SAMPLE GENERATION METHOD AND APPARATUS AND LARGE LANGUAGE MODEL FINE-TUNING METHOD AND APPARATUS

ABSTRACT

Embodiments of this specification provide a sample generation method and apparatus and a large language model fine-tuning method and apparatus. In the method, a dataset is obtained through real-time collection, a question in the dataset is synonymously rewritten, each of a plurality of obtained synonymous questions is constructed into prompt data corresponding to a downstream task, and a plurality of pieces of obtained prompt data are separately input to a large language model, to obtain a plurality of prediction results. Then, a degree of certainty in predicting the first question text by the large language model in the downstream task is determined based on the plurality of prediction results, labeling information of the plurality of pieces of prompt data is determined by using the degree of certainty, and then the labeling information and the plurality of pieces of prompt data can be used to fine-tune the large language model. The collected data includes user privacy data, and privacy protection needs to be performed.

[FIG. 2]