

KNOWLEDGE-DRIVEN AND DATA-DRIVEN BASED GROUPING METHOD AND RELATED DEVICE

The present application claims priority to Chinese Patent Application No. 201910991392.5, filed with the China National Intellectual Property Administration on October 18, 2019 and entitled "KNOWLEDGE-DRIVEN AND DATA-DRIVEN BASED GROUPING METHOD AND RELATED DEVICE", which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

The present application relates to the field of artificial intelligence, and in particular, to a knowledge-driven and data-driven based grouping method and a related device.

BACKGROUND

Personalized precision treatment is not only an important research topic in the field of medical treatment, but also a promising development direction in the field of medical research. Moreover, personalized precision treatment mainly starts from groups and provides personalized treatment based on individual differences of patients. In this process, the greatest difficulty lies in how to determine different groups and provide adequate medical knowledge support for personalized treatment for patients in these groups. In the existing grouping system, a pure data-driven based grouping method or a pure knowledge-driven based grouping method is often used, but either method has many disadvantages.

A core idea of the pure data-driven based grouping method is to learn doctor prescribing rules from data sets based on a machine learning or deep learning algorithm, and apply the learned rules to patients of the same type. A core idea of the pure knowledge-driven based grouping method is to construct rules by referring to existing guidelines, consensus and other relevant authoritative knowledge, and obtain corresponding auxiliary decision-making solutions according to the rules. However, the inventor found that, different from cases in other industries, using the pure data-driven based grouping method cannot guarantee that the medical rules learned by using the algorithm comply with medical knowledge, and cannot provide personalized recommendations of corresponding auxiliary decision-making solutions to different patients in the same group; and using the pure knowledge-driven based grouping method cannot make a reasonable choice from multiple recommended auxiliary decision-making solutions, and there is auxiliary information completely unavailable in the guidelines, consensus and other knowledge, for

example, drug prices. As a result, the accuracy of grouping a to-be-classified crowd into the feature similar crowd is low.

SUMMARY

The present application provides a knowledge-driven and data-driven based grouping method and a related device, so as to create a target knowledge-based grouping decision tree; determine that a to-be-classified crowd belongs to a first feature similar crowd, or create a data-based grouping decision tree and classify the to-be-classified crowd as the first feature similar crowd or a second feature similar crowd based on a target layer number; and perform other steps, thereby improving an accuracy degree of grouping a to-be-classified crowd into a feature similar crowd, and further providing a doctor with reasonable, standardized, comprehensive, and refined evidence-based medical information.

According to a first aspect of the embodiments of the present application, a knowledge-driven and data-driven based grouping method is provided, including: creating an initial knowledge-based grouping decision tree based on medical rule data, and filling sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, where the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group a clinically-similar feature crowd; and the sample data includes clinical result data corresponding to the clinically-similar feature crowd and compliant with a medical knowledge rule; obtaining feature data of a to-be-classified crowd and grouping the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identifying the number of categories of the processing solutions, where the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd; and if one category of the processing solutions is identified, determining that the to-be-classified crowd belongs to a first feature similar crowd, where the first feature similar crowd is a feature similar crowd to which the processing solutions corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree belong; or if at least two categories of the processing solutions are identified, generating a data-based grouping decision tree based on a second grouping rule, and

determining a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, where the second grouping rule indicates classification based on a recursive partitioning algorithm and effectiveness of the processing solutions; wherein, if the target layer number indicates the first layer, the to-be-classified crowd is classified as the first feature similar crowd; or if the target layer number does not indicate the first layer, the to-be-classified crowd is classified as a second feature similar crowd, where the second feature similar crowd is a feature similar crowd to which a processing solution corresponding to the target data node on the data-based grouping decision tree belongs; and each target data node corresponds to a different feature similar crowd.

According to a second aspect of the embodiments of the present application, a knowledge-driven and data-driven based grouping apparatus is provided, including: a first creation unit, configured to create an initial knowledge-based grouping decision tree based on medical rule data, and fill sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, where the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group a clinically-similar feature crowd; and the sample data includes clinical result data corresponding to the clinically-similar feature crowd and compliant with a medical knowledge rule; an identification unit, configured to obtain feature data of a to-be-classified crowd and group the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identify the number of categories of the processing solutions, where the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd; a second creation unit, configured to, if at least two categories of the processing solutions are identified, generate a data-based grouping decision tree based on a second grouping rule, and determine a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, where the second grouping rule indicates classification based on a recursive partitioning algorithm and effectiveness of the processing solutions; a first classification unit, configured to classify the to-be-classified crowd as the first feature similar crowd if the target layer number indicates the first layer; and

a second classification unit, configured to classify the to-be-classified crowd as a second feature similar crowd if the target layer number does not indicate the first layer, where the second feature similar crowd is a feature similar crowd to which a processing solution corresponding to the target data node on the data-based grouping decision tree belongs; and each target data node corresponds to a different feature similar crowd.

According to a third aspect of the embodiments of the present application, a knowledge-driven and data-driven based grouping device is provided, including: a memory, a processor, and a computer program stored in the memory and executable on the processor, and the processor implements the following steps of a knowledge-driven and data-driven based grouping method when executing the computer program: creating an initial knowledge-based grouping decision tree based on medical rule data, and filling sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, where the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group a clinically-similar feature crowd; and the sample data includes clinical result data corresponding to the clinically-similar feature crowd and compliant with a medical knowledge rule; obtaining feature data of a to-be-classified crowd and grouping the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identifying the number of categories of the processing solutions, where the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd; and if one category of the processing solutions is identified, determining that the to-be-classified crowd belongs to a first feature similar crowd, where the first feature similar crowd is a feature similar crowd to which the processing solutions corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree belong; or if at least two categories of the processing solutions are identified, generating a data-based grouping decision tree based on a second grouping rule, and determining a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, where the second grouping rule indicates classification based on a recursive partitioning

algorithm and effectiveness of the processing solutions; wherein, if the target layer number indicates the first layer, the to-be-classified crowd is classified as the first feature similar crowd; or if the target layer number does not indicate the first layer, the to-be-classified crowd is classified as a second feature similar crowd, where the second feature similar crowd is a feature similar crowd to which a processing solution corresponding to the target data node on the data-based grouping decision tree belongs; and each target data node corresponds to a different feature similar crowd.

According to a fourth aspect of the embodiments of the present application, a computer-readable storage medium is provided, including instructions, and when the instructions are run on a computer, the computer is enabled to perform the following steps of a knowledge-driven and data-driven based grouping method: creating an initial knowledge-based grouping decision tree based on medical rule data, and filling sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, where the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group a clinically-similar feature crowd; and the sample data includes clinical result data corresponding to the clinically-similar feature crowd and compliant with a medical knowledge rule; obtaining feature data of a to-be-classified crowd and grouping the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identifying the number of categories of the processing solutions, where the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd; and if one category of the processing solutions is identified, determining that the to-be-classified crowd belongs to a first feature similar crowd, where the first feature similar crowd is a feature similar crowd to which the processing solutions corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree belong; or if at least two categories of the processing solutions are identified, generating a data-based grouping decision tree based on a second grouping rule, and determining a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, where the second grouping rule indicates classification based on a recursive partitioning algorithm and effectiveness of

the processing solutions; wherein if the target layer number indicates the first layer, the to-be-classified crowd is classified as the first feature similar crowd; or if the target layer number does not indicate the first layer, the to-be-classified crowd is classified as a second feature similar crowd, where the second feature similar crowd is a feature similar crowd to which a processing solution corresponding to the target data node on the data-based grouping decision tree belongs; and each target data node corresponds to a different feature similar crowd.

Compared with the prior art, in the technical solutions provided by the embodiments of the present application, a target knowledge-based grouping decision tree is created; the number of categories of processing solutions corresponding to a target end knowledge sub-node on the target knowledge-based grouping decision tree is identified for feature data of a to-be-classified crowd; and when one category is identified, it is determined that the to-be-classified crowd belongs to a first feature similar crowd corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree; or when at least two categories are identified, a data-based grouping decision tree is generated based on a preset second grouping rule, and a target layer number of the data-based grouping decision tree is determined; and the to-be-classified crowd is classified as the first feature similar crowd or a second feature similar crowd corresponding to a target data node on the data-based grouping decision tree based on the target layer number. In the embodiments of the present application, a target knowledge-based grouping decision tree is created; it is determined that a to-be-classified crowd belongs to a first feature similar crowd, or a data-based grouping decision tree is created and the to-be-classified crowd is classified as the first feature similar crowd or a second feature similar crowd based on a target layer number; and other steps are performed, thereby improving an accuracy degree of grouping a to-be-classified crowd into a feature similar crowd, and further providing a doctor with reasonable, standardized, comprehensive, and refined evidence-based medical information.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 is a schematic diagram of a knowledge-driven and data-driven based grouping method according to an embodiment of the present application;

FIG. 2 is a schematic diagram of an output first feature similar crowd according to an embodiment of the present application;

FIG. 3 is a schematic diagram of an output second feature similar crowd according to an

embodiment of the present application;

FIG. 4 is a schematic diagram of grouping by using a knowledge-based grouping decision tree and a data-based grouping decision tree according to an embodiment of the present application;

FIG. 5 is a schematic diagram of a knowledge-driven and data-driven based grouping method according to another embodiment of the present application;

FIG. 6 is a schematic diagram of a knowledge-driven and data-driven based grouping apparatus according to an embodiment of the present application;

FIG. 7 is a schematic diagram of a knowledge-driven and data-driven based grouping apparatus according to another embodiment of the present application; and

FIG. 8 is a schematic diagram of a knowledge-driven and data-driven based grouping device according to another embodiment of the present application.

DESCRIPTION OF EMBODIMENTS

Embodiments of the present application provide a knowledge-driven and data-driven based grouping method and a related device, so as to create a target knowledge-based grouping decision tree; determine that a to-be-classified crowd belongs to a first feature similar crowd, or create a data-based grouping decision tree and classify the to-be-classified crowd into the first feature similar crowd or a second feature similar crowd based on a target layer number; and perform other steps, thereby improving an accuracy degree of grouping a to-be-classified crowd into a feature similar crowd, and further providing a doctor with reasonable, standardized, comprehensive, and refined evidence-based medical information.

To enable a person skilled in the art to better understand the solutions of the present application, the embodiments of the present application are described below with reference to the accompanying drawings in the embodiments of the present application.

FIG. 1 shows a flowchart of a knowledge-driven and data-driven based grouping method according to an embodiment of the present application. The following uses an example to describe the knowledge-driven and data-driven based grouping method provided in the present application. The present application does not limit a type of execution body. The method specifically includes the following steps.

101. Create an initial knowledge-based grouping decision tree based on medical rule data, and fill sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping

decision tree to obtain a target knowledge-based grouping decision tree, where the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group clinically-similar patients; and the sample data includes clinical result data compliant with a medical knowledge rule.

Specifically, data cleansing processing and data reduction processing are performed on the medical disciplinary knowledge data of the medical guidelines, medical consensus, medical files, and expert clinical experience, so as to ensure quality of the medical disciplinary knowledge data. Rule extraction is performed on the medical disciplinary knowledge data that has undergone the cleansing processing and data reduction processing. An extracted rule is used to construct a coarse-grained grouping rule (i.e., medical rule data), and the coarse-grained grouping rule is used to construct the initial knowledge-based grouping decision tree to obtain the target knowledge-based grouping decision tree. For example, as shown in FIG. 2, disease information of a clinically-similar feature crowd grouped corresponding to clinical result data (i.e., sample data) existing in clinical result data and compliant with a medical knowledge rule is filled to an end knowledge sub-node of each branch of the corresponding knowledge-based grouping decision tree, i.e., end knowledge sub-node 1, end knowledge sub-node 2, end knowledge sub-node 3, end knowledge sub-node 4, end knowledge sub-node 5, end knowledge sub-node 6, and end knowledge sub-node 7. Clinical result data outside the knowledge-based grouping decision tree is clinical result data non-compliant with the medical knowledge rule. Therefore, the end knowledge sub-node of each branch of the knowledge-based grouping decision tree (end knowledge sub-node 1, end knowledge sub-node 2, end knowledge sub-node 3, end knowledge sub-node 4, end knowledge sub-node 5, end knowledge sub-node 6, and end knowledge sub-node 7) includes at least one processing solution corresponding to the disease information. The coarse-grained grouping rule is obtained from the medical disciplinary knowledge data of the medical guidelines, medical consensus, medical files, and expert clinical experience and first sample data is filled to the end knowledge sub-node of each branch of the knowledge-based grouping decision tree, so as to construct the knowledge-based grouping decision tree based on the medical disciplinary knowledge data and the clinical result data. The content in FIG. 2 and the example content above are merely used for reference, of which accuracy and actual operations are not considered.

102. Obtain feature data of a to-be-classified crowd and group the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identify the number of categories of the processing solutions, where the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd.

A terminal obtains the feature data of the to-be-classified crowd and groups the to-be-classified crowd based on the feature data and the medical disciplinary knowledge through the target knowledge-based grouping decision tree to obtain the target end knowledge sub-node to which the to-be-classified crowd is grouped on the target knowledge-based grouping decision tree and to obtain processing solutions corresponding to the target end knowledge sub-node for a feature similar crowd; and identifies the number of categories of the processing solutions to determine whether to further group the feature similar crowd of the target end knowledge sub-node. The processing solutions are used to indicate the solution channels for the feature similar crowd, for example, treatment solutions or prescriptions for a disease corresponding to the clinically-similar feature crowd.

103. If one category of the processing solutions is identified, determine that the to-be-classified crowd belongs to a first feature similar crowd, where the first feature similar crowd is a feature similar crowd to which the processing solutions corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree belong.

If the terminal identifies one category of the processing solutions, it classifies the to-be-classified crowd to a feature similar crowd (i.e., the first feature similar crowd) corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree, so as to obtain the first feature similar crowd obtained by grouping the to-be-classified crowd only through the target knowledge-based grouping decision tree.

For example, specifically, analysis may be performed by determining whether a similarity between drug prescriptions in processing solutions is a preset judgment value range, so as to analyze whether there is one category of the processing solutions of the end sub-node of each branch. As shown in FIG. 2, there are three processing solutions on end sub-node 1, which are respectively processing solution A, processing solution B, and processing solution C. The preset judgment value range is a similarity range

of 97%-100%. A similarity between drug prescriptions in processing solution A and processing solution B is 96%, a similarity between drug prescriptions in processing solution A and processing solution C is 98%, and a similarity between drug prescriptions in processing solution B and processing solution C is 97.5%. In this case, there are two categories of the processing solutions on end sub-node 1. When one category of the processing solutions on the end knowledge sub-node of each branch is identified through analysis, a subsequent operation of creating a data-driven grouping model is not performed for the end knowledge sub-node, which indicates that a crowd on the end knowledge sub-node is a feature similar crowd of the same category (i.e., the first feature similar crowd). The example content above is merely used for reference, of which accuracy and actual operations are not considered.

104. If at least two categories of the processing solutions are identified, generate a data-based grouping decision tree based on a second grouping rule, and determine a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, where the second grouping rule indicates classification based on a recursive partitioning algorithm and effectiveness of the processing solutions.

Specifically, when at least two categories of the processing solutions on the end knowledge sub-node of each branch are identified, a subsequent operation of creating a data-based grouping decision tree is performed. The data-based grouping decision tree learns grouping modes from data sets through a machine-learning or deep-learning algorithm, so as to group the clinically-similar feature crowd. The target end knowledge sub-node is used as a parent node. Starting from the parent node, clinical result data of a feature similar crowd corresponding to a data root node is branched based on the second grouping rule that indicates classification based on the recursive partitioning algorithm and effectiveness of the processing solutions, so as to create the data-based grouping decision tree and determine the target layer number of the data sub-node corresponding to each processing solution on the data-based grouping decision tree. Data-driven grouping is performed based on the medical disciplinary knowledge data used in the knowledge-driven grouping to further group the similar crowd, thereby improving accuracy of grouping.

For example, as shown in FIG. 3, a feature similar crowd of the parent node at the first layer is branched based on effectiveness of a processing solution for a feature similar crowd of a data root node at the first layer. At the second layer, a feature similar crowd of data node A is grouped as a second

feature similar crowd with effectiveness of a processing solution less than 90%, and a feature similar crowd of data node B is grouped as a second feature similar crowd with effectiveness of a processing solution greater than 90%. At the third layer, a feature similar crowd of data node C is grouped as a second feature similar crowd with effectiveness of a processing solution from 80% to 85%, and a feature similar crowd of data node D is grouped as a second feature similar crowd with effectiveness of a processing solution from 86% to 90%.

The effectiveness includes, but not limited to, drug effectiveness. The target layer number is used to determine whether to classify the to-be-classified crowd as the first feature similar crowd obtained by grouping the to-be-classified crowd by the target knowledge-based grouping decision tree or the second feature similar crowd obtained by further grouping, by the data-based grouping decision tree, the first feature similar crowd obtained by grouping the to-be-classified crowd by the target knowledge-based grouping decision tree. Starting from a parent node of the data-based grouping decision tree, a data sub-node of each branch at each layer from top to bottom analyzes whether there is a difference between drug effectiveness of a processing solution for a feature similar crowd of the parent node and that of a processing solution for a feature similar crowd of the sub-node. A layer number corresponding to a sub-node having a significant difference in the drug effectiveness of the processing solution is used as a selected layer number of the data-based grouping decision tree, and the target layer number is marked on the end sub-node of the knowledge-based grouping decision tree and corresponding to the parent node. The target layer number is used as a basis for determining whether to invoke the data-based grouping decision tree and to use the data-based grouping decision tree to perform data-driven based grouping on a similar crowd. The drug effectiveness can be analyzed through at least one of a cure rate, a marked improvement rate, an improvement rate, an inefficiency rate, or another drug effectiveness index.

105. Classify the to-be-classified crowd as the first feature similar crowd if the target layer number indicates the first layer.

Specifically, when the target layer number indicates the first layer, i.e., the parent node of the data-based grouping decision tree, the to-be-classified crowd is classified as the first feature similar crowd obtained by grouping the to-be-classified crowd by the target knowledge-based grouping decision tree, and the to-be-classified crowd is not further grouped through the data-based grouping decision tree.

106. Classify the to-be-classified crowd as a second feature similar crowd if the target layer number does not indicate the first layer, where the second feature similar crowd is a feature similar crowd to which a processing solution corresponding to a target data node on the data-based grouping decision tree belongs; and each target data node corresponds to a different feature similar crowd.

Specifically, when the target layer number indicates the first layer, i.e., the parent node of the data-based grouping decision tree, an end sub-node of a layer number of the data-based grouping decision tree and corresponding to the target layer number is obtained, and a feature similar crowd corresponding to the end sub-node is obtained. The feature similar crowd is used as the second feature similar crowd obtained by grouping the to-be-classified crowd by combining the target knowledge-based grouping decision tree and the data-based grouping decision tree. For example, as shown in FIG. 3, if the target layer number is 2, the to-be-classified crowd is classified as a second feature similar crowd corresponding to data node B; and if the target layer number is 3, the to-be-classified crowd is classified as a second feature similar crowd corresponding to node C and node D.

For example, as shown in FIG. 4, the to-be-classified crowd is grouped into a corresponding feature similar crowd by combining the knowledge-based grouping decision tree and the data-based grouping decision tree. It is determined, based on the target layer number, whether to classify the to-be-classified crowd as the first feature similar crowd obtained by grouping the to-be-classified crowd by the target knowledge-based grouping decision tree or the second feature similar crowd obtained by further grouping, by the data-based grouping decision tree, the first feature similar crowd obtained by grouping the to-be-classified crowd by the target knowledge-based grouping decision tree. A medical assistant decision system obtains, based on the first feature similar crowd obtained by using the method of the present application, first evidence-based medical information that corresponds to the first feature similar crowd and that provides medical knowledge and supplements corresponding real-world evidence; and obtains, based on the second feature similar crowd obtained by using the method of the present application, second evidence-based medical information that corresponds to the second feature similar crowd and that provides medical knowledge, supplements corresponding real-world evidence, and implements grouping in combination with the data-driven grouping. Similar patients are defined. Further, personalized evidence-based medical information that complies with the medical disciplinary knowledge is further provided for different patients in the same group, thereby helping a doctor make

decisions.

In the embodiments of the present application, a target knowledge-based grouping decision tree is created; it is determined that a to-be-classified crowd belongs to a first feature similar crowd, or a data-based grouping decision tree is created and the to-be-classified crowd is classified as the first feature similar crowd or a second feature similar crowd based on a target layer number; and other steps are performed, thereby improving an accuracy degree of grouping a to-be-classified crowd into a feature similar crowd, and further providing a doctor with reasonable, standardized, comprehensive, and refined evidence-based medical information.

With reference to FIG. 5, for a method in 501 to 505, reference can be made to 101 and 105 in the embodiments of the present application, and details are not repeated herein. A knowledge-driven and data-driven based grouping method according to another embodiment of the present application includes the following steps.

501. Create an initial knowledge-based grouping decision tree based on medical rule data, and fill sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, where the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group clinically-similar patients; and the sample data includes clinical result data compliant with a medical knowledge rule.

Optionally, before the creating an initial knowledge-based grouping decision tree based on medical rule data, and fill sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, the method further includes: obtaining the medical disciplinary knowledge data and clinical data from a plurality of systems, platforms, and medical systems, and performing data processing on the medical disciplinary knowledge data and the clinical data to obtain the medical rule data, where the data processing includes data cleansing processing, data integration processing, data reduction processing, data transformation processing, feature extraction processing, data desensitization processing, and classification processing. Specifically, the following is included: obtaining the medical disciplinary knowledge data from a plurality of systems and platforms, and obtaining the clinical data from a plurality of medical systems;

performing data pre-processing separately on the medical disciplinary knowledge data and the clinical data, where the data pre-processing includes data cleansing processing, data integration processing, data reduction processing, and data transformation processing; performing feature extraction on the medical disciplinary knowledge data obtained after the data pre-processing to obtain feature information, and performing data desensitization processing and classification on the clinical data obtained after the data pre-processing to obtain first sample data and third sample data; and analyzing the feature information, and classifying the feature information based on a preset classification rule to obtain the medical rule data, where the preset classification rule includes same and/or similar diseases. By performing the data pre-processing including data cleansing processing, data integration processing, data reduction processing, and data transformation processing on the medical disciplinary knowledge data, quality of the medical disciplinary knowledge data can be ensured, thereby facilitating subsequent data analysis and data digging on the medical disciplinary knowledge data. By performing data transformation on some sensitive information in the medical data based on a desensitization rule to provide reliable protection for sensitive private data in the medical data, reliable protection is ensured for the sensitive private data without affecting quality of the medical data. A structured query language (SQL) statement may be executed on a page to ensure that each piece of medical data is desensitized and does not interfere with one another.

502. Obtain feature data of a to-be-classified crowd and group the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identify the number of categories of the processing solutions, where the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd.

Optionally, the identifying the number of categories of the processing solutions specifically further includes: calculating similarities between the processing solutions to obtain a plurality of similarity values, and determining by comparison whether the plurality of similarity values are all in a preset range; if yes, identifying that there is one category of the processing solutions through statistics collection; and if no, identifying that there are multiple categories of the processing solutions through statistics collection. The number of categories of the processing solutions is determined by determining whether

the plurality of similarity values between the processing solutions are all in the preset range. For example, the similarities include, but are not limited to, a similarity between drug prescriptions, a similarity between drug prescriptions in combination with drug usages, and a similarity between disease information in combination with drug prescriptions. By determining the categories of the processing solutions, it is determined whether to construct a data-driven grouping model, so that different operations can be performed based on different input clinic data, thereby reducing unnecessary operations, improving efficiency, and ensuring accurate grouping.

503. If one category of the processing solutions is identified, determine that the to-be-classified crowd belongs to a first feature similar crowd, where the first feature similar crowd is a feature similar crowd to which the processing solutions corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree belong.

504. If at least two categories of the processing solutions are identified, generate a data-based grouping decision tree based on a second grouping rule, and determine a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, where the second grouping rule indicates classification based on a recursive partitioning algorithm and effectiveness of the processing solutions.

Optionally, the generating a data-based grouping decision tree based on a second grouping rule, and determining a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree specifically include: step I: using the target end knowledge sub-node as a data root node of the data-based grouping decision tree, and calculating an effectiveness score of a processing solution corresponding to the data root node; step II: branching the data root node based on an effectiveness score of each processing solution, classifying each processing solution to a data sub-node of each branch in sequence based on the recursive partitioning algorithm, and continuing with effectiveness analysis on a processing solution corresponding to each data sub-node; step III: if the effectiveness score of the processing solution corresponding to a current data sub-node is greater than a first threshold, branching the current data sub-node; step IV: repeating step III until the effectiveness score of the processing solution corresponding to each data sub-node is less than or equal to the first threshold to obtain the data-based grouping decision tree, where the data-based grouping decision tree includes a plurality of data sub-nodes; and each end data leaf sub-node in the plurality of data sub-nodes

corresponds to one processing solution; and step V: determining a target layer number of the end data leaf sub-node corresponding to each processing solution on the data-based grouping decision tree. The data-based grouping decision tree is constructed based on the effectiveness score of the processing solutions corresponding to the target end knowledge sub-node and the recursive partitioning algorithm. For example, the first threshold is 60 points, and the processing solutions of the data root node are branched into processing solutions with effectiveness scores of 95 points or above and those with effectiveness scores less than 95 points (i.e., a feature similar crowd is grouped). If the effectiveness scores less than 95 points are all greater than 60 points, the processing solutions corresponding to the effectiveness scores less than 95 points are branched, and if the effectiveness scores of 95 points or above are all greater than 60 points, the processing solutions corresponding to the effectiveness scores of 95 points or above are branched, and so forth, until a effectiveness score of a processing solution corresponding to a data sub-node corresponding to the last branch is less than or equal to 60 points. In this case, a layer number of an end data leaf sub-node on the data-based grouping decision tree is the target layer number. When the data-based grouping decision tree is branched, the effectiveness scores of the processing solutions may be used for reference. Values in classification ranges (obtained through measurement or detection) of all the effectiveness scores for classifying the processing solutions are all used as observation values. Estimation processing is performed on parameters based on a preset target function and all the observation values to obtain parameter estimates, and stability of the parameters in different classification ranges is evaluated based on the parameter estimates. If an evaluation result is "unstable", classification range variables are analyzed, and a classification range variable with the lowest stability is selected for stability evaluation until the evaluation result is "stable", where the classification range variables are used to indicate regressors of the classification ranges. If the evaluation result is "stable", the stability evaluation on the parameters in the different classification ranges is stopped. Based on the stability of the parameters in the different classification ranges, an optimized region segmentation rule is obtained based on the dichotomy or a variable adaptive method. The dichotomy is used to indicate a bisection method in which the number of the selected sub-nodes is 2. The variable adaptive method is used to indicate a method in which the number of selected sub-nodes is equal to the number of classification ranges. A sub-node corresponding to the region segmentation rule is obtained based on the region segmentation rule. On the sub-node corresponding to the region segmentation rule, stability of

the parameters in the different classification ranges is repeatedly evaluated based on the parameter estimates and an optimized region segmentation rule is repeatedly obtained based on to the dichotomy or variable adaptive method, until the parameters are considered stable in the different classification ranges through the evaluation. In this way, the to-be-classified crowd and the feature similar crowd corresponding to the target end sub-node on the target knowledge-based grouping decision tree are grouped to obtain the data-based grouping decision tree. The data-based grouping decision tree is created by using model-based recursive partitioning and a grouping rule strongly correlated to clinical result data. A target function of a model is used to estimate parameters and segmenting points, and corresponding model-based scoring is performed with respect to parameter instability of each node, so as to determine through evaluation which variable should be used for partitioning, and to help create a non-linear relationship of the data-based grouping decision tree and automatically detect interaction between interpretation variables. In the creation process: each node is associated with the model; stability fluctuation analysis is performed on the parameters to determine whether the node needs to be split; and if any classification range variable is apparently unstable, the node is segmented into local optimal ends, and the process is repeated. The parameters may be estimated through a scoring function or an estimation function. Stability estimates of the parameters in the different classification ranges may be captured through empirical fluctuation.

Optionally, the using the target end knowledge sub-node as a data root node of the data-based grouping decision tree, and calculating an effectiveness score of a processing solution corresponding to the data root node specifically includes: using the target end knowledge sub-node as a data root node of the data-based grouping decision tree; obtaining target feature information of the processing solution corresponding to the data root node, and dividing the target feature information into element factors; calculating association degrees, influence degrees on a preset influence element, and uncontrolled edge association degrees of the element factors on different levels; analyzing the association degrees, the influence degrees, and the edge association degrees through a preset compressibility criterion and comparability criterion to obtain a confounding factor; and calculating an exponential transform value of a regression coefficient of an occurrence probability of the confounding factor and the preset influence element, and using the exponential transform value as a determined score to obtain the effectiveness score of the processing solution corresponding to the data root node, where a calculation formula for the

$$p(y = k|x; w) = \frac{\exp(w_k^T x)}{\sum_{l=1}^K \exp(w_l^T x)}$$

exponential transform value is as follows:

, where $p(y = k|x; w)$ is the

exponential transform value; w is the occurrence probability of the confounding factor and the preset influence element; x is the confounding factor; k is the number of categories of the preset influence element; y is an output predicted value; y is a conditional probability of k ; K is a maximum value of k ; and T is a time period. For example, the target feature information may be disease information in the processing solution on the parent node of the data-based grouping decision tree; the preset influence element is a disease; the influence degree is severity of the disease; the effectiveness is drug effectiveness; the confounding factor is a factor associated with both exposure and the disease, and is an independent risk factor of the studied disease and also definitely a risk factor for a non-exposed group. Drug effectiveness analysis is performed on a processing solution corresponding to a feature similar crowd to obtain a sub-node having statistical analysis significance, so as to rapidly and accurately perform further grouping through the data-based grouping decision tree after the grouping performed through the knowledge-based grouping decision tree. In the process of drug effectiveness analysis, a possible confounding factor is selected first; then the confounding factor is added to multivariate logistic regression analysis; finally an exponential transform value of a regression coefficient of a drug variable is used as a determined final effectiveness score to obtain a plurality of effectiveness scores. A compressibility criterion, a comparability criterion, a counterfactual model, and a graph model may be used as confounding factor identification criteria to obtain the confounding factor.

505. Classify the to-be-classified crowd as the first feature similar crowd if the target layer number indicates the first layer.

506. If the target layer number does not indicate the first layer, determine whether an effectiveness score of a processing solution corresponding to the target layer number is greater than a second threshold, where the second threshold is greater than a first threshold.

When it is identified that the target layer number does not indicate the first layer, the to-be-classified crowd needs to be further grouped through the data-based grouping decision tree. The effectiveness score of the processing solution corresponding to the target layer number is obtained, and whether the effectiveness score is greater than the second threshold is determined to ensure an accuracy degree of grouping the to-be-classified crowd. For example, as shown in FIG. 3, if the target layer number indicates

the second layer, an effectiveness score of a processing solution corresponding to a second feature similar crowd of node B is obtained, and whether the effectiveness score is greater than the second threshold is analyzed.

507. If the effectiveness score of the processing solution corresponding to the target layer number is greater than the second threshold, use a data leaf sub-node on which the processing solution with the effectiveness score greater than the second threshold is located as a target data node, and classify the to-be-classified crowd as a feature similar crowd to which the processing solution corresponding to the target data node belongs, so as to obtain a second feature similar crowd.

When the effectiveness score of the processing solution corresponding to the target layer number is greater than the second threshold, the to-be-classified crowd and the feature similar crowd do not need to be further grouped. At this time, the final feature similar crowd has been obtained by performing grouping strongly correlated to the clinical result data on the to-be-classified crowd by combining knowledge-driven grouping and data-driven grouping. A feature similar crowd corresponding to the target data node (on which the processing solution corresponding to the target layer number and having the effectiveness score greater than the second threshold is located) of the data decision tree is used as the second feature similar crowd obtained by classifying the to-be-classified crowd. For example, the second threshold is 70 points, and the effectiveness score of the processing solution corresponding to the target layer number is 80 points. In this case, a data leaf sub-node corresponding to the processing solution with the effectiveness score of 80 points is used as the target data node, and the to-be-classified crowd is classified as a feature similar crowd (i.e., the second feature similar crowd) to which the processing solution corresponding to the target data node belongs.

508. If the effectiveness score of the processing solution corresponding to the target layer number is less than or equal to the second threshold, analyze an effectiveness score of a processing solution corresponding to an upper-layer number of the target layer number.

When the effectiveness score of the processing solution corresponding to the target layer number is less than or equal to the second threshold, a feature similar crowd corresponding to the target layer number is not the final crowd obtained by classifying the to-be-classified crowd, and the effectiveness score of the processing solution corresponding to the upper-layer number of the target layer number needs to be analyzed. The feature similar crowd corresponding to the processing solution with the

effectiveness score less than or equal to the second threshold has a low matching degree with the to-be-classified crowd. No analysis is further performed on the node corresponding to the processing solution with the effectiveness score less than or equal to the second threshold or a following node, so as to improve operation efficiency and a grouping accuracy degree.

509. If the effectiveness score of the processing solution corresponding to the upper-layer number of the target layer number is greater than the second threshold, use a data leaf sub-node on which the processing solution corresponding to the upper-layer number of the target layer number is located as a target data node, and classify the to-be-classified crowd as a feature similar crowd to which the processing solution corresponding to the target data node belongs, so as to obtain a second feature similar crowd.

If the effectiveness score of the processing solution corresponding to the upper-layer number of the target layer number is greater than the second threshold, the to-be-classified crowd is classified to the second feature similar crowd of the target node (corresponding to the processing solution corresponding to the upper-layer number of the target layer number and having the effectiveness score greater than the second threshold) of the data-based grouping decision tree. The second feature similar crowd corresponding to the node is the final feature similar crowd obtained by performing grouping strongly correlated to the clinical result data on the to-be-classified crowd by combining knowledge-driven grouping and data-driven grouping. If the effectiveness score of the processing solution corresponding to the upper-layer number of the target layer number is less than or equal to the second threshold, it is determined whether an effectiveness score of a processing solution corresponding to an upper-layer number of the upper-layer number of the target layer number is greater than the second threshold, and determining is continued until it is determined through analysis that an effectiveness score of a corresponding processing solution is greater than the second threshold. A feature similar crowd of a target node (corresponding to the processing solution with the effectiveness score greater than the second threshold) on the data-based grouping decision tree is used as the second feature similar crowd obtained by finally classifying the to-be-classified crowd.

Optionally, after the classifying the to-be-classified crowd as a second feature similar crowd if the target layer number does not indicate the first layer, the method further specifically includes: invoking a preset interpretation model to perform interpretation processing and visual chart-based statistical

analysis processing on the first feature similar crowd or the second feature similar crowd to obtain an interpretable feature similar crowd, where the interpretation processing includes effectiveness score sorting processing, matching degree calculation processing, and matching degree sorting processing. Specifically, the preset interpretation model may be invoked to sort effectiveness scores corresponding to the first feature similar crowd or the second feature similar crowd in descending order of values to obtain first sorting information; a matching degree between the to-be-classified crowd and the first feature similar crowd or the second feature similarity crowd is calculated, and the matching degrees are sorted in descending order of values to obtain second sorting information; the first feature similar crowd or the second feature similar crowd is sorted based on the first sorting information and/or the second sorting information, and the first feature similar crowd or the second feature similar crowd is labeled based on the first sorting information and/or second sorting information; and visual chart-based statistical analysis processing is performed on the labeled first feature similar crowd or second feature similar crowd to generate a visual chart, so as to obtain the interpretable feature similar crowd. The preset interpretation model is invoked on the first feature similar crowd or the second feature similar crowd, so as to help a user conveniently obtain various grouping data of the feature similar crowd obtained by classifying the to-be-classified crowd.

After the medical assistant decision system obtains, based on the first feature similar crowd obtained by using the method of the present application, the first evidence-based medical information that corresponds to the first feature similar crowd and that provides medical knowledge and supplements corresponding real-world evidence; and obtains, based on the second feature similar crowd obtained by using the method of the present application, the second evidence-based medical information that corresponds to the second feature similar crowd and that provides medical knowledge, supplements corresponding real-world evidence, and implements grouping in combination with the data-driven grouping, the method specifically includes: invoking the created interpretation model to separately obtain first feature information of the first sample data, second feature information of first evidence-based medical information, and/or third feature information of second evidence-based medical information, where the first sample data indicates clinical result data non-compliant with the medical knowledge rule; analyzing the first feature information and the second feature information to obtain first analysis data, and analyzing the first feature information and the third feature information to obtain

second analysis data; performing statistics collection and analysis on the first feature information, the second feature information, and the third feature information to generate a visual chart; invoking a created knowledge graph to obtain first real object information matching with the second feature information from the knowledge graph and obtain second real object information matching with the third feature information from the knowledge graph; and generating an analysis report with graphs and texts from the first feature information, the second feature information, the third feature information, the first analysis data, the second analysis data, the first real object information, the second real object information, and the visual chart to obtain interpretable data. Interpretative and illustrative data is displayed in the form of graphs and texts, so that a user can better obtain content of the interpretable data. The interpretable data obtained by interpreting the first evidence-based medical information or the second evidence-based medical information can effectively provide more personalized, precise, and effective real-world evidence to help a doctor make decisions.

In the embodiments of the present application, a target knowledge-based grouping decision tree is created; it is determined that a to-be-classified crowd belongs to a first feature similar crowd, or a data-based grouping decision tree is created and the to-be-classified crowd is classified as the first feature similar crowd or a second feature similar crowd based on a target layer number; and other steps are performed, thereby improving an accuracy degree of grouping a to-be-classified crowd into a feature similar crowd, and further providing a doctor with reasonable, standardized, comprehensive, and refined evidence-based medical information.

The knowledge-driven and data-driven based grouping method is described in the embodiments of the present application above. A knowledge-driven and data-driven based grouping apparatus in the embodiments of the present application is described below. With reference to FIG. 6, a knowledge-driven and data-driven based grouping apparatus according to an embodiment of the present application includes:

a first creation unit 601, configured to create an initial knowledge-based grouping decision tree based on medical rule data, and fill sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, where the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical

consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group a clinically-similar feature crowd; and the sample data includes clinical result data corresponding to the clinically-similar feature crowd and compliant with a medical knowledge rule;

an identification unit 602, configured to obtain feature data of a to-be-classified crowd and group the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identify the number of categories of the processing solutions, where the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd;

a determining unit 603, configured to, if one category of the processing solutions is identified, determine that the to-be-classified crowd belongs to a first feature similar crowd, where the first feature similar crowd is a feature similar crowd to which the processing solutions corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree belong;

a second creation unit 604, configured to, if at least two categories of the processing solutions are identified, generate a data-based grouping decision tree based on a second grouping rule, and determine a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, where the second grouping rule indicates classification based on a recursive partitioning algorithm and effectiveness of the processing solutions;

a first classification unit 605, configured to classify the to-be-classified crowd as the first feature similar crowd if the target layer number indicates the first layer; and

a second classification unit 606, configured to classify the to-be-classified crowd as a second feature similar crowd if the target layer number does not indicate the first layer, where the second feature similar crowd is a feature similar crowd to which a processing solution corresponding to the target data node on the data-based grouping decision tree belongs; and each target data node corresponds to a different feature similar crowd.

In the embodiments of the present application, a target knowledge-based grouping decision tree is created; it is determined that a to-be-classified crowd belongs to a first feature similar crowd, or a data-based grouping decision tree is created and the to-be-classified crowd is classified as the first feature similar crowd or a second feature similar crowd based on a target layer number; and other steps are

performed, thereby improving an accuracy degree of grouping a to-be-classified crowd into a feature similar crowd, and further providing a doctor with reasonable, standardized, comprehensive, and refined evidence-based medical information.

With reference to FIG. 7, a knowledge-driven and data-driven based grouping apparatus according to another embodiment of the present application includes:

a first creation unit 601, configured to create an initial knowledge-based grouping decision tree based on medical rule data, and fill sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, where the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group clinically-similar patients; and the sample data includes clinical result data corresponding to the clinically-similar patients and compliant with a medical knowledge rule;

an identification unit 602, configured to obtain feature data of a to-be-classified crowd and group the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identify the number of categories of the processing solutions, where the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd;

a determining unit 603, configured to, if one category of the processing solutions is identified, determine that the to-be-classified crowd belongs to a first feature similar crowd, where the first feature similar crowd is a feature similar crowd to which the processing solutions corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree belong;

a second creation unit 604, configured to, if at least two categories of the processing solutions are identified, generate a data-based grouping decision tree based on a second grouping rule, and determine a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, where the second grouping rule indicates classification based on a recursive partitioning algorithm and effectiveness of the processing solutions;

a first classification unit 605, configured to classify the to-be-classified crowd as the first feature

similar crowd if the target layer number indicates the first layer; and

a judgment unit 606, configured to, if the target layer number does not indicate the first layer, determine whether an effectiveness score of a processing solution corresponding to the target layer number is greater than a second threshold, where the second threshold is greater than the first threshold;

a second classification unit 607, configured to, if the effectiveness score of the processing solution corresponding to the target layer number is greater than the second threshold, use a data leaf sub-node on which the processing solution with the effectiveness score greater than the second threshold is located as a target data node, and classify the to-be-classified crowd as a feature similar crowd to which the processing solution corresponding to the target data node belongs, so as to obtain the second feature similar crowd;

an analysis unit 608, configured to, if the effectiveness score of the processing solution corresponding to the target layer number is less than or equal to the second threshold, analyze an effectiveness score of a processing solution corresponding to an upper-layer number of the target layer number;

a third classification unit 609, configured to, if the effectiveness score of the processing solution corresponding to the upper-layer number of the target layer number is greater than the second threshold, use a data leaf sub-node on which the processing solution corresponding to the upper-layer number of the target layer number is located as a target data node, and classify the to-be-classified crowd as a feature similar crowd to which the processing solution corresponding to the target data node belongs, so as to obtain the second feature similar crowd.

Optionally, the second creation unit 604 is specifically configured to: step I: use the target end knowledge sub-node as a data root node of the data-based grouping decision tree, and calculate an effectiveness score of a processing solution corresponding to the data root node; step II: branch the data root node based on an effectiveness score of each processing solution, classify each processing solution to a data sub-node of each branch in sequence based on the recursive partitioning algorithm, and continue with effectiveness analysis on a processing solution corresponding to each data sub-node; step III: if the effectiveness score of the processing solution corresponding to a current data sub-node is greater than a first threshold, branch the current data sub-node; step IV: repeat step III until the effectiveness score of the processing solution corresponding to each data sub-node is less than or equal to the first threshold to

obtain the data-based grouping decision tree, where the data-based grouping decision tree includes a plurality of data sub-nodes; and each end data leaf sub-node in the plurality of data sub-nodes corresponds to one processing solution; and step V: determine a target layer number of the end data leaf sub-node corresponding to each processing solution on the data-based grouping decision tree.

Optionally, the second creation unit 604 is specifically configured to: use the target end knowledge sub-node as a data root node of the data-based grouping decision tree; obtain target feature information of the processing solution corresponding to the data root node, and divide the target feature information into element factors; calculate association degrees, influence degrees on a preset influence element, and uncontrolled edge association degrees of the element factors on different levels; analyze the association degrees, the influence degrees, and the edge association degrees through a preset compressibility criterion and comparability criterion to obtain a confounding factor; and calculate an exponential transform value of a regression coefficient of an occurrence probability of the confounding factor and the preset influence element, and use the exponential transform value as a determined score to obtain the effectiveness score of the processing solution corresponding to the data root node, where a calculation

$$p(y = k|x; w) = \frac{\exp(w_k^T x)}{\sum_{l=1}^K \exp(w_l^T x)}, \text{ where}$$

formula for the exponential transform value is as follows:

$p(y = k|x; w)$ is the exponential transform value; w is the occurrence probability of the confounding factor and the preset influence element; x is the confounding factor; k is the number of categories of the preset influence element; y is an output predicted value; y is a conditional probability of k ; K is a maximum value of k ; and T is a time period.

Optionally, the knowledge-driven and data-driven based grouping apparatus further includes: a processing unit, configured to obtain the medical disciplinary knowledge data and clinical data from a plurality of systems, platforms, and medical systems, and perform data processing on the medical disciplinary knowledge data and the clinical data to obtain the medical rule data, where the data processing includes data cleansing processing, data integration processing, data reduction processing, data transformation processing, feature extraction processing, data desensitization processing, and classification processing. The processing unit is specifically configured to: obtain the medical disciplinary knowledge data from a plurality of systems and platforms, and obtain the clinical data from

a plurality of medical systems; perform data pre-processing separately on the medical disciplinary knowledge data and the clinical data, where the data pre-processing includes data cleansing processing, data integration processing, data reduction processing, and data transformation processing; perform feature extraction on the medical disciplinary knowledge data obtained after the data pre-processing to obtain feature information, and perform data desensitization processing and classification on the clinical data obtained after the data pre-processing to obtain first sample data and third sample data; and analyze the feature information, and classify the feature information based on a preset classification rule to obtain the medical rule data, where the preset classification rule includes same and/or similar diseases.

Optionally, the identification unit 602 is specifically configured to: calculate similarities between the processing solutions to obtain a plurality of similarity values, and determine by comparison whether the plurality of similarity values are all in a preset range; if yes, identify that there is one category of the processing solutions through statistics collection; and if no, identify that there are multiple categories of the processing solutions through statistics collection.

Optionally, the knowledge-driven and data-driven based grouping apparatus further includes an invoking unit, configured to invoke a preset interpretation model to perform interpretation processing and visual chart-based statistical analysis processing on the first feature similar crowd or the second feature similar crowd to obtain an interpretable feature similar crowd, where the interpretation processing includes effectiveness score sorting processing, matching degree calculation processing, and matching degree sorting processing. The invoking unit is specifically configured to invoke the preset interpretation model to sort the effectiveness corresponding to the first feature similar crowd or the second feature similar crowd in descending order of values to obtain first sorting information; calculate a matching degree between the to-be-classified crowd and the first feature similar crowd or the second feature similarity crowd and sort the matching degrees in descending order of values to obtain second sorting information; sort the first feature similar crowd or the second feature similar crowd based on the first sorting information and/or second sorting information, and label the first feature similar crowd or the second feature similar crowd based on the first sorting information and/or second sorting information; and perform visual chart-based statistical analysis processing on the labeled first feature similar crowd or second feature similar crowd to generate a visual chart, so as to obtain the interpretable feature similar crowd.

In the embodiments of the present application, a target knowledge-based grouping decision tree is created; it is determined that a to-be-classified crowd belongs to a first feature similar crowd, or a data-based grouping decision tree is created and the to-be-classified crowd is classified as the first feature similar crowd or a second feature similar crowd based on a target layer number; and other steps are performed, thereby improving an accuracy degree of grouping a to-be-classified crowd into a feature similar crowd, and further providing a doctor with reasonable, standardized, comprehensive, and refined evidence-based medical information.

The knowledge-driven and data-driven based grouping apparatus in the embodiments of the present application is described in detail above from the perspective of modular functional entities with reference to FIG. 6 and FIG. 7. A knowledge-driven and data-driven based grouping device in an embodiment of the present application is described in detail below from the perspective of hardware processing.

FIG. 8 is a schematic structural diagram of a knowledge-driven and data-driven based grouping device according to an embodiment of the present application. The knowledge-driven and data-driven based grouping device 800 may have relatively large differences due to a configuration or performance difference and may include one or more processors (central processing units (CPUs)) 801, a memory 809, and one or more storage media 808 (such as one or more mass storage devices) that store an application program 807 or data 806. The memory 809 and the storage medium 808 may be used for short-time storage or persistent storage. The program stored in the storage medium 808 may include one or more modules (not shown in the figure), and each module may include a series of operation instructions in the knowledge-driven and data-driven based grouping device. Further, the processor 801 may be configured to communicate with the storage medium 808, and execute a series of operation instructions in the storage medium 808 on the knowledge-driven and data-driven based grouping device 800.

The knowledge-driven and data-driven based grouping device 800 may further include one or more power supplies 802, one or more wired or wireless network interfaces 803, one or more input/output interfaces 804, and/or one or more operating systems 805, such as Windows Server, Mac OS X, Unix, Linux and FreeBSD. It can be understood by a person skilled in the art that the structure of the knowledge-driven and data-driven based grouping device shown in FIG. 8 does not constitute a limitation to the knowledge-driven and data-driven based grouping device, and may include more or

fewer components than those shown, or combine some components, or have different component arrangements. The processor 801 may execute functions of the first creation unit 601, the identification unit 602, the determining unit 603, the second creation unit 604, the first classification unit 605, and the second classification unit 606, or functions of the first creation unit 601, the identification unit 602, the determining unit 603, the second creation unit 604, the first classification unit 605, the judgment unit 606, the second classification unit 607, the analysis unit 608, and the third classification unit 609 in the foregoing embodiments.

Components of the knowledge-driven and data-driven based grouping device are described in detail with reference to FIG. 8:

The processor 801 is a control center of the knowledge-driven and data-driven based grouping device, and may perform processing based on the knowledge-driven and data-driven based grouping method. The processor 801 uses a variety of interfaces and lines to connect parts of the entire knowledge-driven and data-driven based grouping device, and executes various functions of the knowledge-driven and data-driven based grouping device and processes data by running or executing a software program and/or a module stored in the memory 809 and invoking data stored in the memory 809, so as to improve availability of a medical assistant decision system. Both the storage medium 808 and the memory 809 are carriers for storing data. In the embodiments of the present application, the storage medium 808 may be an internal memory with a small storage capacity but a fast speed, while the memory 809 may be an external memory with a large storage capacity but a slow storage speed.

The memory 809 may be configured to store a software program and a module. The processor 801 performs various function applications of the knowledge-driven and data-driven based grouping device 800 and processes data by running the software program and the module stored in the memory 809. The memory 809 may mainly include a program storage area and a data storage area. The program storage area may store an operating system, an application program required by at least one function (obtaining feature data of a to-be-classified crowd, etc.), and the like. The data storage area may store data created for using the knowledge-driven and data-driven based grouping device (feature data is grouped through a target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end sub-node, and the number of categories of the processing solutions is identified, etc.), and the like. In addition, the memory 809 may include a high-speed random access memory, or may include a

nonvolatile memory, such as at least one magnetic disk storage component, a flash memory component, or another volatile solid-state storage component. The program of the knowledge-driven and data-driven based grouping method provided in the embodiments of the present application and received data streams are stored in the memory. The processor 801 invokes the program and the received data streams from the memory 809 when in need.

When computer program instructions are loaded and executed on a computer, some or all of processes or functions according to the embodiments of the present application are generated. The computer may be a general-purpose computer, a dedicated computer, a computer network, or another programmable apparatus. The computer instructions may be stored in a computer-readable storage medium or transmitted from one computer-readable storage medium to another. For example, the computer instructions may be transmitted from one website, computer, server, or data center to another website, computer, server, or data center in a wired mode (for example, a coaxial cable, an optical fiber, or a twisted pair cable) or a wireless mode (for example, infrared, a wireless connection, or a microwave). The computer-readable storage medium may be any medium usable by a computer for storage or a data storage device that integrates one or more usable media, such as a server or a data center. The usable medium may be a magnetic medium (for example, a floppy disk, a hard disk, or magnetic tape), an optical medium (for example, an optical disc), or a semiconductor medium (for example, a solid state disk (SSD)), or the like.

A person skilled in the art can clearly understand that for ease and brevity of description, for specific working processes of the system, apparatus and units described above, reference may be made to the corresponding processes in the foregoing method embodiments. Details are not repeated herein.

CLAIMS

1. A knowledge-driven and data-driven based grouping method, comprising:

creating an initial knowledge-based grouping decision tree based on medical rule data, and filling sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, wherein the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group a clinically-similar feature crowd; and the sample data comprises clinical result data corresponding to the clinically-similar feature crowd and compliant with a medical knowledge rule;

obtaining feature data of a to-be-classified crowd and grouping the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identifying the number of categories of the processing solutions, wherein the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd; and

if one category of the processing solutions is identified, determining that the to-be-classified crowd belongs to a first feature similar crowd, wherein the first feature similar crowd is a feature similar crowd to which the processing solutions corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree belong; or

if at least two categories of the processing solutions are identified, generating a data-based grouping decision tree based on a second grouping rule, and determining a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, wherein the second grouping rule indicates classification based on a recursive partitioning algorithm and effectiveness of the processing solutions;

wherein,

if the target layer number indicates a first layer, the to-be-classified crowd is classified as the first feature similar crowd; or

if the target layer number does not indicate the first layer, the to-be-classified crowd is classified as a second feature similar crowd, wherein the second feature similar crowd is a feature similar crowd to which a processing solution corresponding to the target data node on the data-based grouping decision tree belongs; and each target data node corresponds to a different feature similar crowd.

2. The knowledge-driven and data-driven based grouping method according to claim 1, wherein the generating a data-based grouping decision tree based on a second grouping rule, and determining a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree comprises:

step I: using the target end knowledge sub-node as a data root node of the data-based grouping decision tree, and calculating an effectiveness score of a processing solution in sample data corresponding to the data root node;

step II: branching the data root node based on an effectiveness score of each processing solution, classifying each processing solution to a data sub-node of each branch in sequence based on the recursive partitioning algorithm, and continuing with effectiveness analysis on a processing solution corresponding to each data sub-node;

step III: if the effectiveness score of the processing solution corresponding to a current data sub-node is greater than a first threshold, branching the current data sub-node;

step IV: repeating step III until the effectiveness score of the processing solution corresponding to each data sub-node is less than or equal to the first threshold to obtain the data-based grouping decision tree, wherein the data-based grouping decision tree comprises a plurality of data sub-nodes; and each end data leaf sub-node in the plurality of data sub-nodes corresponds to one processing solution; and

step V: determining a target layer number of the end data leaf sub-node corresponding to each processing solution on the data-based grouping decision tree.

3. The knowledge-driven and data-driven based grouping method according to claim 2, wherein the using the target end knowledge sub-node as a data root node of the data-based grouping decision tree, and calculating an effectiveness score of a processing solution corresponding to the data root node comprises:

using the target end knowledge sub-node as a data root node of the data-based grouping decision tree;

obtaining target feature information of the processing solution in the sample data corresponding to the data root node, and dividing the target feature information into element factors;

calculating association degrees, influence degrees on a preset influence element, and uncontrolled edge association degrees of the element factors on different levels;

analyzing the association degrees, the influence degrees, and the edge association degrees through a preset compressibility criterion and comparability criterion to obtain a confounding factor; and

calculating an exponential transform value of a regression coefficient of an occurrence probability of the confounding factor and the preset influence element, and using the exponential transform value as a determined score to obtain the effectiveness score of the processing solution corresponding to the data root node, wherein a calculation formula for the exponential transform value is as follows:

$$p(y = k|x; w) = \frac{\exp(w_k^T x)}{\sum_{l=1}^K \exp(w_l^T x)},$$

wherein $p(y = k|x; w)$ is the exponential transform value; w is the occurrence probability of the confounding factor and the preset influence element; x is the confounding factor; k is the number of categories of the preset influence element; y is an output predicted value; y is a conditional probability of k ; K is a maximum value of k ; and T is a time period.

4. The knowledge-driven and data-driven based grouping method according to claim 2, wherein the classifying the to-be-classified crowd as a second feature similar crowd if the target layer number does not indicate the first layer comprises:

if the target layer number does not indicate the first layer, determining whether an effectiveness score of a processing solution corresponding to the target layer number is greater than a second threshold, wherein the second threshold is greater than the first threshold; and

if the effectiveness score of the processing solution corresponding to the target layer number is greater than the second threshold, using a data leaf sub-node on which the processing solution with the effectiveness score greater than the second threshold is located as a target data node, and classifying the to-be-classified crowd as a feature similar crowd to which the processing solution corresponding to the target data node belongs, so as to obtain the second feature similar crowd; or

if the effectiveness score of the processing solution corresponding to the target layer number is less than or equal to the second threshold, analyzing an effectiveness score of a processing solution

corresponding to an upper-layer number of the target layer number; and

if the effectiveness score of the processing solution corresponding to the upper-layer number of the target layer number is greater than the second threshold, using a data leaf sub-node on which the processing solution corresponding to the upper-layer number of the target layer number is located as a target data node, and classifying the to-be-classified crowd as a feature similar crowd to which the processing solution corresponding to the target data node belongs, so as to obtain the second feature similar crowd.

5. The knowledge-driven and data-driven based grouping method according to claim 1, wherein before the creating an initial knowledge-based grouping decision tree based on medical rule data, and filling sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, the method further comprises:

obtaining the medical disciplinary knowledge data and clinical data from a plurality of systems, platforms, and medical systems, and performing data processing on the medical disciplinary knowledge data and the clinical data to obtain the medical rule data, wherein the data processing comprises data cleansing processing, data integration processing, data reduction processing, data transformation processing, feature extraction processing, data desensitization processing, and classification processing.

6. The knowledge-driven and data-driven based grouping method according to claim 1, wherein the identifying the number of categories of the processing solutions comprises:

calculating similarities between the processing solutions to obtain a plurality of similarity values, and determining by comparison whether the plurality of similarity values are all in a preset range;

if yes, identifying that there is one category of the processing solutions through statistics collection; and

if no, identifying that there are multiple categories of the processing solutions through statistics collection.

7. The knowledge-driven and data-driven based grouping method according to any one of claims 1 to 6, wherein after the classifying the to-be-classified crowd as a second feature similar crowd if the target layer number does not indicate the first layer, the method further comprises:

invoking a preset interpretation model to perform interpretation processing and visual chart-based

statistical analysis processing on the first feature similar crowd or the second feature similar crowd to obtain an interpretable feature similar crowd, wherein the interpretation processing comprises effectiveness score sorting processing, matching degree calculation processing, and matching degree sorting processing.

8. A knowledge-driven and data-driven based grouping apparatus, comprising:

a first creation unit, configured to create an initial knowledge-based grouping decision tree based on medical rule data, and fill sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, wherein the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group a clinically-similar feature crowd; and the sample data comprises clinical result data corresponding to the clinically-similar feature crowd and compliant with a medical knowledge rule;

an identification unit, configured to obtain feature data of a to-be-classified crowd and group the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identify the number of categories of the processing solutions, wherein the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd;

a determining unit, configured to, if one category of the processing solutions is identified, determine that the to-be-classified crowd belongs to a first feature similar crowd, wherein the first feature similar crowd is a feature similar crowd to which the processing solutions corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree belong;

a second creation unit, configured to, if at least two categories of the processing solutions are identified, generate a data-based grouping decision tree based on a second grouping rule, and determine a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, wherein the second grouping rule indicates classification based on a recursive partitioning algorithm and effectiveness of the processing solutions;

a first classification unit, configured to classify the to-be-classified crowd as the first feature similar

crowd if the target layer number indicates the first layer; and

a second classification unit, configured to classify the to-be-classified crowd as a second feature similar crowd if the target layer number does not indicate the first layer, wherein the second feature similar crowd is a feature similar crowd to which a processing solution corresponding to the target data node on the data-based grouping decision tree belongs; and each target data node corresponds to a different feature similar crowd.

9. A knowledge-driven and data-driven based grouping device, comprising:

a memory, a processor, and a computer program stored in the memory and executable on the processor, wherein the processor implements the following steps of a knowledge-driven and data-driven based grouping method when executing the computer program:

creating an initial knowledge-based grouping decision tree based on medical rule data, and filling sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, wherein the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group a clinically-similar feature crowd; and the sample data comprises clinical result data corresponding to the clinically-similar feature crowd and compliant with a medical knowledge rule;

obtaining feature data of a to-be-classified crowd and grouping the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identifying the number of categories of the processing solutions, wherein the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd; and

if one category of the processing solutions is identified, determining that the to-be-classified crowd belongs to a first feature similar crowd, wherein the first feature similar crowd is a feature similar crowd to which the processing solutions corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree belong; or

if at least two categories of the processing solutions are identified, generating a data-based grouping

decision tree based on a second grouping rule, and determining a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, wherein the second grouping rule indicates classification based on a recursive partitioning algorithm and effectiveness of the processing solutions;

wherein,

if the target layer number indicates a first layer, the to-be-classified crowd is classified as the first feature similar crowd; or

if the target layer number does not indicate the first layer, the to-be-classified crowd is classified as a second feature similar crowd, wherein the second feature similar crowd is a feature similar crowd to which a processing solution corresponding to the target data node on the data-based grouping decision tree belongs; and each target data node corresponds to a different feature similar crowd.

10. The knowledge-driven and data-driven based grouping device according to claim 9, wherein the generating a data-based grouping decision tree based on a second grouping rule, and determining a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree comprises the following steps:

step I: using the target end knowledge sub-node as a data root node of the data-based grouping decision tree, and calculating an effectiveness score of a processing solution in sample data corresponding to the data root node;

step II: branching the data root node based on an effectiveness score of each processing solution, classifying each processing solution to a data sub-node of each branch in sequence based on the recursive partitioning algorithm, and continuing with effectiveness analysis on a processing solution corresponding to each data sub-node;

step III: if the effectiveness score of the processing solution corresponding to a current data sub-node is greater than a first threshold, branching the current data sub-node;

step IV: repeating step III until the effectiveness score of the processing solution corresponding to each data sub-node is less than or equal to the first threshold to obtain the data-based grouping decision tree, wherein the data-based grouping decision tree comprises a plurality of data sub-nodes; and each end data leaf sub-node in the plurality of data sub-nodes corresponds to one processing solution; and

step V: determining a target layer number of the end data leaf sub-node corresponding to each

processing solution on the data-based grouping decision tree.

11. The knowledge-driven and data-driven based grouping device according to claim 10, wherein the using the target end knowledge sub-node as a data root node of the data-based grouping decision tree, and calculating an effectiveness score of a processing solution corresponding to the data root node comprises the following steps:

using the target end knowledge sub-node as a data root node of the data-based grouping decision tree;

obtaining target feature information of the processing solution in the sample data corresponding to the data root node, and dividing the target feature information into element factors;

calculating association degrees, influence degrees on a preset influence element, and uncontrolled edge association degrees of the element factors on different levels;

analyzing the association degrees, the influence degrees, and the edge association degrees through a preset compressibility criterion and comparability criterion to obtain a confounding factor; and

calculating an exponential transform value of a regression coefficient of an occurrence probability of the confounding factor and the preset influence element, and using the exponential transform value as a determined score to obtain the effectiveness score of the processing solution corresponding to the data root node, wherein a calculation formula for the exponential transform value is as follows:

$$p(y = k|x; w) = \frac{\exp(w_k^T x)}{\sum_{l=1}^K \exp(w_l^T x)},$$

wherein $p(y = k|x; w)$ is the exponential transform value; w is the occurrence probability of the confounding factor and the preset influence element; x is the confounding factor; k is the number of categories of the preset influence element; y is an output predicted value; y is a conditional probability of k ; K is a maximum value of k ; and T is a time period.

12. The knowledge-driven and data-driven based grouping device according to claim 10, wherein the classifying the to-be-classified crowd as a second feature similar crowd if the target layer number does not indicate the first layer comprises the following steps:

if the target layer number does not indicate the first layer, determining whether an effectiveness score of a processing solution corresponding to the target layer number is greater than a second threshold, wherein the second threshold is greater than the first threshold; and

if the effectiveness score of the processing solution corresponding to the target layer number is greater than the second threshold, using a data leaf sub-node on which the processing solution with the effectiveness score greater than the second threshold is located as a target data node, and classifying the to-be-classified crowd as a feature similar crowd to which the processing solution corresponding to the target data node belongs, so as to obtain the second feature similar crowd; or

if the effectiveness score of the processing solution corresponding to the target layer number is less than or equal to the second threshold, analyzing an effectiveness score of a processing solution corresponding to an upper-layer number of the target layer number; and

if the effectiveness score of the processing solution corresponding to the upper-layer number of the target layer number is greater than the second threshold, using a data leaf sub-node on which the processing solution corresponding to the upper-layer number of the target layer number is located as a target data node, and classifying the to-be-classified crowd as a feature similar crowd to which the processing solution corresponding to the target data node belongs, so as to obtain the second feature similar crowd.

13. The knowledge-driven and data-driven based grouping device according to claim 9, wherein before the creating an initial knowledge-based grouping decision tree based on medical rule data, and filling sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, the method further comprises the following steps:

obtaining the medical disciplinary knowledge data and clinical data from a plurality of systems, platforms, and medical systems, and performing data processing on the medical disciplinary knowledge data and the clinical data to obtain the medical rule data, wherein the data processing comprises data cleansing processing, data integration processing, data reduction processing, data transformation processing, feature extraction processing, data desensitization processing, and classification processing.

14. The knowledge-driven and data-driven based grouping device according to claim 9, wherein the identifying the number of categories of the processing solutions comprises the following steps:

calculating similarities between the processing solutions to obtain a plurality of similarity values, and determining by comparison whether the plurality of similarity values are all in a preset range;

if yes, identifying that there is one category of the processing solutions through statistics collection;

and

if no, identifying that there are multiple categories of the processing solutions through statistics collection.

15. The knowledge-driven and data-driven based grouping device according to any one of claims 9 to 14, wherein after the classifying the to-be-classified crowd as a second feature similar crowd if the target layer number does not indicate the first layer, the method further comprises the following steps:

invoking a preset interpretation model to perform interpretation processing and visual chart-based statistical analysis processing on the first feature similar crowd or the second feature similar crowd to obtain an interpretable feature similar crowd, wherein the interpretation processing comprises effectiveness score sorting processing, matching degree calculation processing, and matching degree sorting processing.

16. A computer-readable storage medium, comprising instructions, wherein when the instructions are run on a computer, the computer is enabled to perform the following steps of a knowledge-driven and data-driven based grouping method:

creating an initial knowledge-based grouping decision tree based on medical rule data, and filling sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, wherein the medical rule data is used to indicate a first grouping rule obtained by performing rule extraction processing on medical disciplinary knowledge data of archived medical files, medical consensus, medical guidelines, and expert clinical experience; the first grouping rule is used to group a clinically-similar feature crowd; and the sample data comprises clinical result data corresponding to the clinically-similar feature crowd and compliant with a medical knowledge rule;

obtaining feature data of a to-be-classified crowd and grouping the feature data through the target knowledge-based grouping decision tree to obtain processing solutions corresponding to a target end knowledge sub-node, and identifying the number of categories of the processing solutions, wherein the target end knowledge sub-node is a node position of the feature data on the target knowledge-based grouping decision tree; and the processing solutions are used to indicate solution channels for the feature similar crowd; and

if one category of the processing solutions is identified, determining that the to-be-classified crowd

belongs to a first feature similar crowd, wherein the first feature similar crowd is a feature similar crowd to which the processing solutions corresponding to the target end knowledge sub-node on the target knowledge-based grouping decision tree belong; or

if at least two categories of the processing solutions are identified, generating a data-based grouping decision tree based on a second grouping rule, and determining a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, wherein the second grouping rule indicates classification based on a recursive partitioning algorithm and effectiveness of the processing solutions;

wherein,

if the target layer number indicates a first layer, the to-be-classified crowd is classified as the first feature similar crowd; or

if the target layer number does not indicate the first layer, the to-be-classified crowd is classified as a second feature similar crowd, wherein the second feature similar crowd is a feature similar crowd to which a processing solution corresponding to the target data node on the data-based grouping decision tree belongs; and each target data node corresponds to a different feature similar crowd.

17. The computer-readable storage medium according to claim 16, wherein when the instructions for knowledge-driven and data-driven based grouping are executed by the computer to perform the step of generating a data-based grouping decision tree based on a second grouping rule, and determining a target layer number of a data sub-node corresponding to each processing solution on the data-based grouping decision tree, the following steps are comprised:

step I: using the target end knowledge sub-node as a data root node of the data-based grouping decision tree, and calculating an effectiveness score of a processing solution in sample data corresponding to the data root node;

step II: branching the data root node based on an effectiveness score of each processing solution, classifying each processing solution to a data sub-node of each branch in sequence based on the recursive partitioning algorithm, and continuing with effectiveness analysis on a processing solution corresponding to each data sub-node;

step III: if the effectiveness score of the processing solution corresponding to a current data sub-node is greater than a first threshold, branching the current data sub-node;

step IV: repeating step III until the effectiveness score of the processing solution corresponding to each data sub-node is less than or equal to the first threshold to obtain the data-based grouping decision tree, wherein the data-based grouping decision tree comprises a plurality of data sub-nodes; and each end data leaf sub-node in the plurality of data sub-nodes corresponds to one processing solution; and

step V: determining a target layer number of the end data leaf sub-node corresponding to each processing solution on the data-based grouping decision tree.

18. The computer-readable storage medium according to claim 17, wherein when the instructions for knowledge-driven and data-driven based grouping are executed by the computer to perform the step of using the target end knowledge sub-node as a data root node of the data-based grouping decision tree, and calculating an effectiveness score of a processing solution corresponding to the data root node, the following steps are comprised:

using the target end knowledge sub-node as a data root node of the data-based grouping decision tree;

obtaining target feature information of the processing solution in the sample data corresponding to the data root node, and dividing the target feature information into element factors;

calculating association degrees, influence degrees on a preset influence element, and uncontrolled edge association degrees of the element factors on different levels;

analyzing the association degrees, the influence degrees, and the edge association degrees through a preset compressibility criterion and comparability criterion to obtain a confounding factor; and

calculating an exponential transform value of a regression coefficient of an occurrence probability of the confounding factor and the preset influence element, and using the exponential transform value as a determined score to obtain the effectiveness score of the processing solution corresponding to the data root node, wherein a calculation formula for the exponential transform value is as follows:

$$p(y = k|x; w) = \frac{\exp(w_k^T x)}{\sum_{l=1}^K \exp(w_l^T x)},$$

wherein $p(y = k|x; w)$ is the exponential transform value; w is the occurrence probability of the confounding factor and the preset influence element; x is the confounding factor; k is the number of categories of the preset influence element; y is an output predicted value; y is a conditional probability of k ; K is a maximum value of k ; and T is a time period.

19. The computer-readable storage medium according to claim 17, wherein when the instructions for knowledge-driven and data-driven based grouping are executed by the computer to perform the step of classifying the to-be-classified crowd as a second feature similar crowd if the target layer number does not indicate the first layer, the following steps are comprised:

if the target layer number does not indicate the first layer, determining whether an effectiveness score of a processing solution corresponding to the target layer number is greater than a second threshold, wherein the second threshold is greater than the first threshold; and

if the effectiveness score of the processing solution corresponding to the target layer number is greater than the second threshold, using a data leaf sub-node on which the processing solution with the effectiveness score greater than the second threshold is located as a target data node, and classifying the to-be-classified crowd as a feature similar crowd to which the processing solution corresponding to the target data node belongs, so as to obtain the second feature similar crowd; or

if the effectiveness score of the processing solution corresponding to the target layer number is less than or equal to the second threshold, analyzing an effectiveness score of a processing solution corresponding to an upper-layer number of the target layer number; and

if the effectiveness score of the processing solution corresponding to the upper-layer number of the target layer number is greater than the second threshold, using a data leaf sub-node on which the processing solution corresponding to the upper-layer number of the target layer number is located as a target data node, and classifying the to-be-classified crowd as a feature similar crowd to which the processing solution corresponding to the target data node belongs, so as to obtain the second feature similar crowd.

20. The computer-readable storage medium according to claim 16, wherein before the instructions for knowledge-driven and data-driven based grouping are executed by the computer to perform the step of creating an initial knowledge-based grouping decision tree based on medical rule data, and filling sample data to an end knowledge sub-node of each branch of the initial knowledge-based grouping decision tree to obtain a target knowledge-based grouping decision tree, the following steps are further comprised:

obtaining the medical disciplinary knowledge data and clinical data from a plurality of systems, platforms, and medical systems, and performing data processing on the medical disciplinary knowledge

data and the clinical data to obtain the medical rule data, wherein the data processing comprises data cleansing processing, data integration processing, data reduction processing, data transformation processing, feature extraction processing, data desensitization processing, and classification processing.