 p 可表示为

$$p = \frac{l}{L} \quad (9)$$

显然， p 值越大表示模型的预测性能越好。对于正负样本的预测结果，使用 p^+ 和 p^- 来表示。

为了得到稳定可靠的模型，总共对数据集进行100次十折交叉验证，最后取其平均值作为模型性能评估指标。

3 实验结果与分析

3.1 特征选择

特征的好坏在很大程度上决定了分类器的性能。在构建分类模型的过程中，为了得到重要的特征，通常会计算特征重要性，保留重要特征，舍弃不重要甚至无关的特征，这即是特征选择(Feature selection)^[27]。经过特征选择后的模型更有利于提高模型的泛化能力，减少过拟合。在此，本文首先通过分布图绘制出所有特征在正负样本中的分布情况，由此来比较各个特征的分类性能，结果如图2和图3所示。

从图2中可以看出，在 $x_1 \sim x_5$ 这5个基础特征中， x_3 和 x_4 能较大程度上区分出正负样本，表现出一定的分类性能，而在图3中，除了 y_1 和 y_5 这2个特征之外，剩下4个特征在正负样本中的分布都存在着明显的差异。为了进一步了解特征的分类性能，并筛选出重要特征，本文对分类效果较好的6个特征 $x_3, x_4, y_2, y_3, y_4, z_1$ 绘制了箱线图，结果如图4所示。

从图4可以看出，选出来的6个特征都表现出明显的正负样本差异，可以作为分类模型的重要特征。其中 x_3, x_4 与 y_3, y_4 的分布非常相似，但是仔细观察发现， y_3, y_4 表现出的分布差异比 x_3, x_4 更明显。而从图4(f)可知，虽然 z_1 的分类性能不如其他5个特

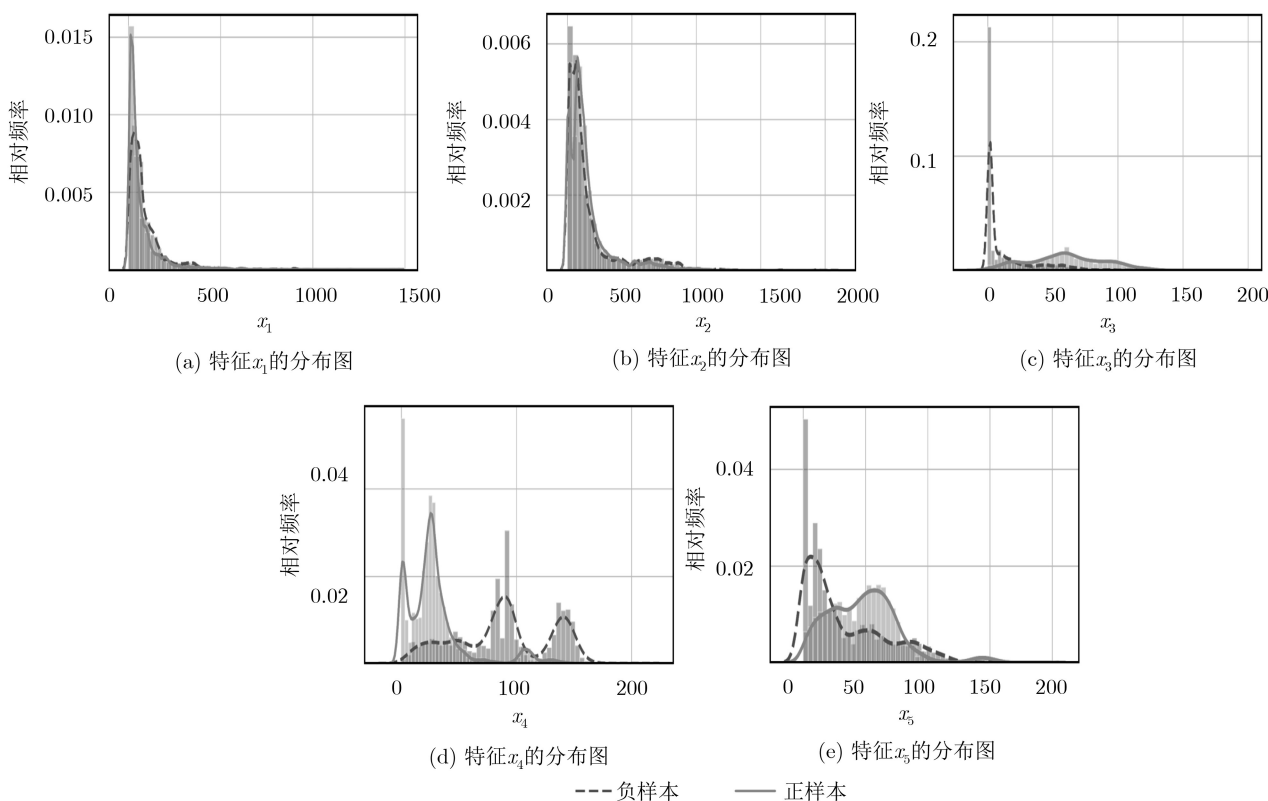


图2 特征 $x_1 \sim x_5$ 在正负样本中的分布

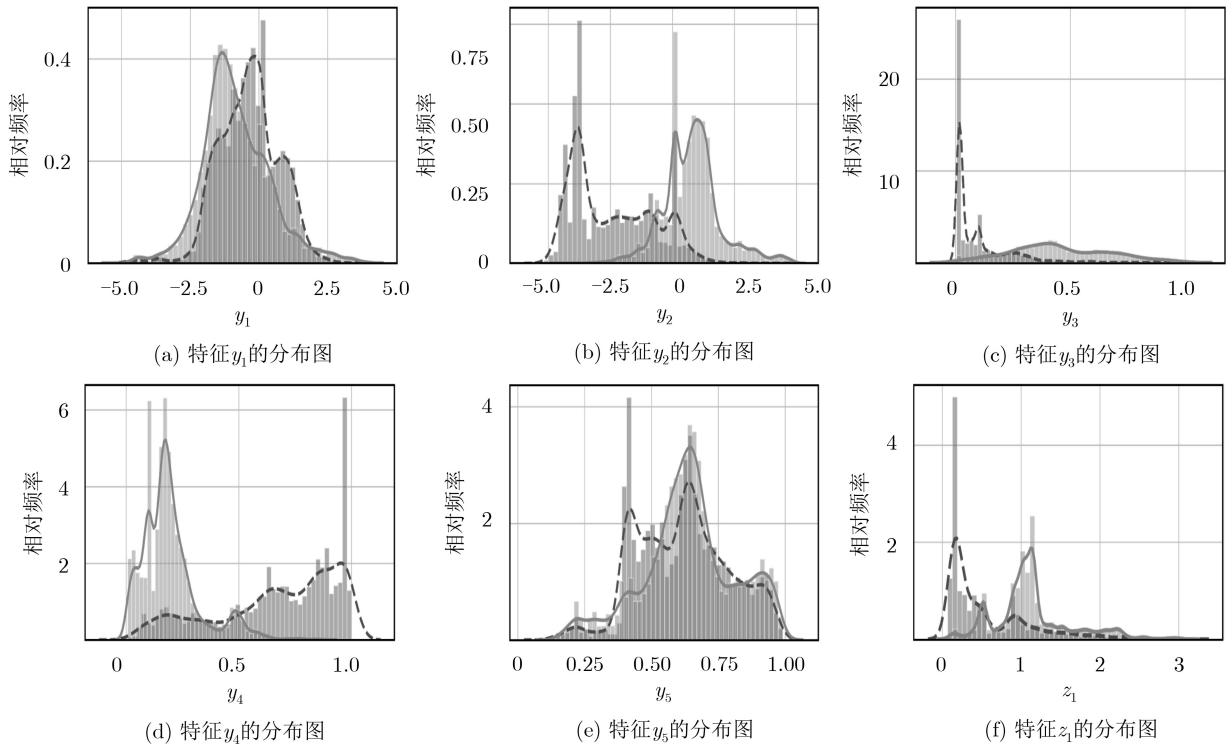


图3 特征 $y_1 \sim y_5, z_1$ 在正负样本中的分布

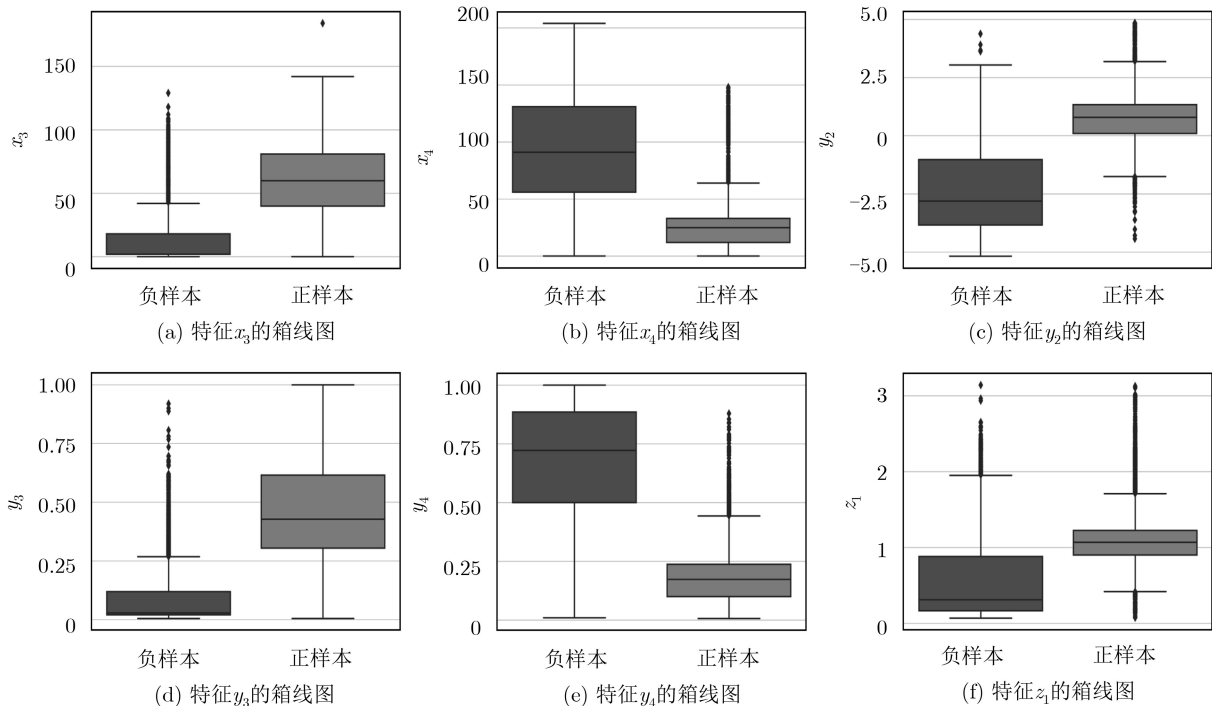


图4 特征 $x_3, x_4, y_2, y_3, y_4, z_1$ 在正负样本中的分布

征，但是它仍然表现出了一定的正负样本差异，作为其他特征的辅助特征也具有一定的意义。此外， y_2 也能够有效区分出正负样本，是一个重要特征，但是由于 y_2 是由 x_3 和 x_4 组合构造而成，为了不重复使用重要特征，本文在之后的分类和预测中不再使用 y_2 。

综上所述，经过特征选择后，本文选出了5个重要特征，分别为 x_3, x_4, y_3, y_4 和 z_1 。

3.2 分类结果分析

为了对比特征的实际分类能力，将选出的5个重要特征进行组合，构建多个分类器，计算它们在数据集上的AUC和分类精度。最后使用十折交叉

重复100次进行验证，将结果的均值作为分类性能度量，并根据结果绘制出ROC(Receiver Operating Characteristic)曲线。图5表示不同特征组合的ROC曲线和分类结果。

从图5(a)中可以看出，基础特征中最优的分类器是由 x_3 、 x_4 和 z_1 构建而成，其AUC和分类精度得分分别为0.9624、0.9098。而 x_3 和 x_4 组成的分类器的AUC和分类精度得分也有0.9538和0.8871，仅比最优的分类器差0.0086和0.0227。这说明在不借助其

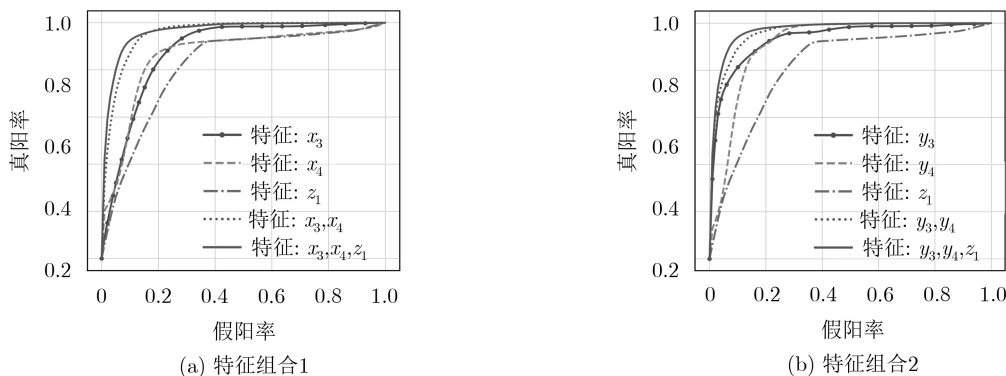


图 5 不同特征组合的ROC曲线

从图5(b)中可以发现，基于 x_3 和 x_4 构造的特征 y_3 和 y_4 ，无论是在单特征分类中还是组合特征分类中，他们的分类性能都要高于 x_3 和 x_4 。其中的原因可能是 x_3 和 x_4 这两个特征容易受到 x_5 的干扰，不能很好地反映他们与所有公共药物之间的关系，而 y_3 和 y_4 则通过标准化的方式消除掉了这种影响，所以效果更好。

3.3 预测结果分析

两个药物之间的相互作用关系可以由邻居节点构造的特征反映，但是如果某些药物邻居节点的信息“丰度”不够，则有可能导致计算出来的特征值成为干扰特征值，影响预测精度。因此，为了提升模型的预测性能，本文首先对测试样本和未知关系样本中进行过滤。假设过滤参数为 f ，保留 $x_3 > f$ 且 $x_4 > f$ 的测试样本和未知关系样本。通过选取不同 f 值，得到不同数量的样本，结果如图6所示。

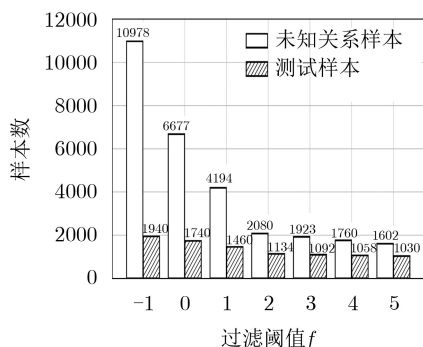


图 6 不同 f 取值对应的预测样本分布情况

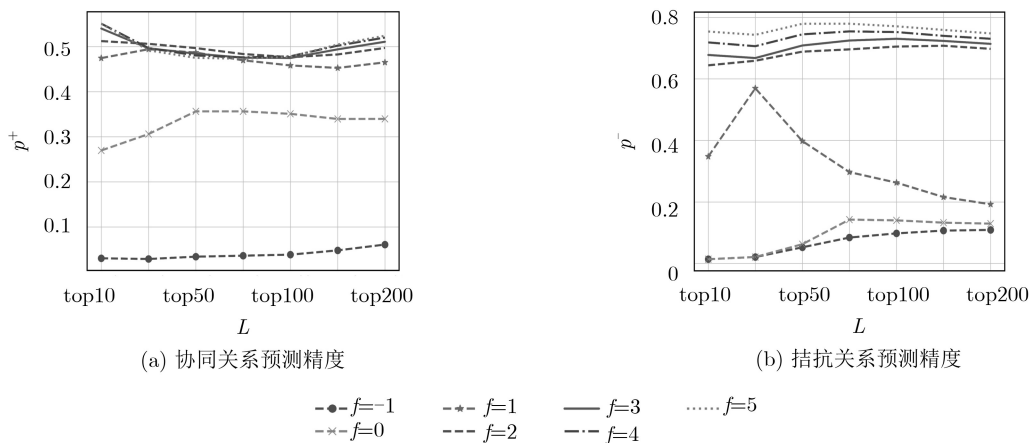
他相似性信息的情况下，药物之间的相互作用可以在很大程度上通过药物与他们公共药物之间的相互作用反映出来。而在加入 z_1 特征之后，对分类精度的进一步提升则说明，公共药物内部的相互作用关系可以进一步反映出药物相互作用关系。此外，只由单个重要特征构建的分类器也具有不错的分类性能，尤其是 x_3 和 x_4 这两个特征的分类精度都达到了0.8以上，而 z_1 特征的分类精度虽然比较低，但也有0.7566。

在图6中，当 $f = -1$ 时，表示不对样本进行过滤。由此可知，当 $f = 0$ 时，被过滤掉的未知样本数最多，共有4301个，占原始样本数近40%；而测试集样本过滤掉200个，仅占原始样本10%。这说明在未知关系样本中存在着大量 $x_3 = 0$ 或 $x_4 = 0$ 的样本，而这些样本值会极大地影响模型的预测性能。当 $f \geq 2$ 时，过滤样本数逐渐减少并趋于平稳。

为了进一步对比模型在不同 L 下的预测性能，本文设置了7组top值 L ，然后在不同过滤阈值 f 的得到的样本中计算正负样本的预测精度 p^+ 和 p^- ，结果如图7所示。

从图7中可知，在协同和拮抗关系预测中，对样本进行过滤都能够有效提高模型的预测精度，而且随着 f 值的增大，预测精度也逐渐增加。部分原因可能是过滤后样本数减少，但是更主要的原因应该是过滤后使得干扰样本和无关样本减少，从而提升了模型的预测性能。

在DDIs预测中，本文的首要目的是获得更高0的预测精度，但是有时为了更好地反映模型对样本的整体预测性能，本文会保留尽量多的样本数。因此，选择合适的 f 和 L 对于DDIs预测至关重要。在图7(a)中，当 $f \geq 2$ 时，协同关系预测精度在0.5左右徘徊，且 L 和 f 的取值对精度的影响不大。而在图7(b)中，当 $f \geq 2$ 时，拮抗关系预测精度走势基本一致，并随着 f 的增大而升高，在 $f=5$ ， $L=50$ 时达到最大值0.7792。因此，在对协同和拮抗关系

图7 不同 f , L 取值下的协同、拮抗关系预测精度

预测中, 本文分别选择 $f=2$ 和 $f=5$ 作为模型的过滤阈值, 而 L 的取值则主要根据预测需求进行选择。此外, 从图7中可以发现, 模型对协同和拮抗的预测精度值存在较大差异, 这可能是原始数据集中正负样本数不平衡导致构造的特征存在正负样本差异。

虽然模型对协同和拮抗关系的预测存在较大差异, 但是从结果来看, 本文的方法能够有效地对未知DDIs进行预测。

4 结束语

本文从DrugBank数据库中提取出ATC分类中“L”类药物的DDIs数据, 以此构建药物互作网络, 通过药物的邻居节点构建特征, 最终使用SVM实现了DDIs的分类和预测。在特征选择中, 本文找出了最重要的几个特征 x_3, x_4, y_3, y_4, z_1 。这些重要特征表明药物互作网络虽然是复杂网络, 但有其自身的分布特征, 通过挖掘出这些特征并用于潜在DDIs预测有一定的实践意义。从分类结果可知, 本文方法具有较高的AUC和分类精度得分, 能有效地对现有DDIs进行分类。而从预测结果可知, 该方法能有效地对潜在的药物协同和拮抗关系进行预测, 尤其是对拮抗关系的预测精度达到了0.75以上。

文本的方法比传统DDIs预测方法简单, 易于实现, 降低了DDIs的预测难度, 但是也存在一些不足之处, 如预测结果容易受到样本的干扰、DDINet网络密度不能过低等等。总体而言, 从DDINet拓扑结构出发预测潜在DDIs无疑是一种研究DDIs的新途径。如何充分挖掘出DDINet中蕴含的信息, 并将其运用到DDIs预测中则有可能成为接下来DDIs研究的一个新方向。

参考文献

[1] VAN ROON E N, FLIKWEERT S, LE COMTE M, *et al.*

Clinical relevance of drug-drug interactions[J]. *Drug Safety*, 2005, 28(12): 1131–1139. doi: 10.2165/00002018-200528120-00007.

[2] CHOU Tingchao. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies[J]. *Pharmacological Reviews*, 2006, 58(3): 621–681. doi: 10.1124/pr.58.3.10.

[3] LEHÁR J, KRUEGER A S, AVERY W, *et al.* Synergistic drug combinations tend to improve therapeutically relevant selectivity[J]. *Nature Biotechnology*, 2009, 27(7): 659–666. doi: 10.1038/nbt.1549.

[4] CHAIT R, CRANEY A, and KISHONY R. Antibiotic interactions that select against resistance[J]. *Nature*, 2007, 446(7136): 668–671. doi: 10.1038/nature05685.

[5] VENKATAKRISHNAN K, VON MOLTKE L L, OBACH R S, *et al.* Drug metabolism and drug interactions: Application and clinical value of in vitro models[J]. *Current Drug Metabolism*, 2003, 4(5): 423–459. doi: 10.2174/1389200033489361.

[6] PIRMOHAMED M and ORME M L. Drug Interactions of Clinical Importance[M]. DAVIES D M, FERNER R E, and DE GLANVILLE H. *Davies's Textbook of Adverse Drug Reactions*. 5th ed. London: Chapman & Hall, 1998: 888–912.

[7] TAKEDA T, HAO Ming, CHENG Tiejun, *et al.* Predicting drug-drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge[J]. *Journal of Cheminformatics*, 2017, 9: 16. doi: 10.1186/s13321-017-0200-8.

[8] FOKOUE A, SADOOGHI M, HASSANZADEH O, *et al.* Predicting drug-drug interactions through large-scale similarity-based link prediction[C]. The 13th International Conference European Semantic Web Conference, Heraklion, Greece, 2016: 774–789. doi: 10.1007/978-3-319-34129-3_47.

[9] VILAR S, HARPAZ R, URIARTE E, *et al.* Drug—drug interaction through molecular structure similarity analysis[J]. *Journal of the American Medical Informatics*

- Association*, 2012, 19(6): 1066–1074. doi: [10.1136/amiajnl-2012-000935](https://doi.org/10.1136/amiajnl-2012-000935).
- [10] VILAR S, URIARTE E, SANTANA L, *et al*. Detection of drug-drug interactions by modeling interaction profile fingerprints[J]. *PLoS One*, 2013, 8(3): e58321. doi: [10.1371/journal.pone.0058321](https://doi.org/10.1371/journal.pone.0058321).
- [11] KASTRIN A, FERK P, and LESKOŠEK B. Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning[J]. *PLoS One*, 2018, 13(5): e0196865. doi: [10.1371/journal.pone.0196865](https://doi.org/10.1371/journal.pone.0196865).
- [12] CHENG Feixiong and ZHAO Zhongming. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties[J]. *Journal of the American Medical Informatics Association*, 2014, 21(e2): e278–e286. doi: [10.1136/amiajnl-2013-002512](https://doi.org/10.1136/amiajnl-2013-002512).
- [13] RYU J Y, KIM H U, and LEE S Y. Deep learning improves prediction of drug-drug and drug-food interactions[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(18): E4304–E4311. doi: [10.1073/pnas.1803294115](https://doi.org/10.1073/pnas.1803294115).
- [14] LUO Heng, ZHANG Ping, HUANG Hui, *et al*. DDI-CPI, a server that predicts drug-drug interactions through implementing the chemical-protein interactome[J]. *Nucleic Acids Research*, 2014, 42(W1): W46–W52. doi: [10.1093/nar/gku433](https://doi.org/10.1093/nar/gku433).
- [15] ZHANG Ping, WANG Fei, HU Jianying, *et al*. Label propagation prediction of drug-drug interactions based on clinical side effects[J]. *Scientific Reports*, 2015, 5: 12339. doi: [10.1038/srep12339](https://doi.org/10.1038/srep12339).
- [16] LIU Lili, CHEN Lei, ZHANG Yuhang, *et al*. Analysis and prediction of drug-drug interaction by minimum redundancy maximum relevance and incremental feature selection[J]. *Journal of Biomolecular Structure and Dynamics*, 2017, 35(2): 312–329. doi: [10.1080/07391102.2016.1138142](https://doi.org/10.1080/07391102.2016.1138142).
- [17] TAKARABE M, SHIGEMIZU D, KOTERA M, *et al*. Network-based analysis and characterization of adverse drug-drug interactions[J]. *Journal of Chemical Information and Modeling*, 2011, 51(11): 2977–2985. doi: [10.1021/ci200367w](https://doi.org/10.1021/ci200367w).
- [18] GOTTLIEB A, STEIN G Y, ORON Y, *et al*. INDI: A computational framework for inferring drug interactions and their associated recommendations[J]. *Molecular Systems Biology*, 2012, 8: 592. doi: [10.1038/msb.2012.26](https://doi.org/10.1038/msb.2012.26).
- [19] WISHART D S, FEUNANG Y D, GUO A C, *et al*. DrugBank 5.0: A major update to the DrugBank database for 2018[J]. *Nucleic Acids Research*, 2017, 46(D1): D1074–D1082. doi: [10.1093/nar/gkx1037](https://doi.org/10.1093/nar/gkx1037).
- [20] CORTES C and VAPNIK V. Support-vector networks[J]. *Machine Learning*, 1995, 20(3): 273–297. doi: [10.1007/bf00994018](https://doi.org/10.1007/bf00994018).
- [21] 陈素根, 吴小俊. 基于特征值分解的中心支持向量机算法[J]. 电子与信息学报, 2016, 38(3): 557–564. doi: [10.11999/JEIT150693](https://doi.org/10.11999/JEIT150693).
- CHEN Sugeng and WU Xiaojun. Eigenvalue proximal support vector machine algorithm based on eigenvalue decomposition[J]. *Journal of Electronics & Information Technology*, 2016, 38(3): 557–564. doi: [10.11999/JEIT150693](https://doi.org/10.11999/JEIT150693).
- [22] 汪廷华, 田盛丰, 黄厚宽. 特征加权支持向量机[J]. 电子与信息学报, 2009, 31(3): 514–518. doi: [10.3724/SP.J.1146.2007.01711](https://doi.org/10.3724/SP.J.1146.2007.01711).
- WANG Tinghua, TIAN Shengfeng, and HUANG Houkuan. Feature weighted support vector machine[J]. *Journal of Electronics & Information Technology*, 2009, 31(3): 514–518. doi: [10.3724/SP.J.1146.2007.01711](https://doi.org/10.3724/SP.J.1146.2007.01711).
- [23] WU Shaomin and FLACH P. A scored AUC metric for classifier evaluation and selection[C]. ICML 2005 Workshop on ROC Analysis in Machine Learning, Bonn, Germany, 2005.
- [24] HOSSIN M and SULAIMAN M N. A review on evaluation metrics for data classification evaluations[J]. *International Journal of Data Mining & Knowledge Management Process*, 2015, 5(2): 1–11. doi: [10.5121/ijdkp.2015.5201](https://doi.org/10.5121/ijdkp.2015.5201).
- [25] ARLOT S and CELISSE A. A survey of cross-validation procedures for model selection[J]. *Statistics Surveys*, 2010, 4: 40–79. doi: [10.1214/09-SS054](https://doi.org/10.1214/09-SS054).
- [26] LIU Weiping and LÜ Linyuan. Link prediction based on local random walk[J]. *EPL (Europhysics Letters)*, 2010, 89(5): 58007. doi: [10.1209/0295-5075/89/58007](https://doi.org/10.1209/0295-5075/89/58007).
- [27] TANG Jiliang, ALELYANI S, and LIU Huan. Feature Selection for Classification: A Review[M]. AGGARWAL C C. *Data Classification: Algorithms and Applications*. New York: CRC Press, 2014: 37.
- 刘文斌: 男, 1969年生, 教授, 研究方向为生物信息学。
陈杰: 男, 1994年生, 硕士生, 研究方向为生物信息学。
方刚: 男, 1969年生, 教授, 研究方向为生物信息学。
石晓龙: 男, 1975年生, 教授, 研究方向为生物信息学。
许鹏: 男, 1986年生, 博士后, 研究方向为生物信息学。