

SOCIAL SECURITY FRAUD BEHAVIORS IDENTIFICATION METHOD, DEVICE, APPARATUS AND COMPUTER-READABLE STORAGE MEDIUM

CROSS-REFERENCE TO RELATED APPLICATIONS

5 **[0001]** The present application is the National Stage of International Application with No. PCT/CN2018/074806, filed January 31, 2018, which claims the benefit of China Patent Application No. 201710091766.9, filed February 20, 2017 with the State Intellectual Property Office and entitled “Social security fraud behaviors Identification Method and Device”, the entirety of which is hereby incorporated
10 herein by reference.

[0002] FIELD

[0003] The present disclosure relates to the field of computer application, and more particularly relates to a social security fraud behaviors identification method, a social security fraud behaviors identification device, and a social security fraud
15 behaviors identification apparatus, as well as a computer-readable storage medium.

[0004] BACKGROUND

[0005] With the improvement of living standards, people has gradually increased awareness of security assurance about personal safety and security. More and more peoples are willing to purchase social security to reduce the economic burden when
20 accidents occur. Consequently, some social security fraud behaviors also generated, for example, making forgery of medical case, changing medical costs, and so on.

[0006] Currently, in most cases, a single rule triggering mechanism is usually used to identify the social security fraud behaviors. However, the single rule triggering mechanism has the defect of using a single evidence chain for identifying the social
25 security fraud behaviors, which is prone to lead to a high rate of misjudgment.

characteristics extracted is input into a preset classification model, in order to identify fraud rate of each node according to the classification model. Specifically, referring to Fig.3, the step S30 includes:

[0053] S31, calculating the similarity of the multi-dimensional group medical treatment characteristics of each node having the same attribute, according to the multi-dimensional group medical treatment characteristics corresponding to each node;

[0054] S32, inputting the calculated similarity of each node into the preset classification model, to calculate fraud rate of each node according to a preset fraud detection formula in the classification model.

[0055] That is to say, after extracting the multi-dimensional group medical treatment characteristics corresponding to each node, the similarity of the multi-dimensional group medical treatment characteristics of each node having the same attribute is calculated. The nodes having the same attribute are such as: a doctor node and another doctor node, or a patient node and another patient node.

[0056] In the exemplary embodiment, the calculating the similarity of the multi-dimensional group medical treatment characteristics of each node having the same attribute preferably takes the following algorithms:

[0057] 1) Jaccard Similarity (representing the generalized similarity):

[0058] $Jaccard(A, B) = |A \cap B| / |A \cup B|$

[0059] wherein, intersect represents intersection, union represents union, A and B represent nodes having the same attribute, for example, A and B both represent the doctor node in Fig.9, or both represent the patient node.

[0060] 2) Euclidean similarity (similarity of Euclidean distance):

[0061] $Euclidean(A, B) = 1 - euclidean_distance(A, B)$

[0062] wherein, A and B represent nodes having the same attribute.

[0063] The two algorithms listed above for calculating the similarity of the multi-dimensional group medical treatment characteristics of each node having the same attribute are merely illustrative. Other algorithms proposed by those skilled in the art by taking use of the technical idea of the present disclosure and according to

treatment characteristics corresponding to each node, the calculating unit 31 calculates the similarity of the multi-dimensional group medical treatment characteristics of each node having the same attribute. The nodes having the same attribute are such as: a doctor node and a doctor node, or a patient node and a patient node.

5 **[0101]** In the exemplary embodiment, the calculating unit 31 calculates the similarity of the multi-dimensional group medical treatment characteristics of each node having the same attribute preferably takes the following algorithms:

[0102] 1) Jaccard Similarity (representing the generalized similarity):

[0103]
$$\text{Jaccard}(A, B) = |A \cap B| / |A \cup B|$$

10 **[0104]** wherein, intersect represents intersection, union represents union, A and B represent nodes having the same attribute, for example, A and B both represent the doctor node in Fig.9, or both represent the patient node.

[0105] 2) Euclidean similarity (similarity of Euclidean distance):

[0106]
$$\text{Euclidean}(A, B) = 1 - \text{euclidean_distance}(A, B)$$

15 **[0107]** wherein, A and B represent nodes having the same attribute.

[0108] The two algorithms listed above for calculating the similarity of the multi-dimensional group medical treatment characteristics of each node with the same attribute are merely illustrative. Other algorithms proposed by those skilled in the art by taking use of the technical idea of the present disclosure and according to their
20 specific requirements are within the protection scope of the present disclosure. Thus, they are not to be detailed herein.

[0109] By the above similarity calculation formulas, the similarity of the multi-dimensional group medical treatment characteristics of any two nodes having the same attribute can be determined.

25 **[0110]** That is to say, after extracting the multi-dimensional group medical treatment characteristics corresponding to each node, the similarity of the multi-dimensional group medical treatment characteristics of each node having the same attribute is calculated. The nodes having the same attribute are such as: a doctor node and another doctor node, or a patient node and another patient node, namely, the
30 nodes with the same type represent the nodes having the same attribute.

[0111] In the exemplary embodiment, the calculating the similarity of the multi-dimensional group medical treatment characteristics of each node having the same attribute preferably takes the following algorithms:

[0112] 1) Jaccard Similarity (representing the generalized similarity):

5 **[0113]** $Jaccard(A, B) = |A \cap B| / |A \cup B|$

[0114] wherein, Intersect represents intersection, Union represents union, A and B represent nodes having the same attribute, for example, A and B both represent nodes of the doctor in Fig.9, or both represent nodes of the patient.

[0115] 2) Euclidean similarity (similarity of Euclidean distance):

10 **[0116]** $Euclidean(A, B) = 1 - euclidean_distance(A, B)$

[0117] wherein, A and B represent nodes having the same attribute.

[0118] After determining the similarity of the multiple-dimensional group medical treatment characteristics of each node having the same attribute, the inputting unit 32 inputs the calculated similarity of each node into a preset classification model, to
15 calculate fraud rate of each node in accordance with a preset fraud detection formula in the classification model. Wherein, the fraud detection formula preferably includes: KNN (k-Nearest Neighbor algorithm, K taking the value 5) algorithm formula; binary Kmeans algorithm formula; Shewhart methods algorithm formula, and so on. For these algorithm formulas are existing formulas, the calculating processes are not to be
20 detailed herein.

[0119] Further, in order to improve accuracy of the classification model calculating fraud rate of nodes, in the exemplary embodiment, the social security fraud behaviors identification device 100 further includes:

[0120] a verifying module, configured to verify the fraud rate of each node, and
25 adding the verification conclusion to the fraud rate of each node;

[0121] a training module, configured to re-input the fraud rate with the verification conclusion into the classification model for training the classification model.

[0122] That is to say, after calculating fraud rate of each node according to the
30 preset fraud detection formula in the classification model, the verifying module can