文章编号：2095-2163(2022)09-0133-06 中图分类号：TP391 文献标志码：A

基于全连接自动编码器的疾病相关 microRNA 预测

徐春旭，玄 萍
（黑龙江大学 计算机科学技术学院，哈尔滨 150080）

摘 要： MicroRNA（miRNA）是一类长度约为 22~24 个核苷酸的非编码 RNA，在细胞的生长和发育过程中发挥着重要的调节作用。识别与疾病相关的 miRNA 候选，有助于探索疾病的发生机理，先前的多数预测方法主要聚焦在整合 miRNA 和疾病相关的多源数据，忽略了 miRNA 家族和聚簇信息，但是相似的疾病通常更可能与具有相似功能和属于相同家族或聚簇的 miRNA 相关。本文根据 miRNA 和疾病节点形成的拓扑结构，结合 miRNA 的家族和聚簇属性，建立了基于全连接自动编码器的 miRNA-疾病关联预测模型。采用了五倍交叉验证进行了实验评估，实验结果表明新的预测模型取得了比其它几个模型更优的性能。

关键词： miRNA-疾病关联；全连接自动编码器；miRNA 家族和聚簇属性

Prediction of disease-related microRNAs based on fully-connected autoencoder

XU Chunxu, XUAN Ping
(School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China)

【Abstract】 MicroRNA (miRNA) is a kind of non-coding RNA with a length of about 22~24 nucleotides, which plays an important regulatory role in the growth and development of cells. Identifying the candidate miRNAs associated with the diseases is helpful for exploring the pathogenesis of diseases. Most of the previous prediction methods focus on integrating miRNA-related and disease-related multi-source data, and ignore the information about miRNA families and clusters. However, similar diseases are usually more likely to associate with the miRNAs that have similar functions and the ones belonging to the same families or clusters. Therefore, according to the topological structure formed by miRNA and disease nodes and combined with the family and cluster information of miRNA nodes, a miRNA-disease association prediction model based on fully-connected autoencoders is established. The paper utilizes five-fold cross-validation to measure the prediction performance, and the experimental results show that the new prediction model is better than other several models.

【Key words】 miRNA-disease associations; fully-connected autoencoder; miRNA family and cluster information

0 引 言

MicroRNA（miRNA）是一种内源的非编码 RNA，具体长度约为 22~24 个核苷酸[1-2]。miRNA 通过靶向信使 RNA 进行剪接或抑制其翻译，进而在动物和植物中发挥重要的调节作用。越来越多的证据表明，miRNA 参与了许多疾病的发生和发展进程[3]。因此，识别可能与疾病相关的候选 miRNA，有助于探索疾病的发生机理。

早期的研究主要通过生物实验获得准确度高的实验结果，但实验成本高、耗时长、成功率低。近年来，研究人员越来越多地利用基于信息学的方法来预测与疾病相关的 miRNA，并取得了良好的效果。这些方法可以分成 2 类。第一类方法主要基于具有相似功能的 miRNA，这些 miRNA 通常与相似的疾病相关[4]。例如，Wang 等人[5]基于与 miRNA 相关的疾病来计算 miRNA 的相似性。Xuan 等人[6]提出了一种基于最相似 miRNA 节点加权 k - 近邻信息的预测方法。这些方法只适用于与已知 miRNA 相关的特定疾病，无法预测无已知相关 miRNA 的新疾病的候选。为了解决这个问题，第二类方法引入了疾病的相似信息。Chen 等人[7]提出了结合 k 近邻和支持向量机的方法，来预测潜在的 miRNA-疾病关联。但上述方法均没有充分整合异构图中蕴含的拓扑信息和 miRNA 所属的家族和聚簇信息。

RFam 和 miRBase 等数据库已收录了 miRNA 的

基金项目：国家自然科学基金(61972135)；黑龙江省自然科学基金项目(LH2019F049)；中国博士后科学基金(2019M650069)；黑龙江省博士后科研启动基金(BHLQ18104)。

作者简介：徐春旭(1998-)，男，硕士研究生，主要研究方向：生物信息学、深度学习；玄 萍(1979-)，女，博士，教授，主要研究方向：复杂生物网络分析、深度学习、数据挖掘。

通讯作者：玄 萍 Email：xuanping@hlju.edu.cn

收稿日期：2022-02-20

家族信息。同一家族中的同源 miRNA 通常具有几乎相同的种子区,同一家族的 miRNA 通常具有类似的功能,比如调节相同的目标,因此就更有可能与相似的疾病相关[8]。此外,许多 miRNA 虽然可能不属于同一家族,但都位于邻近的基因组位点上,形成了 miRNA 聚簇。同一聚簇的 miRNA 通常同步转录并协调表达,并很可能参与了相似的生物过程[9]。因此,miRNA 的家族和聚簇信息可以表示为 miRNA 的节点属性,以构建更完善的 miRNA-疾病异构图,从而更准确地预测 miRNA-疾病关联。

1 miRNA-疾病关联预测方法

本文的主要目的是预测潜在的 miRNA-疾病关联候选。为了整合多种连接和节点属性的信息,首先构建了一个 miRNA-疾病异构图;其次,设计了一个基于全连接自动编码器的预测模型(Fully-connected autoencoders MicroRNA - Disease Associations prediction,FMDA);根据模型可以得到某个 miRNA 与特定疾病之间的关联得分,分数越高,两者之间关联的可能性就越大。

1.1 相关数据集

本文从人类 miRNA-疾病数据库中提取了 miRNA 与疾病的关联,该数据库包含 7 908 个经过实验验证的 miRNA-疾病关联,覆盖 7 93 个 miRNA 和 341 个疾病。本文基于美国国家医学图书馆的疾病术语信息构建相应疾病的有向无环图(DAG),来计算疾病的语义相似性。530 个 miRNA 家族的信息是从 miRNA 数据库 miRBase 中提取的,miRNA 的基因组位点信息也来自 miRBase。将 2 个 miRNA 之间的距离阈值设置为不超过 20 kb,从而提取了 1 309个聚簇。

1.2 建立 miRNA-疾病异构图

本文构建了疾病相似图、miRNA 相似图和 miRNA-疾病关联图,结合 miRNA 家族和聚簇属性建立 miRNA-疾病异构图,如图 1 所示。由图 1 可知,对其中各重要部分拟展开研究分述如下。

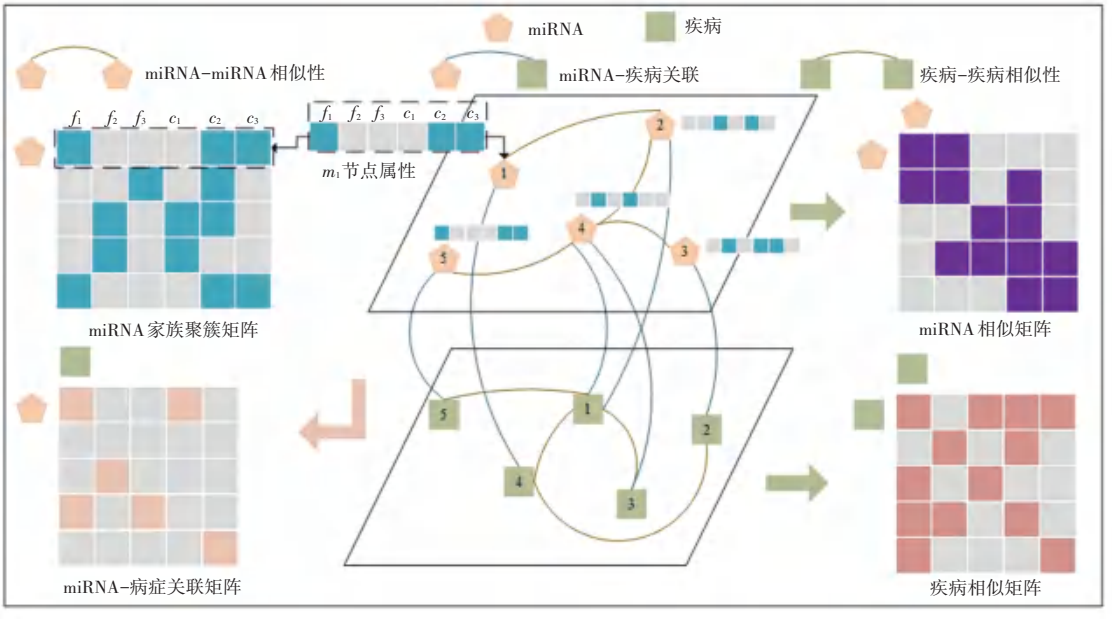


图 1 miRNA-疾病异构图的建立及矩阵表示

Fig. 1 Construction and matrix representation of miRNA-disease heterogeneous graph

(1)疾病相似图:计算疾病之间的相似度是构建疾病相似图的基础,可以从疾病语义的角度来量化相互间的相似程度。Wang 等人[5]根据疾病的有向无环图计算了疾病的语义相似度。疾病的 DAG 由与其相关的所有术语组成。2 个疾病的 DAG 中包含越多相似的术语,彼此间就越相似。本文依据该方法计算了第 i 个疾病与第 j 个疾病间的相似度 S_{ij}^{d}。对此可表示为:

$$S_{ij}^{d}=\frac{\sum_{t\in T_i\cap T_j}(\varphi(i,t)+\varphi(j,t))}{\sum_{t\in T_i}(\varphi(i,t))+\sum_{t\in T_j}(\varphi(j,t))} \tag{1}$$

其中,$\varphi(i,t)$ 为与第 i 个疾病 d_i 相关的第 t 个术语的语义值,$\sum_{t\in T_i}(\varphi(i,t))$ 是与疾病 d_i 相关的术语的语义值之和。$\varphi(i,t)$ 的计算公式可写为:

$$\varphi(i,t)=\begin{cases}1 & t=i\\ \max\{\Delta\varphi(i,t')\mid t'\in M_t\} & t\neq i\end{cases} \quad (2)$$

其中，Δ 是连接了 t 及其子节点 t' 的边的贡献调整因子，M_t 是 t 的所有子节点的集合。

疾病相似图是通过连接所有相似度大于 0 的疾病对来构建的。相似度是一个介于 0 和 1 之间的数，一对疾病之间的相似度就是在图中连接对应的边的权重。该图可以用相似度矩阵 $\boldsymbol{A}=[A_{ij}]\in \mathsf{R}^{N_d\times N_d}$ 来表示，其中 A_{ij} 是疾病 d_i 和 d_j 之间的相似度，N_d 表示疾病的数量。

（2）miRNA 相似图：由于具有相似功能的 miRNA 通常与相似的疾病相关，所以可以通过相关疾病集合的相似性来计算 2 个 miRNA 之间的相似度。例如，假设 miRNA m_a 与疾病 d_1 和 d_2 相关，miRNA m_b 与疾病 d_2、d_5 和 d_7 相关，则可以取 $DT_1=\{d_1,d_2\}$ 和 $DT_2=\{d_2,d_5,d_7\}$ 之间的相似度作为 m_a 和 m_b 间的相似度。本文根据相同方法计算了 miRNA 的相似度。设 miRNA m_a 关联的疾病的集合为 $M_{m_a}=\{d_i^a\mid i=1,\cdots,N_{M_a}\}$，$m_b$ 关联的疾病的集合为 $\boldsymbol{\Phi}_{m_b}=\{d_j^b\mid j=1,\cdots,N_{M_b}\}$，则 m_a 和 m_b 之间的功能相似度 S_{ij}^m 能够根据集合 M_a 和 M_b 之间的相似度来计算。由此推得的计算公式为：

$$S_{ij}^m=\frac{\sum_{q=1}^{N_{M_a}}\max(\varphi(d_i^a,d_*^b))+\sum_{w=1}^{N_{M_b}}max(\varphi(d_j^b,d_*^a))}{M_a+M_b} \quad (3)$$

其中，d_*^a 和 d_*^b 分别表示与 m_a 和 m_b 关联的所有疾病；$\max(\varphi(d_i^a,d_*^b))$ 表示与 m_a 相关的第 i 个相关疾病和 m_b 所有相关疾病间的最大的相似度；N_{M_a} 和 N_{M_b} 分别是集合 M_a 和 M_b 中元素的数量。

每两个 miRNA 节点之间的相似度就是 miRNA 相似图中两点间边的权重。miRNA 相似图可以用相似度矩阵 $\boldsymbol{B}=[B_{ij}]\in \mathsf{R}^{N_m\times N_m}$ 来表示，其中 B_{ij} 是 miRNA m_i 和 m_j 之间的相似度，N_m 表示 miRNA 的数量。

（3）miRNA-疾病关联图：当 miRNA 和疾病节点之间存在已知的关联时，通过连接 miRNA 和疾病节点来构建二分图。根据该二分图，以边来连接 B 中的 N_m 个 miRNA 节点和 $\boldsymbol{A}$ 中的 N_d 个疾病节点，边的集合表示为 $\boldsymbol{E}=[E_{ij}]\in \mathsf{R}^{N_m\times N_d}$。如果 miRNA m_i 与疾病 d_j 相关，则 E_{ij} 的值为 1；如果没有观察到这个关联，则 E_{ij} 的值为 0。

（4）miRNA 节点属性：当 miRNA m_i 和 m_j 属于相同的家族或聚簇，就更可能与相似的疾病相关。因此，miRNA 的家族和聚簇信息在预测 miRNA-疾病关联中起着重要的作用。矩阵 $\boldsymbol{C}\in \mathsf{R}^{N_m\times(N_f+N_c)}$ 用于表示 miRNA 家族和聚簇的信息，C_i 是矩阵 $\boldsymbol{C}$ 的第 i 行，表明第 i 个 miRNA 所属的家族和聚簇信息，$C_{ij}=1$ 表示 miRNA 属于某个家族（或聚簇）。

1.3　miRNA-疾病关联预测模型

建立的 miRNA-疾病异构图的邻接矩阵 $\boldsymbol{W}$，可表示为式（4）：

$$\boldsymbol{W}=\begin{bmatrix}\boldsymbol{B} & \boldsymbol{E}\\ \boldsymbol{E}^{\mathrm{T}} & \boldsymbol{A}\end{bmatrix} \quad (4)$$

其中，$\boldsymbol{B}$ 是 miRNA 相似矩阵；$\boldsymbol{A}$ 是疾病相似矩阵；$\boldsymbol{E}$ 是 miRNA-疾病关联矩阵。

本文将 $\boldsymbol{W}$ 的第 i 行，即 $\boldsymbol{W}_i$ 作为相应 miRNA 或疾病节点的拓扑嵌入向量。此前已经得到了 miRNA 节点属性矩阵 $\boldsymbol{C}$，本文将 $\boldsymbol{C}$ 的第 i 行、即 $\boldsymbol{C}_i$ 作为 miRNA m_i 的节点属性嵌入向量。

为了捕捉 miRNA 和疾病节点之间的多种连接形成的拓扑信息，并整合 miRNA 的家族和聚簇属性，本文将 miRNA m_i 的拓扑嵌入向量 $\boldsymbol{e}_{m_i}$、m_i 的节点属性嵌入向量 $\boldsymbol{c}_i$ 和疾病 d_j 的拓扑嵌入向量 $\boldsymbol{e}_{d_j}$，分别输入到 3 个全连接自动编码器中，以学习低维的拓扑表示和节点属性表示。以 $\boldsymbol{e}_{m_i}$ 为例，本文将编码器中全连接层的权重矩阵定义为 $\boldsymbol{W}_{e1}$ 和 $\boldsymbol{W}_{e2}$，偏置向量定义为 $\boldsymbol{b}_{e1}$ 和 $\boldsymbol{b}_{e2}$，可以根据公式（5）得到编码后的低维特征表示 $\boldsymbol{y}_{m_i}$：

$$\boldsymbol{y}_{m_i}=f(\boldsymbol{W}_{e2}f(\boldsymbol{W}_{e1}\boldsymbol{e}_{m_i}+\boldsymbol{b}_{e1})+\boldsymbol{b}_{e2}) \quad (5)$$

其中，f 表示激活函数 $Relu$。

通过解码器去解码还原得到与输入向量尽可能相似的输出向量，解码器的权重矩阵和偏置分别记作 $\boldsymbol{W}_{d1}$、$\boldsymbol{W}_{d2}$、$\boldsymbol{b}_{d1}$ 和 $\boldsymbol{b}_{d2}$。特别地，本文将权重 $\boldsymbol{W}_{d1}$ 和 $\boldsymbol{W}_{d2}$ 与 $\boldsymbol{W}_{e1}$ 和 $\boldsymbol{W}_{e2}$ 绑定，即 $\boldsymbol{W}_{d1}=\boldsymbol{W}_{e1}^{\mathrm{T}}$，$\boldsymbol{W}_{d2}=\boldsymbol{W}_{e2}^{\mathrm{T}}$ 以加快训练速度，避免模型过拟合。解码器的输出 $\boldsymbol{e}'_{m_i}$ 可以根据公式（6）得到：

$$\boldsymbol{e}'_{m_i}=f(\boldsymbol{W}_{d2}f(\boldsymbol{W}_{d1}\boldsymbol{y}_{m_i}+\boldsymbol{b}_{d1})+\boldsymbol{b}_{d2}) \quad (6)$$

其中，f 表示激活函数 $Relu$。

自动编码器的损失函数是均方误差（MSE），数学定义式见如下：

$$loss=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{e}_{m_i}-\boldsymbol{e}'_{m_i})^2 \quad (7)$$

其中，n 是样本数。

FMDA 模型的框架示意图如图 2 所示。得到 m_i 的低维拓扑表示 $\boldsymbol{y}_{m_i}$ 和低维节点属性表示 $\boldsymbol{y}_{c_i}$ 后，使

用另一个全连接自动编码器来得到 m_i 的融合表示 $\boldsymbol{z}_i$，并使用 *LightGBM* 来预测 m_i 与疾病 d_j 的关联得分；将 $\boldsymbol{z}_i$ 与 $\boldsymbol{y}_{d_j}$ 横向拼接以得到节点对 $m_i - d_j$ 的向量表示 $\boldsymbol{t}$，并将 $\boldsymbol{t}$ 作为 LightGBM 的输入，以得到 $m_i - d_j$ 关联的预测评分。

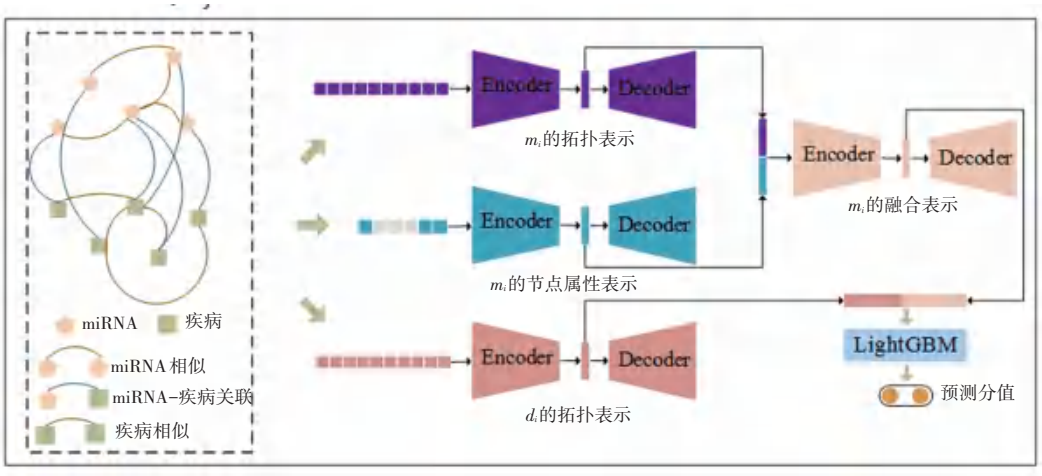


图 2 FMDA 模型的框架示意图

Fig. 2 Framework of the proposed FMDA

2 实验结果与分析

2.1 评价指标

本文进行了五倍交叉验证以充分评估 FMDA 的预测性能。所有已知的 miRNA-疾病关联都被视为正样本，并被随机分成 5 个子集，其中 4 个子集用于训练，1 个子集用于测试。所有未知的 miRNA-疾病关联均被视为负样本。随机选取与正样本数量相同的负样本，并随机分成 5 个子集，其中 4 个用于训练，1 个用于测试。

评估指标包括真阳性率（*TPR*）、假阳性率（*FPR*）、受试者工作特征曲线（*ROC*）下面积（*AUC*）、精确率（*Precision*）- 召回率（*Recall*）曲线下面积（*AUPR*）。给定阈值 θ，若样本的预测得分大于 θ，则认为该样本是正样本，否则认为是负样本。*TPR* 和 *FPR* 的计算公式，具体如下：

$$TPR = \frac{TP}{TP + FN} \tag{8}$$

$$FPR = \frac{FP}{TN + FP} \tag{9}$$

其中，*TP* 和 *TN* 分别为正确识别的正样本和负样本的数量，*FP* 和 *FN* 分别为错误识别的正样本和负样本的数量。

计算出不同 θ 值所对应的 *TPR* 和 *FPR* 后，可以画出 *ROC* 曲线，并以曲线下面积 *AUC* 作为评价性能的标准。

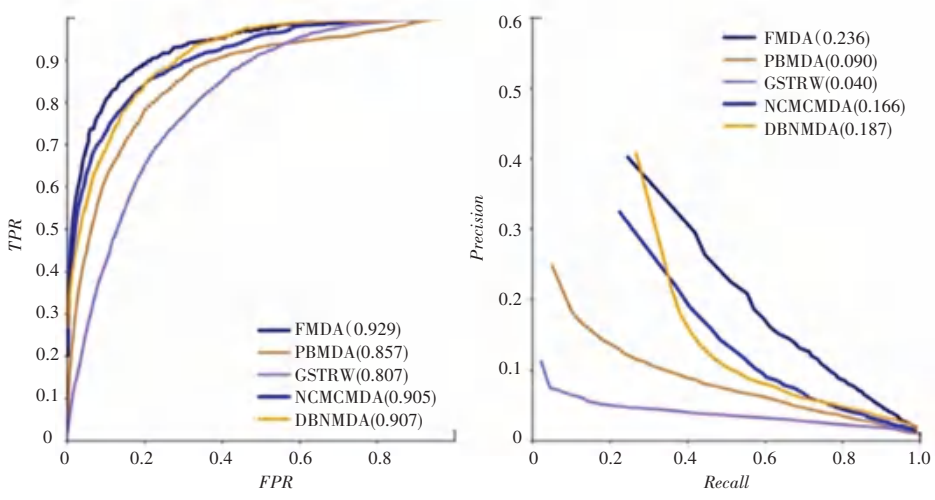
已知的 miRNA-疾病关联（正样本）和尚未被观察到的关联（负样本）的比例约为 1:33，正样本和负样本之间存在着严重的不平衡。在不同类别不平衡的情况下，*PR* 曲线比 *ROC* 曲线更具有参考价值。精确率衡量的是被判定为正样本的样本中真正的正样本的占比，召回率是被正确识别的正样本占所有正样本总数的比例。因此，本文也用 *PR* 曲线和对应面积来评估模型的预测性能，数学定义公式具体如下：

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

2.2 与其他方法的比较

为了更好地评估 FMDA 的预测性能，本文将该模型与 GSTRW[10]、NCMCMDA[11]、PBMDA[12] 和 DBNMDA[13] 进行对比。通过五倍交叉验证，得到了这 5 个模型的 *ROC* 曲线和 *PR* 曲线，如图 3 所示。在所有 341 个被测试的疾病中，FMDA 取得了最高的平均 *AUC* 值（*AUC* = 0.929），比 NCMCMDA 高了 2.4%，比 PBMDA 高 7.2%，比 GSTRW 高 12.2%，比 DBNMDA 高 2.2%。在所有被测试的疾病中，FMDA 的 *PR* 曲线下面积高于其他方法（*AUPR* = 0.236），比 NCMCMDA、PBMDA、GSTRW 和 DBNMDA 分别高 7.0%、14.6%、18.6%和 4.9%。



(a) *ROC*　　(b) *PR*

图 3　不同预测方法的 *ROC* 曲线与 *PR* 曲线

Fig. 3　*ROC* curves and *PR* curves of different methods for prediction

此外，对于 16 个常见疾病，本文列出了这 5 个模型的预测性能，见表 1 和表 2。在这 16 个疾病中，FMDA 取得了 12 个疾病的最高 *AUC*；在 16 个常见疾病中，FMDA 取得了 10 个疾病的最高 *AUPR*。

表 1　所有疾病的平均 *AUC* 与 16 个常见疾病的 *AUC*

Tab. 1　Average *AUC* over all the diseases and *AUC*s of 16 common diseases

疾病名称	*AUC*				
	FMDA	PBMDA	GSTRW	NCMCMDA	DBNMDA
Average AUC on 341 diseases	**0.929**	0.857	0.807	0.905	0.907
Breast Neoplasms	0.965	0.906	0.837	**0.983**	0.982
Hepatocellular Carcinoma	0.957	0.910	0.791	0.967	**0.974**
Glioma	**0.958**	0.882	0.786	0.928	0.940
Acute Myeloid Leukmia	**0.969**	0.885	0.796	0.937	0.968
Lung Neoplasma	**0.973**	0.862	0.813	0.947	0.955
Melanoma	**0.979**	0.849	0.758	0.954	0.962
Osteosarcoma	**0.972**	0.860	0.771	0.968	0.961
Ovarian Neoplasms	**0.980**	0.888	0.844	0.955	0.968
Pancreatic Neoplasms	**0.965**	0.879	0.833	0.904	0.898
Alzheimer Disease	**0.928**	0.833	0.816	0.897	0.901
Carcinoma, Renal Cell	**0.945**	0.856	0.786	0.935	0.799
Diabetes Mellitus, Type 2	**0.964**	0.870	0.870	0.898	0.951
Glioblastoma	**0.950**	0.849	0.759	0.912	0.930
Heart Failure	**0.946**	0.884	0.814	0.899	0.943
Atherosclerosis	0.932	0.891	0.824	**0.961**	0.959

表 2 所有疾病的平均 *AUPR* 与 16 个常见疾病的 *AUPR*

Tab. 2 Average *AUPR* over all the diseases and *AUPR*s of 16 common diseases

疾病名称	*AUPR*				
	FMDA	PBMDA	GSTRW	NCMCMDA	DBNMDA
Average AUPR on 341 diseases	**0.236**	0.090	0.040	0.166	0.187
Breast Neoplasms	0.766	0.718	0.389	0.812	**0.821**
Hepatocellular Carcinoma	0.812	0.767	0.482	0.831	**0.845**
Glioma	**0.443**	0.390	0.225	0.312	0.210
Acute Myeloid Leukmia	**0.396**	0.385	0.123	0.358	0.369
Lung Neoplasma	**0.771**	0.562	0.370	0.685	0.741
Melanoma	**0.591**	0.483	0.205	0.493	0.512
Osteosarcoma	**0.624**	0.357	0.180	0.486	0.603
Ovarian Neoplasms	**0.640**	0.528	0.395	0.480	0.486
Pancreatic Neoplasms	0.581	0.458	0.333	**0.824**	0.531
Alzheimer Disease	0.275	0.136	0.086	0.218	**0.359**
Carcinoma, Renal Cell	**0.336**	0.314	0.136	0.254	0.293
Diabetes Mellitus, Type 2	**0.412**	0.259	0.132	0.399	0.401
Glioblastoma	**0.373**	0.346	0.162	0.293	0.318
Heart Failure	**0.366**	0.301	0.135	0.262	0.289
Atherosclerosis	0.272	0.306	0.084	0.289	**0.310**

3 结束语

本文提出了整合多源数据的相似性、关联以及 miRNA 的家族和聚簇属性的 miRNA-疾病关联预测模型(给出模型的名称),构建了一个异构图以形成拓扑嵌入和节点属性嵌入,并建立了基于全连接自动编码器的框架来编码拓扑表示和 miRNA 节点属性表示。与其他 4 个预测模型比较表明本文的模型在 *AUC* 和 *AUPR* 方面均取得了更好的预测性能。

参考文献

[1] GEBERT L F R, MACRAE I J. Regulation of microRNA function in animals[J]. Nature Reviews Molecular Cell Biology, 2019,20: 21-37.

[2] 程爽,郭茂祖,武雪剑. microRNA 靶基因预测算法的研究与发展[J]. 智能计算机与应用, 2018, 8(01): 1-5,13.

[3] CHEN Xing, XIE Di, ZHAO Qi, et al. MicroRNAs and complex diseases: from experimental results to computational models[J]. Briefings in Bioinformatics, 2019, 20(2): 515-539.

[4] ZHAO Yan, CHEN Xing, YIN Jun. Adaptive boosting-based computational model for predicting potential miRNA-disease associations[J]. Bioinformatics, 2019, 35(22): 4730-4738.

[5] WANG Dong, WANG Juan, LU Ming, et al. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases[J]. Bioinformatics, 2010, 26(13): 1644-1650.

[6] XUAN Ping, HAN Ke, GUO Maozu, et al. Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors[J]. PLoS One, 2013, 8(8): e70204.

[7] CHEN Xing, WU Qiaofeng, YAN Guiying. RKNNMDA: ranking-based KNN for MiRNA-disease association prediction[J]. J RNA Biol, 2017, 14(7): 952-962.

[8] LIM D, LEE S, CHOI M, et al. The conserved microRNA miR-8-3p coordinates the expression of V-ATPase subunits to regulate ecdysone biosynthesis for Drosophila metamorphosis[J]. Faseb Journal, 2020, 34(5): 6449-6465.

[9] GU Changlong, LIAO Bo, LI Xiaoying, et al. Network consistency projection for human miRNA-disease associations inference[J]. Scientific Reports, 2016, 6(1): 36054.

[10] CHEN Min, LIAO Bo, LI Zejun. Global similarity method based on a two-tier random walk for the prediction of microRNA-disease association[J]. Scientific Reports, 2018, 8(1): 1-16.

[11] CHEN Xing, SUN Liangang, ZHAO Yan. NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion[J]. Briefings in Bioinformatics, 2021, 22(1): 485-496.

[12] YOU Zhuhong, HUANG Zhi'an, ZHU Zexuan, et al. PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction[J]. PLoS Computational Biology, 2017, 13(3):e1005455.

[13] CHEN Xing, LI Tianhao, ZHAO Yan, et al. Deep-belief network for predicting potential miRNA-disease associations[J]. Briefings in Bioinformatics, 2021, 22(3): 1-10.