Claims

What is claimed is:

1. A method for risk feature screening, comprising:

training a classification model based on sample events to determine risk events;

acquiring data corresponding to each of a plurality of risk features in the sample events;

calculating, according to the data corresponding to each of the plurality of risk features, a classification accuracy metric of the risk feature corresponding to the classification model;

obtaining a plurality of feature weights according to the plurality of classification accuracy metrics of the plurality of risk features; and

selecting at least a part of the plurality of risk features through screening according to the plurality of feature weights and a predetermined constraint for limiting a length of a message generated based on the risk features.

2. The method according to claim 1, wherein each of the plurality of risk features has a corresponding sub-message word count, and wherein selecting at least a part of the plurality of risk features through screening according to the feature weights and the predetermined constraint comprises:

performing a first sorting on the plurality of risk features according to the feature weights and corresponding sub-message word counts; and

selecting at least a part of the plurality of risk features through screening according to the first sorting result, the sub-message word counts, and the predetermined constraint.

3. The method according to claim 2, wherein performing the first sorting on the plurality of risk features according to the feature weights and the corresponding sub-message word counts comprises:

performing a second sorting on the plurality of risk features according to the feature weights to determine a second sorting result;

selecting at least a part of the plurality of risk features from the plurality of risk features according to the second sorting result; and

performing the first sorting on the selected risk features according to the feature weights and the corresponding sub-message word counts.

4. The method according to claim 2, wherein performing the first sorting on the plurality of risk features according to the feature weights and corresponding sub-message word counts comprises:

calculating unit word count weights corresponding to the risk features based on the

feature weights and the sub-message word counts corresponding to the risk features; and

performing the first sorting on the plurality of risk features according to the unit word count weights.

5. The method according to claim 2, wherein selecting at least a part of the plurality of risk features through screening according to the first sorting result, the sub-message word counts, and the predetermined constraint comprises:

traversing, in a descending order of the unit word count weights, all risk features included in the first sorting result and executing the following for a current risk feature:

adding the current risk feature into a defined set, and determining whether a sum of the word counts of the sub-messages corresponding to risk features included in the defined set satisfies the predetermined constraint; if it is determined that the sum of the word counts satisfies the predetermined constraint, traversing to the next risk feature; otherwise, deleting the current risk feature from the defined set, terminating the traversing process, and using the risk features included in the defined set as the selected risk features.

6. The method according to claim 5, wherein traversing to the next risk feature comprises:

obtaining a value of a classification accuracy metric of the defined set corresponding to the classification model;

determining whether the value of the classification accuracy metric of the defined set is not greater than a value of the classification accuracy metric of the defined set before the addition of the current risk feature; if it is determined that the value of the classification accuracy metric of the defined set is not greater than a value of the classification accuracy metric of the defined set before the addition of the current risk feature, deleting the current risk feature from the defined set and traversing to the next risk feature; otherwise, traversing to the next risk feature.

7. The method according to claim 1, wherein the classification accuracy metric comprises an area under receiver operating characteristic curve (AUC).

8. The method according to claim 1, further comprising:

acquiring an event to be described;

generating a sub-message corresponding to the event to be described with respect to each of the screened at least some risk features; and

generating a descriptive message for the event to be described according to the sub-messages.

9. The method according to claim 8, wherein the event to be described is determined as a risk event by the classification model, and the risk event is a suspected money laundering

transaction.

10. A descriptive message generation method, comprising:

acquiring an event to be described;

determining one or more risk features through screening; and

generating a descriptive message for the event to be described according to the determined one or more risk features,

wherein determining the one or more risk features through screening comprises:

training a classification model based on sample events to determine risk events;

acquiring data corresponding to each of a plurality of risk features in the sample events;

calculating, according to the data corresponding to each of the plurality of risk features, a classification accuracy metric of the risk feature corresponding to the classification model;

obtaining a plurality of feature weights according to the plurality of classification accuracy metrics of the plurality of risk features; and

selecting at least a part of the plurality of risk features through screening according to the plurality of feature weights and a predetermined constraint for limiting a length of a message generated based on the risk features.

11. A risk feature screening device, comprising:

one or more processors; and

a memory storing instructions that, when executed by the one or more processors, cause the device to perform:

training a classification model based on sample events to determine risk events;

acquiring data corresponding to each of a plurality of risk features in the sample events;

calculating, according to the data corresponding to each of the plurality of risk features, a classification accuracy metric of the risk feature corresponding to the classification model;

obtaining a plurality of feature weights according to the plurality of classification accuracy metrics of the plurality of risk features; and

selecting at least a part of the plurality of risk features through screening according to the plurality of feature weights and a predetermined constraint for limiting a length of a message generated based on the risk features.

12. The device according to claim 11, wherein each of the plurality of risk features has a corresponding sub-message word count respectively, and wherein selecting at least a part of the plurality of risk features through screening according to the feature weights and the

predetermined constraint comprises:

performing a first sorting on the plurality of risk features according to the feature weights and corresponding sub-message word counts; and

selecting at least a part of the plurality of risk features through screening according to the first sorting result, the sub-message word counts, and the predetermined constraint.

13. The device according to claim 12, wherein performing the first sorting on the plurality of risk features according to the feature weights and the corresponding sub-message word counts comprises:

performing a second sorting on the plurality of risk features according to the feature weights to determine a second sorting result;

selecting at least a part of the plurality of risk features from the plurality of risk features according to the second sorting result; and

performing the first sorting on the selected risk features according to the feature weights and the corresponding sub-message word counts.

14. The device according to claim 12, wherein performing the first sorting on the plurality of risk features according to the feature weights and corresponding sub-message word counts comprises:

calculating unit word count weights corresponding to the risk features based on the feature weights and the sub-message word counts corresponding to the risk features; and

performing the first sorting on the plurality of risk features according to the unit word count weights.

15. The device according to claim 12, wherein selecting at least a part of the plurality of risk features through screening according to the first sorting result, the sub-message word counts, and the predetermined constraint comprises:

traversing, in a descending order of the unit word count weights, all risk features included in the first sorting result and executing the following for a current risk feature:

adding the current risk feature into a defined set, and determining whether a sum of the word counts of the sub-messages corresponding to risk features included in the defined set satisfies the predetermined constraint; if it is determined that the sum of the word counts satisfies the predetermined constraint, traversing to the next risk feature; otherwise, deleting the current risk feature from the defined set, terminating the traversing process, and using the risk features included in the defined set as the selected risk features.

16. The device according to claim 15, wherein traversing to the next risk feature comprises:

obtaining a value of a classification accuracy metric of the defined set corresponding to the classification model;

determining whether the value of the classification accuracy metric of the defined set is not greater than a value of the classification accuracy metric of the defined set before the addition of the current risk feature; if it is determined that the value of the classification accuracy metric of the defined set is not greater than a value of the classification accuracy metric of the defined set before the addition of the current risk feature, deleting the current risk feature from the defined set and traversing to the next risk feature; otherwise, traversing to the next risk feature.

17. The device according to claim 11, wherein the classification accuracy metric comprises an area under receiver operating characteristic curve (AUC).

18. The device according to claim 11, wherein the memory further comprises instructions, when executed by the one or more processors, cause the device to perform:

acquiring an event to be described;

generating a sub-message corresponding to the event to be described with respect to each of the screened at least some risk features; and

generating a descriptive message for the event to be described according to the sub-messages.