

INVENTION TITLE: METHOD AND APPARATUS FOR CLASSIFYING MEDICAL DATA BASED ON MACHINE LEARNING, COMPUTER DEVICE, AND STORAGE MEDIUM

The present application claims priority to Chinese Patent Application No. 2019101715930, filed with the China National Intellectual Property Administration on March 7, 2019 and entitled "METHOD AND APPARATUS FOR CLASSIFYING MEDICAL DATA BASED ON MACHINE LEARNING, AND COMPUTER DEVICE", which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

The present application relates to the field of computer technologies, and in particular, to a method and apparatus for classifying medical data based on machine learning, a computer device, and a storage medium.

BACKGROUND

Cancers are important health issues. In recent years, the prevalence of cancers has been increasing, and early diagnosis and treatment on cancers can significantly increase the chance of survival. With rapid development of computer technologies and medical technologies, some manners of intelligently classifying massive medical data emerge. For example, structured vocabularies for single medical records are extracted from medical record books, a model of a medical record topic is built, and corresponding categories are obtained through training for the medical record topic. Alternatively, input samples are trained by using prior knowledge, so as to classify cancer types and help reduce labor intensity of medical workers.

In conventional medical data classification manners, most data to be classified and analyzed is existing fixed data, and data sources are limited. Therefore, actual medical history information of users cannot be classified and analyzed, and most medical history information is complex and specific medical history analysis and recorded text. As medical text is special, deviations of terms in medical history information will cause totally inconsistent meanings.

SUMMARY

A method for classifying medical data based on machine learning is executed by a computer device. The method includes:

receiving a medical data classification request sent by a terminal, where the medical data

classification request includes medical history information;

obtaining a preset medical term base, and performing word segmentation on the medical history information based on medical terms in the medical term base to obtain multiple text vectors;

performing feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values;

obtaining a target classifier, and performing traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the target classifier, where the target classifier is obtained by training multiple pieces of medical data;

when a target node corresponding to each of the multiple text vectors is traversed, calculating a category probability corresponding to the text vector based on the target node, and obtaining a category result corresponding to the medical history information based on the category probability; and

pushing the category result corresponding to the medical history information to the terminal.

In an embodiment, the medical history information includes multiple pieces of text data, and the step of performing word segmentation on the medical history information includes: obtaining the preset medical term base, where the medical term base includes multiple medical terms; and performing matching between the multiple pieces of text data in the medical history information and the medical term base, calculating matching degrees between the text data in the medical history information and the multiple medical terms, and extracting text data that reaches a preset matching degree; performing word segmentation on the medical history information based on the matched text data to obtain multiple pieces of text data; and performing vector conversion on the multiple pieces of text data obtained after the word segmentation to obtain the multiple text vectors.

In an embodiment, the step of performing feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values includes: calculating a term frequency and an inverse document frequency of the multiple text vectors; calculating weights of the multiple text vectors by using a preset algorithm based on the term frequency and the inverse document frequency; extracting a text vector whose weight reaches a preset threshold; and calculating a feature dimension value corresponding to the text vector based on a preset algorithm and the weight.

In an embodiment, a step of constructing the target classifier includes: obtaining multiple pieces of medical data, and generating corresponding training set data and verification set data

based on the multiple pieces of medical data; performing clustering analysis on multiple pieces of medical data in the training set data to obtain a clustering result; performing feature extraction on the clustering result to extract multiple feature variables; obtaining a preset neural network model, training the training set data by using the neural network model to obtain feature dimension values and weights corresponding to the multiple feature variables, and constructing an initial classifier based on the feature dimension values and the weights corresponding to the multiple feature variables; and performing further training and verification on the classifier by using the verification set data, and when the quantity of pieces of data that meets a preset threshold in the verification set data reaches a preset proportion, stopping training to obtain a required target classifier.

In an embodiment, the text includes multiple text sentences, the multiple text sentences form a text block, and the step of performing traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the target classifier to obtain the category corresponding to the multiple text vectors includes: calculating a correlation between the multiple text vectors by using the target classifier based on the feature dimension values, calculating the text sentences formed in the text based on the correlation, and calculating sentence vectors of the text sentences; extracting features of the sentence vectors, and calculating a text block vector based on the features of the multiple sentence vectors; and calculating a probability of each category corresponding to the text block vector, extracting a category whose probability reaches a preset probability value, and adding a corresponding category label to the text block.

In an embodiment, the method further includes obtaining multiple pieces of historical medical data from a preset database based on a preset frequency; performing clustering analysis on the multiple pieces of historical medical data to obtain an analysis result; performing feature selection based on the analysis result to obtain multiple feature variables; calculating weights of the multiple feature variables based on a preset algorithm; and performing optimization adjustment on the target classifier based on the multiple feature variables and the corresponding weights.

An apparatus for classifying medical data based on machine learning is provided. The apparatus includes:

a request receiving module, configured to receive a medical data classification request sent by a terminal, where the medical data classification request includes medical history information;

a word segmentation module, configured to obtain a preset medical term base, and perform

word segmentation on the medical history information based on medical terms in the medical term base to obtain multiple text vectors;

a feature extraction module, configured to perform feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values;

a data classification module, configured to obtain a target classifier, and perform traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the target classifier, where the target classifier is obtained by training multiple pieces of medical data; and when a target node corresponding to each of the multiple text vectors is traversed, calculate a category probability corresponding to the text vector based on the target node, and obtain a category result corresponding to the medical history information based on the category probability; and

a data pushing module, configured to push the category result corresponding to the medical history information to the terminal.

In an embodiment, the word segmentation module is further configured to obtain the preset medical term base, where the medical term base includes multiple medical terms; and perform matching between the multiple pieces of text data in the medical history information and the medical term base, calculate matching degrees between the text data in the medical history information and the multiple medical terms, and extract text data that reaches a preset matching degree; perform word segmentation on the medical history information based on the matched text data to obtain multiple pieces of text data; and perform vectorization on the multiple pieces of text data obtained after the word segmentation to obtain the multiple text vectors.

A computer device includes a memory and a processor, where the memory stores at least one computer readable instruction, and the computer readable instruction is loaded by the processor to perform the following steps:

receiving a medical data classification request sent by a terminal, where the medical data classification request includes medical history information;

obtaining a preset medical term base, and performing word segmentation on the medical history information based on medical terms in the medical term base to obtain multiple text vectors;

performing feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values;

obtaining a target classifier, and performing traversal calculation on the multiple text vectors

and the corresponding feature dimension values by using multiple neural network nodes of the target classifier, where the target classifier is obtained by training multiple pieces of medical data;

when a target node corresponding to each of the multiple text vectors is traversed, calculating a category probability corresponding to the text vector based on the target node, and obtaining a category result corresponding to the medical history information based on the category probability; and

pushing the category result corresponding to the medical history information to the terminal.

A non-volatile computer readable storage medium, where the computer readable storage medium stores at least one computer readable instruction, and the computer readable instruction is loaded by a processor to perform the following steps:

receiving a medical data classification request sent by a terminal, where the medical data classification request includes medical history information;

obtaining a preset medical term base, and performing word segmentation on the medical history information based on medical terms in the medical term base to obtain multiple text vectors;

performing feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values;

obtaining a target classifier, and performing traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the target classifier, where the target classifier is obtained by training multiple pieces of medical data;

when a target node corresponding to each of the multiple text vectors is traversed, calculating a category probability corresponding to the text vector based on the target node, and obtaining a category result corresponding to the medical history information based on the category probability; and

pushing the category result corresponding to the medical history information to the terminal.

The details of one or more embodiments of the present application are set forth in the following accompanying drawings and descriptions. Other features and advantages of the present application will become apparent from the specification, accompanying drawings, and claims.

BRIEF DESCRIPTION OF DRAWINGS

The following briefly describes the accompanying drawings required for describing the embodiments. Clearly, the accompanying drawings in the following description show merely some

embodiments of the present application, and a person of ordinary skill in the art may derive other drawings from these accompanying drawings without creative efforts.

FIG. 1 is an application scenario diagram of a method for classifying medical data based on machine learning according to an embodiment;

FIG. 2 is a schematic flowchart of a method for classifying medical data based on machine learning according to an embodiment;

FIG. 3 is a schematic flowchart of a step of performing word segmentation on medical history information according to an embodiment;

FIG. 4 is a schematic flowchart of a step of constructing a target classifier according to an embodiment;

FIG. 5 is a structural block diagram of an apparatus for classifying medical data based on machine learning according to an embodiment; and

FIG. 6 is a diagram of an internal structure of a computer device according to an embodiment.

DESCRIPTION OF EMBODIMENTS

To make the technical solutions and advantages of the present application clearer and more comprehensible, the following further describes the present application in detail with reference to the accompanying drawings and embodiments. It should be understood that the specific embodiments described herein are merely used to explain the present application but are not intended to limit the present application.

A method for classifying medical data based on machine learning provided in the present application may be applied to an application environment shown in FIG. 1. A terminal 102 communicates with a server 104 over a network. A medical worker may send a medical data classification request to the server 104 by using the corresponding terminal 102. The medical data classification request includes medical history information. After receiving the medical data classification request sent by the terminal 102, the server 104 performs word segmentation on the medical history information to obtain multiple text vectors. The server 104 further performs feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values. The server 104 further obtains a target classifier, where the target classifier is obtained by training multiple pieces of medical data; and performs classification analysis on the obtained multiple text vectors and corresponding feature dimension values by using multiple neural network nodes of the target classifier, so as to effectively obtain a category result

corresponding to the medical history information. In addition, the server 104 pushes the category result corresponding to the medical history information to the corresponding terminal 102. Effective word segmentation and feature extraction are performed on the medical history information, and the extracted text data is classified by using a pre-trained and pre-constructed classifier, thereby effectively improving classification accuracy of the medical history information. The terminal 102 may be but is not limited to various personal computers, laptops, smartphones, tablets, and portable wearable devices, and the server 104 may be implemented by a standalone server or a server cluster including multiple servers.

In an embodiment, as shown in FIG. 2, a method for classifying medical data based on machine learning is provided. An example in which the method is applied to the server in FIG. 1 is used for description. The method includes the following steps:

Step 202: Receive a medical data classification request sent by a terminal, where the medical data classification request includes medical history information.

The medical history information may include an identity, basic information, medical history record information, historical diagnosis information, and the like of a patient. When diagnosing the patient, a medical worker may obtain the medical history information of the patient by using the corresponding terminal. The medical history information may include information entered by the medical worker, or may include medical history information obtained from a database based on the identity of the patient. After obtaining the medical history information of the patient, the terminal sends the medical data classification request to the server based on the medical history information. The medical data classification request includes the medical history information and the identity.

Further, the server may further obtain, from a third-party database, historical medical history information of the patient based on the identity of the patient, for example, medical history information of the patient about hospitalization in another place, so as to effectively obtain complete medical history information corresponding to the patient.

Step 204: Obtain a preset medical term base, and perform word segmentation on the medical history information based on medical terms in the medical term base to obtain multiple text vectors.

Before performing word segmentation on the medical history information, the server may obtain a large amount of medical data, and perform semantic analysis on the large amount of medical data obtained. For example, the server may perform semantic analysis on the large amount of medical data by using a preset semantic analysis model, to obtain multiple types of medical

terms. The server further generates a medical term base corresponding to each type in the medical field by using the medical terms obtained by analysis.

After receiving the medical data classification request sent by the terminal, the server performs word segmentation on the medical history information. Specifically, the server obtains the preset medical term base, where the medical term base includes a large quantity of medical terms and corresponding vectors. The server performs matching between multiple pieces of text data in the medical history information and multiple medical terms in the medical term base. Specifically, the server may calculate the similarities between the text data in the medical history information and the medical terms by using a preset distance algorithm, so as to further calculate matching degrees between the text data in the medical history information and the medical terms. The server further extracts text data that reaches a preset matching degree. The server performs word segmentation on the medical history information based on the matched text data to obtain multiple pieces of text data. The server further performs vectorization on the multiple pieces of text data obtained after word segmentation, and converts the text data into corresponding vectorization information to obtain multiple text vectors corresponding to the multiple pieces of text data.

Step 206: Perform feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values.

After performing word segmentation on the text vectors corresponding to the medical history information to obtain the multiple text vectors, the server further performs feature extraction on the text data. The server calculates, based on a preset algorithm, weights of the multiple text vectors obtained after word segmentation. For example, the server may calculate term frequency (TF) values and inverse document frequency (IDF) values of the multiple text vectors by using an TF-IDF algorithm. TF indicates the occurrence frequency of a text vector in a document. IDF is a measure of universal importance of a term. In addition, the server may calculate multiple corresponding weights based on TF values and IDF values of multiple terms. For example, the server may obtain the weights corresponding to the text vectors by calculating products of the TF values and the IDF values. Further, the server performs feature extraction on the text vectors based on the weights of the text vectors to extract a text vector whose weight reaches a preset threshold.

After extracting the text vector whose weight reaches the preset threshold, the server calculates the feature dimension value of the text vector based on the preset algorithm and the weight of the text vector. A feature dimension value may represent a feature dimension to which a text vector belongs. By calculating the weights of the text vectors to filter the text vectors by weight,

feature extraction can be effectively performed on the text vectors and the feature dimension values corresponding to the text vectors can be obtained.

Step 208: Obtain a target classifier, and perform traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the target classifier, where the target classifier is obtained by training multiple pieces of medical data.

Step 210: When a target node corresponding to each of the multiple text vectors is traversed, calculate a category probability corresponding to the text vector based on the target node, and obtain a category result corresponding to the medical history information based on the category probability.

Before obtaining the target classifier, the server may further pre-construct and pre-train the target classifier. Specifically, the server may obtain a large amount of medical data from a local database or a third-party database in advance, and generate corresponding training set data and verification set data based on the multiple pieces of medical data. The server performs vectorization on multiple fields of data corresponding to the medical data to obtain feature vectors corresponding to the multiple pieces of text data, and converts the feature vectors into corresponding feature variables. The server further performs clustering analysis on feature variables corresponding to the training set data by using a preset clustering algorithm, to extract feature variables that reach a preset threshold. Then, the server obtains a preset neural network model, trains the training set data by using the neural network model to obtain feature dimension values and weights corresponding to the multiple feature variables, and constructs an initial classifier based on the feature dimension values and the weights corresponding to the multiple feature variables. The server performs further training and verification on the classifier by using the verification set data, and when the quantity of pieces of data that meets a preset threshold in the verification set data reaches a preset proportion, stops training to obtain a required target classifier.

After the server performs feature extraction on the text data to obtain multi-dimensional vectors corresponding to the multiple pieces of text data, the server obtains the trained target classifier, and inputs the multiple text vectors and the corresponding dimension feature values into the target classifier. The target classifier includes multiple preset neural network layer nodes and corresponding node weights. Traversal calculation is performed on the multiple text vectors and the corresponding dimension feature values by using preset loss functions of the multiple nodes in the target classifier. When the target node corresponding to each of the multiple text vectors is

obtained, the category probability corresponding to the text vector is calculated based on the target node, and a category result corresponding to the text vector is obtained based on the category probability, so as to obtain the category result corresponding to the medical history information.

Step 212: Push the category result corresponding to the medical history information to the terminal.

After the server classifies the medical history information by using the target classifier to obtain the category result corresponding to the medical history information, the server pushes the category result corresponding to the medical history information to the corresponding terminal. By performing effective word segmentation and feature extraction on the medical history information, and classifying the extracted text information by using the pre-trained and pre-constructed target classifier, classification accuracy of medical history information can be effectively improved. In this way, the medical worker can perform effective diagnosis based on the pushed category result corresponding to the medical history information, thereby effectively improving diagnosis efficiency of medical workers.

For example, the medical history information includes historical medical history information corresponding to the patient, which includes data such as multiple historical symptom descriptions, historical prescription information, and historical diagnosis information. After the medical history information is filtered multiple times and text is extracted, classification analysis is performed on the extracted text by using the pre-trained target classifier. After classification analysis is performed on all data in the medical history information of the patient, the category result corresponding to the medical history information is obtained. For example, when the patient is suffering from a cancer, a specific cancer category may be obtained.

In the foregoing method for classifying medical data based on machine learning, after receiving the medical data classification request sent by the terminal, the server performs word segmentation on the medical history information carried in the medical data classification request, so as to effectively obtain the multiple text vectors by performing word segmentation based on a medical field. The server further performs feature extraction on the multiple text vectors to effectively obtain the multiple text vectors and the corresponding feature dimension values. The server further obtains the target classifier. The target classifier is obtained by training the multiple pieces of medical data. Then, traversal calculation is performed on the multiple text vectors and the corresponding feature dimension values by using the multiple neural network nodes of the target classifier. When the target node corresponding to each of the multiple text vectors is traversed, the category probability corresponding to the text vector is calculated based on the target

node, and the category result corresponding to the medical history information is obtained based on the category probability. In this way, the category result corresponding to the medical history information can be effectively obtained, and the extracted text data is classified by using the pre-trained and pre-constructed classifier, thereby effectively improving classification accuracy of the medical history information. The server pushes the category result corresponding to the medical history information to the terminal. Therefore, the medical worker can effectively make a decision based on the pushed category result corresponding to the medical history information, and accurately classify the medical history information, thereby effectively improving medical data processing efficiency.

In an embodiment, as shown in FIG. 3, the medical history information includes multiple pieces of text data, and the step of performing word segmentation on the medical history information includes the following:

Step 302: Obtain the preset medical term base, where the medical term base includes multiple medical terms; and perform matching between the multiple pieces of text data in the medical history information and the medical term base, calculate matching degrees between the text data in the medical history information and the multiple medical terms, and extract text data that reaches a preset matching degree.

Step 304: Perform word segmentation on the medical history information based on the matched text data to obtain multiple pieces of text data.

Step 306: Perform vector conversion on the multiple pieces of text data obtained after the word segmentation to obtain the multiple text vectors.

Before processing the medical data, the server can set up a medical term base in advance. Specifically, the server may obtain a large amount of medical data, and perform semantic analysis on the large amount of medical data obtained. For example, the server may perform semantic analysis on the large amount of medical data by using a preset semantic analysis model, to obtain multiple types of medical terms. The server further generates a medical term base corresponding to each type in the medical field by using the medical terms obtained by analysis.

A medical worker may send a medical data classification request to the server by using a corresponding terminal. The medical data classification request includes medical history information. After receiving the medical data classification request sent by the terminal, the server performs word segmentation on the medical history information. Specifically, the server obtains the preset medical term base, where the medical term base includes a large quantity of medical

terms and corresponding vectors. The server performs matching between multiple pieces of text data in the medical history information and multiple medical terms in the medical term base. Specifically, the server may calculate a similarity between the text data in the medical history information and the medical terms by using a preset distance algorithm, so as to further calculate matching degrees between the text data in the medical history information and the medical terms. The server further extracts text data that reaches a preset matching degree. The server performs word segmentation on the medical history information based on the matched text data to obtain multiple pieces of text data.

The server further performs vectorization on the multiple pieces of text data obtained after word segmentation, and converts the text data into corresponding vectorization information to obtain multiple text vectors corresponding to the multiple pieces of text data. For example, Doc2Vec and Word2Vec algorithms can be used to perform word vectorization and paragraph vectorization on the multiple pieces of text data obtained after word segmentation, so as to obtain corresponding text vectors. The text vectors may include a word vector, a phrase vector, and a sentence vector.

After obtaining the text vectors corresponding to the multiple pieces of text data, the server calculates feature dimension values of the text vectors based on a preset algorithm, and performs feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values. The server further obtains a preset classifier; and performs classification analysis on the multiple text vectors and the corresponding feature dimension values by using the classifier, so as to effectively obtain a category result corresponding to the medical history information. In addition, the server pushes the category result corresponding to the medical history information to the corresponding terminal. By performing effective word segmentation and feature extraction on the medical history information, and classifying the extracted text information by using the pre-trained and pre-constructed classifier, classification accuracy of medical history information can be effectively improved. In this way, the medical worker can perform effective diagnosis based on the pushed category result corresponding to the medical history information.

In an embodiment, the step of performing feature extraction on the multiple pieces of text data to obtain multi-dimensional vectors corresponding to the multiple text vectors includes: calculating a term frequency and an inverse document frequency of the multiple text vectors; calculating weights of the multiple text vectors by using a preset algorithm based on the term frequency and the inverse document frequency; extracting a text vector whose weight reaches a preset threshold; and calculating a feature dimension value corresponding to the text vector based

on a preset algorithm and the weight.

A medical worker may send a medical data classification request to the server by using a corresponding terminal. The medical data classification request includes medical history information. After receiving the medical data classification request sent by the terminal, the server performs word segmentation on the medical history information in the medical data classification request to obtain multiple text vectors.

After obtaining the multiple text vectors corresponding to the medical history information, the server calculates, based on a preset algorithm, weights of the multiple text vectors obtained after word segmentation. For example, the server may calculate TF values and IDF values of the multiple text vectors by using an TF-IDF algorithm. TF indicates the occurrence frequency of a text vector. IDF may represent a measure of universal importance of a term. In addition, the server may calculate multiple corresponding weights based on TF values and IDF values of multiple terms. For example, the server may obtain weights corresponding to the text data by calculating products of the TF values and the IDF values.

For example, the server may calculate TF values of the multiple text vectors by using the following formula:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

A formula for calculating IDF values of the text vectors may be as follows:

$$idf_i = \lg \frac{|D|}{\left| \left\{ j : t_i \in d_j \right\} \right|}$$

A formula for calculating the weights of the text vectors may be as follows:

$$tfidf_{i,j} = tf_{i,j} \times idf_{i,j}$$

If fewer documents contain a text vector t , n is smaller and an IDF is higher. In this case, the text vector t has a good category differentiation capability. If m documents in a certain category of documents C contain a term t , and the total quantity of documents containing t in other categories of documents is k , it is clear that the total quantity of documents containing t is $n = m + k$. Therefore, a larger m leads to a larger n , and an IDF value obtained by using the IDF formula is smaller. That

is, a category differentiation capability of the term t is poor. If a term occurs frequently in a category of documents, it indicates that the term can well represent a feature of this category of text, and the term has a relatively large weight. By calculating the product of the TF and the IDF and further calculating the weights of the text vectors, the server further performs feature extraction on the text vectors based on the weights of the text vectors to extract a text vector whose weight reaches a preset threshold.

After extracting the text vector whose weight reaches the preset threshold, the server calculates the feature dimension values of the multiple text vectors based on the preset algorithm and the weights of the text vectors. A feature dimension value may represent a feature dimension to which a text vector belongs. The text vector may include multiple feature dimensions. After calculating the weights of the text vectors, the server may use the weights to calculate the importance of the feature dimensions of the text vectors, so as to obtain the feature dimension values corresponding to the text vectors. By calculating the weights of the text vectors to filter the text vectors by weight, feature extraction can be effectively performed on the text vectors and the feature dimension values corresponding to the text vectors can be obtained.

In an embodiment, as shown in FIG. 4, before the target classifier is obtained, the method further includes a step of constructing the target classifier. This step specifically includes the following content:

Step 402: Obtain multiple pieces of medical data, and generate corresponding training set data and verification set data based on the multiple pieces of medical data.

The target classifier further needs to be constructed and trained before being obtained. Specifically, the server may obtain a large amount of medical data from a local database or a third-party database in advance. The medical data may include medical diagnosis information, clinical data, research data, and the like. Then, the server generates training set data and verification set data based on the large amount of medical data. The training set data may be manually labeled data.

Step 404: Perform clustering analysis on multiple pieces of medical data in the training set data to obtain a clustering result.

Step 406: Perform feature extraction on the clustering result to extract multiple feature variables.

Step 408: Obtain a preset neural network model, train the training set data by using the neural network model to obtain feature dimension values and weights corresponding to the multiple

feature variables, and construct an initial classifier based on the feature dimension values and the weights corresponding to the multiple feature variables.

Step 410: Perform further training and verification on the classifier by using the verification set data, and when the quantity of pieces of data that meets a preset threshold in the verification set data reaches a preset proportion, stop training to obtain a required target classifier.

The server first performs data cleansing and data preprocessing on medical data in the training set data. Specifically, the server performs vectorization on multiple fields of data corresponding to the medical data to obtain feature vectors corresponding to the multiple pieces of text data, and converts the feature vectors into corresponding feature variables. The server further performs derivation processing on the feature variables to obtain multiple processed feature variables. For example, the server fills missing values in the feature variables, extracts abnormal values from the feature variables and replaces them, and the like.

The server further performs clustering analysis on feature variables corresponding to the training set data by using a preset clustering algorithm. For example, the preset clustering algorithm may be the k-means (k-means algorithm) clustering method. The server obtains multiple clustering results by performing clustering on the feature variables multiple times. The server calculates similarities between multiple feature variables based on a preset algorithm, and extracts feature variables whose similarities reach a preset threshold.

For example, the server may combine feature variables in the multiple clustering results to obtain multiple combined feature variables. The server may obtain a target variable and use the target variable to perform correlation verification on the multiple combined feature variables. When the verification succeeds, the server may add interactive labels to the combined feature variables. The server may parse the corresponding feature variables by using the combined feature variables with the interactive labels added. The combined feature variables with the interactive labels added may be the feature variables whose similarity reaches the preset threshold. In this case, the server extracts the feature variables whose similarities reach the preset threshold. By performing feature processing and feature extraction on the feature variables, valuable feature variables can be extracted effectively.

The server obtains a preset machine learning model, for example, the decision tree-based Xgboost machine learning model. For example, the machine learning model includes multiple neural network models. A neural network model may include an input layer, multiple Long Short Term Memory (LSTM) layers, a dropout layer, and an output layer that are preset. The neural

network model includes multiple network nodes. A drop rate of each layer of network node may be 0.2. The LSTM layers of the neural network model include an activation function and a loss function. A fully connected artificial neural network that is output through an LSTM layer also includes a corresponding activation function. The neural network model further includes a calculation method for determining an error, for example, a mean square error algorithm, and further includes an iteration update method for determining a weight parameter, for example, the RMSprop algorithm. The neural network model may further include a common neural network layer, used for output result dimension reduction.

After obtaining the preset neural network model, the server further inputs the medical data in the training set data into the neural network model for learning and training. After training a large amount of medical data in the training set, the server may obtain the feature dimension values and the weights corresponding to the multiple feature variables, and constructs the initial classifier based on the feature dimension values and the weights corresponding to the multiple feature variables.

After obtaining the initial classifier, the server obtains the verification set data, and trains and verifies the constructed initial classifier by using a large amount of medical data in the verification set data. When the quantity of pieces of data that meets a preset threshold in the verification set data reaches a preset proportion, the server stops training to obtain a trained target classifier. By training and learning a large amount of medical data, a classifier with high prediction accuracy can be constructed effectively, thereby effectively improving classification accuracy of medical data.

In an embodiment, the text includes multiple text sentences, the multiple text sentences form a text block, and the step of performing traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the classifier to obtain the category corresponding to the multiple text vectors includes: calculating a correlation between the multiple text vectors by using the target classifier based on the feature dimension values, calculating the text sentences formed in the text based on the correlation, and calculating sentence vectors of the text sentences; extracting features of the sentence vectors, and calculating a text block vector based on the features of the multiple sentence vectors; and calculating a probability of each category corresponding to the text block vector, extracting a category whose probability reaches a preset probability value, and adding a corresponding category label to the text block.

A medical worker may send a medical data classification request to the server by using a corresponding terminal. The medical data classification request includes medical history

information. After receiving the medical data classification request sent by the terminal, the server performs word segmentation on the medical history information in the medical data classification request to obtain multiple text vectors corresponding to multiple pieces of text data. The server further performs feature extraction on the text vectors to obtain multiple text vectors and corresponding feature dimension values.

After extracting the multiple text vectors and the corresponding feature dimension values, the server obtains a target classifier and uses the multiple text vectors and the corresponding feature dimension values as input of the target classifier. The target classifier includes multiple preset neural network layer nodes and corresponding node weights. Traversal calculation is performed on the multiple text vectors and the corresponding feature dimension values by using the multiple neural network layer nodes of the target classifier. Specifically, the text may include multiple words and phrases, that is, text sentences. The text vectors may include word vectors and phrase vectors. The server may first calculate a correlation between multiple text vectors in the text based on the text vectors and the corresponding dimension feature values, then calculate the text sentences formed in the text based on the correlation, and calculate sentence vectors corresponding to the text sentences. Then, the server extracts features of the sentence vectors, and calculates a text block vector based on the features of the multiple sentence vectors. The text block includes multiple text sentences, and the text block vector may include multiple sentence vectors. The server calculates a probability that a text block vector belongs to each category by using the preset loss functions of the multiple neural network layer nodes, and inputs multiple text block vectors into the next neural network layer node based on the category probability for calculation, until a target node corresponding to the multiple text block vectors is obtained. Further, category probabilities corresponding to the multiple text block vectors are obtained through calculation based on the target node, and a category result with the highest category probability is obtained. In this way, the category result to which the multiple text block vectors belong is obtained. The text vectors in the medical history information are classified through the target classifier obtained by training the large amount of data. In this way, a category of the medical history information can be effectively and accurately obtained, thereby improving classification accuracy of the medical history information.

In an embodiment, the method further includes obtaining multiple pieces of historical medical data from a preset database based on a preset frequency; performing clustering analysis on the multiple pieces of historical medical data to obtain an analysis result; performing feature selection based on the analysis result to obtain multiple feature variables; calculating weights of the multiple

feature variables based on a preset algorithm; and performing optimization adjustment on the classifier based on the multiple feature variables and the corresponding weights.

After obtaining the target classifier through training, the server may further perform optimization adjustment on the classifier based on the preset frequency. Specifically, the server may obtain a large amount of historical medical data from a local database or a third-party database. For example, the preset frequency may be one month, three months, or sixth months. Then, the server may obtain historical medical data in the past one month, three months, or sixth months. The historical medical data may include medical diagnosis information, clinical data, research data, and the like.

The server first performs data cleansing and data preprocessing on the large amount of obtained historical medical data. Specifically, the server performs vectorization on multiple fields of data corresponding to the historical medical data to obtain feature variables corresponding to the multiple fields of data, and performs derivation processing on the feature variables to obtain multiple processed feature variables. For example, the server fills missing values in the feature variables, extracts abnormal values from the feature variables and replaces them, and the like.

The server further performs clustering analysis on feature variables corresponding to the training set data by using a preset clustering algorithm. For example, the preset clustering algorithm may be the k-means (k-means algorithm) clustering method. The server obtains multiple clustering results by performing clustering on the feature variables multiple times. The server calculates similarities between multiple feature variables based on a preset algorithm, and extracts feature variables whose similarities reach a preset threshold.

For example, the server may combine feature variables in the multiple clustering results to obtain multiple combined feature variables. The server may obtain a target variable and use the target variable to perform correlation verification on the multiple combined feature variables. When the verification succeeds, the server may add interactive labels to the combined feature variables. The server may parse the corresponding feature variables by using the combined feature variables with the interactive labels added. The combined feature variables with the interactive labels added may be the feature variables that reach the preset threshold. In this case, the server extracts the feature variables that reach the preset threshold. By performing feature processing and feature extraction on the feature variables, valuable feature variables can be extracted effectively.

The server further calculates weights of the multiple feature variables based on a preset algorithm, and further performs optimization adjustment on the target classifier based on the

multiple feature variables and the corresponding weights. Specifically, the server may adjust parameters of the target classifier based on the multiple feature variables and the corresponding weights, thereby effectively performing parameter adjustment and optimization on the target classifier.

It should be understood that although the steps in the flowcharts in FIG. 2 to FIG. 4 are shown in a sequence indicated by arrows, these steps are not necessarily performed in the sequence indicated by the arrows. Unless expressly specified in this specification, these steps are not performed in a strict sequence, and can be executed in another sequence. In addition, at least some steps in FIG. 2 to FIG. 4 may include multiple substeps or multiple stages that are not necessarily performed at the same time, but may be performed at different time. These substeps or stages are not necessarily performed in sequence, but may be performed in turn or alternately with at least some of other steps or substeps or stages of other steps.

In an embodiment, as shown in FIG. 5, an apparatus for classifying medical data based on machine learning is provided, including a request receiving module 502, a word segmentation module 504, a feature extraction module 506, a data classification module 508, and a data pushing module 510, where

the request receiving module 502 is configured to receive a medical data classification request sent by a terminal, where the medical data classification request includes medical history information;

the word segmentation module 504 is configured to obtain a preset medical term base, and perform word segmentation on the medical history information based on medical terms in the medical term base to obtain multiple text vectors;

the feature extraction module 506 is configured to perform feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values;

the data classification module 508 is configured to obtain a target classifier, and perform traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the target classifier, where the target classifier is obtained by training multiple pieces of medical data; and when a target node corresponding to each of the multiple text vectors is traversed, calculate a category probability corresponding to the text vector based on the target node, and obtain a category result corresponding to the medical history information based on the category probability; and

the data pushing module 510 is configured to push the category result corresponding to the medical history information to the terminal.

In an embodiment, the medical history information includes multiple pieces of text data, and the word segmentation module 504 is further configured to obtain the preset medical term base, where the medical term base includes multiple medical terms; and perform matching between the multiple pieces of text data in the medical history information and the medical term base, calculate matching degrees between the text data in the medical history information and the multiple medical terms, and extract text data that reaches a preset matching degree; perform word segmentation on the medical history information based on the matched text data to obtain multiple pieces of text data; and perform vectorization on the multiple pieces of text data obtained after the word segmentation to obtain the multiple text vectors.

In an embodiment, the feature extraction module 506 is further configured to calculate a term frequency and an inverse document frequency of the multiple text vectors; calculate weights of the multiple text vectors by using a preset algorithm based on the term frequency and the inverse document frequency; extract a text vector whose weight reaches a preset threshold; and calculate a feature dimension value corresponding to the text vector based on a preset algorithm and the weight.

In an embodiment, the apparatus further includes a target classifier construction module, configured to: obtain multiple pieces of medical data, and generate corresponding training set data and verification set data based on the multiple pieces of medical data; perform clustering analysis on multiple pieces of medical data in the training set data to obtain a clustering result; perform feature extraction on the clustering result to extract multiple feature variables; obtain a preset neural network model, train the training set data by using the neural network model to obtain feature dimension values and weights corresponding to the multiple feature variables, and construct an initial classifier based on the feature dimension values and the weights corresponding to the multiple feature variables; and perform further training and verification on the classifier by using the verification set data, and when the quantity of pieces of data that meets a preset threshold in the verification set data reaches a preset proportion, stop training to obtain a required target classifier.

In an embodiment, the text includes multiple text sentences, the multiple text sentences form a text block, and the data classification module 508 is further configured to calculate a correlation between the multiple text vectors by using the target classifier based on the feature dimension values, calculate the text sentences formed in the text based on the correlation, and calculate

sentence vectors of the text sentences; extract features of the sentence vectors, and calculate a text block vector based on the features of the multiple sentence vectors; and calculate a probability of each category corresponding to the text block vector, extract a category whose probability reaches a preset probability value, and add a corresponding category label to the text block.

In an embodiment, the apparatus further includes a target classifier optimization module, configured to obtain multiple pieces of historical medical data from a preset database based on a preset frequency; perform clustering analysis on the multiple pieces of historical medical data to obtain an analysis result; perform feature selection based on the analysis result to obtain multiple feature variables; calculate weights of the multiple feature variables based on a preset algorithm; and perform optimization adjustment on the target classifier based on the multiple feature variables and the corresponding weights.

For specific limitations of the apparatus for classifying medical data based on machine learning, refer to the limitations on the method for classifying medical data based on machine learning. Details are omitted herein for simplicity. The modules in the apparatus for classifying medical data based on machine learning can be implemented in whole or in part by software, hardware, or a combination thereof. The modules may be built in or independent of a processor in a computer device in a hardware form, or may be stored in a memory in a computer device in a software form, so that the processor invokes and executes operations corresponding to the modules.

In an embodiment, a computer device is provided, which may be a server. FIG. 6 shows a diagram of an internal structure of the computer device. The computer device includes a processor, a memory, a network interface, and a database that are connected through a system bus. The processor of the computer device is configured to provide computing and control capabilities. The memory of the computer device includes a non-volatile storage medium and an internal storage. The non-volatile storage medium stores an operating system, a computer readable instruction, and a database. The internal storage provides an environment for running the operating system and the computer readable instruction in the non-volatile storage medium. The database of the computer device is configured to store medical data, medical history information, and other data. The network interface of the computer device is configured to communicate with an external terminal through a network connection. The computer readable instruction is executed by the processor to implement the steps of the method for classifying medical data based on machine learning provided in any of the embodiments of the present application.

A person skilled in the art may understand that the structure shown in FIG. 6 is merely a block diagram of a partial structure related to the solutions of the present application, and does not

constitute a limitation on the computer device to which the solutions of the present application are applied. A specific computer device may include more or fewer components than those shown in the figure, or combine some components, or have different component arrangements.

A person of ordinary skill in the art may understand that all or some of the procedures for implementing the methods in the foregoing embodiments may be completed by the computer readable instruction by instructing related hardware. The computer readable instruction may be stored in a non-volatile computer readable storage medium. When executed, the computer readable instruction may include the procedures in the embodiments of the foregoing methods. Any reference to the memory, the storage, the database, or other media used in the embodiments provided in the present application may include a non-volatile memory and/or a volatile memory. The non-volatile memory may include a read-only memory (ROM), a programmable ROM (PROM), an electrically programmable ROM (EPROM), an electrically erasable programmable ROM (EEPROM), or a flash memory. The volatile memory may include a random access memory (RAM) or an external cache memory. As an illustration rather than a limitation, the RAM is available in a variety of forms, such as a static RAM (SRAM), a dynamic RAM (DRAM), a synchronous DRAM (SDRAM), a double data rate SDRAM (DDRSDRAM), an enhanced SDRAM (ESDRAM), a synchronous link (Synchlink) DRAM (SLDRAM), a RAM bus (Rambus) direct RAM (RDRAM), a direct Rambus dynamic RAM (DRDRAM), and a Rambus dynamic RAM (RDRAM).

The technical features of the foregoing embodiments may be randomly combined. For brevity of description, not all possible combinations of the technical features of the foregoing embodiments are described. However, provided that the combinations of these technical features are not contradictory, it should be considered that they fall within the scope of this specification.

The foregoing embodiments represent only several implementations of the present application, and descriptions thereof are relatively specific and detailed, but cannot be construed as a limitation on the scope of the present invention. It should be noted that a person of ordinary skill in the art can further make some improvements and variations without departing from the idea of the present application, and the improvements and variations shall fall within the protection scope of the present application. Therefore, the protection scope of the present application shall be subject to the appended claims.

CLAIMS

1. A method for classifying medical data based on machine learning, wherein the method comprises:

receiving a medical data classification request sent by a terminal, wherein the medical data classification request comprises medical history information;

obtaining a preset medical term base, and performing word segmentation on the medical history information based on medical terms in the medical term base to obtain multiple text vectors;

performing feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values;

obtaining a target classifier, and performing traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the target classifier, wherein the target classifier is obtained by training multiple pieces of medical data;

when a target node corresponding to each of the multiple text vectors is traversed, calculating a category probability corresponding to the text vector based on the target node, and obtaining a category result corresponding to the medical history information based on the category probability; and

pushing the category result corresponding to the medical history information to the terminal.

2. The method according to claim 1, wherein the medical history information comprises multiple pieces of text data, and the step of performing word segmentation on the medical history information comprises:

obtaining the preset medical term base, wherein the medical term base comprises multiple medical terms; and performing matching between the multiple pieces of text data in the medical history information and the medical term base, calculating matching degrees between the text data in the medical history information and the multiple medical terms, and extracting text data that reaches a preset matching degree;

performing word segmentation on the medical history information based on the matched text data to obtain multiple pieces of text data; and

performing vector conversion on the multiple pieces of text data obtained after the word segmentation to obtain the multiple text vectors.

3. The method according to claim 1, wherein the step of performing feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values comprises:

calculating a term frequency and an inverse document frequency of the multiple text vectors;

calculating weights of the multiple text vectors by using a preset algorithm based on the term frequency and the inverse document frequency;

extracting a text vector whose weight reaches a preset threshold; and

calculating a feature dimension value corresponding to the text vector based on a preset algorithm and the weight.

4. The method according to claim 1, wherein a step of constructing the target classifier comprises:

obtaining multiple pieces of medical data, and generating corresponding training set data and verification set data based on the multiple pieces of medical data;

performing clustering analysis on multiple pieces of medical data in the training set data to obtain a clustering result;

performing feature extraction on the clustering result to extract multiple feature variables;

obtaining a preset neural network model, training the training set data by using the neural network model to obtain feature dimension values and weights corresponding to the multiple feature variables, and constructing an initial classifier based on the feature dimension values and the weights corresponding to the multiple feature variables; and

performing further training and verification on the classifier by using the verification set data, and when the quantity of pieces of data that meets a preset threshold in the verification set data reaches a preset proportion, stopping training to obtain a required target classifier.

5. The method according to any one of claims 1 to 4, wherein the text comprises multiple text sentences, the multiple text sentences form a text block, and the step of performing traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the target classifier to obtain the category corresponding to the multiple text vectors comprises:

calculating a correlation between the multiple text vectors by using the target classifier based on the feature dimension values, calculating the text sentences formed in the text based on the

correlation, and calculating sentence vectors of the text sentences;

extracting features of the sentence vectors, and calculating a text block vector based on the features of the multiple sentence vectors; and

calculating a probability of each category corresponding to the text block vector, extracting a category whose probability reaches a preset probability value, and adding a corresponding category label to the text block.

6. The method according to claim 1, wherein the method further comprises:

obtaining multiple pieces of historical medical data from a preset database based on a preset frequency;

performing clustering analysis on the multiple pieces of historical medical data to obtain an analysis result;

performing feature selection based on the analysis result to obtain multiple feature variables;

calculating weights of the multiple feature variables based on a preset algorithm; and

performing optimization adjustment on the target classifier based on the multiple feature variables and the corresponding weights.

7. An apparatus for classifying medical data based on machine learning, wherein the apparatus comprises:

a request receiving module, configured to receive a medical data classification request sent by a terminal, wherein the medical data classification request comprises medical history information;

a word segmentation module, configured to obtain a preset medical term base, and perform word segmentation on the medical history information based on medical terms in the medical term base to obtain multiple text vectors;

a feature extraction module, configured to perform feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values;

a data classification module, configured to obtain a target classifier, and perform traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the target classifier, wherein the target classifier is obtained by training multiple pieces of medical data; and when a target node corresponding to each of the

multiple text vectors is traversed, calculate a category probability corresponding to the text vector based on the target node, and obtain a category result corresponding to the medical history information based on the category probability; and

a data pushing module, configured to push the category result corresponding to the medical history information to the terminal.

8. The apparatus according to claim 7, wherein the word segmentation module is further configured to obtain the preset medical term base, wherein the medical term base comprises multiple medical terms; and perform matching between the multiple pieces of text data in the medical history information and the medical term base, calculate matching degrees between the text data in the medical history information and the multiple medical terms, and extract text data that reaches a preset matching degree; perform word segmentation on the medical history information based on the matched text data to obtain multiple pieces of text data; and perform vectorization on the multiple pieces of text data obtained after the word segmentation to obtain the multiple text vectors.

9. The apparatus according to claim 7, wherein the feature extraction module is further configured to calculate a term frequency and an inverse document frequency of the multiple text vectors; calculate weights of the multiple text vectors by using a preset algorithm based on the term frequency and the inverse document frequency; extract a text vector whose weight reaches a preset threshold; and calculate a feature dimension value corresponding to the text vector based on a preset algorithm and the weight.

10. The apparatus according to claim 7, wherein the apparatus further comprises a classifier construction module, configured to: obtain multiple pieces of medical data, and generate corresponding training set data and verification set data based on the multiple pieces of medical data; perform clustering analysis on multiple pieces of medical data in the training set data to obtain a clustering result; perform feature extraction on the clustering result to extract multiple feature variables; obtain a preset neural network model, train the training set data by using the neural network model to obtain feature dimension values and weights corresponding to the multiple feature variables, and construct an initial classifier based on the feature dimension values and the weights corresponding to the multiple feature variables; and perform further training and verification on the classifier by using the verification set data, and when the quantity of pieces of data that meets a preset threshold in the verification set data reaches a preset proportion, stop training to obtain a required target classifier.

11. The apparatus according to claim 7, wherein the text comprises multiple text sentences, the multiple text sentences form a text block, and the data classification module is further configured to calculate a correlation between the multiple text vectors by using the target classifier based on the feature dimension values, calculate the text sentences formed in the text based on the correlation, and calculate sentence vectors of the text sentences; extract features of the sentence vectors, and calculate a text block vector based on the features of the multiple sentence vectors; and calculate a probability of each category corresponding to the text block vector, extract a category whose probability reaches a preset probability value, and add a corresponding category label to the text block.

12. The apparatus according to claim 7, wherein the apparatus further comprises a model optimization module, configured to obtain multiple pieces of historical medical data from a preset database based on a preset frequency; perform clustering analysis on the multiple pieces of historical medical data to obtain an analysis result; perform feature selection based on the analysis result to obtain multiple feature variables; calculate weights of the multiple feature variables based on a preset algorithm; and perform optimization adjustment on the target classifier based on the multiple feature variables and the corresponding weights.

13. A computer device, comprising a memory and a processor, wherein the memory stores at least one computer readable instruction, and the computer readable instruction is loaded by the processor to perform the following steps:

receiving a medical data classification request sent by a terminal, wherein the medical data classification request comprises medical history information;

obtaining a preset medical term base, and performing word segmentation on the medical history information based on medical terms in the medical term base to obtain multiple text vectors;

performing feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values;

obtaining a target classifier, and performing traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the target classifier, wherein the target classifier is obtained by training multiple pieces of medical data;

when a target node corresponding to each of the multiple text vectors is traversed, calculating a category probability corresponding to the text vector based on the target node, and obtaining a category result corresponding to the medical history information based on the category probability; and

pushing the category result corresponding to the medical history information to the terminal.

14. The computer device according to claim 13, wherein the medical history information comprises multiple pieces of text data, and when executing the computer readable instruction, the processor further performs the following steps: obtaining the preset medical term base, wherein the medical term base comprises multiple medical terms; and performing matching between the multiple pieces of text data in the medical history information and the medical term base, calculating matching degrees between the text data in the medical history information and the multiple medical terms, and extracting text data that reaches a preset matching degree; performing word segmentation on the medical history information based on the matched text data to obtain multiple pieces of text data; and performing vector conversion on the multiple pieces of text data obtained after the word segmentation to obtain the multiple text vectors.

15. The computer device according to claim 13, wherein when executing the computer readable instruction, the processor further performs the following steps: calculating a term frequency and an inverse document frequency of the multiple text vectors; calculating weights of the multiple text vectors by using a preset algorithm based on the term frequency and the inverse document frequency; extracting a text vector whose weight reaches a preset threshold; and calculating a feature dimension value corresponding to the text vector based on a preset algorithm and the weight.

16. The computer device according to claim 13, wherein the text comprises multiple text sentences, the multiple text sentences form a text block, and when executing the computer readable instruction, the processor further performs the following steps: calculating a correlation between the multiple text vectors by using the target classifier based on the feature dimension values, calculating the text sentences formed in the text based on the correlation, and calculating sentence vectors of the text sentences; extracting features of the sentence vectors, and calculating a text block vector based on the features of the multiple sentence vectors; and calculating a probability of each category corresponding to the text block vector, extracting a category whose probability reaches a preset probability value, and adding a corresponding category label to the text block.

17. A non-volatile computer readable storage medium, wherein the computer readable storage medium stores at least one computer readable instruction, and the computer readable instruction is loaded by a processor to perform the following steps:

receiving a medical data classification request sent by a terminal, wherein the medical data classification request comprises medical history information;

obtaining a preset medical term base, and performing word segmentation on the medical history information based on medical terms in the medical term base to obtain multiple text vectors;

performing feature extraction on the multiple text vectors to obtain multiple text vectors and corresponding feature dimension values;

obtaining a target classifier, and performing traversal calculation on the multiple text vectors and the corresponding feature dimension values by using multiple neural network nodes of the target classifier, wherein the target classifier is obtained by training multiple pieces of medical data;

when a target node corresponding to each of the multiple text vectors is traversed, calculating a category probability corresponding to the text vector based on the target node, and obtaining a category result corresponding to the medical history information based on the category probability; and

pushing the category result corresponding to the medical history information to the terminal.

18. The storage medium according to claim 17, wherein the medical history information comprises multiple pieces of text data, and when executing the computer readable instruction, the processor further performs the following steps: obtaining the preset medical term base, wherein the medical term base comprises multiple medical terms; and performing matching between the multiple pieces of text data in the medical history information and the medical term base, calculating matching degrees between the text data in the medical history information and the multiple medical terms, and extracting text data that reaches a preset matching degree; performing word segmentation on the medical history information based on the matched text data to obtain multiple pieces of text data; and performing vector conversion on the multiple pieces of text data obtained after the word segmentation to obtain the multiple text vectors.

19. The storage medium according to claim 17, wherein when executing the computer readable instruction, the processor further performs the following steps: calculating a term frequency and an inverse document frequency of the multiple text vectors; calculating weights of the multiple text vectors by using a preset algorithm based on the term frequency and the inverse document frequency; extracting a text vector whose weight reaches a preset threshold; and calculating a feature dimension value corresponding to the text vector based on a preset algorithm and the weight.

20. The storage medium according to claim 17, wherein the text comprises multiple text sentences, the multiple text sentences form a text block, and when executing the computer readable instruction, the processor further performs the following steps: calculating a correlation between

the multiple text vectors by using the target classifier based on the feature dimension values, calculating the text sentences formed in the text based on the correlation, and calculating sentence vectors of the text sentences; extracting features of the sentence vectors, and calculating a text block vector based on the features of the multiple sentence vectors; and calculating a probability of each category corresponding to the text block vector, extracting a category whose probability reaches a preset probability value, and adding a corresponding category label to the text block.