Claims

1. A method for determining a degree of gene association, characterized by comprising:

determining a disease description entry in a case description text;

for each of a plurality of preset association databases:

determining record data of target association record(s) in the preset association database according to the disease description entry and gene identifiers each corresponding to one of a plurality of genes, wherein each of the preset association databases stores association records of the disease description entry and gene identifiers corresponding to at least one of the plurality of genes;

inputting the record data into a preset entry-gene association matrix, to determine an association score of the disease description entry with respective to each of the plurality of genes in the preset association database; and

determining degree of association between the disease description entry and each of the plurality of genes according to association scores of the disease description entry with respect to the gene identifiers corresponding to the plurality of genes in the plurality of preset association databases; the determining record data of target association record(s) in the preset association database according to the disease description entry and gene identifiers each corresponding to one of a plurality of genes comprises:

for any gene of the plurality of genes, determining record data of the target association record(s) of the disease description entry and a gene identifier corresponding to the gene in the preset association database respectively, wherein the record data of the target association record(s) comprises a first number of association record(s) including both the gene identifier of the gene and the disease description entry, a second number of association record(s) including the disease description entry, a third number of association record(s) including the gene identifier of the gene, and a total number of association record(s) in the preset association database; the preset entry-gene association matrix is:

$$M=\frac{T_1}{T_2}\times\log\left(\frac{T_s}{T_3+1}\right)$$

wherein, M is the association score of the disease description entry with respect to the gene identifier corresponding to the gene in the preset association database; T_1 is the first number, T_2 is the second number, T_3 is the third number; T_S is the total number.

2. The method according to any one of claims 1 to 3, characterized in that the determining degree of association between the disease description entry and each of the plurality of genes according to association scores of the disease description entry with respect to the gene identifiers corresponding to the plurality of genes in the plurality of preset association databases comprises, for any gene of the plurality of genes:

weighting the association scores of the disease description entry with respect to the gene identifier corresponding to the gene in the plurality of preset association databases to obtain an association value between the disease description entry and the gene; and

determining the degree of association between the disease description entry and the gene according to the association value.

3. The method according to claim 2, characterized in that the determining the degree of association between the disease description entry and the gene according to the association value comprises:

determining the degree of association between the disease description entry and the gene according to $L=\frac{1}{\left(1+\mathrm{e}^{-z}\right)}$, wherein L is the degree of association between the disease description entry and the gene, e is the natural constant, and z is the association value.

4. The method according to claim 1, characterized in that a training process of the preset entry-gene association matrix comprises:

inputting, as training information, record data of target association record(s) including a plurality of disease description entries and gene identifiers of a plurality of genes in the plurality of preset association databases into a current entry-gene association training

matrix, and determining predicted association scores corresponding to the training information and the plurality of genes in the preset information gene association databases respectively;

determining predicted degrees of association between the plurality of disease description entries and the plurality of genes respectively according to the determined predicted association scores;

determining errors between the predicted degrees of association and actual degrees of association of the plurality of disease description entries and the plurality of genes;

determining whether the errors are below a preset error threshold;

if yes, determining that the current entry-gene association training matrix is the entry-gene association matrix; and

if not, adjusting matrix parameters in the current entry-gene association training matrix through a gradient descent algorithm, using the adjusted entry-gene association training matrix as the current entry-gene association training matrix, and turning to the step of inputting, as training information, record data of target association records including a plurality of disease description entries and gene identifiers of a plurality of genes in the plurality of preset association databases into a current entry-gene association training matrix, and determining predicted association scores corresponding to the training information and the plurality of genes in the preset association databases respectively.

5. A device for determining a degree of gene association, characterized by comprising a disease description entry determination unit, an association score determination unit, and an association degree determination unit;

wherein the disease description entry determination unit is used for determining a disease description entry in a case description text;

the association score determination unit is used for, for each of a plurality of preset association databases: determining record data of target association record(s) in the preset association database according to the disease description entry and gene identifiers each corresponding to one of a plurality of genes, wherein each of the preset association databases stores association records of the disease description entry and

gene identifiers corresponding to at least one of the plurality of genes; and inputting the record data into a preset entry-gene association matrix, to determine an association score of the disease description entry with respective to each of the plurality of genes in the preset association database; and

the association degree determination unit is used for determining degree of association between the disease description entry and each of the plurality of genes according to association scores of the disease description entry with respect to the gene identifiers corresponding to the plurality of genes in the plurality of preset association databases; wherein the association score determination unit is specifically used for, for any gene of the plurality of genes, respectively determining record data of target association record(s) including the disease description entry and a gene identifier corresponding to the gene in the preset association database, the record data of the target association record(s) comprising: a first number of association record(s) including both the gene identifier of the gene and the disease description entry, a second number of association record(s) including the disease description entry, a third number of association record(s) including the gene identifier of the gene, and a total number of association record(s) in the preset association database; the preset entry-gene association matrix is:

$$M=\frac{T_1}{T_2}\times\log\left(\frac{T_s}{T_3+1}\right)$$

wherein, M is the association score of the disease description entry with respect to the gene identifier corresponding to the gene in the preset association database; T_1 is the first number, T_2 is the second number, T_3 is the third number; T_S is the total number.

6. The device according to claim 5, characterized in that the association degree determination unit comprises an association value obtaining subunit and an association degree determination subunit,

wherein the association value obtaining subunit is used for, for any gene of the plurality of genes, weighting the association scores of the disease description entry with

respect to the gene identifier corresponding to the gene in the plurality of preset association databases to obtain an association value between the disease description entry and the gene;

the association degree determination subunit is used for determining the degree of association between the disease description entry and the gene according to the association value.

7. The device according to claim 6, characterized in that the association degree determination subunit is used for determining the degree of association between the disease description entry and the gene according to $L = \frac{1}{\left(1+e^{-z}\right)}$, wherein L is the degree of association between the disease description entry and the gene, e is the natural constant, and z is the association value.

8. The device according to claim 5, characterized in further comprising: a predicted association score determination unit, a predicted association degree determination unit, an error determination unit, an error comparison unit, an entry-gene association matrix determination unit, and a matrix parameter adjustment unit,

wherein the predicted association score determination unit is used for inputting, as training information, record data of target association records including a plurality of disease description entries and gene identifiers of a plurality of genes in the plurality of preset association databases into a current entry-gene association training matrix, and determining predicted association scores corresponding to the training information and the plurality of genes in the preset association databases respectively;

the predicted association degree determination unit is used for determining predicted degrees of association between the plurality of disease description entries and the plurality of genes respectively according to the determined predicted association scores;

the error determination unit is used for determining errors between the predicted degrees of association and actual degrees of association of the plurality of disease description entries and the plurality of genes;

the error comparison unit is used for determining whether the errors are below a

preset error threshold, if yes, triggering the entry-gene association matrix determination unit, or if not, triggering the matrix parameter adjustment unit;

the entry-gene association matrix determination unit is used for determining the current entry-gene association training matrix as the entry-gene association matrix;

the matrix parameter adjustment unit is used for adjusting matrix parameters in the current entry-gene association training matrix through a gradient descent algorithm, using the adjusted entry-gene association training matrix as the current entry-gene association training matrix, and triggering the predicted association score determination unit.

9. A storage medium, characterized in that a program is stored in the storage medium, which when executed by a processor implements the method for determining the degree of gene association according to any one of claims 1 to 4.

10. A processor, characterized in that the processor is used to execute a program, wherein the program when executed performs the method for determining the degree of gene association according to any one of claims 1 to 4.

11. An electronic device, characterized in comprising at least one processor, at least one memory coupled to the processor, and a bus;

wherein, the processor communicates with the memory through the bus, and the processor is used for invoking program instructions in the memory to perform the method for determining the degree of gene association according to any one of claims 1 to 4.

12. A computer program product, characterized in that when executed on an electronic device, the computer program product is adapted to execute a program for initializing the steps of the method for determining the degree of gene association according to any one of claims 1 to 4.