Claims

1. A method for chronic disease comorbidity pattern recognition with graph convolution constraint-enhanced low rank representation, characterized by comprising:

   obtaining medical data resources, wherein the medical data resources include data records of chronic disease incidence sorted out from testing, diagnosis, treatment, prescription, electronic medical records, and clinical data resources;
   constructing a chronic disease comorbidity network using medical data resources, wherein the chronic disease comorbidity network is used to describe the correlation and influence between diseases, and can be constructed based on multi-source data by calculating the correlation between diseases or patients, the multi-source data includes patient basic information as well as testing, diagnosis, treatment, prescription data, electronic medical records, clinical data, and omics big data, among them, the chronic disease comorbidity network is represented by graphs $G = (V, E)$ and $V = \{v_i | i \in \{1, \dots, n\}\}$ as a set containing disease nodes, and $E = \{e_{ij} | i, j \in \{1, \dots, m\}\}$ as a set containing connected edges; the adjacency matrix $A = [a_{ij}]$ is used to represent and store the target comorbidity network$G$, where a_{ij} represents the interaction relationship between disease nodes $v_i$and v_j, and its value is equal to $I(s_i, s_j)$;
   perform low rank representation learning on the chronic disease comorbidity network to mine the community structure output, wherein the low rank representation learning uses a low rank representation method with graph convolution constraints to represent the comorbidity network;
   using community segmentation mechanisms to identify community structures in comorbidity networks and explore comorbidity patterns.

2. The method according to claim 1, characterized in that the construction method of the chronic disease comorbidity network comprises:

   sort out chronic disease records, and extract data records from testing, diagnosis, treatment, prescription, electronic medical records, and clinical data;
   count the frequency of disease occurrence, and calculate the frequency $Q(s_i)$ of each disease occurrence and the frequency $Q(s_i, s_j)$ of disease combination occurrence from the data record, if a certain disease s_i appears in one disease record, count it once, if two diseases s_i and s_j appear together in one disease record, count it once, $i, j = 1,2, \dots, n$ represents the number of diseases of interest;
   calculate the mutual information value between two diseases using the following method:

$$I\left(s_i,s_j\right)=\log\frac{p\left(s_i,s_j\right)}{p\left(s_i\right)p\left(s_j\right)}$$

among them, $p(s_i) = Q(s_i)/F$ and $p(s_j) = Q(s_j)/F$ respectively calculate the observation frequency of diseases and, where F represents the total observation frequency, and $p(s_i, s_j) = Q(s_i, s_j)/F$ calculates the frequency of simultaneous observation of diseases s_i and s_j;

build a comorbidity network to describe the complex impact relationships of chronic diseases.

3. The method according to claim 2, characterized in that the low rank representation method using graph convolution constraints to characterize the comorbidity network comprises:

design the objective function, which is as follows:

$$\begin{aligned}&J_{GCGS}=\|A-UU^{\mathrm{T}}\|_F^2+\lambda_1\mathrm{Tr}\left(U^{\mathrm{T}}LU\right),\\&s.t.\ U=\sigma\left(\bar{A}\sigma\left(\bar{A}HW^{(1)}\right)W^{(2)}\right),\\&\quad U,H,W^{(1)},W^{(2)}\geq 0\end{aligned}$$

among them, A is the adjacency matrix of the comorbidity network, U is the feature matrix of size $n\times K$, and n and K are the number of nodes and communities in the comorbidity network, respectively; $\|\cdot\|_F^2$ is the F-norm of the computational matrix; Tr(·) is the trace of the computed matrix; U^{T} represents the transpose of matrix U; $L=D-A$ is the Laplacian matrix of A, where $D=\sum_l A_{il}$ is the degree matrix of A, and A_{il} is the element value at the corresponding index in A; $H=\{h_1, \ldots, h_i, \ldots, h_n\}\in\mathbb{R}^{n\times d}$ representing the hidden layer feature matrix of the GCN module, where h_i corresponds to the feature vector of node v_i and d represents the feature dimension; $W^{(1)}\in\mathbb{R}^{d\times K}$ and $W^{(2)}\in\mathbb{R}^{K\times K}$ represent the weight matrix of the GCN module; $\bar{A}=\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$ is a normalized adjacency matrix, where $\tilde{A}$ is an adjacency matrix with self connections added, $\tilde{D}=\sum_l\tilde{A}_{il}$ is the degree matrix of $\tilde{A}$, and $\tilde{A}_{il}$ is the value of the element at the corresponding index in $\tilde{A}$; λ_1 is the graph regularization coefficient; $\sigma(\cdot)$ is the nonlinear activation function of the GCN module;

transform the graph convolution constraint into the following regularized form:

$$\begin{aligned}&J_{GCGS}=\|A-UU^{\mathrm{T}}\|_F^2+\lambda_1\mathrm{Tr}\left(U^{\mathrm{T}}LU\right)+\lambda_2\left\|U-\sigma\left(\bar{A}\sigma\left(\bar{A}HW^{(1)}\right)W^{(2)}\right)\right\|_F^2,\\&s.t.\ U,H,W^{(1)},W^{(2)}\geq 0\end{aligned}$$

solve the following optimization problem to obtain the parameter matrix U、H、

$W^{(1)}$ and $W^{(2)}$:

$$\left(U,H,W^{(1)},W^{(2)}\right)\leftarrow\arg\min J_{GCGS},\ \ s.t.\ \ U,H,W^{(1)},W^{(2)}\geq 0$$

based on the Lagrange function method and the KKT condition with inequality constraints, derive the learning rules for the optimization problem:

$$X\leftarrow X\odot\left(\left(1-\beta\right)+\beta\frac{\left(1+\lambda_1\right)AX+\lambda_2 g_2}{XX^{\mathrm{T}}X+\lambda_1 DX+\lambda_2 X}\right)$$

$$H\leftarrow H\odot\frac{\bar{A}^{\mathrm{T}}\bar{A}^{\mathrm{T}}\{X\odot g_2'\odot g_1'\}(W^{(2)})^{\mathrm{T}}(W^{(1)})^{\mathrm{T}}}{\bar{A}^{\mathrm{T}}\bar{A}^{\mathrm{T}}\{g_2\odot g_2'\odot g_1'\}(W^{(2)})^{\mathrm{T}}(W^{(1)})^{\mathrm{T}}}$$

$$W^{(1)}\leftarrow W^{(1)}\odot\frac{(\bar{A}H)^{\mathrm{T}}\bar{A}^{\mathrm{T}}\{X\odot g_2'\odot g_1'\}(W^{(2)})^{\mathrm{T}}}{(\bar{A}H)^{\mathrm{T}}\bar{A}^{\mathrm{T}}\{g_2\odot g_2'\odot g'\}(W^{(2)})^{\mathrm{T}}}$$

$$W^{(2)}=W^{(2)}\odot\frac{\left(\bar{A}g_1\right)^{\mathrm{T}}\{X\odot g_2'\}}{\left(\bar{A}g_1\right)^{\mathrm{T}}\{g_2\odot g_2'\}}$$

among them, λ_1 is the graph regularization coefficient, λ_2 is the graph convolution regularization coefficient, g_1 and g_1' (derivative of g_1), representing $\sigma(\bar{A}HW^{(1)})$ and$\sigma'(\bar{A}HW^{(1)})$; g_2 and g_2' (derivative of g_2), representing $\sigma(\sigma(\bar{A}HW^{(1)})W^{(2)})$ and $\sigma'(\sigma(\bar{A}HW^{(1)})W^{(2)})$, respectively; $\sigma'(\cdot)$ is the derivative of $\sigma(\cdot)$; β is the linear adjustment coefficient of the learning scale factor of the control feature factor matrix U, which takes values in the interval (0,1); $\odot$ represents the Hadamard product of a matrix; according to the learning rules, iteratively update the feature and weight matrices U、H、$W^{(1)}$ and $W^{(2)}$ until the convergence condition is reached, and stop, the final feature matrix U、H and weight $W^{(1)}$、$W^{(2)}$ obtained are used as the solution to the optimization problem, and the output feature matrix U is used as the indicator matrix for social division.

4. The method according to claim 3, characterized in that the use of community partitioning mechanism to identify community structures in the comorbidity network and discover comorbidity patterns comprises:

after training the feature parameters U and H, use U as the indicator matrix for dividing the node disease comorbidity module ownership relationship, and judge each node in the comorbidity network one by one, assigning them to the corresponding comorbidity module, the assignment rule is as follows:

$$\forall v_i\in V:\ \ v_i\in C_s,\ \mathrm{if}\, s=\arg\max_k u_{ik}, k=\{1,2,\ldots,K\}$$

among them, V represents the set of nodes in the comorbidity network, and v_i

represents any one of the nodes; C_s is the comorbidity combination s, $s \in \{1, 2, \ldots, K\}$, K is the total number of comorbidity combinations; u_{ik} is an element in the low rank matrix U, representing the probability of node v_i being assigned to the comorbidity combination k; if u_{is} is the maximum value among them, then node v_i has the highest possibility of being assigned to the comorbidity combination s, after assigning comorbidity combinations to all nodes in the target comorbidity network, all identified comorbidity combinations may be $C=\{C_1, C_2, \ldots, C_K\}$.

5. A device for chronic disease comorbidity pattern recognition with graph convolution constraint-enhanced low rank representation, characterized by comprising:

an acquisition unit used to obtain medical data resources, wherein the medical data resources include data records of chronic disease incidence sorted out from inspection, diagnosis, treatment, prescription, electronic medical records, and clinical data resources; building a unit that utilizes medical data resources to construct a chronic disease comorbidity network, wherein the chronic disease comorbidity network is used to describe the correlation and influence between diseases, and can be constructed based on multi-source data by calculating the correlation between diseases or patients, the multi-source data includes patient basic information as well as testing, diagnosis, treatment, prescription data, electronic medical records, clinical data, and omics big data, among them, the chronic disease comorbidity network is represented by graphs $G = (V, E)$ and $V = \{v_i | i \in \{1, \ldots, n\}\}$ as a set containing n disease nodes, and $E = \{e_{ij} | i, j \in \{1, \ldots, m\}\}$ as a set containing m connected edges; the adjacency matrix $A = [a_{ij}]$ is used to represent and store the target comorbidity network G, where a_{ij} represents the interaction relationship between disease nodes v_i and v_j, and its value is equal to $I(s_i, s_j)$;
a characterization unit is used to perform low rank representation learning on the chronic disease comorbidity network and mine the community structure in it, the low rank representation learning uses a low rank representation method that integrates graph convolution constraints to represent the comorbidity network;
partition units are used to identify community structures in the comorbidity network and explore comorbidity patterns using community partitioning mechanisms.

6. An electronic equipment characterized by comprising:

processor;

a memory for storing executable instructions of the processor;
the processor is used to read the executable instructions from the memory and execute the executable instructions to implement any one of the methods described in claims 1 to 4.

7. A computer-readable storage medium, characterized in that the computer-readable storage medium stores a computer program for executing the method according to any one of claims 1 to 4.