

CLAIMS

What is claimed is:

1. A risk identification model building method, comprising:

extracting user state records from data of a given user within predetermined duration before a specific event occurs, wherein the user state records comprise user operation behavior and/or system events, and the specific event is an event whose risk type is determined;

sorting the extracted user state records based on occurrence time, and generating a user state sequence for the specific event based on a sorting result;

mining a frequent sequence item set from multiple user state sequences;

for each frequent sequence item in the frequent sequence item set, determining a feature value corresponding to the frequent sequence item;

converting the generated user state sequence into a sequence feature comprising:

determining, based on the frequent sequence item set, a frequent sequence item comprised in the generated user state sequence; and

determining the sequence feature corresponding to the user state sequence based on the frequent sequence item comprised in the user state sequence and the predetermined feature value corresponding to each frequent sequence item; and

building risk identification sample data by using the obtained sequence feature as an feature value and using the risk type of the event as a label value, and building a risk identification model by using at least one piece of sample data.

2. The method according to claim 1, wherein the generating a user state sequence for the specific event based on a sorting result comprises:

in the sorting result, converting, based on an interval between the i th user state and the $(i+1)$ th user state, the i th user state into a user state that comprises information about the interval.

3. The method according to claim 1, wherein the generating a user state sequence for the specific event based on a sorting result comprises:

in the sorting result, converting, based on a pre-obtained evaluation result of the i th user state, the i th user state into a user state that comprises information about the evaluation result.

4. The method according to claim 1, wherein the generating a user state sequence

for the specific event based on a sorting result comprises:

filtering user states in the sorting result based on predetermined filtering rules.

5. The method according to claim 1, wherein the converting the generated user state sequence into a sequence feature comprises:

encoding the generated user state sequence as a sequence vector.

6. The method according to claim 5, wherein the encoding the generated user state sequence into a sequence vector comprises:

encoding each state in the user state sequence as a state vector based on a first encoding rule; and

encoding a sequence consisting of state vectors into the sequence vector by using a neural network.

7. The method according to claim 1, further comprising for each frequent sequence item in the frequent sequence item set, determining the feature value corresponding to the frequent sequence item based on a black and white sample distribution status corresponding to the frequent sequence item.

8. A computer device, comprising:

a processor; and

a memory, configured to store a processor executable instruction, wherein the processor is configured to:

extract user state records within predetermined duration from data of a given user before a specific event occurs, wherein the user state records comprise user operation behavior and/or system events, and the specific event is an event whose risk type is determined;

sort the extracted user state records based on occurrence time, and generate a user state sequence for the specific event based on a sorting result;

mine a frequent sequence item set from multiple user state sequences;

for each frequent sequence item in the frequent sequence item set, determine a feature value corresponding to the frequent sequence item;

convert the generated user state sequence into a sequence feature comprising:

determine, based on the frequent sequence item set, a frequent sequence item comprised in the generated user state sequence; and

determine the sequence feature corresponding to the user state sequence based on the frequent sequence item comprised in the user state sequence and the

predetermined feature value corresponding to each frequent sequence item; and build risk identification sample data by using the obtained sequence feature as a feature value and using the risk type of the event as a label value, and build a risk identification model by using at least one piece of sample data.

9. The computer device according to claim 8, wherein the processor is further configured to:

in the sorting result, convert, based on an interval between the i th user state and the $(i+1)$ th user state, the i th user state into a user state that comprises information about the interval.

10. The computer device according to claim 8, wherein the processor is further configured to:

in the sorting result, convert, based on a pre-obtained evaluation result of the i th user state, the i th user state into a user state that comprises information about the evaluation result.

11. The computer device according to claim 8, wherein the processor is further configured to:

filter user states in the sorting result based on predetermined filtering rules.

12. The computer device according to claim 8, wherein the processor is further configured to:

encode the generated user state sequence as a sequence vector.

13. The computer device according to claim 12, wherein the processor is further configured to:

encode each state in the user state sequence as a state vector based on a first encoding rule; and

encode a sequence consisting of state vectors into the sequence vector by using a neural network.

14. The computer device according to claim 8, wherein the processor is further configured to:

determine the feature value corresponding to the frequent sequence item based on a black and white sample distribution status corresponding to the frequent sequence item.