

CLAIMS

What is claimed is:

1. A URL abnormality locating method, comprising:

performing field segmentation on a URL to obtain a multiple-instance bag comprising all instances corresponding to all fields;

inputting the multiple-instance bag into a multiple-instance learning-based URL abnormality locating model to predict an abnormal instance, wherein the URL abnormality locating model is trained based on a plurality of URL samples by:

segmenting each of the plurality of URL samples to obtain a corresponding multiple-instance bag comprising all instances of all fields in the URL sample;

for an abnormal URL sample in the plurality of URL samples, randomly selecting one or more instances in the multiple-instance bag corresponding to the abnormal URL sample and assigning positive labels to the one or more instances;

for a normal URL sample in the plurality of URL samples, assigning negative labels to all instances in the multiple-instance bag corresponding to the normal URL sample; and

training the URL abnormality locating model based on the labeled instances in the multiple-instance bags of the plurality of URL samples; ~~and~~ by:

updating labels of instances in the multiple-instance bags based on predictions generated by a classifier; and

training a new classifier based on the updated labels of the instances in the multiple-instance bags

locating a corresponding abnormal field according to the abnormal instance.

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

training the URL abnormality locating model based on a multiple-instance learning algorithm, to obtain the URL abnormality locating model.

3. The method according to claim 2, wherein the training the URL abnormality locating model based on the multiple-instance learning algorithm, to obtain the URL abnormality locating model comprises:

performing abnormal instance and normal instance classification training on the multiple-instance bag set based on the multiple-instance learning algorithm; and

obtaining the URL abnormality locating model based on the classification training.

~~4. The method according to claim 1, wherein the training the URL abnormality locating model based on the multiple instance bags of the plurality of URL samples comprises:
updating labels of instances in the multiple instance bags based on predictions generated by a classifier; and
training a new classifier based on the updated labels of the instances in the multiple instance bags.~~

54. The method according to claim 1, wherein the inputting the multiple-instance bag into a preset URL abnormality locating model to predict an abnormal instance comprises:
predicting the value of the abnormality label of each instance in the multiple-instance bag according to the URL abnormality locating model, so as to determine whether each instance in the multiple-instance bag is an abnormal instance.

65. The method according to any of claims 1 to 54, further comprising:
determining whether the URL is abnormal:
if the multiple-instance bag corresponding to the URL comprises an abnormal instance, determining that the URL is abnormal; and
if the multiple-instance bag corresponding to the URL comprises no abnormal instance, determining that the URL is normal.

76. The method according to any of claims 1 to 54, wherein the instance is represented by a feature vector of a corresponding field.

87. The method according to any of claims 1 to 54, wherein the field is a parameter request field in the URL.

98. A URL abnormality locating training method, comprising:
collecting a URL sample set comprising a plurality of URL samples;
performing field segmentation on each URL sample in the URL sample set, and for each URL sample, obtaining a multiple-instance bag comprising all instances corresponding to all fields;
collecting multiple-instance bags of all of the URL samples;

for an abnormal URL sample in the plurality of URL samples, randomly selecting one or more instances in the multiple-instance bag corresponding to the abnormal URL sample and assigning positive labels to the one or more instances;

for a normal URL sample in the plurality of URL samples, assigning negative labels to all instances in the multiple-instance bag corresponding to the normal URL sample;

performing abnormal instance and normal instance classification training on the labeled instances in the multiple-instance bags based on a multiple-instance learning algorithm; and

obtaining the URL abnormality locating model based on the classification training.

~~109~~. The method according to claim ~~98~~, wherein

the performing abnormal instance and normal instance classification training on the labeled instances in the multiple-instance bags comprises:

performing iterative learning on the value of the abnormality label to update a final value of the abnormality label of each instance.

~~110~~. A URL abnormality locating device, comprising:

a segmenting unit configured to perform field segmentation on a URL to obtain a multiple-instance bag comprising all instances corresponding to all fields;

a predicting unit configured to input the multiple-instance bag into a multiple-instance learning-based URL abnormality locating model to predict an abnormal instance, wherein the URL abnormality locating model is trained based on a plurality of URL samples by:

segmenting each of the plurality of URL samples to obtain a corresponding multiple-instance bag comprising all instances of all fields in the URL sample;

for an abnormal URL sample, randomly selecting one or more instances in the multiple-instance bag corresponding to the abnormal URL sample and assigning positive labels to the one or more instances;

for a normal URL sample, assigning negative labels to all instances in the multiple-instance bag corresponding to the normal URL sample; ~~and~~

training the URL abnormality locating model based on the labeled instances in the multiple-instance bags of the plurality of URL samples by:

updating labels of instances in the multiple-instance bags based on predictions generated by a classifier; and

training a new classifier based on the updated labels of the instances in the multiple-instance bags; and

a locating unit configured to locate a corresponding abnormal field according to the abnormal instance.

~~42~~11. The device according to claim ~~44~~10, further comprising:

a model training unit configured to train the URL abnormality locating model based on a multiple-instance learning algorithm, to obtain the URL abnormality locating model.

~~43~~12. The device according to claim ~~42~~11, wherein the model training unit comprises:

a sample segmenting subunit configured to perform field segmentation on each URL sample in the plurality of URL samples, and for each URL sample, obtain a multiple-instance bag comprising all instances corresponding to all fields;

an instance bag collecting subunit configured to collect multiple-instance bags of all of the URL samples to obtain a multiple-instance bag set; and

a training subunit configured to perform abnormal instance and normal instance classification training on the multiple-instance bag set based on the multiple-instance learning algorithm, so as to obtain the URL abnormality locating model.

~~44~~13. The device according to claim ~~44~~10, wherein the predicting unit is specifically configured to: predict the value of the abnormality label of each instance in the multiple-instance bag according to the URL abnormality locating model, so as to determine whether each instance is an abnormal instance.

~~45~~14. The device according to any of claims ~~44~~10 to ~~44~~13, further comprising:

an abnormal URL determining unit configured to determine whether the URL is abnormal: if the multiple-instance bag corresponding to the URL comprises an abnormal instance, determine that the URL is abnormal; and if the multiple-instance bag corresponding to the URL comprises no abnormal instance, determine that the URL is normal.

~~46~~15. The device according to any of claims ~~44~~10 to ~~44~~13, wherein the instance is represented by a feature vector of a corresponding field.

~~47~~16. The device according to any of claims ~~44~~10 to ~~44~~13, wherein the field is a parameter request field in the URL.

~~18~~17. A URL abnormality locating training device, comprising:

a sample acquiring unit configured to collect a URL sample set comprising a plurality of URL samples;

a sample segmenting unit configured to perform field segmentation on each URL sample in the URL sample set, and for each URL sample, obtain a multiple-instance bag comprising all instances corresponding to all fields;

an instance bag collecting unit configured to collect multiple-instance bags of all of the URL samples to obtain a multiple-instance bag set; and

a training unit configured to:

for an abnormal URL sample in the plurality of URL samples, randomly select one or more instances in the multiple-instance bag corresponding to the abnormal URL sample and assign positive labels to the one or more instances;

for a normal URL sample in the plurality of URL samples, assign negative labels to all instances in the multiple-instance bag corresponding to the normal URL sample; and

train the URL abnormality locating model based on the labeled instances in the multiple-instance bags of the plurality of URL samples by:

updating labels of instances in the multiple-instance bags based on predictions generated by a classifier; and
training a new classifier based on the updated labels of the instances in the multiple-instance bags.

~~19~~18. A server comprising a memory, a processor, and a computer program stored on the memory and executable on the processor, when the processor executes the program, steps of the method according to any of claims 1 to ~~10~~9 being implemented.

~~20~~19. A computer-readable storage medium storing a computer program, when the program is executed by a processor, steps of the method according to any of claims 1 to ~~10~~9 being implemented.