



基于多模态深度自编码器网络嵌入和正负样本学习的药物相互作用预测

Yang Zhang^a, Yang Qiu^a, Yuxin Cui^a, Shichao Liu^{a,✉}, Wen Zhang^{a,✉}

中国武汉华中农业大学信息学院, 邮编 430070

文章信息

关键词: 药物相互作用 网络嵌入 深度自编码器 缺失链路预测 正负样本学习

摘要

药物相互作用 (DDIs) 对公共健康和患者安全至关重要, 这在学术界和工业界引起了广泛关注。现有的药物相互作用预测计算方法主要分为四类: 基于文献提取的、基于相似性的、基于矩阵运算的和基于网络的。近期的多项研究揭示, 整合异质性药物特征对于开发高精度预测模型具有重要意义。同时, 缺乏某些特征的药物可以利用其他特征来学习表示。然而, 这也带来了一些新的挑战, 如数据不完整、非线性关系和异质性属性。在本文中, 我们提出了一种基于多模态深度自编码器的药物表示学习方法, 名为 DDI-MDAE, 用于从大规模、嘈杂且稀疏的数据中预测药物相互作用。我们的方法旨在使用多模态深度自编码器同时从多个药物特征网络中学习统一的药物表示。然后, 我们对学习到的药物嵌入应用四种运算来表示药物对, 并采用随机森林分类器训练模型以预测药物相互作用。实验结果表明, 我们提出的药物相互作用 (DDI) 预测方法是有效的, 并且相较于其他最先进的基准方法有显著的改进。此外, 我们在正负未标记 (PU) 学习设置中应用了专门的随机森林分类器, 以提高预测的准确性。实验结果表明, 在 3 折交叉验证 (3-CV) 和 5 折交叉验证 (5-CV) 中, 通过 PU 学习改进的模型在 AUPR 指标上分别比原始方法 DDI-MDAE 提高了 7.1% 和 6.2%。在 F 值指标上, 改进后的模型在 3-CV 和 5-CV 中分别比 DDI-MDAE 提高了 10.4% 和 8.4%。通过案例研究进一步证明了 DDI-MDAE 的实用性。

1. 简介

过去十年间, 新药研发数量迅速增加, 联合用药疗法在疾病治疗方面前景广阔。同时使用两种或两种以上药物可能会产生意想不到的相互影响, 这种现象被称为药物相互作用 (DDIs)。药物相互作用对受影响药物的反应可分为三类: 协同作用、拮抗作用和无反应。协同作用是患者治疗的最佳结果, 意味着联合用药的疗效优于单独用药。无反应类型既不会增强也不会削弱药物的疗效。拮抗作用不仅会降低疗效, 还可能导致不良反应, 甚至危及患者生命。因此, 检测药物相互作用在公共卫生安全和药物安全性监测方面引起了广泛关注。

通过传统临床试验和体外方法来识别药物相互作用 (DDIs) 的成本高且耗

时长。然而, 技术的快速发展有利于收集不同类型的药物数据, 这为计算药物发现和药物安全性研究领域带来了广阔前景。因此, 提出了大量的计算预测方法, 有助于在药物开发过程中为患者提供更安全、更有效的处方, 并有助于减少临床试验的时间和成本。诸如医疗报告、文件、科学期刊等非结构化生物医学文献以及免费的特定药物数据库提供了丰富的最新药物信息。因此, 药物数据主要来自文献和数据库。

由于自然语言处理相关技术的成功应用, 基于文献提取药物间相互作用 (DDIs) 的计算方法通过统计和文本挖掘技术实现了准确预测。FBK-irst 团队[2]提出了一种应用混合核的基于关系抽取 (RE) 分类器的方法, 在 2013 年 DDIEtraction 挑战赛中取得了 0.651 的 F 值, 排名首位。Kim 等人[3]开发了一种两阶段方法, 基于线性支持向量机 (SVM) 并利用丰富的特征, 获得了 0.670 的 F 值。

✉ Corresponding authors.

E-mail addresses: scliu@mail.hzau.edu.cn (S. Liu), zhangwen@mail.hzau.edu.cn (W. Zhang).

<https://doi.org/10.1016/j.ymeth.2020.05.007>

Received 9 February 2020; Received in revised form 6 May 2020; Accepted 13 May 2020 1046-2023/ © 2020 Elsevier Inc. All rights reserved.

在相同的语料库上提取特征。深度卷积神经网络的应用前景也很广阔。Sun 等人[1]提出了 DCNN，它能够通过具有多层小卷积的深度架构自动生成合适的特征集。Zhang 等人[4]采用分层 RNNs 模型将句子序列和最短依存路径（SDP）进行拼接。Jiang 等人[5]考虑了骨架-LSTM 网络以提取 DDI 的内在结构。Xu 等人[6]利用用户生成内容（UGC）资源，通过全注意力机制获取新鲜且丰富的信息。

基于相似性的方法利用数据库中的数据，其基于这样一种假设：相似的药物可能会与相同的药物发生相互作用。Vilar 等人[7]提出了一种简单且实用的矩阵计算方法，利用分子结构相似性来预测药物相互作用。Gottlieb 等人[8]引入了“推断药物相互作用”（INDI）方法，该方法考虑了七种药物间相似性，从而能够推断出药代动力学和药效学方面的药物相互作用。Ma 等人[9]提出了多视图图自编码器（GAE），将每种类型的特征视为一个视图，并在整合药物相似性时采用注意力机制，然后将该模型扩展到半监督和归纳设置。

基于矩阵运算的方法能够有力地识别新药的药物相互作用（DDIs），即便这些新药没有任何已知的相互作用。张等人[10]提出了结合四种药物特征和已知药物相互作用的具有线性邻域正则化的稀疏特征学习集成方法（SFLN）。还有一些方法被提出以进一步预测全面的药物相互作用。例如，石等人[11]提出了使用三重矩阵分解和统一预测模型的 TMFUF 方法。于等人[12]开发了基于半非负矩阵分解的 DDINMF 方法。

基于网络的方法通过数据库中的网络结构推断潜在的相互作用。Zhang 等人[13]开发了一种考虑高阶相似性的标签传播方法来进行预测。Park 等人[14]通过带有重启的随机游走算法在蛋白质-蛋白质相互作用网络上模拟信号传播。Sridhar 等人[15]提出了一种概率方法，从基于多种药物的相似性和已知相互作用的网络中联合推断缺失的相互作用。Huang 等人[16]开发了一种称为“S 分数”的度量标准，用于衡量网络连接强度以进行预测。

此外，集成学习以某种方式结合了多个机器学习模型，通常能比单独的模型取得更好的性能。迪皮卡等人[17]采用了一种半监督学习框架，结合网络表示学习和元学习，从四个药物数据集中预测药物相互作用。张等人[18]提出了一种稀疏特征学习集成方法，结合基于相似性的方法和矩阵扰动方法。

尽管这些计算方法表现良好，但近期研究揭示，整合异质性药物特征对于药物相互作用（DDI）预测具有重要意义[19-21]，这给从多个生物学数据源中学习带来了巨大挑战。首先，不同药物数据集的统计特性各异。因此，浅层模型在从多个数据源中提取高级概念时可能表现不佳。其次，生物学事件之间的关系在各类药物特征中通常是非线性的。第三，某些药物数据集的不完整性（例如标签缺失、噪声数据）可能会对预测模型产生不利影响。此外，大多数现有方法将已知存在相互作用的药物对视为正例，而将未知药物对视为负例[21,22]。然而，可能存在大量尚未发现的真实正例（假阴性）。因此，如果直接将从正例和负例中学习的方法应用于正例和未标记数据，预测模型的性能可能会受到负面影响。

为解决上述问题，我们提出了一种基于多模态深度自编码器的药物表征学习方法用于药物相互作用预测，命名为 DDI-MDAE。在本文中，我们使用了五个药物数据源。

这些网络被视为药物特征网络的不同形式，包括药物相互作用网络、药物化学亚结构网络、药物靶点网络、药物酶网络和药物通路网络。然后，我们通过多模态深度自编码器同时从这五个药物特征网络中学习药物的统一表示。我们对学习到的药物嵌入应用若干操作来表示药物对，然后采用随机森林分类器训练模型以预测准确的药物相互作用。此外，我们通过 PU 学习改进了我们的方法。总之，我们的研究做出了四个主要贡献：

我们引入了 DDI-MDAE，这是一种高效的基于多模态深度自编码器的药物表征学习方法，为同时利用多种药物特征学习药物的统一表征提供了新的思路。此外，即使面对大规模、噪声大且稀疏的数据，DDI-MDAE 也能够对具有不完整特征的药物进行预测。

我们对学习到的药物嵌入采用多种运算符来表示药物对，并训练随机森林分类器来预测潜在的药物相互作用。

我们在多个任务上对模型进行了全面评估，实验结果表明，即使使用稀疏的药物特征进行训练，DDI-MDAE 也具有非常高的效率和鲁棒性。我们通过在正负未标记学习环境中应用专门的随机森林分类器来改进所提出的 DDI-MDAE，以提高预测准确性。实验结果表明，在大多数情况下，改进后的方法优于原始的 DDI-MDAE 方法。

2. 背景

在本节中，我们将介绍所提出方法的背景，包括网络嵌入、深度神经网络和正负样本学习。

2.1. 网络嵌入

网络嵌入（又称图嵌入）学习旨在自动学习低维节点表示，近年来在生物学网络分析中引起了越来越多的关注[23]。

一些早期的方法，如线性嵌入[24]、等距映射[25]，是基于特征向量构建亲和图，并将亲和图嵌入到低维空间中。最近，DeepWalk[26]将图结构采样为随机游走序列，用其训练跳字模型来预测随机游走的路径。LINE[27]定义了一个损失函数来捕获 1 步和 2 步的局部结构信息。Node2vec[28]提出了一种有偏随机游走，结合广度优先搜索和深度优先搜索策略来探索邻域，从而对网络进行采样。还有几种基于深度学习的网络嵌入方法。SDNE[29]利用深度自编码器来保留网络的一阶和二阶邻近性。GCN[30]设计了卷积神经网络的变体，以编码局部图结构和节点特征。

此外，还有一些关于异构网络[31-33]和属性网络[34-36]中表示的研究。这些模型将具有多种模态的复杂网络嵌入到一个共享的低维空间中。受先驱工作的启发，本研究利用结构网络嵌入从多个药物数据集中学习药物的统一表示。

2.2. 深度神经网络

深度神经网络（DNN）已在众多领域成功应用于特征学习[37-39]。与以往的工作不同，

我们采用深度神经网络来管理网络表示学习。像自动编码器这样的深度神经网络能够学习高度非线性的映射，将数据从高维空间映射到低维空间。自动编码器是一种非线性的无监督神经网络模型，已成为常用的构建模块之一[40]。自动编码器由两部分组成，包括编码器和解码器。在编码器部分，确定性映射 f 将输入向量 x 转换为隐藏表示 y 。是一个非线性激活函数（例如 S 型函数）。解码器将得到的隐藏表示 y 映射回重构维度 x 。

$\theta = (W, b)$ 和 $\hat{\theta} = (\hat{W}, \hat{b})$ 是待确定的参数集。

训练过程。

$$\begin{aligned} y &= f_{\theta}(x) = \sigma(Wx + b) \\ \hat{x} &= g_{\hat{\theta}}(y) = \sigma(\hat{W}y + \hat{b}) \end{aligned} \quad (1)$$

自编码器的目标是将输入与输出的重构误差最小化。自编码器的损失函数如下所示：

$$L(\hat{x}, x) = \sum_{i=1}^n \left\| \hat{x}_i - x_i \right\|_2^2 \quad (2)$$

其中 n 表示输入的维度大小。参数 θ 和 $\hat{\theta}$ 可以通过梯度下降法进行优化。堆叠式自动编码器包含多层这样的自动编码器，并使用逐层训练方法[41]来学习每一层的合适表示。

2.3. 正负未标记学习

在机器学习中，传统的二分类器是从正样本和负样本中学习得到的。然而，在现实领域中，负样本通常难以获取。面对现实中的问题：从包含正负样本的大量未标记数据中识别出正样本，一种名为正样本未标记样本学习（PU 学习）的先进学习技术应运而生，以解决这一难题。

在 PU 学习任务中，存在两组样本：正样本集 P 和未标记样本集 U 。其最终目标是从 U 中识别出合适的正样本。在 PU 学习环境中学习分类器，基本上提出了三种方法：直接法、两步法和 PU 学习的装袋法。

至于直接方法，刘等人[42]提出了一种基于支持向量机（SVM）有偏公式的方法。埃尔坎等人[43]提出了 LPU 方法，旨在使从正样本和未标记样本中学习与从正样本和负样本中学习本质上等同。还有一些其他方法旨在按照一定规则为未标记样本分配权重，并将这些未标记样本视为加权负样本以训练分类器[44,45]。例如，李等人[44]提出通过将所有未标记样本视为负样本并应用线性函数，利用加权逻辑回归算法从含噪样本中学习。

大多数方法采用两步策略。第一步是从未标记样本中提取一个子集作为可靠的负样本。第二步是使用迭代分类算法构建分类器。在这两步中可以采用不同的策略。对于第一步，1-DNF [46]、间谍技术 [47]、Rocchio [48] 和朴素贝叶斯 [42] 被广泛使用。对于第二步，通常采用期望最大化和支持向量机进行迭代学习。

袋装 PU 学习方法分为两类：归纳式 PU 学习和推导式 PU 学习 [49]。归纳式 PU 学习包括对从 U 中随机抽取的小样本训练出的分类器进行聚合，然后对预测结果取平均值。推导式 PU 学习是归纳式 PU 学习的一种扩展，它聚合了在排除当前样本的子样本上训练的分类器的预测结果。

上述大多数传统方法在文本分类中得到了广泛应用。令人欣喜的是，PU 学习技术已在一些生物信息学研究领域成功应用，例如疾病基因识别[50-52]、药物靶点相互作用预测[53]以及药物相互作用预测[54,55]。例如，Mordelet 等人[52]提出了一种基于正样本和未标记样本的集成方法 ProDiGe 用于疾病基因预测。Deepika 等人[17]在其研究中采用了一种基于 PU 学习的分类器——袋式支持向量机分类器作为元分类器。Hameed 等人[55]考虑采用自组织映射聚类方法来预测药物相互作用。Zheng 等人[54]采用两步策略，结合两种技术（即单类支持向量机和 k 近邻）来预测药物相互作用。Lan 等人[53]采用了三种策略（带重启的随机游走、k 近邻和热核扩散）将未标记样本分为两类：可靠负样本和可能负样本。然后，他们采用不同的惩罚因子来预测药物靶点相互作用。

在我们的论文中，我们从 DrugBank [56] 收集数据，该数据库提供了可获取的标注数据。然而，DrugBank 提供的信息大多是关于正样本的，而可用的负样本数量则明显较少。因此，可以采用正负未标注学习来解决这一问题。我们采用了上述提到的一种典型直接方法，即 LPU 方法。这种方法的依据在于，在“完全随机选择”的假设下，基于正样本和未标注样本训练的分类器所预测的概率与正样本的真实条件概率仅相差一个常数因子 [43]。我们在正负未标注学习的框架下构建了一个专门的随机森林分类器来预测潜在的药物相互作用。更多细节将在下一节中阐述。

3. 材料与方法

3.1. 数据集

DrugBank [56] 是一个全面的公共数据库，包含广泛的药物信息，例如药物相互作用、药物亚结构、药物靶点、药物酶和药物转运蛋白。在本文中，DDI-MDAE 所关注的应用场景是从大规模、嘈杂且稀疏的数据中预测药物相互作用。因此，我们从 DrugBank（2018 年 4 月发布的 5.1.0 版本）中收集了至少包含一种特征或一种相互作用的药物来构建我们的数据集。药物靶点被映射到 KEGG 数据库 [57] 以获取药物通路。如表 1 所示，我们的数据集包含 2367 种药物，这些药物由四个不同维度的特征来表征：699 种化学亚结构、2411 个靶点、285 种酶和 314 条通路。

不同的药物特征对药物相互作用（DDI）预测的影响程度各不相同。因此，我们的论文整合了多种药物特征来学习药物表示，并构建了一个有效的模型来进行 DDI 预测。每个药物数据集都被视为药物特征网络的一个视角。同时，从 DrugBank 中提取的已知药物相互作用也被视为一个药物特征网络。因此，我们生成了五个药物特征网络。这些网络的详细信息如表 2 所示。例如，药物 - 靶点网络包含与原始数据集相关的药物节点和靶点节点，以及表示药物与靶点之间关联的边。特别地，我们获得了 209,494 个独特的药物相互作用。

表 1
关于药物数据集的描述。

数据类型	数据库	描述
药物	DrugBank [56]	2367 种药物类型
目标	DrugBank [56]	2411 个目标类型
酶	DrugBank [56]	285 种酶类型
路	KEGG [57]	314 种路径类型
子结构	PubChem [58]	699 种子结构类型

表 2
五个特征网络的属性。

网络类型	节点	边缘	平均度数	密度
DDI网络	2,367	209,494	219.251	0.075
药物-靶点网络	4,347	10,885	5.008	0.001
药物-酶网络	1,501	4,199	5.595	0.004
药物-通路网络	1,892	20,633	21.811	0.012
药物-结构网络	2,806	240,695	171.557	0.061

对于这 2367 种药物，有 2590667 种可能的药物组合（未标注实例）。已知的药物相互作用与未标注的药物相互作用的比例约为 1 比 12 (1:12)。

3.2. 方法概述

在本节中，我们介绍了从五个药物特征网络进行药物相互作用（DDI）预测的深度学习架构。图 1 展示了我们所提出模型的整体架构。为了捕捉多种非线性网络结构，我们提出了一种使用多模态深度自编码器学习药物表示的结构网络嵌入模型，命名为 DDI-MDAE。在我们的模型中，每个作为输入的药物特征网络，都通过一个深度自编码器通道进行训练（在图 1 中分别用不同颜色标记）。我们使用不同的颜色和形状来表示不同特征网络中的节点。隐藏层中药物的表示是共享的，以实现同步学习，这可能有助于捕捉不同网络之间的相关性。然后，我们对学习到的药物表示采用四种操作，再将其输入随机森林分类器以训练模型。此外，我们通过在正负样本学习设置中采用随机森林分类器开发了 DDI-MDAE（PU 学习），以改进我们用于 DDI 预测的原始方法。因此，在我们的深度架构中学习药物表示至关重要。我们将在下一节介绍 DDI-MDAE 方法的更多细节。

3.3. 多模态深度自编码器

如图 1 所示，DDI-MDAE 由多个自动编码器通道和一个共享隐藏层组成

，旨在从五个药物特征网络中学习统一的表示。经过训练过程，期望在共享隐藏层中获得低维空间。DDI-MDAE 方法的目标函数由三部分约束：

• 多个自编码器的重构误差

由于药物特征网络的稀疏性，我们对非零元素的重构误差施加的惩罚要多于对零元素的惩罚，这与文献[29]中的建议不同。因此，我们对重构误差的损失函数进行了修正，具体如下：

$$H_1 = \sum_{t=1}^T \sum_{i=1}^{n_t} \left\| \left(\hat{x}_i^t - x_i^t \right) \odot b_i \right\|_2^2 \quad (3)$$

其中 T 表示输入特征网络的数量。这里， $T = 5$ 。我们将邻接矩阵 $A = \{A(1), A(2), \dots, A(T)\}$ 作为我们模型的输入。 n_t 是 $A(t)$ 的维度大小。 x_i^t 是自动编码器的输入， \hat{x}_i^t 是自动编码器重构的数据。是哈达玛积，且 $b_i = \{b_{ij}\}_{j=1}^{n_t}$ 。如果节点 v_i 和 v_j 相连（即 $a_{ij} = 1$ ），则 $b_{ij} > 1$ ，否则 $b_{ij} = 1$ 。如果这些节点具有相似的邻域结构，那么在隐藏层中节点应具有相似的表示，以最小化 H_1 损失。

• 已知药物相互作用的限制

我们的模型旨在从包含少量已知药物相互作用的五个药物特征网络中学习药物表示。受[59,60]的启发，我们假设存在相互作用的药物在隐藏嵌入空间中应被映射到较近的位置。损失函数可描述如下：

$$H_2 = \sum_{(i,j) \in A(ddd)} a_{i,j} \left\| y_i^{(K)} - y_j^{(K)} \right\|_2^2 \quad (4)$$

其中 $y_i^{(K)}$ 是药物 v_i 的第 K 层隐藏表示。 K 是投影层的长度。 $A(ddd)$ 表示药物-药物相互作用网络的邻接矩阵。如果药物 v_i 和 v_j 相连，则 $a_{i,j} = 1$ ，否则 $a_{i,j} = 0$ 。

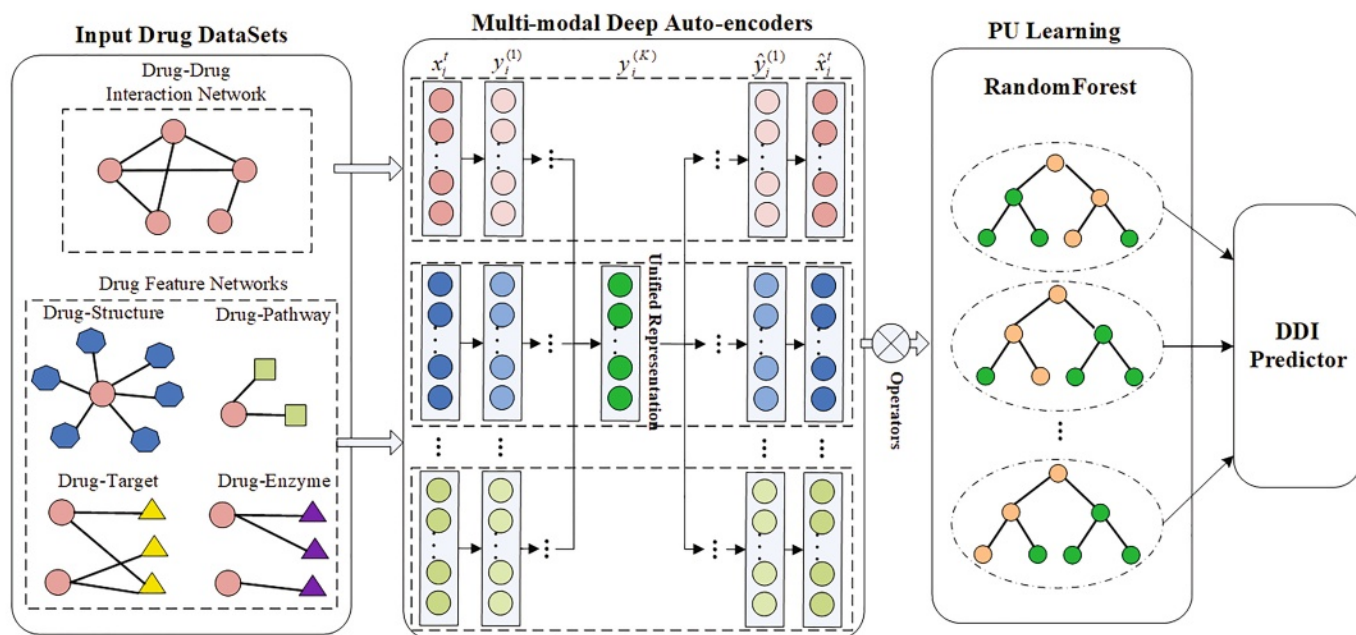


图 1. 基于 PU 学习的 MDAE 概述。

· 药物特征网络的相似性约束

值得注意的是，可能存在一些未被观察到关联的药物之间潜在的相互作用，因为它们具有某些相似的药物特征。与关注已知药物相互作用约束的损失函数 H_2 不同，我们关注的是药物相似性在药物对上的约束。我们首先根据五个网络（即 $didd, didt, dide, didp$ 和 $dids$ ）分别为每种药物 v_i 构建五个二进制向量。然后将这五个向量连接起来，得到一个综合向量 $D_i = [didd, didt, dide, didp, dids]$ 来表示包含五个药物特征网络的药物 v_i 。药物特征向量的维度等于所有节点类型的总和（即药物、靶点、酶、通路和子结构）。我们使用 Jaccard 公式基于特征向量计算药物相似性：

$$\mathcal{J}(D_i, D_j) = \frac{S_{11}}{S_{01} + S_{10} + S_{11}} \quad (5)$$

其中 S_{11} 是 D_i 和 D_j 均为 1 的维度数量； S_{01} 是 D_i 为 0 且 D_j 为 1 的维度数量； S_{10} 是 D_i 为 1 且 D_j 为 0 的维度数量。此目标的损失函数可定义如下：

$$H_3 = \sum_{i,j=1}^{n_d} \mathcal{J}(D_i, D_j) \left\| y_i^{(K)} - y_j^{(K)} \right\|_2^2 \quad (6)$$

其中 n_d 表示药物的数量， $\mathcal{J}(D_i, D_j)$ 表示药物 v_i 和 v_j 的杰卡德相似度。

为了同时保留上述三个目标函数，我们为 DDI-MDAE 定义了一个统一的损失函数，如下所示：

$$L_r = \sum_{k=1}^K \left(\left\| W^{(k)} \right\|_2 + \left\| \hat{W}^{(k)} \right\|_2 \right) \quad (8)$$

其中 K 表示投影层的长度。 $W^{(k)}$ 表示自编码器编码部分第 k 层的权重矩阵，而 $\hat{W}^{(k)}$ 则表示解码部分第 k 层的权重矩阵。

为了使用参数集 $\theta = (W, b)$ 和 $\hat{\theta} = (\hat{W}, \hat{b})$ 将总损失 H 最小化，我们采用 Tensorflow [45] 中的 Adam 优化方法对目标函数 (7) 进行优化。我们将在实验设置部分描述我们方法的超参数。

3.4. 预测药物相互作用

DDI 预测任务可被视为一个二分类问题，将药物组合视为正例或负例。我们直接从 DDI-MDAE 的共享隐藏层获取药物表示。因此，我们设计了四种操作符来表示药物组合。给定两种药物 v_i, v_j 及其表示 $(v_i), (v_j)$ ，通过所展示的四种操作符，我们可以分别获得药物组合的不同表示。

Φ Φ

表 3
表示药物组合的四种运算符的描述。

操作员	维度	描述
平均值	d	$\Phi(v_i, v_j) = \frac{1}{2} (\Phi(v_i) + \Phi(v_j))$
哈达玛	d	$\Phi(v_i, v_j) = (v_i \odot v_j) \odot \Phi$
L1 范数	d	$\Phi(v_i, v_j) = (v_i \oplus v_j) - \Phi$
连接	2d	$\Phi(v_i, v_j) = (v_i \oplus v_j) \oplus \Phi$

表 3 中的第 5 行，其中 d 是药物表示的维度。我们打算表示每一对药物，即便它们之间没有边相连，因为我们的测试集包含在构建的分类器中的正边和负边。

在计算出药物对的表示形式后，我们训练一个随机森林分类器 [61] 来预测潜在的药物相互作用。随机森林是一种元估计器，它在数据集的不同子样本上拟合许多决策树分类器，并通过平均来提高预测的准确性并控制过拟合问题。

为了更详细地说明我们的方法，我们在算法 1 中描述了 DDI-MDAE 方法。

Algorithm 1 The algorithm of DDI-MDAE

Input: Drug feature networks with adjacency matrix $A = \{A^{(1)}, A^{(2)}, \dots, A^{(T)}\}$
the parameters dimension $d, \alpha, \beta, \gamma, \lambda$, learning rate
Output: newly predicted DDIs

- 1: Initialize parameter sets $\theta = (W, b)$, $\hat{\theta} = (\hat{W}, \hat{b})$.
- 2: Learn drug representations Φ .
- 3: **repeat**
- 4: Compute the loss function H based on Eq. 7.
- 5: Calculate $\partial \theta, \partial \hat{\theta}$ and use Adam optimizer to update $\theta, \hat{\theta}$.
- 6: **until** convergence
- 7: Obtain the representations of drugs according to Eq. 1.
- 8: **for** each drug pairs (v_i, v_j) **do**
- 9: Representations $\Phi(v_i, v_j) = \Phi(v_i)$ operators $\Phi(v_j)$.
- 10: **end for**
- 11: Build the random forest classifier based on the representations of drug pairs and then predict unobserved DDIs.

3.5. 利用 PU 学习提升药物相互作用预测

对于由正例和大量未标记实例组成的训练数据，大多数先前的工作以及所提出的原始方法都将所有未标记实例视为负例，这会对预测模型的性能产生不利影响。在我们的数据集中，正向药物相互作用（DDI）与未标记 DDI 的比例约为 1 比 12（1:12）。因此，我们通过将正例-未标记学习设置中应用专门的随机森林分类器来改进 DDI-MDAE，以提高预测性能。为了尝试使从正例和未标记实例中学习在某种程度上等同于从正例和负例中学习，我们采用了 LPU 方法，其核心概念是：在正例和未标记示例上训练的分类器所预测的概率与正例的真实条件概率仅相差一个常数因子 [43]。从正例和未标记实例中学习的功能定义如下：

$$g(x) = p(\hat{y} = 1|x) \quad (9)$$

其中 x 表示实例。如果实例 x 被标注，则 $y = 1$ ，否则 $y = 0$ 。
我们的目标是学习一个函数 $f(x) = p(y = 1|x)$ 。其核心概念可以表述为：

$$\begin{aligned} p(\hat{y} = 1|x) &= p(y = 1 \cap \hat{y} = 1|x) \\ &= p(y = 1|x)p(\hat{y} = 1|y = 1, x) \\ &= p(y = 1|x)p(\hat{y} = 1|y = 1) \\ &= f(x)p(\hat{y} = 1|y = 1) \end{aligned} \quad (10)$$

将等式两边同时除以 $p(y = 1|y = 1)$ ，结果即得。因此，常数概率 $c = p(y = 1|y = 1)$ 。

我们将药物对的表示和标签矩阵（包括正例和未标记实例）输入到随机森林分类器 [61] 中进行训练，训练时随机移除 10% 的正例。设 P 为所选的正例集合。然后将 P 中的实例输入到训练好的随机森林分类器中，为每个实例获取 $g(x)$ 。 c 的估计值按如下方式计算：

$$c = \frac{1}{n} \sum_{x \in P} g(x) \quad (11)$$

其中 n 是集合 P 的基数。最后，我们将训练好的分类器应用于测试集，然后根据公式 11 生成 $f(x)$ 。

4. 实验

在本节中，我们将详细说明实验设置，并展示所提出模型与其他基准方法的性能对比。

4.1. 实验设置

在本文中，我们使用 k 折交叉验证（即 3 折交叉验证、5 折交叉验证）来评估 DDI 预测模型的性能。我们将已知的药物相互作用（DDI）随机等分为 k 个子集。然后，我们选择一个子集作为测试集，其余的作为训练集。为了实现 DDI 预测，隐藏测试集中现有的链接，使用剩余网络来训练 DDI-MDAE 模型。然后，我们利用获得的药物表示来预测未观察到的链接。我们分别对所有方法应用 3 折交叉验证和 5 折交叉验证，以评估我们的深度架构与其他模型相比的性能。

采用了几个广泛使用的评估指标来衡量预测模型的性能，包括受试者工作特征曲线下的面积（AUC）、准确率和召回率、准确度（ACC）、F 值（F1）以及精确率-召回率曲线下的面积（AUPR）。

4.2. 参数讨论

我们调整基准方法的参数以实现最佳性能。在此，我们考虑了几个关键参数的不同组合。DDI-MDAE 模型包含 3 个隐藏层。

$d \in \{32, 64, 128, 192, 256\}$ 是共享层的嵌入尺寸，
 $\{0.01, 0.1, 0.2, 0.3, 0.4\}$, $\{2, 5, 10, 15, 20\}$, $\{0, 0.001, 0.002, 0.005, 0.01\}$
 正则化项

$$\alpha \in \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\} \text{ 以及学习率}$$

设置学习率为 $\{0.005, 0.01, 0.02, 0.03, 0.04\}$ 。至于随机设置方面对于森林分类器，我们选择保留 `sklearn.ensemble.RandomForestClassifier` [62] 的默认设置来进行药物相互作用预测。在实验中，我们对药物对的表示形式评估了这四种运算符。

我们研究了不同参数组合对所提出方法性能的影响。我们以初始值 $d = 128, \alpha = 0.01, \beta = 10, \gamma = 1e-5$ 以及学习率为 0.01 为起点，寻找合适的参数。在分析某个测试参数的影响时，其余参数均固定为默认值。我们报告了在 3 折交叉验证（3-CV）下不同超参数组合的 AUPR 指标结果。如图 2 所示，当 $d = 128$ 或 256 时，DDI-MDAE 获得了最佳的 AUPR 分数。参数 β 被设计用来衡量已知药物相互作用（DDIs）对 DDI-MDAE 的影响。我们的

当 $\alpha = 0$ 时，该方法仅受 H_1 和 H_3 的约束。可以观察到

使用 $\gamma = 0.01$ 时模型的表现优于

α 等于 0 且当 α 增大时，AUPR 分数降低，这表明对于参数 α 而言，它是对零元素重建进行控制的关键，对于我们的方法至关重要，应当谨慎确定。至于参数 β ，它用于衡量多个药物特征网络（即第 3.3 节中定义的 H_3 ）中药物相似性的效果。当 β 从 2 增加到 10 时，AUPR 分数随之上升，在 $\beta = 10$ 时达到峰值，之后则下降，因此，我们选择 $\beta = 10$ 作为实验评估。参数 γ 旨在衡量多个药物特征网络（即第 3.3 节中定义的 H_3 ）中药物相似性的效果。当 $\gamma = 0.001$ 时，DDI-MDAE 的 AUPR 分数比 $\gamma = 0$ 时有所提高。对于参数 γ 和学习率，它们对我们的方法影响较小。可以推断，在我们选择的参数范围内，我们的模型对 γ 和学习率具有一定的鲁棒性。

经过上述分析，在接下来的评估实验中，我们采用 $d = 128, \alpha = 0.01, \beta = 10, \gamma = 1e-5$ 以及学习率为 0.01。

4.3. 基准方法

在本节中，我们将介绍几种最先进的药物相互作用预测方法以作比较。

协同过滤 [7] 是一种基于相似药物可能与相同药物发生相互作用这一假设的直观方法。我们在每个药物特征网络上应用协同过滤，并获得了五个协同过滤模型，即基于结构的协同过滤、基于靶点的协同过滤、基于酶的协同过滤、基于通路的协同过滤和基于药物相互作用的协同过滤。

- **随机游走** [14] 采用带有重启策略的随机游走方法来模拟药物特征网络中的信号传播，以捕捉它们可能存在的远距离干扰。与协同过滤方法相同，我们构建了五个随机游走模型，即基于结构的 RW、基于靶点的 RW、基于酶的 RW、基于通路的 RW 和基于药物相互作用的 RW。

- **MP** [18] 提出了一种基于网络结构不会因随机移除少量链路而改变这一假设的矩阵扰动方法。在接下来的稀疏性实验中，将使用 MP 方法与我们提出的模型进行比较，该实验将在下一部分进行描述。

- **LPMS** [13] 提出了一种用于药物相互作用预测的综合标签传播框架，该框架结合了药物化学子结构、标签副作用和非标签副作用这三种相似性。

- **SFLLN** [10] 是一种具有线性邻域正则化的稀疏特征学习集成方法，它结合了多种药物特征，包括化学亚结构、靶点、酶、通路以及已知的药物相互作用，用于药物相互作用的预测。

- **Node2vec** [28] 是一种通过训练一系列可调参数的游走来进行网络嵌入的方法。我们采用此方法进行药物表征学习，并使用随机森林分类器进行预测。

4.4. 结果

在本节中，我们首先报告了在 3 折和 5 折交叉验证下，我们的方法与基准方法的实证结果。我们采用了四种操作符（即平均、哈达玛积、L1 范数和拼接），并评估了它们对模型性能的影响。如表 4 所示，在药物表示中使用拼接操作符的方法非常稳定，并且在 3 折和 5 折交叉验证中均优于其他三种操作符。表 5 显示，在大多数情况下，我们的方法（使用拼接操作符）比其他基准方法表现更好，在 3 折交叉验证中，F 值指标比 SFLLN 方法提高了 16.6%，AUPR 指标比基于 DDI 的 CF 方法提高了 12.4%。如表 6 所示，在 5 折交叉验证中，F 值指标比 node2vec 方法提高了 19.5%，AUPR 指标比 node2vec 方法提高了 11.8%。从结果可以看出，基于药物特征（不包括 DDI 特征）的协同过滤（CF）方法、LPMS 方法以及随机游走方法表现不佳。这是因为表 2 中呈现的药物特征网络非常嘈杂且稀疏。

我们开展实验以分析不同阈值对 DDI-MDAE（PU 学习）性能的影响。如表 7 所示，我们发现当阈值等于 0.5 时，DDI-MDAE（PU 学习）在大多数评估指标上都能取得最佳预测性能。我们还将 DDI-MDAE（PU 学习）与原始方法 DDI-MADE 进行了比较。如表 8 所示，DDI-MDAE（PU 学习）的性能优于 DDI-MADE。DDI-MDAE（PU 学习）

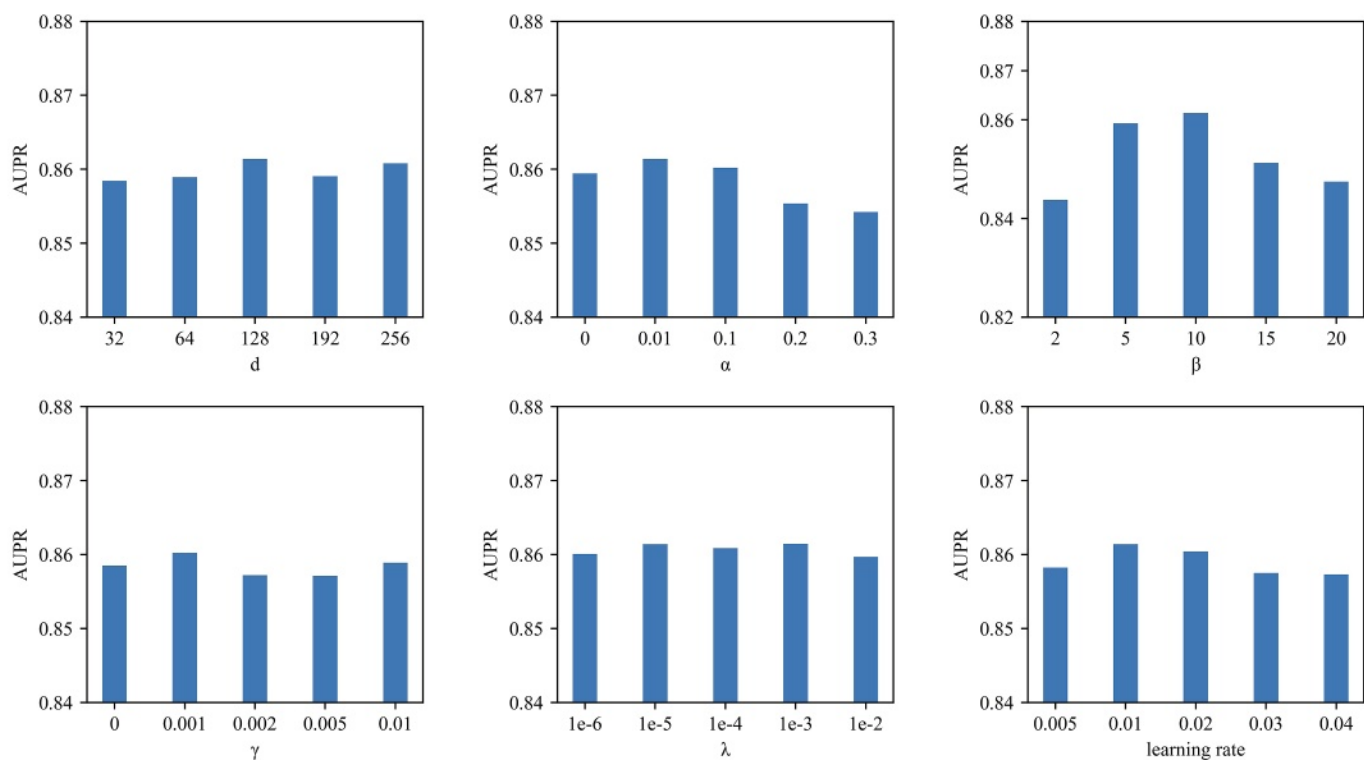


图 2. 参数敏感性。

在 3 折交叉验证中, DDI-MDAE (PU 学习) 在 AUPR 指标上提高了 7.1%, 在 F 值指标上提高了 10.4%。在 5 折交叉验证中, DDI-MDAE (PU 学习) 在 AUPR 指标上提高了 6.2%, 在 F 值指标上提高了 8.4%。因此, PU 学习策略能够提升我们模型的预测性能。

为了评估我们基于深度学习架构的方法在上述稀疏且有噪声的药物特征网络中的表现, 我们设计了一个实验, 通过移除网络中一定比例的链接来评估其在不同稀疏度下的性能。图 3 报告了在 3 折和 5 折交叉验证中不同预测方法的结果。如图所示, 当移除的链接比例在 10% 到 50% 之间时, DDI-MDAE 显著优于其他基准方法。即使移除了其一半的链接, 我们的方法仍能获得高于 0.5 的 AUPR 分数。在该任务中, 我们发现通过 PU 学习将 MDAE 的阈值提高到约 0.7 或 0.8, 在一定程度上能够提升性能。可以看出, 通过 PU 学习改进的方法在 3 折和 5 折交叉验证中均优于原始的 DDI-MDAE 方法。我们的方法和 Node2vec 都是基于网络嵌入的方法, 可以观察到, 当网络更稀疏时, 基于网络嵌入的方法能够取得更好的性能, 这可能是因为

表 5
不同模型在 3 折交叉验证上的实验结果。

方法	AUPR	AUC	F1	准确性	精度	召回
基于结构的协同过滤	0.3274	0.8512	0.3662	0.9416	0.3327	0.4077
基于目标的协同过滤	0.4310	0.8200	0.4386	0.9513	0.4193	0.4597
基于酶的 CF	0.3157	0.7485	0.3526	0.9574	0.4760	0.2804
基于通路的协同过滤	0.3895	0.8359	0.3958	0.9553	0.4491	0.3538
基于药物相互作用的协同过滤	0.7667	0.9758	0.7000	0.9773	0.7741	0.6390
基于结构的随机游走	0.3250	0.8501	0.3642	0.9407	0.3282	0.4100
基于目标的随机游走	0.4274	0.8022	0.4378	0.9516	0.4214	0.4555
基于酶的 RW	0.3139	0.7446	0.3524	0.9573	0.4741	0.2805
基于路径的随机游走	0.3846	0.8320	0.3926	0.9546	0.4407	0.3540
基于 DDI 的关系抽取	0.7517	0.9736	0.6880	0.9764	0.7586	0.6296
LPMS	0.3113	0.8440	0.3533	0.9391	0.3160	0.4015
SFLN	0.7574	0.9502	0.7149	0.9535	0.7070	0.7237
Node2vec 算法	0.7460	0.7394	0.6465	0.9863	0.9995	0.4789
DDI-MDAE	0.8614	0.8582	0.8339	0.9925	0.9989	0.7165

在稀疏网络中, 它成为了一种有效的药物相互作用预测策略。至于我们之前的工作 SFLN [10], 实验结果表明其 AUPR 分数并不理想。需要注意的是, SFLN 的应用场景与当前情况并非完全一致。

表 4
不同操作符在药物表示上的实验结果, 基于 3 折和 5 折交叉验证。

方法	AUPR	AUC	F1	准确性	精度	召回
3-CV						
DDI-MDAE (平均值)	0.7921	0.7875	0.7271	0.9888	0.9980	0.5750
DDI-MDAE (哈达玛)	0.7950	0.7903	0.7316	0.9890	0.9983	0.5807
DDI-MDAE (加权 L1)	0.6466	0.6382	0.4287	0.9810	0.9978	0.2764
DDI-MDAE (串联)	0.8614	0.8582	0.8339	0.9925	0.9989	0.7165
5-CV						
DDI-MDAE (平均值)	0.8149	0.8138	0.7683	0.9940	0.9963	0.6275
DDI-MDAE (哈达玛)	0.8196	0.8185	0.7752	0.9942	0.9964	0.6371
DDI-MDAE (加权 L1)	0.6685	0.6652	0.4917	0.9893	0.9959	0.3304
DDI-MDAE (串联)	0.8714	0.8703	0.8498	0.9958	0.9980	0.7407

表 6
不同模型在五折交叉验证上的实验结果。

方法	AUPR	AUC	F1	准确性	精度	召回
基于结构的协同过滤	0.2478	0.8514	0.3006	0.9601	0.2699	0.3393
基于目标的协同过滤	0.3665	0.8232	0.3807	0.9695	0.3908	0.3713
基于酶的 CF	0.2587	0.7483	0.3186	0.9724	0.4230	0.2556
基于通路的协同过滤	0.3187	0.8363	0.3522	0.9711	0.4063	0.3111
基于药物相互作用的协同过滤	0.7045	0.9760	0.6643	0.9846	0.7416	0.6016
基于结构的随机游走	0.2456	0.8503	0.2987	0.9602	0.2693	0.3356
基于路径的随机游走	0.3642	0.8046	0.3794	0.9689	0.3826	0.3764
基于酶的 RW	0.2572	0.7446	0.3178	0.9724	0.4229	0.2546
基于路径的随机游走	0.3139	0.8324	0.3487	0.9711	0.4060	0.3060
基于 DDI 的关系抽取	0.6870	0.9738	0.6506	0.9840	0.7276	0.5884
LPMS	0.2337	0.8442	0.2877	0.9595	0.2591	0.3236
SFLN	0.6286	0.8792	0.5834	0.9056	0.5982	0.5693
Node2vec 算法	0.7793	0.7763	0.7109	0.9929	0.9988	0.5527
DDI-MDAE	0.8714	0.8703	0.8498	0.9958	0.9980	0.7407

表 7
DDI-MDAE (PU 学习) 在 5 折交叉验证中采用不同阈值的实验结果。

门檻	AUPR	AUC	F1	准确性	精度	召回
0.1	0.6197	0.9689	0.4047	0.9539	0.2547	0.9844
0.2	0.8401	0.8376	0.8060	0.9948	0.9998	0.6751
0.3	0.9084	0.9590	0.9076	0.9970	0.8958	0.9197
0.4	0.9104	0.9372	0.9096	0.9096	0.8971	0.9225
0.5	0.9252	0.9604	0.9215	0.9976	0.9736	0.8748
0.6	0.9055	0.9070	0.8950	0.9970	0.9940	0.8140
0.7	0.8775	0.8761	0.8581	0.9960	0.9989	0.7521
0.8	0.8744	0.8729	0.8540	0.9959	0.9990	0.7459
0.9	0.8344	0.8318	0.7977	0.9946	0.9998	0.6636

表 8
改进模型与原始模型在 3 折和 5 折交叉验证下的实验结果。

方法	AUPR	AUC	F1	准确性	精度	召回
3-CV						
DDI-MDAE	0.8614	0.8582	0.8339	0.9925	0.9989	0.7165
DDI-MDAE (无监督学习)	0.9229	0.9486	0.9210	0.9960	0.9446	0.8985
5-CV						
DDI-MDAE DDI-MDAE (PU 学习)	0.8714	0.8703	0.8498	0.9958	0.9980	0.7407
	0.9252	0.9604	0.9215	0.9976	0.9736	0.8748

SFLN 采用了 DDI-MDAE 的方法。SFLN 集成了多种药物特征，并且假定每种药物都具备所有特征。如前所述，本文所用的数据集规模大且稀疏。

4.5. 案例研究

在实验中，我们的数据集包含 2367 种药物以及 2800161 对药物间的相互作用。具体而言，已知的药物间相互作用有 209494 对，还有 2590667 对未标注的药物组合，其中可能包含未被观察到的药物间相互作用。我们应用 DDI-MDAE 通过所有已知相互作用进行训练来预测未被观察到的药物间相互作用，未标注药物组合的得分越高，表明其存在相互作用的可能性越大。表 9 展示了我们方法预测出的前 20 种药物间相互作用，这些在我们的数据集中并不存在。我们查找了新预测出的药物间相互作用的证据，发现其中很大一部分（20 种中有 16 种）在最新的在线版本 5.1.4 的 DrugBank 数据库中得到了证实。例如，药物“氯氮平”（DB00363）与药物“兰索拉唑”（DB00448）之间的相互作用描述为“氯氮平与兰索拉唑合用时，其代谢可能会降低。”这些案例研究证明了我们的方法在检测潜在药物间相互作用方面的有效性。

5. 结论

在本文中，我们探讨了如何利用多个药物数据集来准确预测药物间的相互作用。我们提出了一种名为 DDI-MDAE 的深度架构，旨在捕捉多种非线性网络结构。DDI-MDAE 旨在通过多模态深度自编码器学习统一的药物表示，其中每个药物特征网络都由一个深度自编码器通道进行训练，并且隐藏层中的药物表示被共享以实现同步学习。此外，我们对学习到的药物嵌入应用四种运算符来表示药物对。我们采用随机森林分类器来进行药物相互作用预测。在基准数据集上的评估表明，与现有的其他药物相互作用预测模型相比，我们的方法在性能上有显著提升。特别是，我们的模型也适用于非常稀疏的网络。此外，我们通过 PU 学习设置中应用专门的随机森林分类器来改进所提出的 DDI-MDAE 方法，以提高预测准确性。

在本文中，我们的模型仅从药物特征网络的结构拓扑中学习药物表示。在我们未来的工作中，我们考虑将网络拓扑和语义信息结合起来用于药物表示，这可能会提高预测性能。

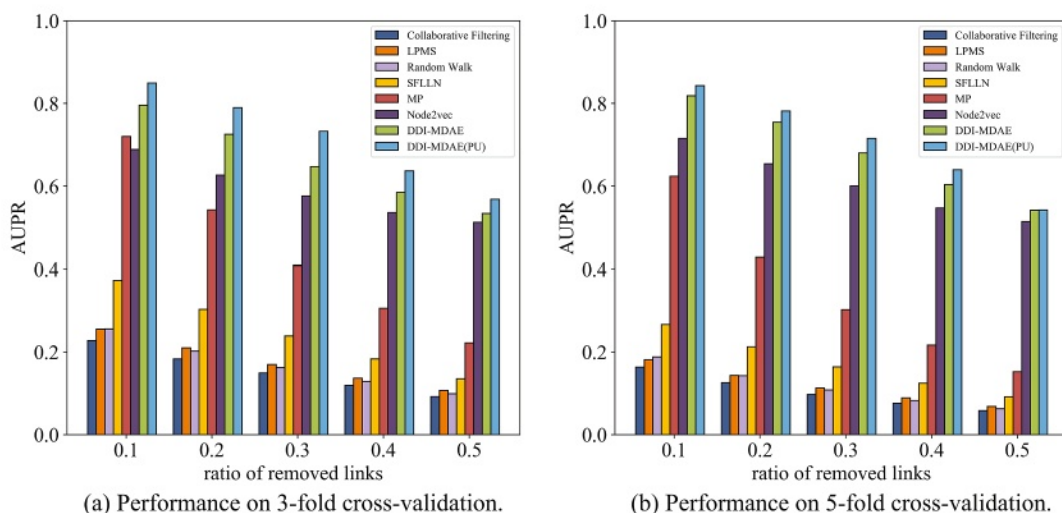


图 3. 在不同稀疏度的网络上的表现。

表9

我们方法预测出的前 20 个药物相互作用 (DDI)。

等级	药物 1	药物 2	描述
1	氯氮平	兰索拉唑	氯氮平与兰索拉唑合用时, 其代谢可能会减慢。
2	甲基麦角新碱	西酞普兰	无可用数据。
3	斯皮拉普利	舒芬太尼	舒芬太尼可能会降低螺旋哌啶的降压作用。
4	氯苯酚	氟哌啶醇	当氟哌啶醇与氯苯酚合用时, 其血清浓度会升高。
5	卡马西平	马尼地平	卡马西平与氯氮地平合用时, 其代谢可能会减慢。
6	乙酰水杨酸	氯氮平	乙酰水杨酸可能会降低氯氮平的排泄率, 从而导致血清水平升高。
7	沙奎那韦	泰普替啶	当与沙奎那韦合用时, 泰普替啶的代谢可能会减缓。
8	吗氯贝胺	帕唑帕尼	帕唑帕尼与吗氯贝胺合用时, 其代谢可能会减缓。
9	氟伏沙明	利福喷丁	氟伏沙明与利福喷丁合用时, 其代谢可能会加快。
10	西酞普兰	扎鲁司特	西酞普兰与扎鲁司特合用时, 其血清浓度可能会升高。
11	塞利洛尔	西尼地平	西尼地平与塞利洛尔合用时, 可能会增加心动过缓的风险或严重程度。
12	昂达特罗	舍曲林	昂达特罗与舍曲林合用时, 其血清浓度会升高。
13	尼麦角林	泮托拉唑	无可用数据。
14	托莫西汀	扑米酮	当托莫西汀与扑米酮合用时, 不良反应的风险或严重程度可能会增加。
15	奥希替尼	匹伐他汀	奥希替尼与匹伐他汀合用时, 其血清浓度会升高。
16	西比灵	沙奎那韦	当西比灵与沙奎那韦合用时, 可能会增加 QTc 间期延长的风险或严重程度。
17	地拉罗司	丙戊酸	丙戊酸与地拉罗司合用时, 丙戊酸的代谢可能会减慢。
18	拉帕替尼	替格瑞洛	替格瑞洛与拉帕替尼合用时, 其代谢可能会减缓。
19	雷迪帕韦	泰利霉素	无可用数据。
20	比马前列素	雌酮	无可用数据。

N.A.: 证据未提供。

CRediT 作者贡献声明

杨章: 概念化、方法论、软件开发。**杨秋:** 数据整理、原始草稿撰写。**崔雨欣:** 软件开发、验证。**刘世超:** 概念化、方法论、监督、项目管理。**张文:** 审校与编辑。

致谢

本研究得到了国家自然科学基金 (61772381, 61572368)、国家重点研发计划 (2018YFC0407904)、华中农业大学自主科技创新基金、中央高校基本科研业务费专项资金 (2042017kf0219, 2042018kf0249, 2662019QD011) 的支持。资助方在研究设计、数据收集、数据分析、数据解读或论文撰写方面均未发挥任何作用。

附录 A. 补充数据

与本文相关的补充数据可在在线版本中获取, 网址为: <https://doi.org/10.1016/j.ymeth.2020.05.007>。

参考文献

- [1] X. Sun, L. Ma, X. Du, J. Feng, and K. Dong, "Deep convolution neural networks for drug-drug interaction extraction," in 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2018, pp. 1662–1668.
- [2] M.F.M. Chowdhury, A. Lavelli, Fbk-irst: A multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information, Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval), 2013, pp. 351–355.
- [3] S. Kim, H. Liu, L. Yeganova, W.J. Wilbur, Extracting drug–drug interactions from literature using a rich feature-based linear kernel approach, J. Biomed. Inform. 55 (2015) 23–30.
- [4] Y. Zhang, W. Zheng, H. Lin, J. Wang, Z. Yang, M. Dumontier, Drug–drug interaction extraction via hierarchical rnns on sequence and shortest dependency paths, Bioinformatics 34 (5) (2017) 828–835.
- [5] Z. Jiang, L. Gu, Q. Jiang, "Drug drug interaction extraction from literature using a skeleton long short term memory neural network," in 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2017, pp. 552–555.
- [6] B. Xu, X. Shi, Z. Zha, W. Zheng, H. Lin, Z. Yang, J. Wang, F. Xia, "Full-attention based drug drug interaction extraction exploiting user-generated content," in 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2018, pp. 560–565.
- [7] S. Vilar, R. Harpaz, E. Uriarte, L. Santana, R. Rabadan, C. Friedman, Drug–drug interaction through molecular structure similarity analysis, J. Am. Med. Inform.

Assoc. 19 (6) (2012) 1066–1074.

- [8] A. Gottlieb, G.Y. Stein, Y. Oron, E. Ruppim, R. Sharan, "Indi: a computational framework for inferring drug interactions and their associated recommendations," Mol. Syst. Biol., vol. 8, no. 1, 2012.
- [9] T. Ma, C. Xiao, J. Zhou, and F. Wang, "Drug similarity integration through attentive multi-view graph auto-encoders," arXiv preprint arXiv:1804.10850, 2018.
- [10] W. Zhang, K. Jing, F. Huang, Y. Chen, B. Li, J. Li, J. Gong, Sfln: A sparse feature learning ensemble method with linear neighborhood regularization for predicting drug–drug interactions, Inf. Sci. 497 (2019) 189–201.
- [11] J.-Y. Shi, H. Huang, J.-X. Li, P. Lei, Y.-N. Zhang, K. Dong, S.-M. Yiu, Tmfuf: a triple matrix factorization-based unified framework for predicting comprehensive drug–drug interactions of new drugs, BMC Bioinform. 19 (14) (2018) 411.
- [12] H. Yu, K.-T. Mao, J.-Y. Shi, H. Huang, Z. Chen, K. Dong, S.-M. Yiu, Predicting and understanding comprehensive drug–drug interactions via semi-nonnegative matrix factorization, BMC Syst. Biol. 12 (1) (2018) 14.
- [13] P. Zhang, F. Wang, J. Hu, R. Sorrentino, Label propagation prediction of drug–drug interactions based on clinical side effects, Sci. Rep. 5 (2015) 12339.
- [14] K. Park, D. Kim, S. Ha, D. Lee, Predicting pharmacodynamic drug–drug interactions through signaling propagation interference on protein–protein interaction networks, PLoS one 10 (10) (2015) e0140816.
- [15] D. Sridhar, S. Fakhraei, L. Getoor, A probabilistic approach for collective similarity-based drug–drug interaction prediction, Bioinformatics 32 (20) (2016) 3175–3182.
- [16] J. Huang, C. Niu, C.D. Green, L. Yang, H. Mei, J.-D.J. Han, Systematic prediction of pharmacodynamic drug–drug interactions through protein–protein–interaction network, PLoS Comput. Biol. 9 (3) (2013) e1002998.
- [17] S. Deepika, T. Geetha, A meta-learning framework using representation learning to predict drug–drug interaction, J. Biomed. Inform. 84 (2018) 136–147.
- [18] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, X. Li, Predicting potential drug–drug interactions by integrating chemical, biological, phenotypic and network data, BMC Bioinform. 18 (1) (2017) 18.
- [19] T. Takeda, M. Hao, T. Cheng, S.H. Bryant, Y. Wang, Predicting drug–drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge, J. Cheminform. 9 (1) (2017) 16.
- [20] P. Zhang, F. Wang, J. Hu, "Towards drug repositioning: a unified computational framework for integrating multiple aspects of drug similarity and disease similarity," in AMIA Annual Symposium Proceedings, vol. 2014. American Medical Informatics Association, 2014, p. 1258.
- [21] F. Cheng, Z. Zhao, Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties, J. Am. Med. Inform. Assoc. 21 (e2) (2014) e278–e286.
- [22] A. Cami, S. Manzi, A. Arnold, B.Y. Reis, Pharmacointeraction network models predict unknown drug–drug interactions, PLoS one 8 (4) (2013) e61468.
- [23] X. Yue, Z. Wang, J. Huang, S. Parthasarathy, S. Moosavinasab, Y. Huang, S.M. Lin, W. Zhang, P. Zhang, H. Sun, Graph embedding on biomedical networks: methods, applications and evaluations, Bioinformatics 36 (4) (2020) 1241–1251.
- [24] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
- [25] J.B. Tenenbaum, V. De Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
- [26] B. Perozzi, R. Al-Rfou, S. Skiena, Deepwalk: Online learning of social representations, Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2014, pp. 701–710.
- [27] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, Q. Mei, "Line: Large-scale information network embedding, Proceedings of the 24th International Conference on World

- Wide Web. International World Wide Web Conferences Steering Committee, 2015, pp. 1067–1077.
- [28] A. Grover, J. Leskovec, "node2vec: Scalable feature learning for networks, Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2016, pp. 855–864.
- [29] D. Wang, P. Cui, W. Zhu, "Structural deep network embedding, Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2016, pp. 1225–1234.
- [30] T.N. Kipf, M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.
- [31] S. Chang, W. Han, J. Tang, G.-J. Qi, C.C. Aggarwal, T.S. Huang, "Heterogeneous network embedding via deep architectures," in, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 119–128.
- [32] Y. Dong, N.V. Chawla, A. Swami, "metapath2vec: Scalable representation learning for heterogeneous networks, Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, ACM, 2017, pp. 135–144.
- [33] Y. Ma, Z. Ren, Z. Jiang, J. Tang, D. Yin, "Multi-dimensional network embedding with hierarchical structure, Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, ACM, 2018, pp. 387–395.
- [34] C. Yang, Z. Liu, D. Zhao, M. Sun, E.Y. Chang, "Network representation learning with rich text information, Proceedings of the 24th International Joint Conference on Artificial Intelligence, Buenos Aires Argentina, 2015, pp. 2111–2117.
- [35] L. Liao, X. He, H. Zhang, T.-S. Chua, Attributed social network embedding, IEEE Trans. Knowl. Data Eng. 30 (12) (2018) 2257–2270.
- [36] S. Liu, S. Zhai, L. Zhu, F. Zhu, Z.M. Zhang, W. Zhang, "Efficient network representations learning: An edge-centric perspective," in International Conference on Knowledge Science, Engineering and Management. Springer, 2019, pp. 373–388.
- [37] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2015) 295–307.
- [38] G.E. Dahl, D. Yu, L. Deng, A. Acero, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, IEEE Trans. Audio, Speech, Language Process. 20 (1) (2011) 30–42.
- [39] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, B. Kingsbury et al., "Deep neural networks for acoustic modeling in speech recognition," IEEE Signal processing magazine, vol. 29, 2012.
- [40] G.E. Hinton, R.S. Zemel, Autoencoders, minimum description length and helmholtz free energy, Adv. Neural Inform. Process. Syst. (1994) 3–10.
- [41] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, Stacked denoising autoencoders: Learning useful representations in a deep network with a local de-noising criterion, J. Mach. Learn. Res. 11 (Dec) (2010) 3371–3408.
- [42] B. Liu, Y. Dai, X. Li, W.S. Lee, P.S. Yu, "Building text classifiers using positive and unlabeled examples," in Third IEEE International Conference on Data Mining, IEEE, 2003, pp. 179–186.
- [43] C. Elkan, K. Noto, "Learning classifiers from only positive and unlabeled data, Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2008, pp. 213–220.
- [44] W.S. Lee, B. Liu, "Learning with positive and unlabeled examples using weighted logistic regression," in ICML, vol. 3, 2003, pp. 448–455.
- [45] Z. Liu, W. Shi, D. Li, Q. Qin, "Partially supervised classification: based on weighted unlabeled samples support vector machine, Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications, IGI Global, 2008, pp. 1216–1230.
- [46] H. Yu, J. Han, K.-C. Chang, Pebl: Web page classification without negative ex-amples, IEEE Trans. Knowl. Data Eng. 16 (1) (2004) 70–81.
- [47] B. Liu, W.S. Lee, P.S. Yu, X. Li, "Partially supervised classification of text documents," in ICML, vol. 2, Citeseer, 2002, pp. 387–394.
- [48] X. Li, B. Liu, Learning to classify texts using positive and unlabeled data, IJCAI 3 (2003) 587–592.
- [49] F. Mordelet, J.-P. Vert, A bagging svm to learn from positive and unlabeled ex-amples, Pattern Recogn. Lett. 37 (2014) 201–209.
- [50] P. Yang, X. Li, H.-N. Chua, C.-K. Kwok, S.-K. Ng, Ensemble positive unlabeled learning for disease gene identification, PloS one 9 (5) (2014) pp.
- [51] P. Yang, X.-L. Li, J.-P. Mei, C.-K. Kwok, S.-K. Ng, Positive-unlabeled learning for disease gene identification, Bioinformatics 28 (20) (2012) 2640–2647.
- [52] F. Mordelet, J.-P. Vert, Prodigy: Prioritization of disease genes with multitask machine learning from positive and unlabeled examples, BMC Bioinform. 12 (1) (2011) 389.
- [53] W. Lan, J. Wang, M. Li, J. Liu, Y. Li, F.-X. Wu, Y. Pan, Predicting drug–target interaction using positive-unlabeled learning, Neurocomputing 206 (2016) 50–57.
- [54] Y. Zheng, H. Peng, X. Zhang, Z. Zhao, X. Gao, J. Li, Ddi-pulearn: a positive-un-labeled learning method for large-scale prediction of drug-drug interactions, BMC Bioinform. 20 (19) (2019) 1–12.
- [55] P.N. Hameed, K. Verspoor, S. Kusljic, S. Halgamuge, Positive-unlabeled learning for inferring drug interactions based on heterogeneous attributes, BMC Bioinform. 18 (1) (2017) 1–15.
- [56] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, "Drugbank: a comprehensive resource for in silico drug discovery and exploration," Nucleic acids research, vol. 34, no. suppl_1, pp. D668–D672, 2006.
- [57] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, M. Hirakawa, "Kegg for representation and analysis of molecular networks involving diseases and drugs," Nucleic acids research, vol. 38, no. suppl_1, pp. D355–D360, 2009.
- [58] Y. Wang, J. Xiao, T.O. Suzek, J. Zhang, J. Wang, S.H. Bryant, "Pubchem: a public information system for analyzing bioactivities of small molecules," Nucleic acids research, vol. 37, no. suppl_2, pp. W623–W633, 2009.
- [59] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
- [60] M. Jamali, M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks," in, Proceedings of the fourth ACM conference on Recommender systems, ACM, 2010, pp. 135–142.
- [61] A. Liaw, M. Wiener, et al., Classification and regression by randomforest, R news 2 (3) (2002) 18–22.
- [62] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., "Scikit-learn: Machine learning in python," Journal of machine learning research, vol. 12, no. Oct, pp. 2825–2830, 2011.