ABSTRACT

Disclosed are a data protection method and apparatus, and a server and a medium. A particular embodiment of the method comprises: acquiring gradient associated information, which respectively corresponds to a target sample that belongs to a binary classification sample set with unbalanced distribution and a reference sample that belongs to the same batch as the target sample; generating information of data noise to be added; according to the information of said data noise, correcting an initial gradient transfer value corresponding to the target sample, such that corrected gradient transfer information corresponding to samples in the sample set that belong to different types is consistent; and sending the gradient transfer information to a passive party of a joint training model. By means of the embodiment, there is no significant difference between corrected gradient transfer information corresponding to positive and negative samples, thereby effectively protecting the security of data.

FIG.2