

ABSTRACT

Provided are a method for processing model parameters, and an apparatus. The method comprises: a model parameter set to be sharded is obtained (201), wherein the model
5 parameter set comprises a multi-dimensional array corresponding to a feature embedding; attribute information for a storage system used for storing the model parameter set to be sharded is obtained (202), wherein the storage system used for storing the model parameter set to be sharded differs from a system on which a model corresponding to the model parameter set to be sharded is located when operating; the model parameter set to be sharded
10 is stored in the storage system according to the attribute information (203). The present method implements separation of a system upon which a model operates from a model parameter set that occupies a relatively large amount of space, thereby optimizing a means for parameter storage for large-scale feature embedding and improving resource utilization efficiency.