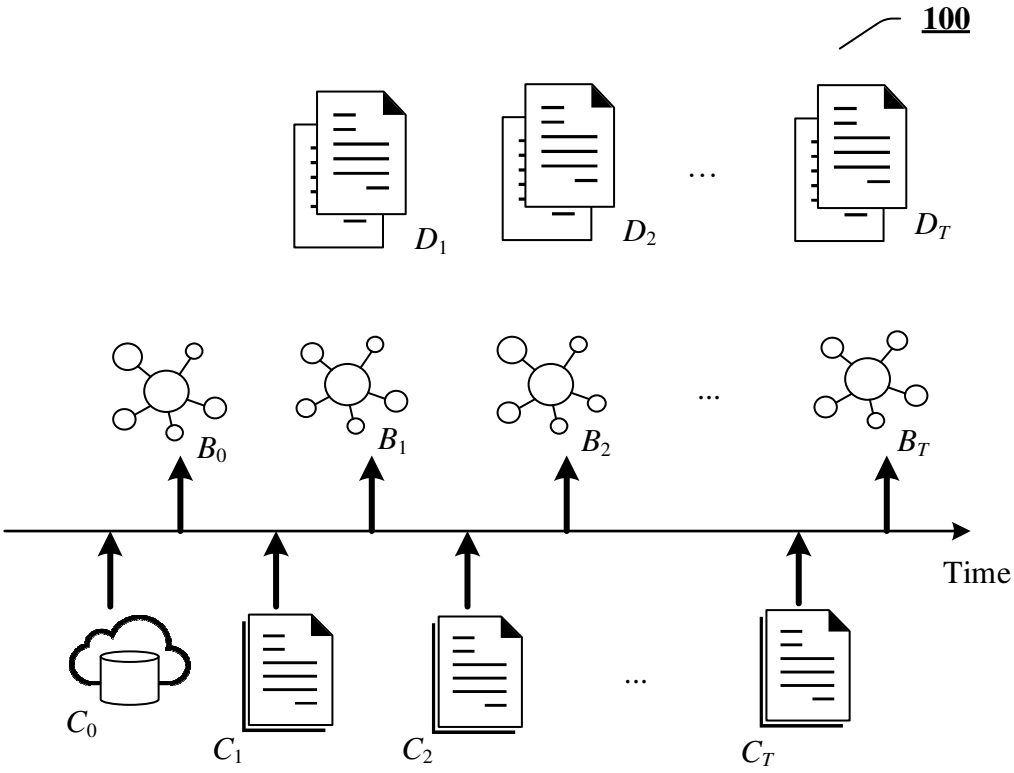
100
D_1
D_2
...
D_T
B_0
B_1
B_2
...
B_T
Time
C_0
C_1
C_2
...
C_T

FIG. 1

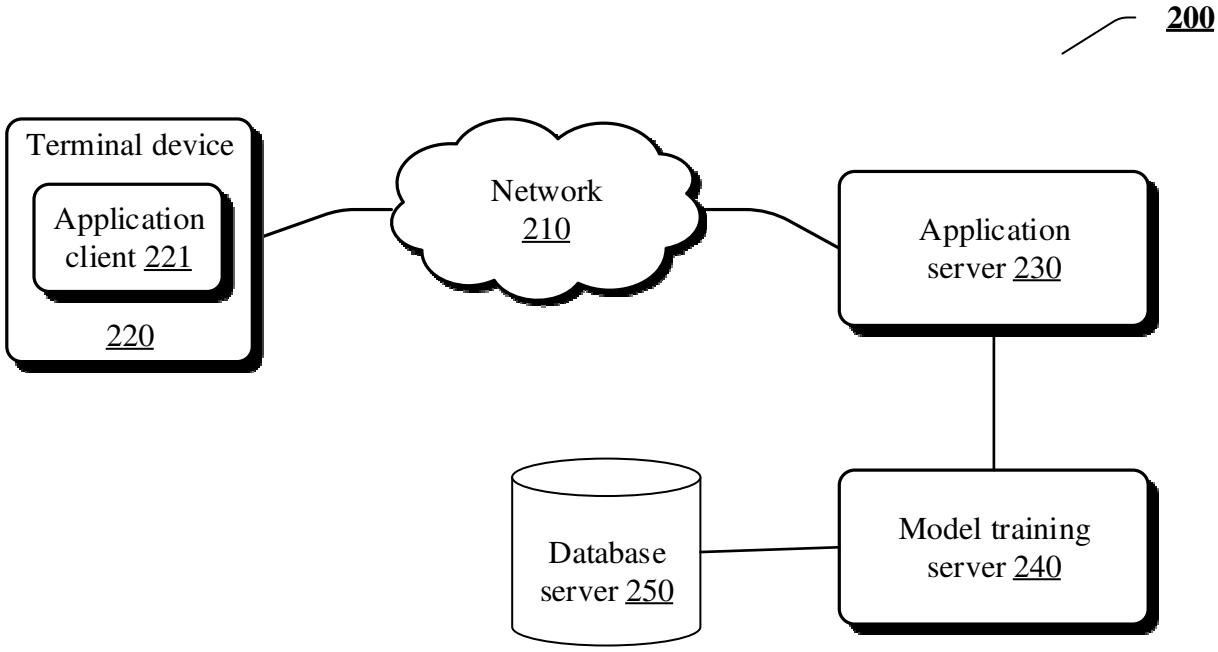
200
Terminal device
Application client 221
220
Network 210
Application server 230
Database server 250
Model training server 240

FIG. 2

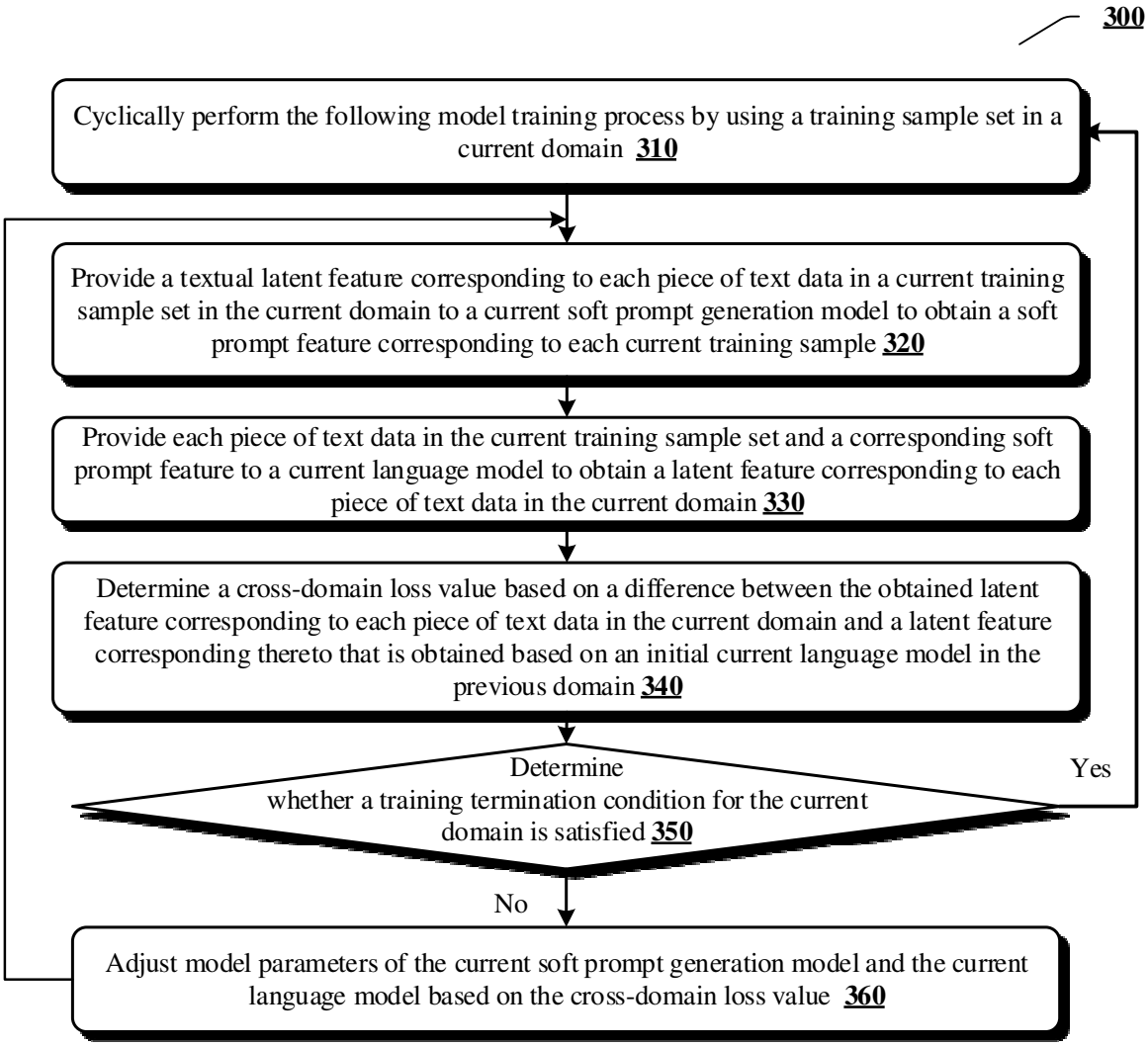
300
Cyclically perform the following model training process by using a training sample set in a current domain 310
Provide a textual latent feature corresponding to each piece of text data in a current training sample set in the current domain to a current soft prompt generation model to obtain a soft prompt feature corresponding to each current training sample 320
Provide each piece of text data in the current training sample set and a corresponding soft prompt feature to a current language model to obtain a latent feature corresponding to each piece of text data in the current domain 330
Determine a cross-domain loss value based on a difference between the obtained latent feature corresponding to each piece of text data in the current domain and a latent feature corresponding thereto that is obtained based on an initial current language model in the previous domain 340
Determine whether a training termination condition for the current domain is satisfied 350
Yes
No
Adjust model parameters of the current soft prompt generation model and the current language model based on the cross-domain loss value 360

FIG. 3

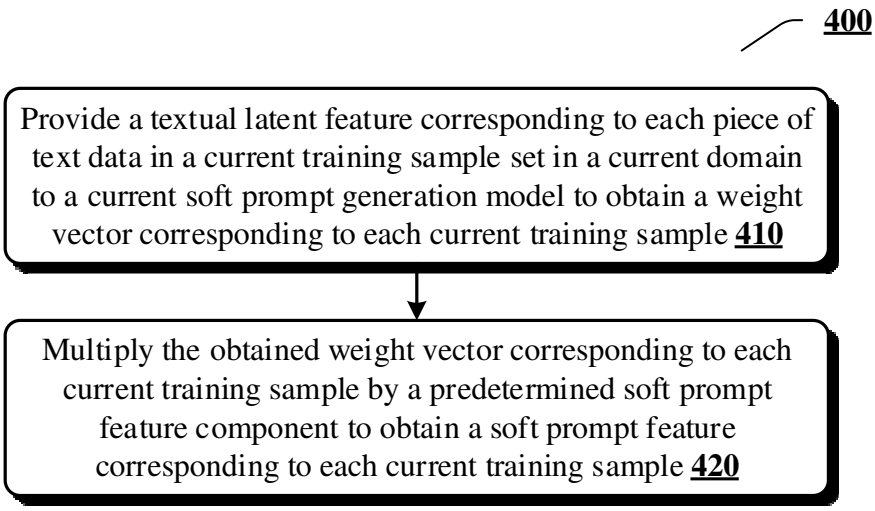
400
Provide a textual latent feature corresponding to each piece of text data in a current training sample set in a current domain to a current soft prompt generation model to obtain a weight vector corresponding to each current training sample 410
Multiply the obtained weight vector corresponding to each current training sample by a predetermined soft prompt feature component to obtain a soft prompt feature corresponding to each current training sample 420

FIG. 4

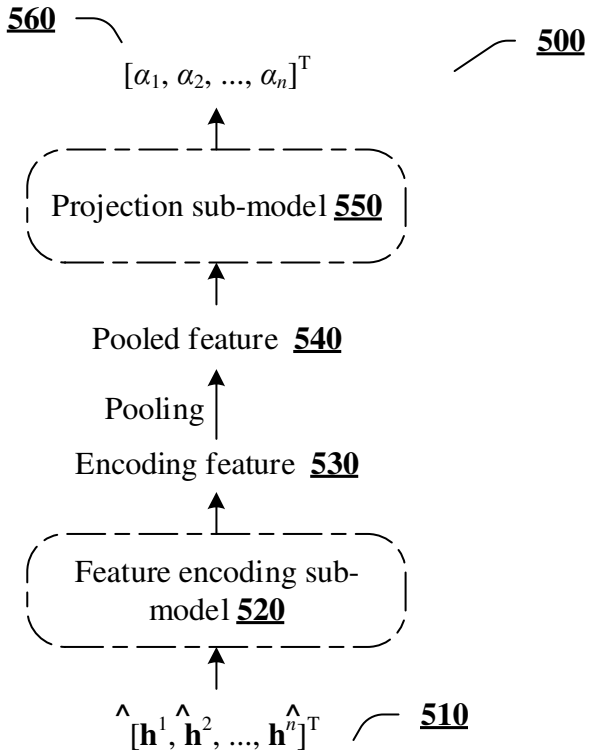
560
$[\alpha_1, \alpha_2, ..., \alpha_n]^T$
500
Projection sub-model 550
Pooled feature 540
Pooling
Encoding feature 530
Feature encoding sub-model 520
$\hat{[}\mathbf{h}^1, \hat{\mathbf{h}}^2, ..., \mathbf{h}^{\hat{n}}]^T$
510

FIG. 5

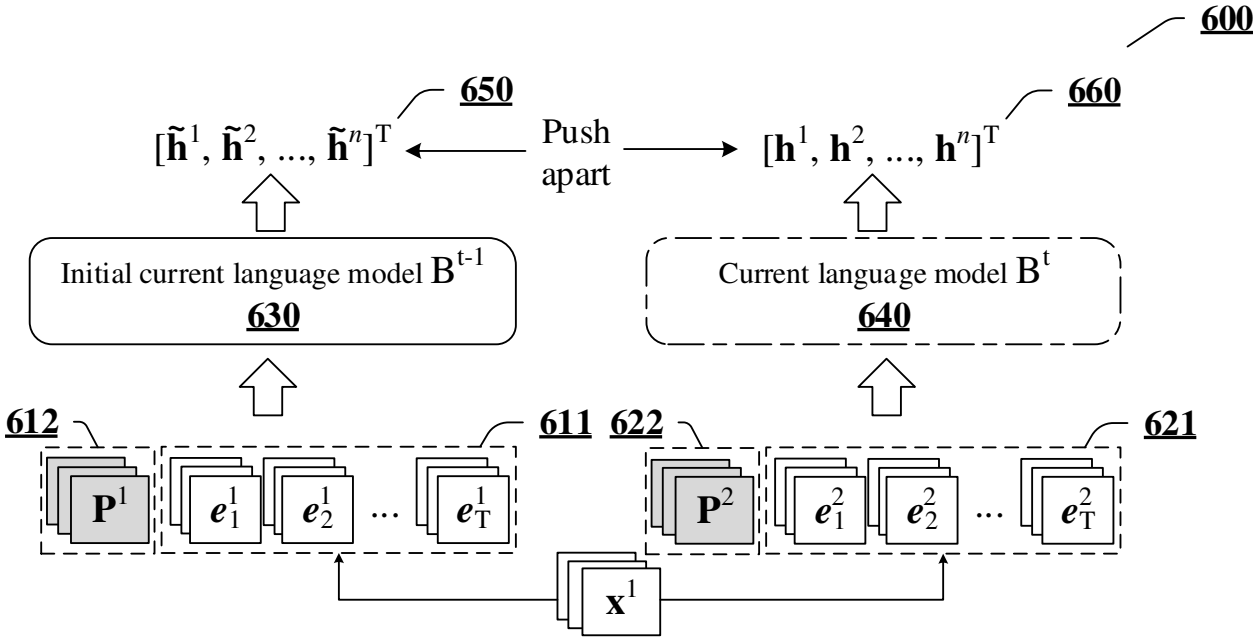
600
650
660
$[\tilde{\mathbf{h}}^1, \tilde{\mathbf{h}}^2, ..., \tilde{\mathbf{h}}^n]^T$
Push apart
$[\mathbf{h}^1, \mathbf{h}^2, ..., \mathbf{h}^n]^T$
Initial current language model B^{t-1}
630
Current language model B^t
640
612
611
622
621
$\mathbf{P}^1$
$\boldsymbol{e}_1^1$ $\boldsymbol{e}_2^1$ ... $\boldsymbol{e}_T^1$
$\mathbf{P}^2$
$\boldsymbol{e}_1^2$ $\boldsymbol{e}_2^2$ ... $\boldsymbol{e}_T^2$
$\mathbf{x}^1$

FIG. 6

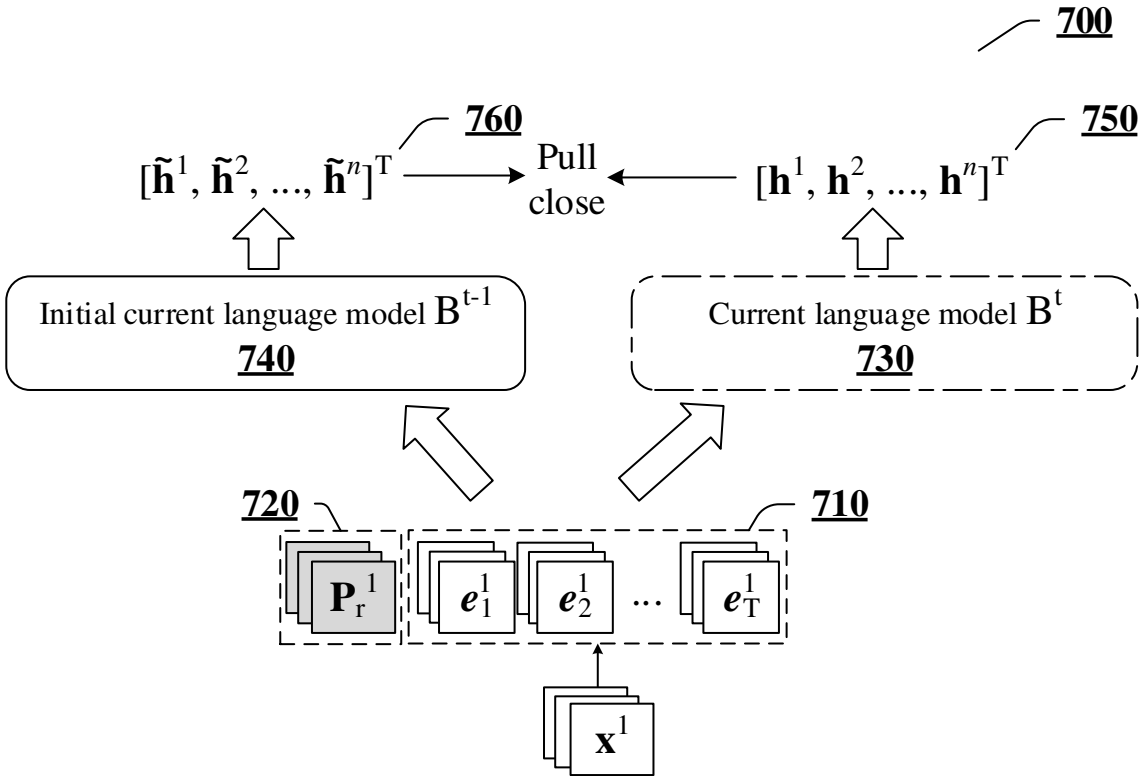
700
760
750
$[\tilde{\mathbf{h}}^1, \tilde{\mathbf{h}}^2, ..., \tilde{\mathbf{h}}^n]^T$
Pull close
$[\mathbf{h}^1, \mathbf{h}^2, ..., \mathbf{h}^n]^T$
Initial current language model B^{t-1}
740
Current language model B^t
730
720
710
$\mathbf{P}_r^1$
$\boldsymbol{e}_1^1$ $\boldsymbol{e}_2^1$ ... $\boldsymbol{e}_T^1$
$\mathbf{x}^1$

FIG. 7

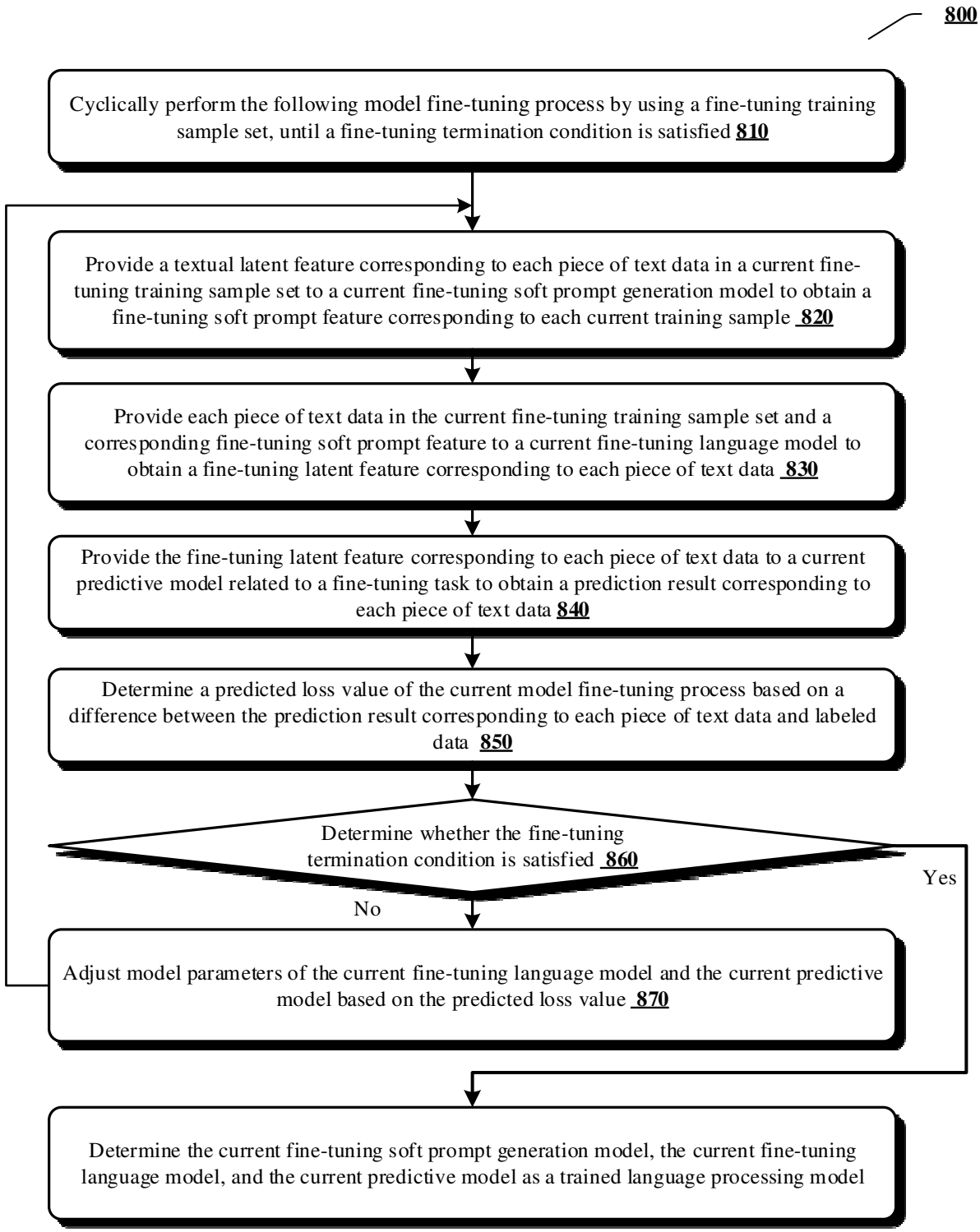
800
Cyclically perform the following model fine-tuning process by using a fine-tuning training sample set, until a fine-tuning termination condition is satisfied 810
Provide a textual latent feature corresponding to each piece of text data in a current fine-tuning training sample set to a current fine-tuning soft prompt generation model to obtain a fine-tuning soft prompt feature corresponding to each current training sample 820
Provide each piece of text data in the current fine-tuning training sample set and a corresponding fine-tuning soft prompt feature to a current fine-tuning language model to obtain a fine-tuning latent feature corresponding to each piece of text data 830
Provide the fine-tuning latent feature corresponding to each piece of text data to a current predictive model related to a fine-tuning task to obtain a prediction result corresponding to each piece of text data 840
Determine a predicted loss value of the current model fine-tuning process based on a difference between the prediction result corresponding to each piece of text data and labeled data 850
Determine whether the fine-tuning termination condition is satisfied 860
Yes
No
Adjust model parameters of the current fine-tuning language model and the current predictive model based on the predicted loss value 870
Determine the current fine-tuning soft prompt generation model, the current fine-tuning language model, and the current predictive model as a trained language processing model

FIG. 8

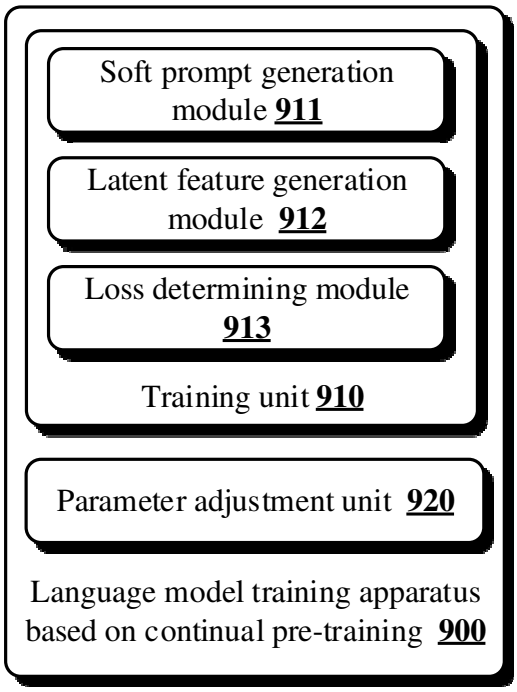
Soft prompt generation module 911
Latent feature generation module 912
Loss determining module 913
Training unit 910
Parameter adjustment unit 920
Language model training apparatus based on continual pre-training 900

FIG. 9

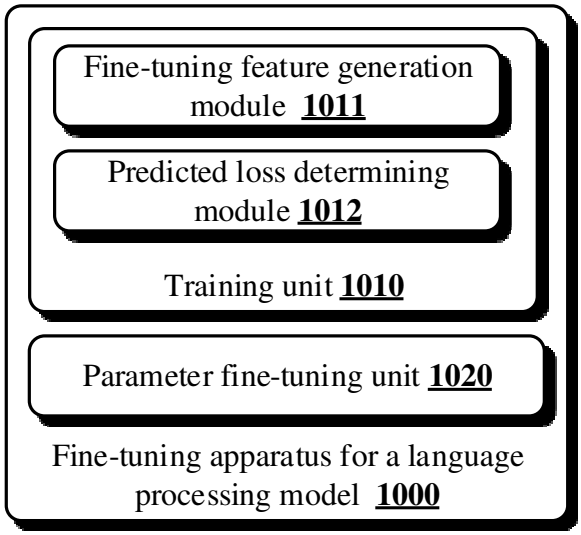
Fine-tuning feature generation module 1011
Predicted loss determining module 1012
Training unit 1010
Parameter fine-tuning unit 1020
Fine-tuning apparatus for a language processing model 1000

FIG. 10

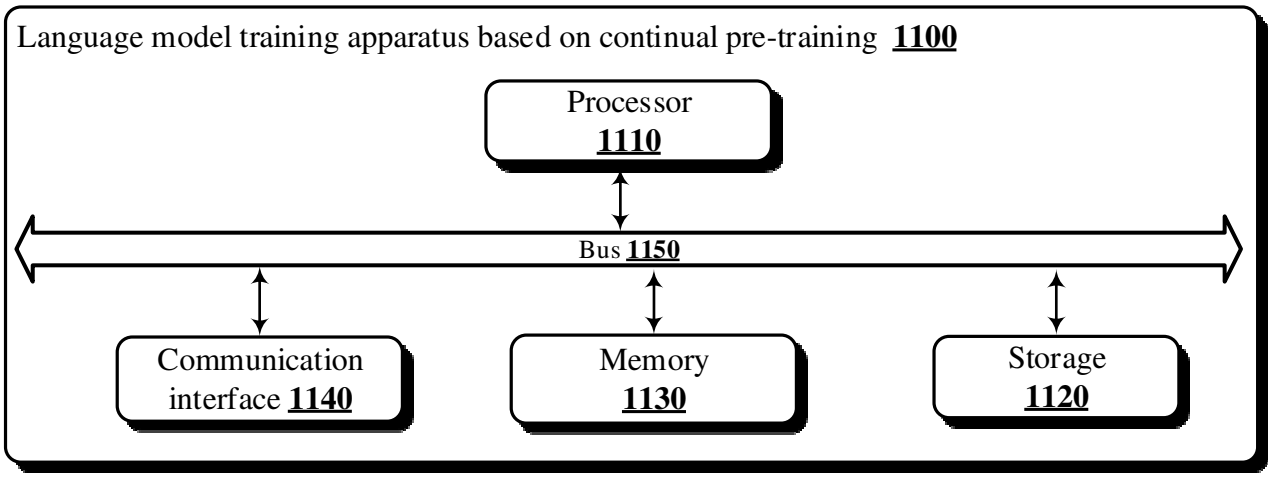
Language model training apparatus based on continual pre-training 1100
Processor
1110
Bus 1150
Communication interface 1140
Memory
1130
Storage
1120

FIG. 11

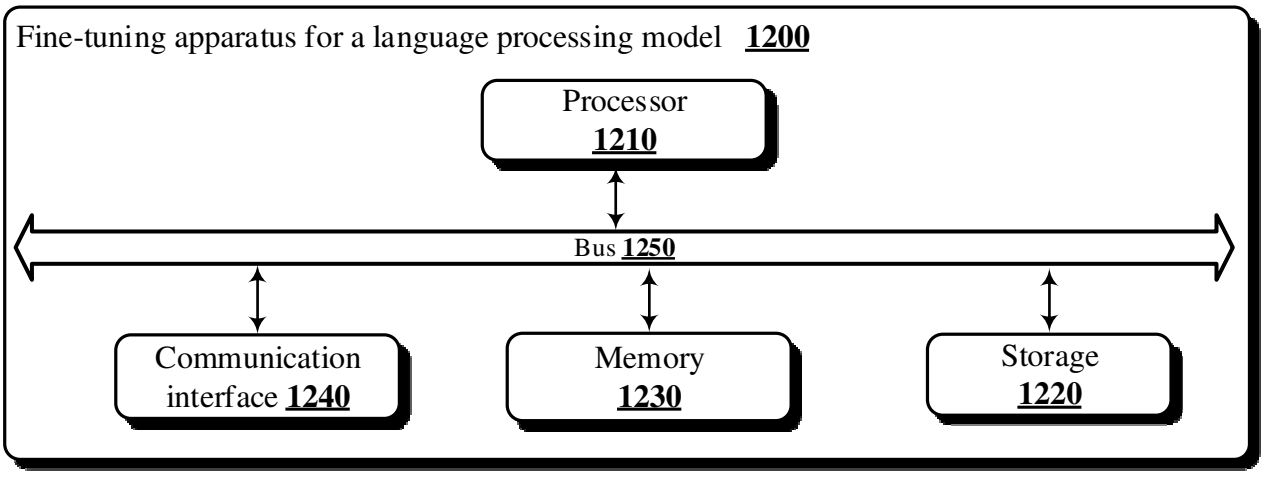
Fine-tuning apparatus for a language processing model 1200
Processor
1210
Bus 1250
Communication interface 1240
Memory
1230
Storage
1220

FIG. 12