

ABSTRACT

Provided in the present disclosure are a large-model training method, apparatus, device, and storage medium. A large model includes an adaptor header, a model main body and an adaptor tail, where the adaptor header includes a first cloud part and a first local part, and the adaptor tail includes a second cloud part and a second local part. During a process of training the large model, the first cloud part, the model main body, and the second cloud part are deployed on a cloud terminal, and the first local part and the second local part are deployed on a local client terminal. In the method, merely the adaptor header and the adaptor tail are distributed to the cloud terminal and the client terminal. Encrypted client data is used to perform model training for the adaptor part, while the major structure of the model remains on the cloud terminal, and cloud computing power is used to perform training and inference, thereby ensuring the computing power performance. In addition, an adaptor fine-tuning mechanism is effectively utilized, thereby realizing effective training of the large model under the premise of secure and full utilization of the computing power.

[Fig. 4]