CLAIMS

What is Claimed is:

1. A device for training natural-language-understanding (NLU) models online, comprising:

a communication interface;

a processor, operably connected to the communication interface; and

a memory, operably connected to the processor, wherein the memory stores a plurality of NLU models, a plurality of applications, and instructions, wherein the plurality of NLU models are configured to provide natural language understanding for the plurality of applications, and wherein the instructions, when executed by the processor, cause the processor to perform following operations:

calling one of the plurality of NLU models from the memory, creating a plurality of intents within the NLU model, and defining intent attributes for each of the plurality of intents, wherein the intent attributes comprise at least an intent priority and a downgrading strategy configured to resolve multi-intent ambiguity during data maintenance, and forming a model training corpus data structure based on the intent attributes of the plurality of intents;

obtaining dialog content through the communication interface, performing data maintenance based on the model training corpus data structure and linguistic context of the dialog content, thereby generating a corpus dataset, wherein performing the data maintenance comprises:

extracting entries of corpus data from the dialog content and maintaining each of the entries of corpus data in accordance with the model training corpus data structure, such that each of the entries of corpus data is associated with corresponding intent-attribute information, the corresponding intent-attribute information comprising at least the intent priority and the downgrading strategy; and

when one corpus datum among the entries of corpus data corresponds to two or more intents among the plurality of intents, determining a weight of each of the two or more intents according to the linguistic environment in which the corpus datum is embedded, selecting one intent of the two or more intents having the highest weight, and downgrading the remaining intents of the two or more

intents;

performing online training of the NLU model based on the corpus dataset, to obtain a trained NLU model, and testing the trained NLU model to obtain a tested NLU model; and

deploying the tested NLU model to an external electronic device through the communication interface, for understanding, analyzing, and processing natural language acquired in real time.

2. The device for training the NLU models online according to claim 1, wherein the intent attributes further comprise one or more of: an intent name and an intent example.

3. The device for training the NLU models online according to claim 1, wherein testing the trained NLU model comprises: inputting a test dataset to the trained NLU model, calculating multi-categorization model evaluation indexes based on predicted values output by the trained NLU model, and determining whether the trained NLU model is qualified based on the multi-categorization model evaluation indexes; wherein the multi-categorization model evaluation indexes comprise one or more of an accuracy index, a precision index, a recall index, and an F1 score index.

4. The device for training the NLU models online according to claim 1, wherein online training of the NLU model comprises: obtaining abnormal corpus data and unconnected data generated during operation of the NLU model; and iteratively training the NLU model using the abnormal corpus data and unconnected data as input data to the NLU model.

5. The device for training the NLU models online according to claim 1, wherein the plurality of NLU models comprise one of a BERT model, a Transformer model, and a GPT model.

6. The device for training the NLU models online according to claim 1, wherein the processor is further caused to input one or more of the plurality of intents into a ChatGPT model and to pre-describe definitions of the one or more of the plurality of intents to generate the corpus dataset by applying the ChatGPT model.

7. The device for training the NLU models online according to claim 1, wherein before

online training of the NLU model based on the corpus dataset, the processor is further caused to perform error correction on the corpus dataset by applying the ChatGPT model to correct incorrect labels.

8. The device for training the NLU models online according to claim 1, wherein the processor is further caused to generate the dialog content by engaging the ChatGPT model in a conversation with a speech robot, wherein the processor then generates the corpus dataset based on the dialog content.

9. A method for training an NLU model online, comprising:

calling one of a plurality of NLU models configured to provide natural language understanding for a plurality of applications, creating a plurality of intents within the NLU model, and defining intent attributes for each of the plurality of intents, wherein the intent attributes comprise at least an intent priority and a downgrading strategy configured to resolve multi-intent ambiguity during data maintenance, and forming a NLU model training corpus data structure based on the intent attributes of the plurality of intents;

obtaining a dialog content, performing data maintenance based on the NLU model training corpus data structure and linguistic context of the dialog content, thereby generating a corpus dataset, wherein performing the data maintenance comprises:

extracting entries of corpus data from the dialog content and maintaining each of the entries of corpus data in accordance with the NLU model training corpus data structure, such that each of the entries of corpus data is associated with corresponding intent-attribute information, the corresponding intent-attribute information comprising at least the intent priority and the downgrading strategy; and

when one corpus datum among the entries of corpus data corresponds to two or more intents among the plurality of intents, determining a weight of each of the two or more intents according to the linguistic environment in which the corpus datum is embedded, selecting one intent of the two or more intents having the highest weight, and downgrading the remaining intents of the two or more intents;

performing online training of the NLU model based on the corpus dataset, to

obtain a trained NLU model, and testing the trained NLU model to obtain a tested NLU model; and

deploying the tested NLU model to an external electronic device, for understanding, analyzing, and processing natural language acquired in real time.

10. A non-transitory computer-readable storage medium having a computer program stored thereon, wherein the computer program, when executed by a processor, implements the method for training the NLU model online according to claim 9.

11. An electronic terminal comprising: a processor and a memory;

wherein the memory comprises a storage for a computer program; and

wherein the processor executes the computer program to cause the electronic terminal to perform the method for training the NLU model online according to claim 9.