

ABSTRACT

A model optimization method and device, an information processing method and device are provided. The model optimization method includes: inputting multiple segmentation units into a mixture of experts model, and calculating a target entropy corresponding to each of the multiple segmentation units by a router in the mixture of experts model; dividing, based on the target entropy, the multiple segmentation units into global segmentation units and local segmentation units; broadcasting the global segmentation units to all sub expert models in the mixture of experts model for processing, and broadcasting the local segmentation units to sub expert models that match the local segmentation units in the mixture of experts model for processing; and tuning parameters of the mixture of experts model based on processing results until obtaining a target mixture of experts model that meets an optimization stopping condition.

(Figure 1)