WHAT IS CLAIMED IS:

1. A data protection method, comprising:

acquiring gradient correlation information respectively corresponding to a target sample and a reference sample belonging to the same batch as the target sample, wherein a sample set to which the target sample belongs pertains to a binary classification sample set that is imbalanced in distribution;

generating information of to-be-added data noise according to a comparison result of the gradient correlation information corresponding to the target sample and the gradient correlation information corresponding to the reference sample;

correcting an initial gradient transfer value corresponding to the target sample according to the information of the to-be-added data noise, to generate corrected gradient transfer information, so as to enable gradient transfer information corresponding to samples belonging to different categories in the sample set to be consistent, wherein the initial gradient transfer value is used for indicating that a trained joint training model is adjusted according to the target sample; and

sending the gradient transfer information to a passive party of the joint training model, so as to enable the passive party to adjust a parameter of the joint training model according to the gradient transfer information.

2. The method according to claim 1, wherein the gradient correlation information comprises a sample label used for characterizing a sample category and a sample prediction probability, wherein the sample prediction probability is used for characterizing a probability that a sample is predicted to be in a preset category label by the joint training model, the preset category label comprises a sample label with a small proportion in the sample set, and the reference sample comprises a sample of which a sample prediction probability meets a preset probability condition and a sample label is consistent with the preset category label; and

generating the information of the to-be-added data noise according to the comparison result of the gradient correlation information corresponding to the target sample and the gradient correlation information corresponding to the reference sample comprises:

generating a standard deviation of to-be-added data noise matched with the target sample according to whether the sample label corresponding to the target sample is consistent with the preset category label.

3. The method according to claim 2, wherein generating the standard deviation of the to-be-added data noise matched with the target sample according to whether the sample label corresponding to the target sample is consistent with the preset category label comprises:

in response to determining that the sample label corresponding to the target sample is inconsistent with the preset category label, generating a first standard deviation as the standard deviation of the to-be-added data noise matched with the target sample, wherein the first standard deviation is in negative correlation with a sample prediction probability corresponding to the target sample.

4. The method according to claim 2, wherein generating the standard deviation of the to-be-added data noise matched with the target sample according to whether the sample label corresponding to the target sample is consistent with the preset category label comprises:

in response to determining that the sample label corresponding to the target sample is consistent with the preset category label and a sample prediction probability corresponding to the target sample is greater than a sample prediction probability of the reference sample, generating a second standard deviation as the standard deviation of the to-be-added data noise matched with the target sample, wherein the second standard deviation is in positive correlation with the sample prediction probability corresponding to the target sample.

5. The method according to any one of claims 2-4, wherein the reference sample comprises a sample with a minimum sample prediction probability in samples that are in the same batch as the target sample and have sample labels consistent with the preset category label.

6. The method according to claim 1, wherein the gradient correlation information comprises a gradient modulus used for indicating a gradient modulus of a loss function, corresponding to a sample, on a parameter of a model part belonging to the passive party; and

generating the information of the to-be-added data noise according to the comparison result of the gradient correlation information corresponding to the target sample and the gradient correlation information corresponding to the reference sample comprises:

in response to determining that a gradient modulus corresponding to the target sample is smaller than a gradient modulus corresponding to the reference sample, generating a third standard deviation as the standard deviation of the to-be-added data noise matched with the target sample, wherein the third standard deviation is in negative correlation with the gradient modulus corresponding to the target sample.

7. The method according to claim 6, wherein the reference sample comprises a sample with a

maximum gradient modulus in the samples in the same batch as the target sample.

8. A server, comprising:

one or more processors;

a storage device storing one or more programs therein;

when the one or more programs are executed by the one or more processors, cause the one or more processors to implement the method of claim 1 or 2.

9. A non-transitory computer readable medium storing a computer program therein, wherein when the computer program is executed by a processor, cause the processor to implement the method of any one of claims 1-7.