

CLAIMS

1. A computer implemented method for identifying money laundering risks, the method comprising:

obtaining (602) a to-be-identified dataset, wherein the to-be-identified dataset comprises at least two pieces of to-be-identified data;

5 constructing (604) a first relational network of the at least two pieces of to-be-identified data, and representing the at least two pieces of to-be-identified data by using nodes of the first relational network,

wherein each node in the first relational network is connected to a set of adjacent nodes through a set of outgoing edges, each outgoing edge comprising
10 a weight value, wherein the weight value represents an amount of money transferred between accounts;

obtaining (606) a vectorized representation of each node in the first relational network based on a predetermined network embedding learning model, comprising:

for each node in the first relational network,

15 calculating a transition probability for each outgoing edge of the node based on i) the weight value of the outgoing edge and ii) a sum of weight values of the set of outgoing edges;

generating a first node sequence using a random walk process based on the transition probability of each outgoing edge; and

20 vectorizing each node in the first node sequence based on the predetermined network embedding learning model; and

determining (608), based on a pre-trained classification model, a data label corresponding to the vectorized representation of each node in the first relational network,

25 wherein training the pre-trained classification model comprises obtaining a training sample dataset based on a predetermined time interval.

2. The method according to claim 1, wherein obtaining (606) a vectorized representation of each node in the first relational network based on a predetermined
30 network embedding learning model comprises:

performing sequence sampling on each node in the first relational network by

using a random walk algorithm, and generating a first node sequence; and
vectorizing each node in the first node sequence based on the predetermined network embedding learning model.

5 3. The method according to claim 2, wherein vectorizing each node in the first node sequence based on the predetermined network embedding learning model comprises:

vectorizing (708) each node in the first node sequence based on a Node2vec network embedding learning model.

10

4. The method according to claim 2, wherein vectorizing each node in the first node sequence based on the predetermined network embedding learning model comprises:

15 vectorizing each node in the first node sequence based on a DeepWalk network embedding learning model.

5. The method according to claim 3, wherein vectorizing (708) each node in the first node sequence based on a Node2vec network embedding learning model comprises:

20 vectorizing, by the Node2vec network embedding learning model, each node in the first node sequence based on the SkipGram framework of Word2vec.

6. The method according to claim 1, wherein the data label comprises a white data label and a black data label, the method preferably further comprising:

25 if the data label is a black data label, performing subsequent processing on the vectorized representation, corresponding to the data label, of the node in the first relational network.

7. The method according to claim 1, wherein the to-be-identified dataset comprises a set of to-be-identified accounts that determines, by using the classification model,
30 whether there is a money laundering behavior.

8. The method according to one of the preceding claims, wherein the training of the pre-trained classification model comprises:

obtaining (202) the training sample dataset, wherein the training sample dataset

comprises at least two pieces of sample data and a sample label corresponding to each piece of sample data;

constructing (204) a second relational network of the at least two pieces of sample data, and representing the at least two pieces of sample data by using nodes of the

5 second relational network;

obtaining (206) a vectorized representation of each node in the second relational network based on a predetermined network embedding learning model; and

training (208) a classification model by using the training sample dataset to obtain the classification model, wherein the classification model associates the sample label
10 with the vectorized representation of each node in the second relational network.

9. The method according to claim 8, wherein obtaining (206) a vectorized representation of each node in the second relational network based on a predetermined network embedding learning model comprises:

15 performing (302) sequence sampling on each node in the second relational network by using a random walk algorithm, and generating a second node sequence; and

vectorizing (304) each node in the second node sequence based on the predetermined network embedding learning model.

20

10. The method according to claim 9, wherein vectorizing (304) each node in the second node sequence based on the predetermined network embedding learning model comprises:

vectorizing (510) each node in the second node sequence based on a Node2vec
25 network embedding learning model.

11. The method according to claim 9, wherein vectorizing (304) each node in the second node sequence based on the predetermined network embedding learning model comprises:

30 vectorizing each node in the second node sequence based on a DeepWalk network embedding learning model.

12. The method according to claim 10, wherein vectorizing (510) each node in the second node sequence based on a Node2vec network embedding learning model

comprises:

vectorizing, by the Node2vec network embedding learning model, each node in the second node sequence based on the SkipGram framework of Word2vec.

5 13. The method according to claim 8, wherein the classification model comprises an XGBoost model, a random forest model, a support vector machine model, or a logistic regression model.

10 14. A computing device, comprising a memory, a processor, and computer instructions that are stored in the memory and that can run on the processor, wherein steps of the method according to any one of the preceding claims are implemented when the processor executes the instructions.