

Multimodal Large Language Model Training Method, Relevance Calculation and Label Generation Method

ABSTRACT

The present disclosure provides a multimodal large language model training method, a relevance calculation method, and a label generation method, belonging to the field of artificial intelligence technology. The method comprises: training a pre-trained multimodal large language model based on a sample text feature vector, a sample image feature vector, and a sample first multimodal feature vector obtained by processing image information and text description information of a sample product using the pre-trained multimodal large language model, to obtain a multimodal large language model; and training the multimodal large language model based on a sample search term feature vector, a first multimodal feature vector, and a sample second multimodal feature vector obtained by processing a sample search term, image information, and text description information using the multimodal large language model, to obtain a trained multimodal large language model. The multimodal large language model trained by the present disclosure is capable of simultaneously learning image features associated with both the search term and the text description information, and the generated second multimodal feature vector is more accurate.

(Fig. 6)