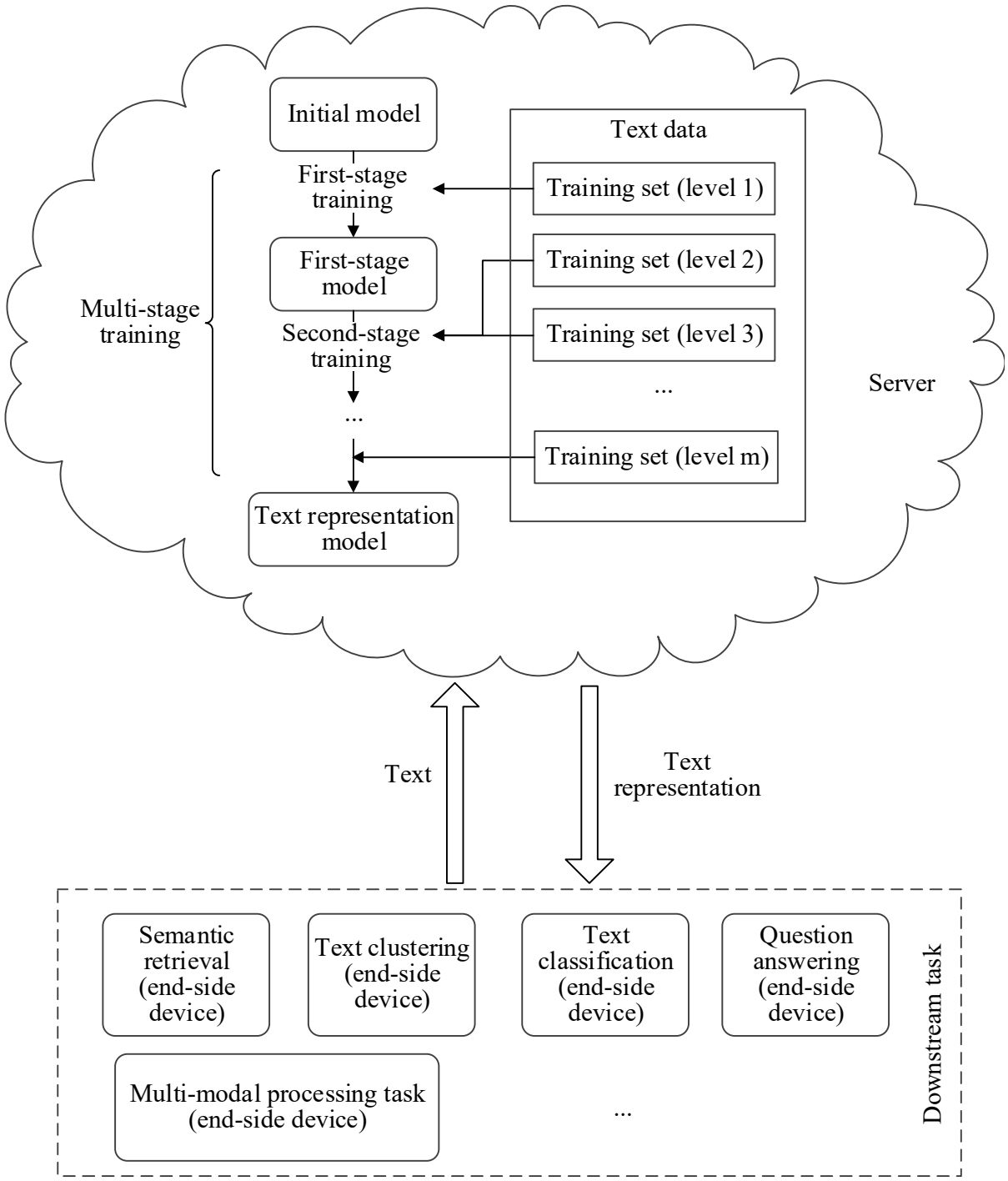

Initial model
First-stage training
First-stage model
Second-stage training
...
Text representation model
Multi-stage training
Text data
Training set (level 1)
Training set (level 2)
Training set (level 3)
...
Training set (level m)
Server
Text
Text representation
Semantic retrieval (end-side device)
Text clustering (end-side device)
Text classification (end-side device)
Question answering (end-side device)
Multi-modal processing task (end-side device)
...
Downstream task

FIG. 1

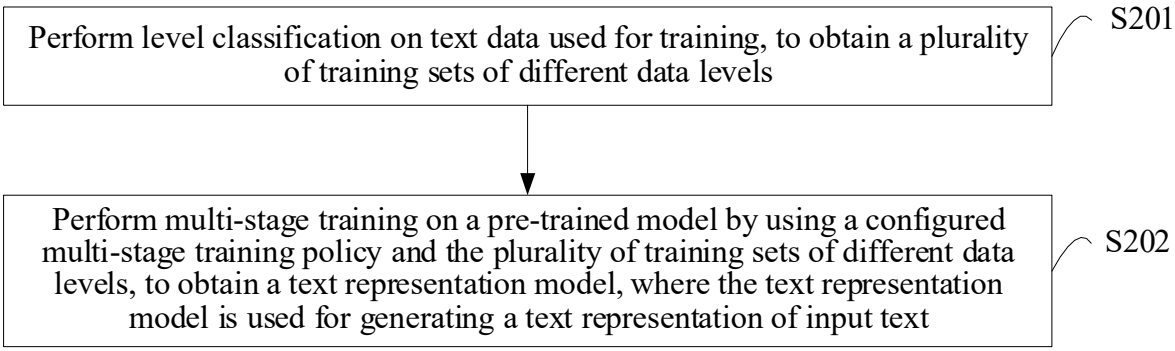

Perform level classification on text data used for training, to obtain a plurality of training sets of different data levels
S201
Perform multi-stage training on a pre-trained model by using a configured multi-stage training policy and the plurality of training sets of different data levels, to obtain a text representation model, where the text representation model is used for generating a text representation of input text
S202

FIG. 2

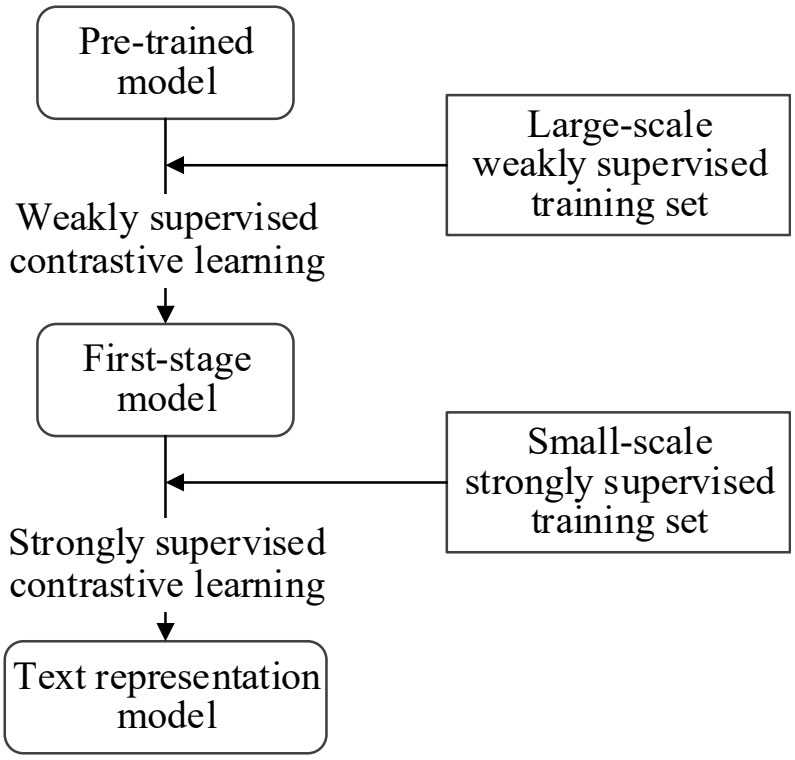

Pre-trained model
Large-scale weakly supervised training set
Weakly supervised contrastive learning
First-stage model
Small-scale strongly supervised training set
Strongly supervised contrastive learning
Text representation model

FIG. 3

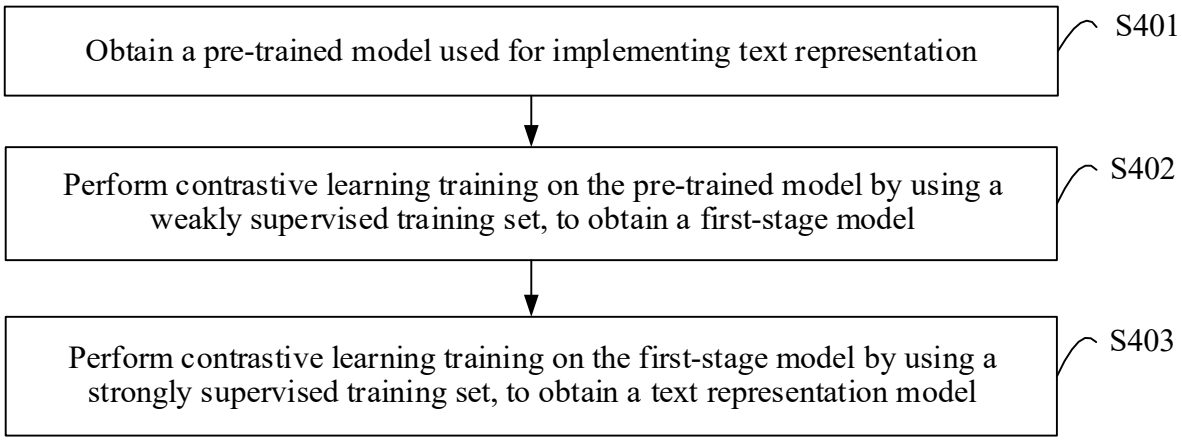


FIG. 4

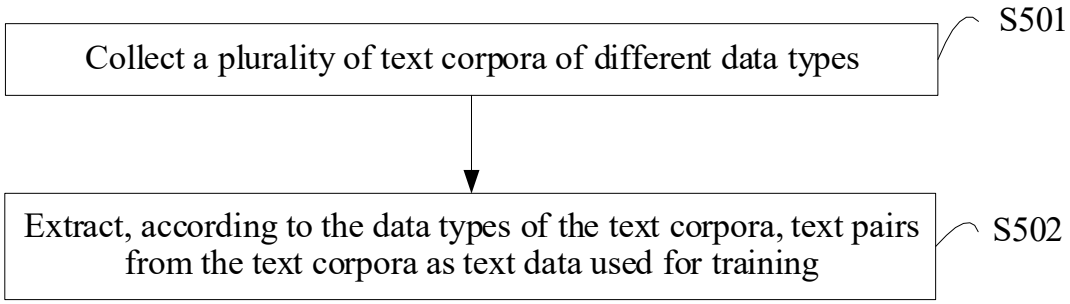


FIG. 5

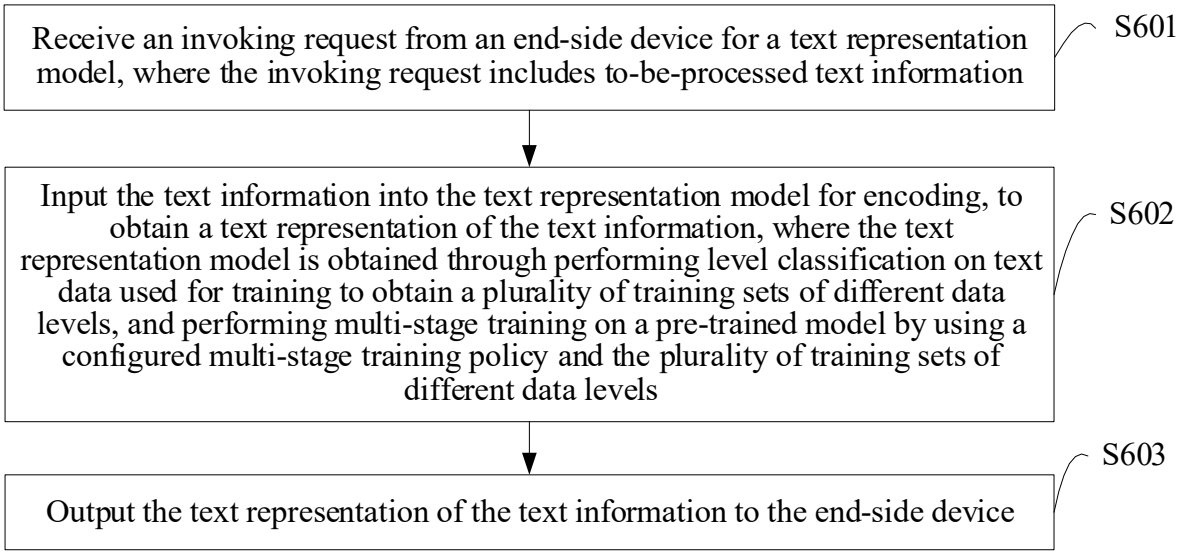


FIG. 6

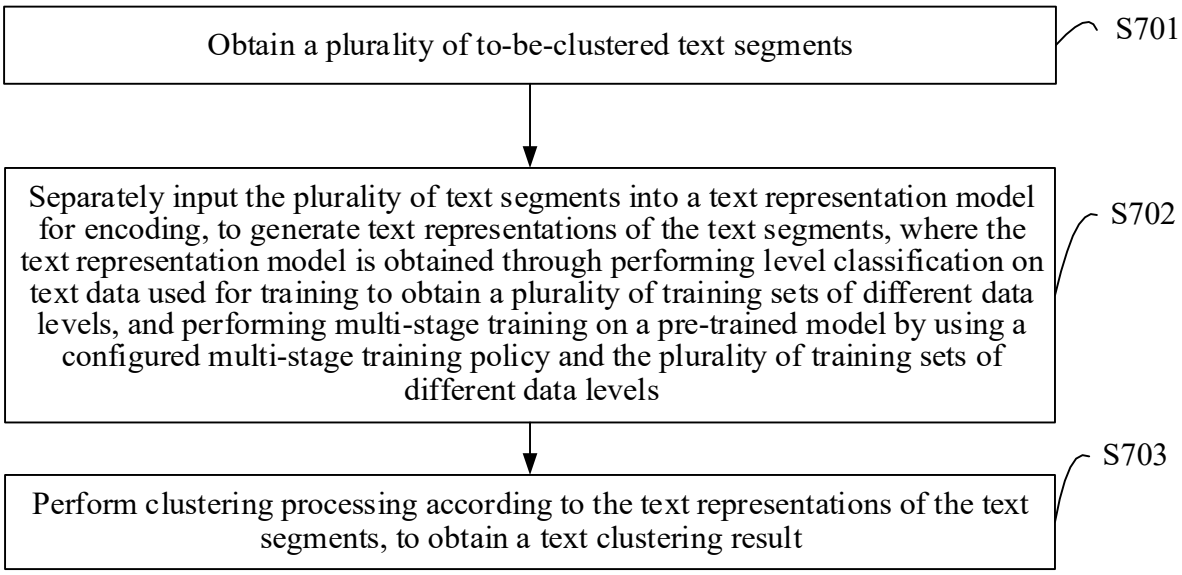

Obtain a plurality of to-be-clustered text segments
S701
Separately input the plurality of text segments into a text representation model for encoding, to generate text representations of the text segments, where the text representation model is obtained through performing level classification on text data used for training to obtain a plurality of training sets of different data levels, and performing multi-stage training on a pre-trained model by using a configured multi-stage training policy and the plurality of training sets of different data levels
S702
Perform clustering processing according to the text representations of the text segments, to obtain a text clustering result
S703

FIG. 7

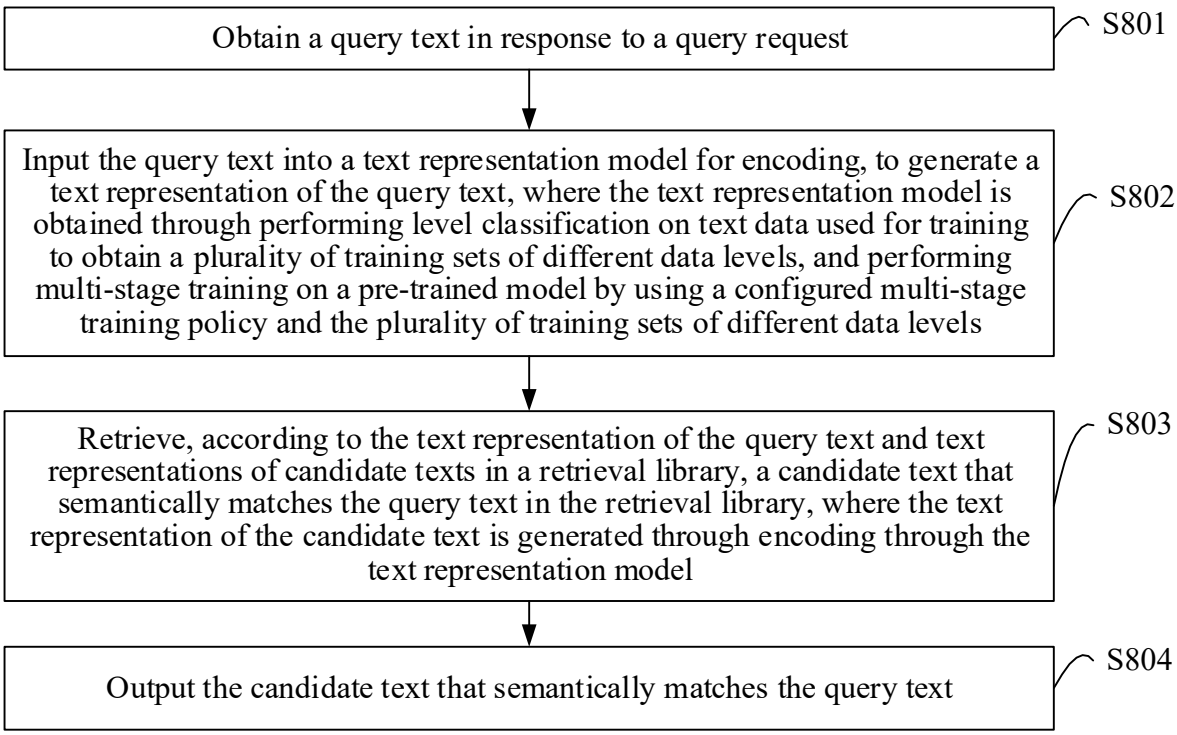

Obtain a query text in response to a query request
S801
Input the query text into a text representation model for encoding, to generate a text representation of the query text, where the text representation model is obtained through performing level classification on text data used for training to obtain a plurality of training sets of different data levels, and performing multi-stage training on a pre-trained model by using a configured multi-stage training policy and the plurality of training sets of different data levels
S802
Retrieve, according to the text representation of the query text and text representations of candidate texts in a retrieval library, a candidate text that semantically matches the query text in the retrieval library, where the text representation of the candidate text is generated through encoding through the text representation model
S803
Output the candidate text that semantically matches the query text
S804

FIG. 8

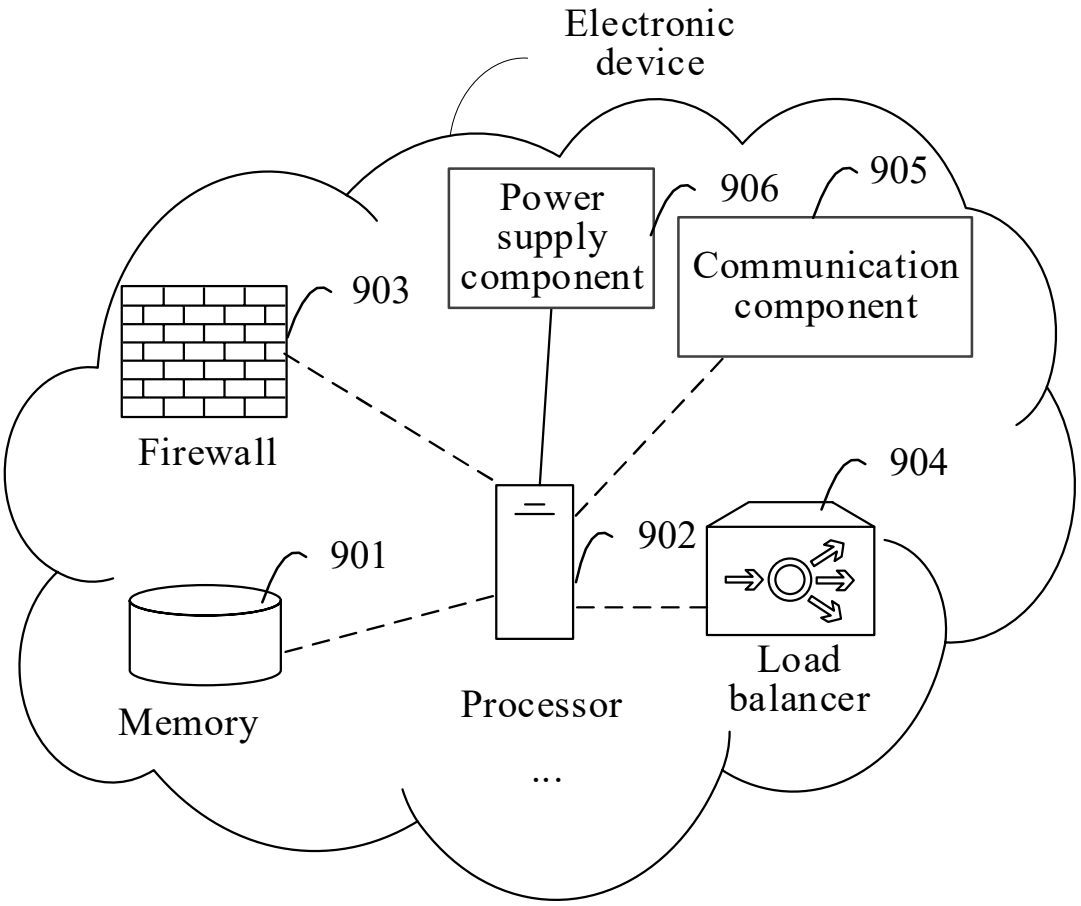

Electronic device
Power supply component
906
905
Communication component
903
Firewall
904
902
901
Load balancer
Memory
Processor
...

FIG. 9