CLAIMS

What is claimed is:

1. A graphical structure model-based method for risk control, comprising:

obtaining a graphical structure model trained by using a labeled sample, wherein the graphical structure model is defined based on an enterprise relationship network, the enterprise relationship network is composed of one or more nodes among enterprises, accounts, executives, civil servants, and legal persons and a relationship between the nodes, and the graphical structure model is used to calculate embedding vectors of the node after a plurality of iterations in hidden feature space based on original features of the node and/or edge features associated with the nodes, and wherein the graphical structure model is further used to calculate a predicted probability of the node based on the embedding vector, wherein the predicted probability indicates a probability that the node is a risk node;

calculating an embedding vector corresponding to a sample to be tested by using the trained graphical structure model; and

performing risk control on the sample to be tested based on the embedding vector corresponding to the sample to be tested, wherein

the labeled sample is labeled whether a node corresponding to the labeled sample is a risk node, the performing risk control on the sample to be tested based on the embedding vector corresponding to the sample to be tested comprising:

calculating a predicted probability corresponding to the sample to be tested by using the trained graphical structure model based on the embedding vector corresponding to the sample to be tested; and

performing risk control on the sample to be tested based on the predicted probability corresponding to the sample to be tested.

2. The method according to claim 1, wherein the enterprise relationship network comprises one or a combination of the following networks: a shareholding relationship network between enterprises, a shareholding relationship network between a legal person and an enterprise, an employment relationship network between an executive and an enterprise, an association relationship network between an executive and a civil servant, an association relationship network between an executive and a civil servant, a belonging relationship network between an executive

and an account, a family relationship network between accounts, a friend relationship network between accounts, a fund transfer network between accounts, and a labeling network between accounts.

3. The method according to claim 1, wherein the edge feature comprises at least one type of the following data related to the edge-connected nodes: shareholding relationship data, subordinate relationship data, interpersonal relationship data, service relationship data, and evaluation data.

4. The method according to claim 1, wherein an embedding vector of the node after the tth iteration in the hidden feature space is calculated based on the original feature of the node, edge features associated with the node and another node, and an embedding vector of the node after the $(t-1)$th iteration in the hidden feature space.

5. The method according to claim 1, wherein the calculating embedding vectors of the node after a plurality of iterations in hidden feature space based on original features of the node and edge features associated with the nodes comprises:

calculating the embedding vectors of the node after the plurality of iterations in the hidden feature space based on the following equation:

$$H_i^t = \sigma\left(W_1 F_i + W_2 \sum_{j \in N_i} E_{ij} + W_3 \sum_{j \in N_i} \alpha_{ij} H_i^{t-1} \right), \text{ wherein}$$

H_i^t represents an embedding vector of the ith node after the tth iteration in the hidden feature space, σ represents a nonlinear transformation function, W_1, W_2, and W_3represent weight matrices, F_i represents an original feature of the ith node, E_{ij} represents edge features associated with the ith node and the jth node, N_i represents a set of adjacent nodes of the ith node, and α_{ij} represents a weight matrix for the ith node and the jth node.

6. The method according to claim 5, wherein the calculating a predicted probability of the node based on the embedding vector comprises:

calculating the predicted probability of the node based on the following equation:

$$pred_i = softmax\left(W_4 * H_i^T \right), \text{ wherein}$$

$pred_i$ represents a predicted probability of the ith node after T iterations, and W_4 represents a weight matrix.

7. The method according to claim 1, wherein training the graphical structure model by using the labeled sample comprises:

training the graphic structure by using the labeled sample, to improve consistency between the predicted probability and a corresponding sample labeling result.

8. The method according to claim 6, wherein training the graphical structure model by using the labeled sample comprises:

improving $\arg\min_{w_1,w_2,w_3,w_4} \sum_i corss_entrep(pred_i, y_i)$ by using a back propagation algorithm and the labeled sample, to obtain optimal W_1, W_2, W_3, and W_4, wherein

$corss_entrep$ represents a function used to calculate cross entropy.

9. A graphical structure model-based apparatus for risk control, comprising:

an acquisition module, configured to obtain a graphical structure model trained by using a labeled sample, wherein the graphical structure model is defined based on an enterprise relationship network,

the enterprise relationship network is composed of one or more nodes among enterprises, accounts, executives, civil servants, and legal persons and a relationship between the nodes, and the graphical structure model is used to calculate embedding vectors of the node after a plurality of iterations in hidden feature space based on original features of the node and/or edge features associated with the nodes, and wherein the graphical structure model is further used to calculate a predicted probability of the node based on the embedding vector, wherein the predicted probability indicates a probability that the node is a risk node;

a calculation module, configured to calculate an embedding vector corresponding to a sample to be tested by using the trained graphical structure model; and

a risk control module, configured to perform risk control on the sample to be tested based on the embedding vector corresponding to the sample to be tested, wherein

the labeled sample is labeled whether a node corresponding to the labeled sample is a risk node, wherein the risk control module is configured to perform risk control on the sample to be tested based on the embedding vector corresponding to the sample to be tested comprises:

the risk control module is configured to calculate a predicted probability

corresponding to the sample to be tested by using the trained graphical structure model based on the embedding vector corresponding to the sample to be tested; and

perform risk control on the sample to be tested based on the predicted probability corresponding to the sample to be tested.

10. The apparatus according to claim 9, wherein the enterprise relationship network comprises one or a combination of the following networks: a shareholding relationship network between enterprises, a shareholding relationship network between a legal person and an enterprise, an employment relationship network between an executive and an enterprise, an association relationship network between an executive and a civil servant, an association relationship network between an executive and a civil servant, a belonging relationship network between an executive and an account, a family relationship network between accounts, a friend relationship network between accounts, a fund transfer network between accounts, and a labeling network between accounts.

11. The apparatus according to claim 9, wherein the edge feature comprises at least one type of the following data related to the edge-connected nodes: shareholding relationship data, subordinate relationship data, interpersonal relationship data, service relationship data, and evaluation data.

12. The apparatus according to claim 9, wherein an embedding vector of the node after the tth iteration in the hidden feature space is calculated based on the original feature of the node, edge features associated with the node and another node, and an embedding vector of the node after the $(t-1)$th iteration in the hidden feature space.

13. The apparatus according to claim 9, wherein the calculating embedding vectors of the node after a plurality of iterations in hidden feature space based on original features of the node and edge features associated with the nodes comprises:

calculating, by the calculation module, the embedding vectors of the node after the plurality of iterations in the hidden feature space based on the following equation:

$$H_i^t = \sigma\left(W_1 F_i + W_2 \sum_{j \in N_i} E_{ij} + W_3 \sum_{j \in N_i} \alpha_{ij} H_i^{t-1} \right), \text{ wherein}$$

H_i^t represents an embedding vector of the ith node after the tth iteration in the hidden feature space, σ represents a nonlinear transformation function, W_1,

W_2, and W_3 represent weight matrices, F_i represents an original feature of the ith node, E_{ij} represents edge features associated with the ith node and the jth node, N_i represents a set of adjacent nodes of the ith node, and α_{ij} represents a weight matrix for the ith node and the jth node.

14. The apparatus according to claim 13, wherein the calculating a predicted probability of the node based on the embedding vector comprises:

calculating, by the calculation module, the predicted probability of the node based on the following equation:

$pred_i = softmax\left(W_4 * H_i^T\right)$, wherein

$pred_i$ represents a predicted probability of the ith node after T iterations, and W_4 represents a weight matrix.

15. The apparatus according to claim 9, wherein training the graphical structure model by using the labeled sample comprises:

training the graphic structure by using the labeled sample, to improve consistency between the predicted probability and a corresponding sample labeling result.

16. The apparatus according to claim 14, wherein training the graphical structure model by using the labeled sample comprises:

improving $\arg\min_{w_1,w_2,w_3,w_4} \sum_i corss_entrep\left(pred_i, y_i\right)$ by using a back propagation algorithm and the labeled sample, to obtain optimal W_1, W_2, W_3, and W_4, wherein

$corss_entrep$ represents a function used to calculate cross entropy.

17. A graphical structure model-based device for risk control, comprising:

at least one processor; and

at least one memory, connected to and communicating with the at least one processor, wherein

the at least one memory stores an instruction executable by the at least one processor, and when the instruction is executed by the at least one processor, the at least one processor is enabled to perform the following steps:

obtaining a graphical structure model trained by using a labeled sample, wherein the graphical structure model is defined based on an enterprise relationship network,

the enterprise relationship network is composed of one or more nodes among enterprises, accounts, executives, civil servants, and legal persons and a relationship between the nodes, and the graphical structure model is used to calculate embedding vectors of the node after a plurality of iterations in hidden feature space based on original features of the node and/or edge features associated with the nodes, and wherein the graphical structure model is further used to calculate a predicted probability of the node based on the embedding vector, wherein the predicted probability indicates a probability that the node is a risk node;

calculating an embedding vector corresponding to a sample to be tested by using the trained graphical structure model; and

performing risk control on the sample to be tested based on the embedding vector corresponding to the sample to be tested, wherein

the labeled sample is labeled whether a node corresponding to the labeled sample is a risk node, the performing risk control on the sample to be tested based on the embedding vector corresponding to the sample to be tested comprising:

calculating a predicted probability corresponding to the sample to be tested by using the trained graphical structure model based on the embedding vector corresponding to the sample to be tested; and

performing risk control on the sample to be tested based on the predicted probability corresponding to the sample to be tested.