frontiers in Microbiology

ORIGINAL RESEARCH
published: 27 October 2020
doi: 10.3389/fmicb.2020.592430

RNMFMDA: A Microbe-Disease Association Identification Method Based on Reliable Negative Sample Selection and Logistic Matrix Factorization With Neighborhood Regularization

*Lihong Peng, Ling Shen, Longjie Liao, Guangyi Liu and Liqian Zhou**

School of Computer Science, Hunan University of Technology, Zhuzhou, China

OPEN ACCESS

Edited by:
Qi Zhao,
University of Science and Technology Liaoning, China

Reviewed by:
Wen Zhang,
Huazhong Agricultural University, China
Jiyuan An,
Queensland University of Technology, Australia

***Correspondence:**
Liqian Zhou
zhoulq11@163.com

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *07 August 2020*
Accepted: *17 September 2020*
Published: *27 October 2020*

Citation:
Peng L, Shen L, Liao L, Liu G and Zhou L (2020) RNMFMDA: A Microbe-Disease Association Identification Method Based on Reliable Negative Sample Selection and Logistic Matrix Factorization With Neighborhood Regularization. Front. Microbiol. 11:592430. doi: 10.3389/fmicb.2020.592430

Microbes with abnormal levels have important impacts on the formation and development of various complex diseases. Identifying possible Microbe-Disease Associations (MDAs) helps to understand the mechanisms of complex diseases. However, experimental methods for MDA identification are costly and time-consuming. In this study, a new computational model, RNMFMDA, was developed to find possible MDAs. RNMFMDA contains two main processes. First, Reliable Negative MDA samples were selected based on Positive-Unlabeled (PU) learning and random walk with restart on the heterogeneous microbe-disease network. Second, Logistic Matrix Factorization with Neighborhood Regularization (LMFNR) was developed to compute the association probabilities for all microbe-disease pairs. To evaluate the performance of the proposed RNMFMDA method, we compared RNMFMDA with five state-of-the-art MDA prediction methods based on five-fold cross-validations on microbes, diseases, and MDAs. As a result, RNMFMDA obtained the best AUCs of 0.6332, 0.8669, and 0.9081, respectively for the three five-fold cross validations, significantly outperforming other models. The promising prediction performance may be attributed to the following three features: highly quality negative MDA sample selection, LMFNR-based MDA prediction model, and various biological information integration. In addition, a few predicted microbe-disease pairs with high association scores are worthy of further experimental validation.

Keywords: microbe-disease associations, reliable negative samples, positive-unlabeled learning, random walk with restart, logistic matrix factorization with neighborhood regularization

1. INTRODUCTION

Microbes are the most abundant microscopic organisms on Earth and control many major biological and chemical processes (Ley et al., 2006; Qu J. et al., 2019; Sachdeva et al., 2019). Normal microbial flora are beneficial for the host health (McFarland, 2000; Langella and Martín, 2019; Qu J. et al., 2019). Beneficial microbes including biotherapeutic agent, probiotics and synbiotics have been reported as effective therapeutic clues when normal microflora are disrupted (McFarland, 2000; Langella and Martín, 2019).

More importantly, microorganisms have an important affect on infectious diseases and non-infectious diseases (Findley et al., 2013; Ding and Schloss, 2014; Abu-Ali et al., 2018; Byrd et al., 2018; Liu et al., 2019). The human body is possible to get sick when foreign microorganisms invade or a microbial community is imbalanced (Zhu et al., 2018; Qu K. et al., 2019). For example, there are more abundant *Fusobacterium* in asthmatic patients than healthy people (Davis-Richardson et al., 2014). Lecithinase-negative *Clostridium* and *Lactobacillus* are much more in colorectal carcinoma patients (Heavey and Rowland, 2004). Increased *Lactobacillus* can result in tertiary lymphoid (Sze et al., 2012). All the above reports suggested that there are close associations between microbes and human diseases. Therefore, finding new Microbe-Disease Associations (MDAs) helps to provide diagnostic and therapeutic clues for clinical researches Chen et al. (2017).

Experimental methods to predict possible MDAs are costly and time-consuming. Computational methods are thus gradually developed to find potential MDAs. Ma et al. (2016) collected published MDA data from literatures and constructed Human Microbe-Disease Association Database (HMDAD). Various computational models are then exploited based on known MDA data, Gaussian Interaction Profile Kernel (GIP) similarity for diseases and microbes. Chen et al. (2017) assumed that functionally similar microbes are likely to associate with similar non-infectious diseases and presented the first tool (KATZHMDA) to predict potential MDAs based on the KATZ measure. Huang et al. (2017) proposed a neighbor and graph-based recommendation model (NGRHMDA). Bao et al. (2017) designed a Network Consistency Projection-based MDA prediction method (NCPHMDA). Luo and Long (2018) constructed a heterogeneous network and presented a Network Topological Similarity-based human MDA prediction model (NTSHMDA). Wang et al. (2017) developed a semi-supervised learning framework (LRLSHMDA) to prioritize microbe candidates for all interested diseases based on Laplacian Regularized Least Squares. Peng et al. (2018b) exploited a adaptive boosting-based method to compute association scores for human microbe-disease pairs based on a strong classification model. Zhang et al. (2018) proposed a bi-direction similarity integration label propagation method (BDSILP) for identifying MDAs. Shi et al. (2018) assumed that observed incomplete microbe-non-infectious disease association matrix is composed of a parameterized matrix and a noise matrix, and then developed a Binary Matrix Completion-based model (BMCMDA) to infer possible microbe-non-infectious disease associations. Qu J. et al. (2019) presented a human MDA model (MDLPHMDA) based on matrix decomposition and label propagation.

The above methods were effectively applied to MDA identification and captured a few MDAs, however, the prediction performance remains to be improved. More importantly, in MDA identification problem, negative training examples are missing. Therefore, most of models randomly extracted negative MDAs from unknown microbe-disease pairs, which may contain positive MDAs, thereby severely affecting the prediction accuracy. Learning from Positive and Unlabeled examples (PU learning) (Li et al., 2014) is one type of methods used to learn the models from numerous positive and unlabeled examples. PU learning has been widely applied to text mining and obtained better performance.

In this study, we developed a computational model, RNMFMDA, to predict human MDA candidates. RNMFMDA integrated Reliable Negative MDA selection based on PU learning and random walk with restart, Logistic Matrix Factorization with Neighborhood Regularization (LMFNR), and multiple heterogeneous data. RNMFMDA first computed disease similarity and microbe similarity. Credible negative MDAs were then selected based on PU learning and random walk with restart. LMFNR was finally developed to identify MDA candidates. RNMFMDA was compared to five state-of-the-art MDA prediction methods, MDLPHMDA (Qu J. et al., 2019), NGRHMDA (Huang et al., 2017), NTSHMDA (Luo and Long, 2018), LRLSHMDA (Wang et al., 2017), and KATZHMDA (Chen et al., 2017). To evaluate our proposed RNMFMDA, we conducted five-fold Cross Validations (CVs) on microbes, diseases, and MDAs. The results showed that RNMFMDA obtained the best AUCs under the above three CVs. In addition, we further performed the experiments to find possible microbes/diseases associate with a known disease/microbe. The experimental result analysis suggested that RNMFMDA is a powerful MDA identification method.

2. MATERIALS AND EQUIPMENT

Assume that the ith microbe is represented as $m_i(i = 1, 2, \ldots, n)$, and the jth disease is denoted as $d_j(j = 1, 2, \ldots, m)$. The associations between n microbes and m diseases are denoted as a binary matrix $\boldsymbol{Y}_{(n\times m)}$ where

$$y_{ij} = \begin{cases} 1 & if \quad m_i \text{ } associates \text{ } with \text{ } d_j \\ 0 & \qquad\qquad\qquad otherwise \end{cases} \tag{1}$$

The non-zero elements in $\boldsymbol{Y}$ are called "MDA pairs" and considered as positive observations. The zero elements in $\boldsymbol{Y}$ are called "unknown microbe-disease pairs" and considered as unlabeled observations. The microbe similarity matrix and the disease similarity matrix are represented as $\boldsymbol{S}_M \in \Re^{n\times n}$ and $\boldsymbol{S}_D \in \Re^{m\times m}$, respectively.

Our objective is to select reliable negative MDAs based on PU learning and random walk with restart on the heterogeneous network, and then compute the association probability score for each microbe-disease pair by LMFNR, finally rank candidate microbe-disease pairs according to the scores in descending order, so that the top microbe-disease pairs are the most likely to be MDAs.

We collected confirmed MDAs from HMDAD (Ma et al., 2016) (http://www.cuilab.cn/hmdad). The database provides 483 MDAs between 292 microbes and 39 diseases from 61 previous works. We deleted the same MDAs based on different evidences and finally obtained 450 MDAs from these microbes and diseases.

3. METHODS

3.1. Microbe GAP Similarity

Motivated by the similarity computation method provided by van Laarhoven et al. (2011), we computed microbe Gaussian Association Profile (GAP) similarity based on known MDA matrix. Given a microbe $m(i)$, its GAP $\boldsymbol{AP}(m(i))$ can be represented as the ith row of $\boldsymbol{Y}$. The GAP similarity between two microbes $m(i)$ and $m(j)$ can be computed by Equation (2):

$$\boldsymbol{S}_M(m(i), m(j)) = \exp(-\gamma_m ||\boldsymbol{AP}(m(i)) - \boldsymbol{AP}(m(j))||^2) \qquad (2)$$

where $\gamma_m = \gamma_m{}'/(\frac{1}{n}\sum_{k=1}^{n} ||\boldsymbol{AP}(m(k))||^2)$ denotes the normalized kernel bandwidth with bandwidth parameter $\gamma_m{}'$. The microbe similarity matrix $\boldsymbol{S}_{M(n\times n)}$ can be obtained based on the GAP similarity.

3.2. Disease Similarity

3.2.1. Disease GAP Similarity

For a disease $d(i)$, its GAP $\boldsymbol{AP}(d(i))$ can be represented as the ith column of $\boldsymbol{Y}$. The GAP similarity between two diseases $d(i)$ and $d(j)$ can be calculated by Equation (3):

$$\boldsymbol{S}_G(d(i), d(j)) = \exp(-\gamma_d ||\boldsymbol{AP}(d(i)) - \boldsymbol{AP}(d(j))||^2) \qquad (3)$$

where $\gamma_d = \gamma_d{}'/(\frac{1}{m}\sum_{k=1}^{m} ||\boldsymbol{AP}(d(k))||^2)$ denotes the normalized kernel bandwidth with bandwidth parameter $\gamma_d{}'$.

3.2.2. Disease Symptom Similarity

Inspired by the similarity measure method provided by Zhou et al. (2014), we computed disease symptom similarity matrix $\boldsymbol{S}_S$.

Finally, the disease similarity matrix $\boldsymbol{S}_{D(m\times m)}$ can be computed by Equation (4):

$$\boldsymbol{S}_D(d(i), d(j)) = \boldsymbol{S}_G(d(i), d(j)) + \gamma \boldsymbol{S}_S(d(i), d(j)) \qquad (4)$$

where γ is a parameter used to weigh the importance between the GAP similarity and the symptom similarity.

3.3. Reliable Negative MDA Selection

There exists a few known MDAs and numerous unobserved microbe-disease pairs in the HMDAD database (Ma et al., 2016). There are no negative MDA samples because of the limitations of experimental methods. High-quality negative MDAs can boost the performance of MDA prediction models. Therefore, most of machine learning-based methods have to randomly select negative examples from unknown microbe-disease pairs. However, this part of randomly selected negative examples probably contains positive MDAs, thereby severely affecting the performance of MDA identification algorithms. Therefore, we developed a negative sample selection method to extract reliable negative MDA data based on PU learning and random walk with restart. The pipeline mainly contains two basic processes: computing the association probability for each microbe-disease pair based on random walk with restart and extracting high-quality negative MDA samples based on PU learning and the computed association scores.

3.3.1. Random Walk With Restart on the Heterogeneous Microbe-Disease Network

Inspired by the method proposed by Chen et al. (2012), we consider microbe similarity network, disease similarity network, and MDA network to construct a heterogeneous microbe-disease network. We used microbe similarity matrix $\boldsymbol{S}_{M(n\times n)}$, disease similarity matrix $\boldsymbol{S}_{D(m\times m)}$, and MDA matrix $\boldsymbol{Y}_{(n\times m)}$ as the adjacency matrices of the above three networks, respectively. And the adjacency matrix on the heterogeneous network can be denoted as:

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{S}_M & \boldsymbol{Y} \\ \boldsymbol{Y}^T & \boldsymbol{S}_D \end{bmatrix} \qquad (5)$$

where $\boldsymbol{Y}^T$ denotes the transpose of $\boldsymbol{Y}$.

We then calculate different transition probabilities of random walk with restart on the heterogeneous graph. Assume that $\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_{MM} & \boldsymbol{H}_{MD} \\ \boldsymbol{H}_{DM} & \boldsymbol{H}_{DD} \end{bmatrix}$ represent the transition probability matrix, where $\boldsymbol{H}_{MM}$ and $\boldsymbol{H}_{DD}$ represent the walks within microbe-microbe similarity network and disease-disease similarity network, respectively, $\boldsymbol{H}_{MD}$ and $\boldsymbol{H}_{DM}$ represent the skips between networks. Given a microbe/disease, if there exist a bipartite association between the microbe/disease and diseases/microbes, the particle will either skip between the four networks or stay in the current network with a transition probability $\lambda \in [0, 1]$.

We predict MDA candidates from a perspective of microbes. Assume that a particle be situated on the i-th microbe node $m_i \in M$, it will walk to a microbe node $m_j \in M$ with the transition probability $\boldsymbol{H}_{MM}(i, j)$:

$$\boldsymbol{H}_{MM}(i,j) = \begin{cases} \boldsymbol{S}_M(i,j)/\sum_{k=1}^{n} \boldsymbol{S}_M(i,k), & \text{if} \sum_{k=1}^{m} \boldsymbol{Y}(i,k) = 0 \\ (1-\lambda)\boldsymbol{S}_M(i,j)/\sum_{k=1}^{n} \boldsymbol{S}_M(i,k), & otherwise \end{cases} \qquad (6)$$

or skip to a disease $d_j \in D$ based on a bipartite association with d_j with the transition probability $\boldsymbol{H}_{MD}(i, j)$:

$$\boldsymbol{H}_{MD}(i,j) = \begin{cases} \lambda \boldsymbol{Y}(i,j)/\sum_{k=1}^{m} \boldsymbol{Y}(i,k), & \text{if} \sum_{k=1}^{m} \boldsymbol{Y}(i,k) \neq 0 \\ 0, & otherwise \end{cases} \qquad (7)$$

Similarly, we can find possible MDAs from a perspective of diseases. Assume that a particle be situated on the ith disease node $d_i \in D$. It will walk to a disease node $d_j \in D$ with the transition probability $\boldsymbol{H}_{DD}(i, j)$:

$$\boldsymbol{H}_{DD}(i,j) = \begin{cases} \boldsymbol{S}_D(i,j)/\sum_{k=1}^{m} \boldsymbol{S}_D(i,k), & \text{if} \sum_{k=1}^{n} \boldsymbol{Y}(k,i) = 0 \\ (1-\lambda)\boldsymbol{S}_D(i,j)/\sum_{k=1}^{m} \boldsymbol{S}_D(i,k), & otherwise \end{cases} \qquad (8)$$

or skip to a microbe $m_j \in M$ based on a bipartite association with m_j with a transition probability $\boldsymbol{H}_{DM}(i,j)$:

$$\boldsymbol{H}_{DM}(i,j) = \begin{cases} \lambda \boldsymbol{Y}(j,i) / \sum\limits_{k=1}^{n} \boldsymbol{Y}(k,i), & \text{if } \sum\limits_{k=1}^{n} \boldsymbol{Y}(k,i) \neq 0 \\ 0, & otherwise \end{cases} \tag{9}$$

Therefore, we describe random walk with restart on the heterogeneous network as:

$$\boldsymbol{P}(t+1) = (1-\theta)\boldsymbol{H}^T * \boldsymbol{P}(t) + \theta \boldsymbol{P}(0) \tag{10}$$

where $\boldsymbol{P}(t)$ denotes a probability matrix used to represent the association scores of all unobserved microbe-disease pairs at the t-th step random walk, $\boldsymbol{H}^T$ denotes the transpose of $\boldsymbol{H}$, and θ represents the restarting probability. The particle will return to either a seed microbe or a seed disease. More importantly, it is possible to differentiate the relative important of each network based on the initial probability $p_i(0) = \begin{bmatrix} (1-\eta)v_i \\ \eta s_i \end{bmatrix}$, where v_i and s_i denote the initial probability distributions on disease-disease similarity network and microbe-microbe similarity network starting from their seed nodes, respectively. The parameter $\eta \in [0,1]$ is used to control the restarting probability in these two similarity networks. If $\eta < 0.5$, the particle will more tend to restart from one of the seed microbes than from one of the seed diseases.

3.3.2. Reliable Negative MDA Extraction

We took known MDAs as initial positive sample set P, observed microbe-disease pairs as initial unlabeled sample set U and developed a reliable negative MDA selection based on PU learning. The method contains the following five steps:

```
Step 1. Randomly selecting positive sample subset $S$ from $P$ and adding $S$ into $U$;
Step 2. Taking $P - S$ as positive samples, $U + S$ as negative samples;
Step 3. Computing the association score matrix $\boldsymbol{AM}$ based on random walk with restart on the heterogeneous microbe-disease network;
Step 4. Ranking microbe-disease pairs in $S$ based on $\boldsymbol{AM}$ and finding the minimum score $\boldsymbol{AM}_{min}$ in $S$;
Step 5. For every sample $x$ in $U$:
        if $\boldsymbol{AM}_x$ satisfying $\boldsymbol{AM}_x < \boldsymbol{AM}_{min}$
        then $RN = RN \cup x$
```

We can obtain reliable negative MDA example set RN with the above negative selection method.

3.4. MDA Prediction Based on LMFNR

The logistic matrix factorization method has widely applied to the area of various association prediction and obtained better performance (Liu et al., 2016, 2020). Inspired by the logistic matrix factorization method provided by Liu et al. (2016) and Liu et al. (2020), we developed an MDA prediction method (RNMFMDA) by integrating the Reliable Negative MDA sample selection method and the LMFNR method.

Suppose that both microbes and diseases are mapped into r-dimensional shared latent spaces where $r \ll n, m$. The properties of a microbe m_i / disease d_j is represented by a latent vector $\boldsymbol{a}_i \in \Re^{1\times t}$ / $\boldsymbol{b}_i \in \Re^{1\times t}$. Then, the association probability p_{ij} between m_i and d_j can be computed by Equation (11):

$$p_{ij} = \frac{\exp(\boldsymbol{a}_i \boldsymbol{b}_j^T)}{1+\exp(\boldsymbol{a}_i \boldsymbol{b}_j^T)} \tag{11}$$

The latent vectors of all microbes / diseases can be denoted as $\boldsymbol{A} \in \Re^{n\times r}$ / $\boldsymbol{B} \in \Re^{m\times r}$, where $\boldsymbol{a}_i$ / $\boldsymbol{b}_j$ is the ith / jth row in $\boldsymbol{A}$/$\boldsymbol{B}$.

In MDA identification tasks, the observed MDAs have been experimentally validated and are more reliable than unknown microbe-disease pairs. To more accurately find MDA candidates, we assigned higher confidence scores to known MDAs than unknown pairs. Particularly, each MDA is considered as $c(c \geq 1)$ positive training samples, and each reliable negative MDA is considered as a single negative training sample. c is a constant to measure the importance of observations. The importance weighting technique has been effectively applied to the area of informatics. And we built the following MDA prediction model:

$$\begin{aligned} p(\boldsymbol{Y}|\boldsymbol{A},\boldsymbol{B}) = &\Big(\prod_{1\leq i\leq n, 1\leq j\leq m, y_{ij}=1} [p_{ij}^{y_{ij}}(1-p_{ij})^{(1-y_{ij})}]^c\Big) \\ &\times\Big(\prod_{1\leq i\leq n, 1\leq j\leq m, y_{ij}=0} [p_{ij}^{y_{ij}}(1-p_{ij})^{(1-y_{ij})}]\Big) \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m} p_{ij}^{cy_{ij}}(1-p_{ij})^{(1-y_{ij})} \end{aligned} \tag{12}$$

The above model can represented as the following optimization function considering the probability distribution based on a Bayesian inference:

$$\begin{aligned} \min_{\boldsymbol{A},\boldsymbol{B}} \sum_{i=1}^{m}\sum_{j=1}^{n} &(1+cy_{ij}-y_{ij})\log[1+\exp(\boldsymbol{a}_i\boldsymbol{b}_j^T)] \\ &-cy_{ij}\boldsymbol{a}_i\boldsymbol{b}_j^T + \tfrac{\lambda_m}{2}||\boldsymbol{A}||_F^2 + \tfrac{\lambda_d}{2}||\boldsymbol{B}||_F^2 \end{aligned} \tag{13}$$

where λ_m and λ_d are parameters, $||\boldsymbol{A}||_F$ and $||\boldsymbol{B}||_F$ denote the Frobenius norm of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively.

The nearest neighborhood information of biological entities in the association network can improve the prediction performance (Zhang et al., 2019a,b,c). For example, Zhang et al. (2019a), Zhang et al. (2019b), and Zhang et al. (2019c) used neighborhood information and effectively found microRNA-disease associations, drug-drug interactions and long non-coding RNA-miRNA interactions. Therefore, we integrated neighborhood information to the above optimization model and built the final LMFNR model by Equation (14):

$$\begin{aligned} \min_{A,B} \sum_{i=1}^{m}\sum_{j=1}^{n} &(1+cy_{ij}-y_{ij})\ln[1+\exp(\boldsymbol{a}_i\boldsymbol{b}_j^T)] - cy_{ij}\boldsymbol{a}_i\boldsymbol{b}_j^T \\ &+\tfrac{1}{2}tr[\boldsymbol{A}^T(\lambda_m\boldsymbol{I}+\alpha\boldsymbol{L}_m)\boldsymbol{A} + \tfrac{1}{2}tr[\boldsymbol{B}^T(\lambda_d\boldsymbol{I}+\alpha\boldsymbol{L}_d)\boldsymbol{B} \end{aligned} \tag{14}$$

where $tr(\cdot)$ denotes the trace of a matrix, $\boldsymbol{L}_m$ and $\boldsymbol{L}_d$ were defined as the same to Liu et al. (2016).

We can obtain $\boldsymbol{A}$ and $\boldsymbol{B}$ by solving with the optimization problem by Equation (14) with an alternating gradient ascent procedure.

Finally, the association probability matrix $\boldsymbol{Y}_p$ for all unknown microbe-disease pairs can be represented as:

$$\boldsymbol{Y}_p = \boldsymbol{A}\boldsymbol{B}^T \tag{15}$$

4. RESULTS

4.1. Experimental Settings and Evaluation

The experiment was performed under 100 trials of five-fold Cross Validation. An average performance was finally computed to reduce the prediction bias. For an MDA matrix $Y_{n\times m}$, CVs were conducted under three different experimental settings as follows.

- Five-fold Cross Validation 1 (CV1): CV on microbes, that is, random rows in Y (i.e., microbes) were masked for testing.
- Five-fold Cross Validation 2 (CV2): CV on diseases, that is, random columns in Y (i.e., diseases) were masked for testing.
- Five-fold Cross Validation 3 (CV3): CV on microbe-disease pairs, that is, random entries in Y (i.e., microbe-disease pairs) were masked for testing.

Under CV1, in each round, 80% of rows in Y was used as training set and the remaining was used as test set. Under CV2, in each round, 80% of columns in Y was used as training set and the remaining was used as test set. Under CV3, in each round, 80% of entries in Y was used as training set and the remaining was used as test set. These three CVs refer to MDA prediction for (1) new (unknown) microbes, (2) new diseases, and new microbe-disease pairs, respectively.

Sensitivity, specificity, accuracy, and AUC were used to evaluate the performances. AUC is the average area under the receiver operating characteristics (ROC) curve. The curve can be plotted by the ratio of True Positive Rate (TPR) to False Positive Rate (FPR) according to different thresholds. TPR and FPR can be computed by Equations (16, 17). High AUC value represents better performance. In our experiments, AUC was computed in each round of CV and final AUC was averaged over the five rounds for 100 times.

TABLE 1 | Confusion matrix of a binary classifier.

	True class = 1	True class = −1
Predicted class = 1	True positive (TP)	False positive (FP)
Predicted class = −1	False negative (FN)	True negative (TN)

$$TPR = \frac{TP}{TP + FN} = \frac{TP}{T} \tag{16}$$

$$FPR = \frac{FP}{FP + TN} = \frac{FP}{F} \tag{17}$$

where the definitions of TP, FP and FN are as shown in **Table 1**.

λ is used to determine the probability of jumping between nodes. θ is the restart rate. η denotes the restarting probability in microbe similarity network and disease similarity network. c is the importance level of positive samples to negative samples. K denotes the number of neighborhood. For the parameters λ, θ, η, c, and K, we conducted grid search to find the optimal values. RNMFMDA obtained the best performance when these five parameters are set as $\lambda = 0.9$, $\theta = 0.5$, $\eta = 0.9$, $c = 8$, and $K = 5$. So we set the above five parameters as the corresponding values. Parameters $\gamma_m{}'$, $\gamma_d{}'$, and γ are set the same values in previous works, that is, $\gamma_m{}' = 1$, $\gamma_d{}' = 1$, and $\gamma = 0.9$. For other parameters, we set the corresponding values according to the method provided by Liu et al. (2016). When $||P(t+1) - P(t)||_F \leq 10e - 12$, the iteration for random walk will stop. The ratio of extracted negative MDAs to positive MDAs is set as 1:1, this is to say, the number of negative MDAs is 450. The parameters in other five methods were set as the same values provided by the corresponding papers.

4.2. Performance Comparison of RNMFMDA With Other Five Methods

In this section, we compared our proposed RNMFMDA method with five state-of-the-art MDA prediction models, MDLPHMDA (Qu J. et al., 2019), NGRHMDA (Huang et al., 2017), NTSHMDA (Luo and Long, 2018), LRLSHMDA (Wang et al., 2017), and KATZHMDA (Chen et al., 2017). **Tables 2–4** showed the performance of RNMFMDA with other five methods. The best performance is described in boldface in **Tables 2–4**.

As shown in **Tables 2–4**, RNMFMDA performed more efficiently than other five methods. Compared with MDLPHMDA, NGRHMDA, and NTSHMDA, RNMFMDA obtained a more remarkable improvement over four evaluation metrics under three CVs. KATZHMDA and LRLSHMDA are

TABLE 2 | Performance comparison of RNMFMDA with other five methods under CV1.

Method	Sensitivity	Specificity	Accuracy	AUC
KATZHMDA	0.2772	0.6690	0.6653	0.3646
LRLSHMDA	0.3286	**0.7538**	**0.7496**	0.4364
NGRHMDA	0.0777	0.3423	0.4817	0.4156
MDLPHMDA	0.3273	0.6890	0.6855	0.4022
NTSHMDA	0.1899	0.6177	0.6138	0.3042
RNMFMDA	**0.4938**	0.6278	0.6274	**0.6332**

These bolded values represent the best values for the different methods under the same evaluation.

TABLE 3 | Performance comparison of RNMFMDA with other five methods under CV2.

Method	Sensitivity	Specificity	Accuracy	AUC
KATZHMDA	**0.8317**	0.6487	0.6501	0.8662
LRLSHMDA	0.6944	0.7333	0.7330	0.8086
NGRHMDA	0.3800	0.3285	0.7403	0.8224
MDLPHMDA	0.7318	0.6653	0.6658	0.8178
NTSHMDA	0.7913	0.5905	0.5921	0.8292
RNMFMDA	0.5850	**0.8304**	**0.8283**	**0.8669**

These bolded values represent the best values for the different methods under the same evaluation.

TABLE 4 | Performance comparison of RNMFMDA with other five methods under CV3.

Method	Sensitivity	Specificity	Accuracy	AUC
KATZHMDA	**0.8262**	0.6503	0.6518	0.8571
LRLSHMDA	0.7971	0.7412	0.7416	0.8794
NGRHMDA	0.4207	0.3308	0.7796	0.9025
MDLPHMDA	0.8268	0.6729	0.6741	0.8938
NTSHMDA	0.8545	0.5904	0.5926	0.8896
RNMFMDA	0.5810	**0.8818**	**0.8793**	**0.9081**

These bolded values represent the best values for the different methods under the same evaluation.

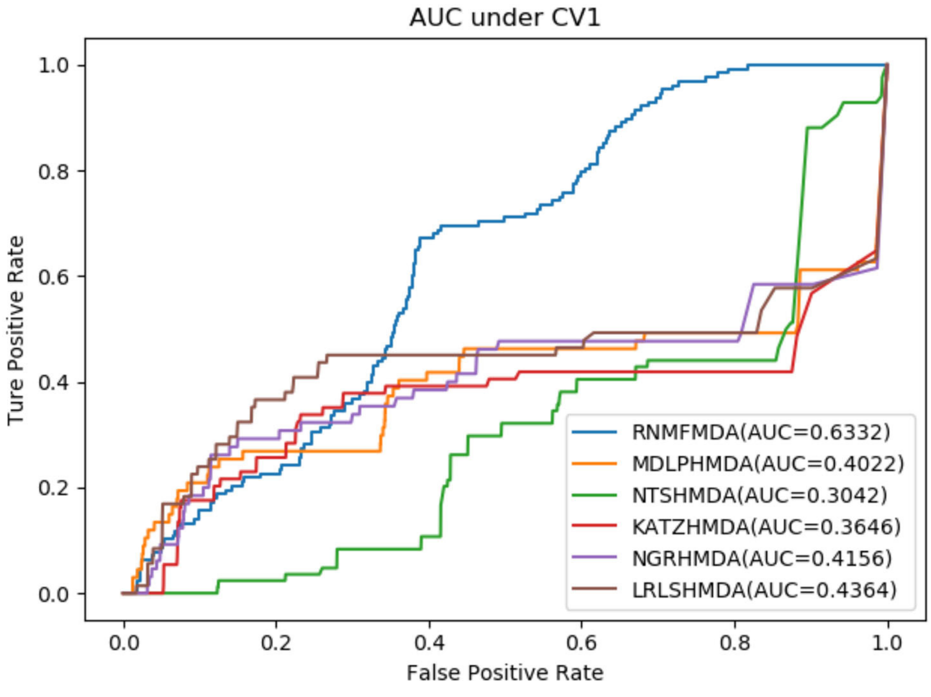

FIGURE 1 | Performance comparison of RNMFMDA with other five methods under CV1.

two classic MDA prediction methods. Under CV1, KATZHMDA and LRLSHMDA computed better specificity and accuracy than RNMFMDA. Under CV2 and CV3, these two methods obtained better sensitivity than RNMFMDA. Although KATZHMDA and LRLSHMDA obtained relatively better specificity and accuracy than RNMFMDA under individual CVs, RNMFMDA computed the best AUCs under three CVs. For example, the AUC values in RNMFMDA increased by 42.42, 31.08, 36.48, 34.37, and 51.96% compared with those in KATZHMDA, LRLSHMDA, MDLPHMDA, NGRHMDA, and NTSHMDA under CV1; the corresponding values increased by 0.08, 6.73, 5.66, 5.13, and 4.35%, respectively, under CV2; the values also increased by 5.62, 3.16, 1.57, 0.62, and 2.04%, respectively, under CV3. **Figures 1–3** showed the AUCs of these six methods. AUC is a more important measurement compared with other three evaluation metrics. Based on the comprehensive measure of the experimental results, RNMFMDA showed the optimal performance.

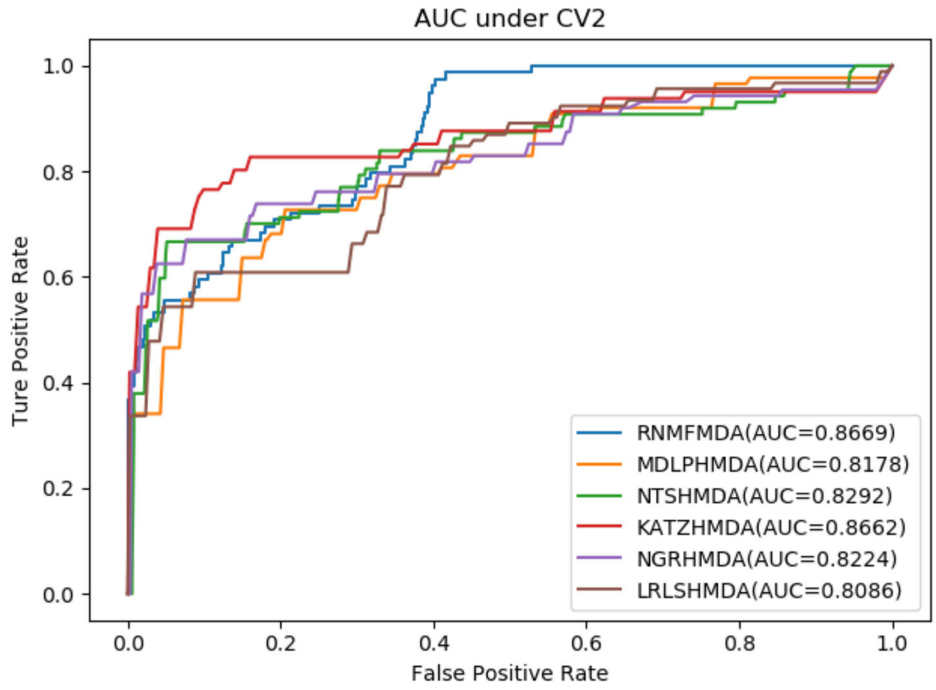

FIGURE 2 | Performance comparison of RNMFMDA with other five methods under CV2.

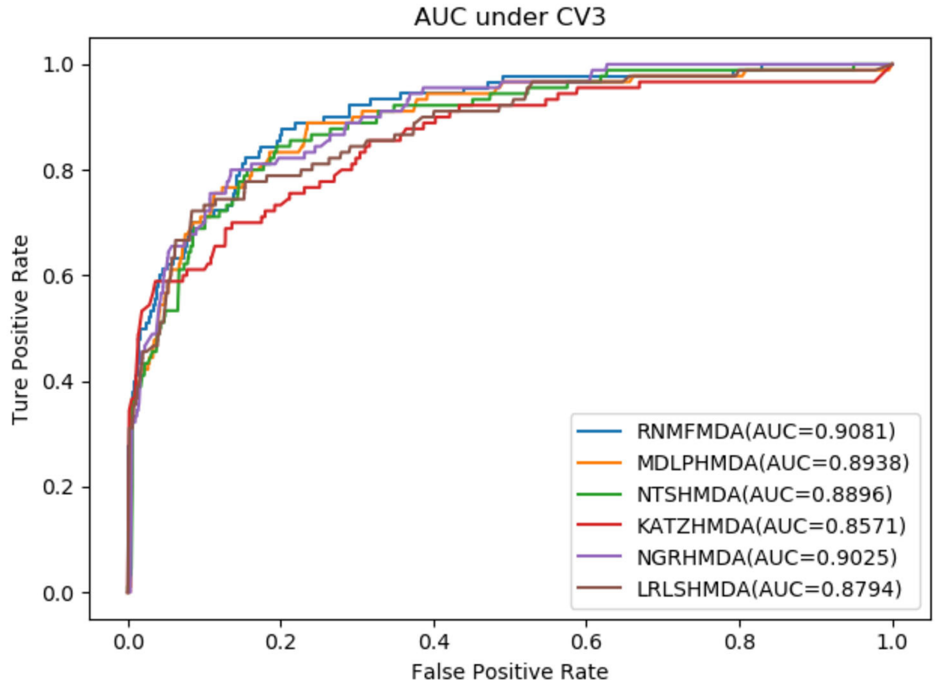

FIGURE 3 | Performance comparison of RNMFMDA with other five methods under CV3.

In addition, these six methods showed different advantages under different CVs. These variation in improvement can be attributed to differences in data structures under different CVs. In particular, RNMFMDA is more suitable to find possible microbes associated with a given disease.

TABLE 5 | Performance comparison considering the number of negative sample CV1.

NMDAR	Sensitivity	Specificity	Accuracy	AUC
0.0	0.4941	0.5696	0.5696	0.5262
0.1	**0.4945**	0.5933	0.5931	0.5920
0.2	0.4943	0.6131	0.6128	0.6123
0.5	0.4941	**0.6282**	**0.6278**	0.6279
1.0	0.4938	**0.6278**	**0.6274**	**0.6332**
2.0	0.4938	0.6226	0.6223	0.6226
3.0	0.4931	0.6218	0.6216	0.6115
4.0	0.4926	0.6067	0.6066	0.6057
5.0	0.4923	0.5674	0.5676	0.5650

These bolded values represent the best values for the different methods under the same evaluation.

TABLE 6 | Performance comparison considering the number of negative sample CV2.

NMDAR	Sensitivity	Specificity	Accuracy	AUC
0	0.5439	0.7621	0.7603	0.7673
0.1	0.5707	0.7676	0.7660	0.8088
0.2	0.6040	0.7678	0.7664	0.8220
0.5	**0.6178**	0.7830	0.7816	0.8410
1.0	0.5850	0.8304	0.8283	0.8669
2.0	0.5581	0.8547	0.8521	0.8791
3.0	0.5560	**0.8575**	**0.8547**	**0.8756**
4.0	0.5492	0.8563	0.8533	0.8782
5.0	0.5461	0.8515	0.8483	0.8734

These bolded values represent the best values for the different methods under the same evaluation.

TABLE 7 | Performance comparison considering the number of negative sample CV3.

NMDAR	Sensitivity	Specificity	Accuracy	AUC
0.0	0.5437	0.8559	0.8533	0.8662
0.1	0.5668	0.8565	0.8541	0.8827
0.2	0.6012	0.8532	0.8511	0.8886
0.5	**0.6206**	0.8560	0.8541	0.8970
1.0	0.5810	0.8818	0.8793	0.9081
2.0	0.5612	0.8916	0.8887	**0.9121**
3.0	0.5559	**0.8935**	**0.8904**	0.9096
4.0	0.5527	0.8912	0.8879	0.9099
5.0	0.5459	0.8842	0.8807	0.9026

These bolded values represent the best values for the different methods under the same evaluation.

4.3. Performance Comparison Considering PU Learning

In this section, we performed extensive experiments to analyze the influence of different negative MDA selection ratios on prediction performance. **Tables 5–7** described the comparison results. NMDAR represents the ratio of selected negative MDA samples to known positive MDA samples.

As shown in **Tables 5–7**, RNMFMDA did not extract negative MDAs when NMDAR is 0, and selected negative MDAs according to different NMDARs of 10, 20, 50%, 1, 2, 3, 4, and 5. When NMDAR is 1, RNMFMDA obtained promising performance under three CVs. Compared with the situation without negative MDA samples, when NMDAR is 1, the AUC values of RNMFMDA respectively increased 16.90, 11.49, and 4.61% under three CVs. Taken as a whole, RNMFMDA with NMDAR of 1 obtained better performance. To reduce overfitting of the experimental results, we selected NMDAR as 1, that is, we extracted negative MDA examples with the same number of positive MDA examples.

Figures 4–6 showed the AUC values obtained by RNMFMDA under different NMDARs. The results suggested that our proposed negative example extraction method helps to improve MDA prediction.

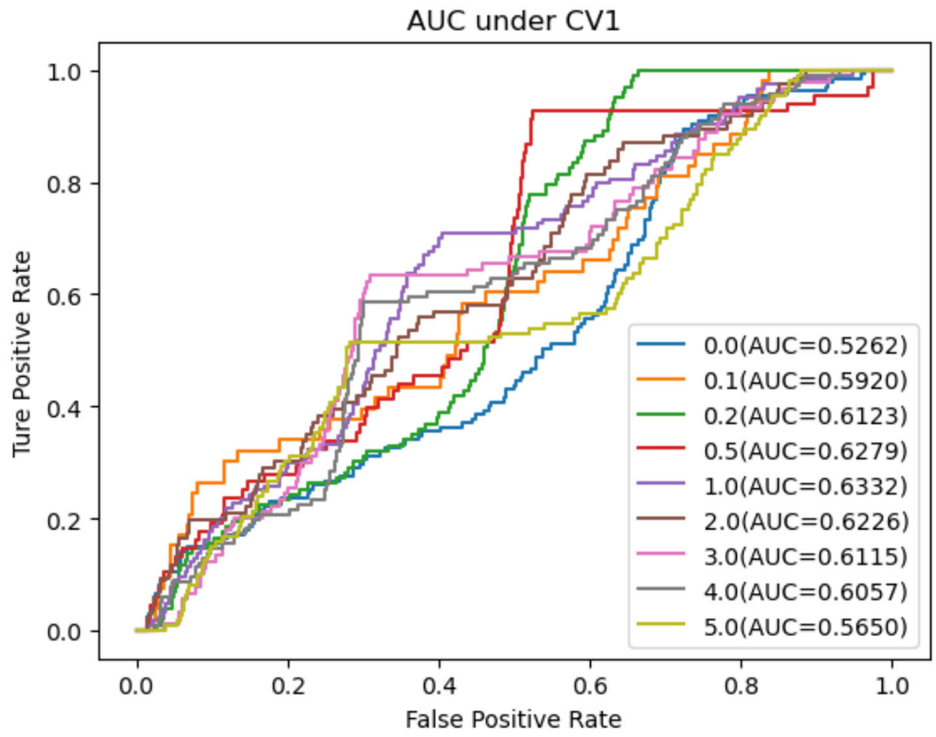

FIGURE 4 | The performance comparison under different negative MDA selection ratios under CV1.

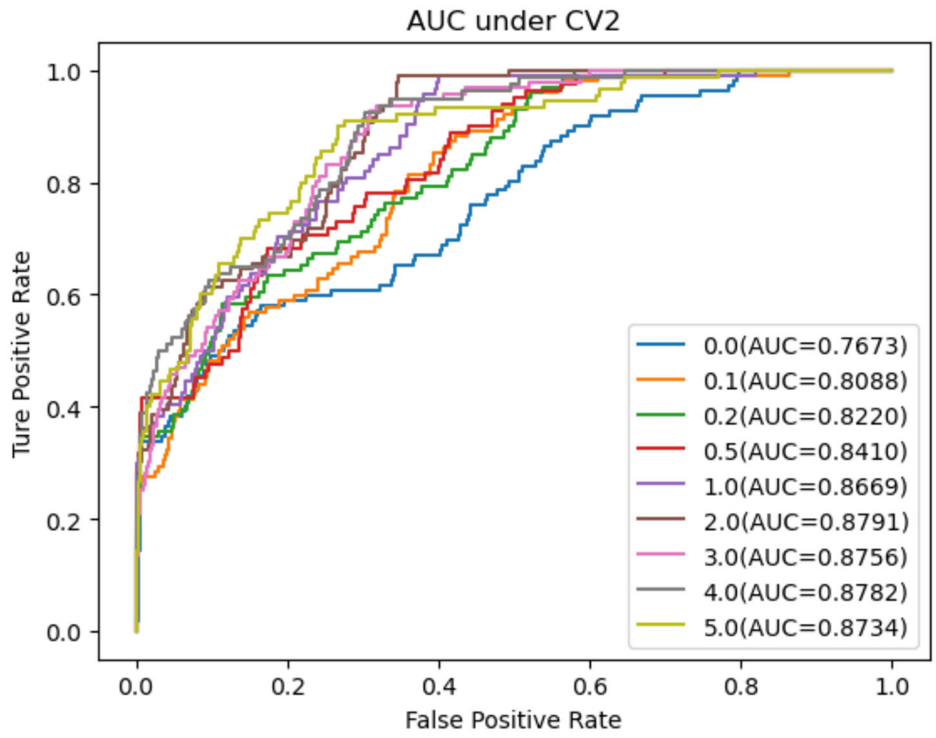

FIGURE 5 | The performance comparison under different negative MDA selection ratios under CV2.

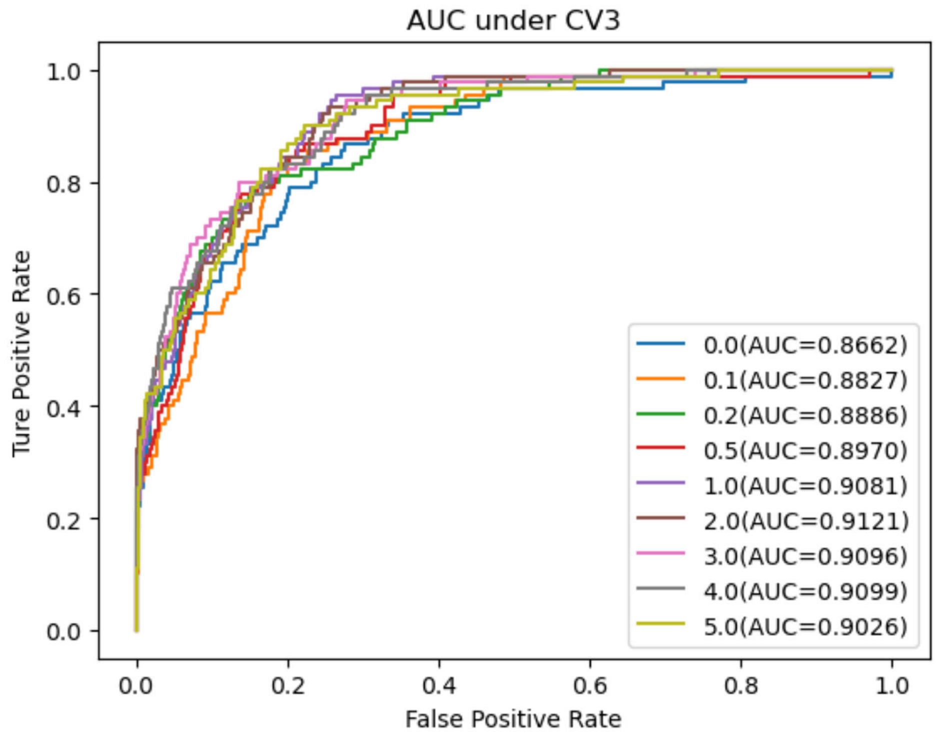

FIGURE 6 | The performance comparison under different negative MDA selection ratios under CV3.

TABLE 8 | The top 20 microbes associated with asthma.

Rank	Disease	Evidence
1	Clostridium difficile	PMID:21872915
2	Firmicutes	PMID:23265859
3	Bacteroides	PMID:18822123
4	Veillonella	PMID:25329665
5	Clostridia	Unconfirmed
6	Clostridium coccoides	PMID:21477358
7	Actinobacteria	PMID:28947029
8	Collinsella aerofaciens	Unconfirmed
9	Lachnospiraceae	PMID:26220531
10	Lactobacillus	PMID:20592920
11	Enterobacteriaceae	Unconfirmed
12	Staphylococcus aureus	Unconfirmed
13	Streptococcus	PMID:17950502
14	Fusobacterium	DOI:10.4167/jbv.2013.43.4.270
15	Burkholderia	PMID:24451910
16	Enterococcus	PMID:29788027
17	Bifidobacterium	PMID:24735374
18	Klebsiella	PMID:29788027
19	Faecalibacterium prausnitzii	Unconfirmed
20	Pseudomonas	PMID:13268970

4.4. Case Study

We further evaluated the prediction performance of our proposed RNMFMDA on the confirmed 450 MDAs by two case studies. Asthma is a disease with considerable global morbidity. Over the past 10 years, little improvement in asthma has been observed despite of escalating treatment costs (Pavord et al., 2018). In the first class, we mask all associated information for asthma to find possible microbes. The results are shown in **Table 8**. Among the predicted top 10 and 20 microbe-asthma association pairs, 8 and 15 microbes have been reported to associate with asthma by related publications, respectively.

Inflammatory Bowel Disease (IBD) is a periodic inflammation. It may be produced by a deregulated immune response to gut microbiome dysbiosis (Halfvarson et al., 2017). In the second class, we mask all association information for IBD to find possible microbes. The results are shown in **Table 9**. Among the predicted top 10 and 20 microbe-IBD association pairs, there are 9 and 17 microbes that are validated to associate with IBD by recent works, respectively.

5. DISCUSSION

There are numerous microbes in the human body. They play an important role in various biological processes. Many human diseases including gastrointestinal diseases are reported to be closely associated with microorganisms. Therefore, identifying the associations between microbes and diseases helps to understand the pathogenic mechanisms of these diseases and further develop new drugs.

Traditional experimental methods applied to validate possible associations between microbes and diseases are expensive

TABLE 9 | The top 20 microbes associated with asthma.

Rank	Disease	Evidence
1	Helicobacter pylori	PMID:22221289
2	Clostridium difficile	PMID:27698615
3	Bacteroidetes	PMID:25307765
4	Firmicutes	PMID:25307765
5	Prevotella	PMID:25307765
6	Clostridium coccoides	PMID:19235886
7	Bacteroides	PMID:25307765
8	Veillonella	PMID:28842640
9	Clostridia	Unconfirmed
10	Collinsella aerofaciens	PMID:26848182
11	Staphylococcus aureus	PMID:24117882
12	Enterobacteriaceae	Unconfirmed
13	Staphylococcus	PMID:30246806
14	Haemophilus	PMID:24013298
15	Lactobacillus	PMID:26340825
16	Bifidobacterium	Unconfirmed
17	Enterococcus	PMID:24629344
18	Burkholderia	PMID:24325678
19	Streptococcus	PMID:23679203
20	Klebsiella	PMID:29573336

and time-consuming, computational methods are developed to solve with this problem. However, the performance of existing computational models need to further improve. More importantly, lacking of reliable negative MDA examples affects prediction performance. Therefore, RNMFMDA is developed to find possible MDAs.

RNMFMDA obtained the optimal performance under three CVs. We analyzed the reason that RNMFMDA obtained excellent performance and thought that it may be contributed to the following three features. First, we developed a high-quality negative MDA extraction method based on PU learning and random walk with restart. Second, LMFNR is a optimal model in predicting associations between two entities. Finally, we integrated various heterogeneous biological information. Multiple heterogeneous data integration efficiently reflected the biological features of MDAs.

In the future, we will construct a multi-partite network by integrating MDAs, disease-gene associations (Tran et al., 2020), miRNA-disease associations (Peng et al., 2018a; Huang et al., 2019), long non-coding RNA-protein interactions (Zhao et al., 2018; Peng et al., 2019), and long non-coding RNA-disease associations (Chen et al., 2018; Li et al., 2019). More importantly, we will still develop more robust models, for example, ensemble strategy (Hu et al., 2018) and deep learning-based models (Min et al., 2017; Peng L. et al., 2018) to improve MDA prediction.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

LP, LS, and LZ: conceptualization. LP: funding acquisition, project administration, writing—original draft, and writing—review and editing. LP and LZ: investigation. LP and LS: methodology. LS, LL, and GL: software. LS: validation. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Natural Science Foundation of China (Grant 61803151) and the Natural Science Foundation of Hunan province (Grant 2018JJ3570).

ACKNOWLEDGMENTS

We would like to thank all authors of the cited references.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.592430/full#supplementary-material

REFERENCES

Abu-Ali, G. S., Mehta, R. S., Lloyd-Price, J., Mallick, H., Branck, T., Ivey, K. L., et al. (2018). Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol*. 3:356. doi: 10.1038/s41564-017-0084-4

Bao, W., Jiang, Z., and Huang, D.-S. (2017). Novel human microbe-disease association prediction using network consistency projection. *BMC Bioinformatics* 18:543. doi: 10.1186/s12859-017-1968-2

Byrd, A. L., Belkaid, Y., and Segre, J. A. (2018). The human skin microbiome. *Nat. Rev. Microbiol*. 16:143. doi: 10.1038/nrmicro.2017.157

Chen, L., Zhang, Y.-H., Huang, G., Pan, X., Wang, S., Huang, T., et al. (2018). Discriminating cirRNAs from other lncRNAs using a hierarchical extreme learning machine (H-ELM) algorithm with feature selection. *Mol. Genet. Genomics* 293, 137–149. doi: 10.1007/s00438-017-1372-7

Chen, X., Huang, Y.-A., You, Z.-H., Yan, G.-Y., and Wang, X.-S. (2017). A novel approach based on Katz measure to predict associations of human microbiota with non-infectious diseases. Bioinformatics 33, 733–739. doi: 10.1093/bioinformatics/btw715

Chen, X., Liu, M.-X., and Yan, G.-Y. (2012). Drug-target interaction prediction by random walk on the heterogeneous network. *Mol. Biosyst*. 8, 1970–1978. doi: 10.1039/c2mb00002d

Davis-Richardson, A. G., Ardissone, A. N., Dias, R., Simell, V., Leonard, M. T., Kemppainen, K. M., et al. (2014). *Bacteroides dorei* dominates gut microbiome prior to autoimmunity in Finnish children at high risk for type 1 diabetes. *Front. Microbiol*. 5:678. doi: 10.3389/fmicb.2014.00678

Ding, T., and Schloss, P. D. (2014). Dynamics and associations of microbial community types across the human body. *Nature* 509:357. doi: 10.1038/nature13178

Findley, K., Oh, J., Yang, J., Conlan, S., Deming, C., Meyer, J. A., et al. (2013). Topographic diversity of fungal and bacterial communities in human skin. *Nature* 498:367. doi: 10.1038/nature12171

Halfvarson, J., Brislawn, C. J., Lamendella, R., Vázquez-Baeza, Y., Walters, W. A., Bramer, L. M., et al. (2017). Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* 2, 1–7. doi: 10.1038/nmicrobiol.2017.4

Heavey, P. M., and Rowland, I. R. (2004). Gastrointestinal cancer. *Best Pract. Res. Clin. Gastroenterol.* 18, 323–336. doi: 10.1016/j.bpg.2003.10.003

Hu, H., Zhang, L., Ai, H., Zhang, H., Fan, Y., Zhao, Q., et al. (2018). HLPI-ensemble: prediction of human lncRNA-protein interactions based on ensemble strategy. *RNA Biol.* 15, 797–806. doi: 10.1080/15476286.2018.1457935

Huang, F., Yue, X., Xiong, Z., Yu, Z., and Zhang, W. (2019). Tensor decomposition with relational constraints for predicting multiple types of microRNA-disease associations. *arXiv preprint arXiv:1911.05584*. doi: 10.1093/bib/bbaa140

Huang, Y.-A., You, Z.-H., Chen, X., Huang, Z.-A., Zhang, S., and Yan, G.-Y. (2017). Prediction of microbe-disease association from the integration of neighbor and graph with collaborative recommendation model. *J. Transl. Med.* 15:209. doi: 10.1186/s12967-017-1304-7

Langella, P., and Martín, R. (2019). Emerging health concepts in the probiotics field: streamlining the definitions. *Front. Microbiol.* 10:1047. doi: 10.3389/fmicb.2019.01047

Ley, R. E., Turnbaugh, P. J., Klein, S., and Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature* 444:1022. doi: 10.1038/4441022a

Li, G., Luo, J., Liang, C., Xiao, Q., Ding, P., and Zhang, Y. (2019). Prediction of lncRNA-disease associations based on network consistency projection. *IEEE Access* 7, 58849–58856. doi: 10.1109/ACCESS.2019.2914533

Li, H., Chen, Z., Liu, B., Wei, X., and Shao, J. (2014). "Spotting fake reviews via collective positive-unlabeled learning," in *2014 IEEE International Conference on Data Mining* (Shenzhen: IEEE), 899–904. doi: 10.1109/ICDM.2014.47

Liu, H., Han, M., Li, S. C., Tan, G., Sun, S., Hu, Z., et al. (2019). Resilience of human gut microbial communities for the long stay with multiple dietary shifts. *Gut* 68, 2254–2255. doi: 10.1136/gutjnl-2018-317298

Liu, H., Ren, G., Chen, H., Liu, Q., Yang, Y., and Zhao, Q. (2020). Predicting lncRNA-miRNA interactions based on logistic matrix factorization with neighborhood regularized. *Knowl. Based Syst.* 191:105261. doi: 10.1016/j.knosys.2019.105261

Liu, Y., Wu, M., Miao, C., Zhao, P., and Li, X.-L. (2016). Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. *PLoS Comput. Biol.* 12:e1004760. doi: 10.1371/journal.pcbi.1004760

Luo, J., and Long, Y. (2018). NTSHMDA: prediction of human microbe-disease association based on random walk by integrating network topological similarity. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 17, 1341–51. doi: 10.1109/TCBB.2018.2883041

Ma, W., Zhang, L., Zeng, P., Huang, C., Li, J., Geng, B., et al. (2016). An analysis of human microbe-disease associations. *Brief. Bioinform.* 18, 85–97. doi: 10.1093/bib/bbw005

McFarland, L. V. (2000). Beneficial microbes: health or hazard? *Eur. J. Gastroenterol. Hepatol.* 12, 1069–1071. doi: 10.1097/00042737-200012100-00001

Min, S., Lee, B., and Yoon, S. (2017). Deep learning in bioinformatics. *Brief. Bioinform.* 18, 851–869. doi: 10.1093/bib/bbw068

Pavord, I. D., Beasley, R., Agusti, A., Anderson, G. P., Bel, E., Brusselle, G., et al. (2018). After asthma: redefining airways diseases. *Lancet* 391, 350–400. doi: 10.1016/S0140-6736(17)30879-6

Peng, L., Liu, F., Yang, J., Liu, X., Meng, Y., Deng, X., et al. (2019). Probing lncRNA-protein interactions: data repositories, models, and algorithms. *Front. Genet.* 10:1346. doi: 10.3389/fgene.2019.01346

Peng, L., Peng, M., Liao, B., Huang, G., Li, W., and Xie, D. (2018). The advances and challenges of deep learning application in biological big data processing. *Curr. Bioinform.* 13, 352–359. doi: 10.2174/1574893612666170707095707

Peng, L.-H., Sun, C.-N., Guan, N.-N., Li, J.-Q., and Chen, X. (2018a). HNMDA: heterogeneous network-based miRNA-disease association prediction. *Mol. Genet. Genomics* 293, 983–995. doi: 10.1007/s00438-018-1438-1

Peng, L.-H., Yin, J., Zhou, L., Liu, M.-X., and Zhao, Y. (2018b). Human microbe-disease association prediction based on adaptive boosting. *Front. Microbiol.* 9:2440. doi: 10.3389/fmicb.2018.02440

Qu, J., Zhao, Y., and Yin, J. (2019). Identification and analysis of human microbe-disease associations by matrix decomposition and label propagation. *Front. Microbiol.* 10:291. doi: 10.3389/fmicb.2019.00291

Qu, K., Guo, F., Liu, X., Lin, Y., and Zou, Q. (2019). Application of machine learning in microbiology. *Front. Microbiol.* 10:827. doi: 10.3389/fmicb.2019.00827

Sachdeva, R., Campbell, B. J., and Heidelberg, J. F. (2019). Rare microbes from diverse earth biomes dominate community activity. *bioRxiv* 636373. doi: 10.1101/636373

Shi, J.-Y., Huang, H., Zhang, Y.-N., Cao, J.-B., and Yiu, S.-M. (2018). BMCMDA: a novel model for predicting human microbe-disease associations via binary matrix completion. *BMC Bioinformatics* 19:169. doi: 10.1186/s12859-018-2274-3

Sze, M. A., Dimitriu, P. A., Hayashi, S., Elliott, W. M., McDonough, J. E., Gosselink, J. V., et al. (2012). The lung tissue microbiome in chronic obstructive pulmonary disease. *Am. J. Respirat. Crit. Care Med.* 185, 1073–1080. doi: 10.1164/rccm.201111-2075OC

Tran, V. D., Sperduti, A., Backofen, R., and Costa, F. (2020). Heterogeneous networks integration for disease-gene prioritization with node kernels. *Bioinformatics* 36, 2649–2656. doi: 10.1093/bioinformatics/btaa008

van Laarhoven, T., Nabuurs, S. B., and Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* 27, 3036–3043. doi: 10.1093/bioinformatics/btr500

Wang, F., Huang, Z.-A., Chen, X., Zhu, Z., Wen, Z., Zhao, J., et al. (2017). LRLSHMDA: Laplacian regularized least squares for human microbe-disease association prediction. *Sci. Rep.* 7, 1–11. doi: 10.1038/s41598-017-08127-2

Zhang, W., Jing, K., Huang, F., Chen, Y., Li, B., Li, J., et al. (2019a). SFLLN: a sparse feature learning ensemble method with linear neighborhood regularization for predicting drug-drug interactions. *Inform. Sci.* 497, 189–201. doi: 10.1016/j.ins.2019.05.017

Zhang, W., Li, Z., Guo, W., Yang, W., and Huang, F. (2019b). A fast linear neighborhood similarity-based network link inference method to predict microRNA-disease associations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 1. doi: 10.1109/TCBB.2019.2931546

Zhang, W., Tang, G., Zhou, S., and Niu, Y. (2019c). LncRNA-miRNA interaction prediction through sequence-derived linear neighborhood propagation method with information combination. *BMC Genomics* 20:946. doi: 10.1186/s12864-019-6284-y

Zhang, W., Yang, W., Lu, X., Huang, F., and Luo, F. (2018). The bi-direction similarity integration method for predicting microbe-disease associations. *IEEE Access* 6, 38052–38061. doi: 10.1109/ACCESS.2018.2851751

Zhao, Q., Yu, H., Ming, Z., Hu, H., Ren, G., and Liu, H. (2018). The bipartite network projection-recommended algorithm for predicting long non-coding RNA-protein interactions. *Mol. Therapy Nucleic Acids* 13, 464–471. doi: 10.1016/j.omtn.2018.09.020

Zhou, X., Menche, J., Barabási, A.-L., and Sharma, A. (2014). Human symptoms-disease network. *Nat. Commun.* 5:4212. doi: 10.1038/ncomms5212

Zhu, Y.-G., Gillings, M., Simonet, P., Stekel, D., Banwart, S., and Penuelas, J. (2018). Human dissemination of genes and microorganisms in Earth's critical zone. *Glob. Change Biol.* 24, 1488–1499. doi: 10.1111/gcb.14003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.