



← Back to results Reverse time;

## Intelligent man-machine conversation system of closed domain

CN108415923B

China

[Download PDF](#)

[Find Prior Art](#)

[Similar](#)

**Other languages:** [Chinese](#)

**Inventor:** 鄂海红, 宋美娜, 胡莺夕, 赵文骏, 王昕睿, 赵鑫禄, 陈忠富, 祝一帆

**Current Assignee :** Beijing University of Posts and Telecommunications

### Worldwide applications

2017 [CN](#)

### Application CN201710973047.XA events

**2017-10-18** Application filed by Beijing University of Posts and Telecommunications

**2017-10-18** Priority to CN201710973047.XA

**2018-08-17** Publication of CN108415923A

**2020-12-11** Application granted

**2020-12-11** Publication of CN108415923B

**Status** Active

**2037-10-18** Anticipated expiration

**Info:** [Patent citations \(4\)](#), [Non-patent citations \(3\)](#), [Cited by \(22\)](#), [Legal events](#), [Similar documents](#), [Priority and Related Applications](#)

**External links:** [Espacenet](#), [Global Dossier](#), [Discuss](#)

### Abstract

The invention provides an intelligent man-machine conversation system of a closed domain, which comprises: the first modeling module is used for constructing a multi-feature fusion depth intention recognition model based on a bidirectional long-time and short-time memory network and a convolutional neural network; the second modeling module is used for constructing a dialogue state tracking model based on MC-BLSTM-MSCNN by adopting a mode of jointly modeling the current state input and the context statement of the man-machine dialogue state system; and the third modeling module is used for constructing a Bi-LSTM matching model based on the domain-outside recovery mechanism of the shift attention mechanism, so that the identified user intention and the user slot value are input into the shift network for weight distribution of the attention mechanism, and the coding of the conversation state and the matching of conversation control are realized. The invention has higher intention recognition accuracy, conversation state tracking accuracy and conversation control stability, thereby improving the cognitive intelligence capability of a man-machine system.

## Classifications

■ **G06F16/35** Clustering; Classification

*View 5 more classifications*

## Claims (9)

Hide Dependent ^

1. A closed domain intelligent human-machine dialog system, comprising:

the first modeling module is used for constructing a multi-feature fusion depth intention recognition model based on a bidirectional long-time memory network BiLSTM and a convolutional neural network CNN so as to extract and fuse features of short texts input by a user with different granularities and dimensions;

the second modeling module is used for constructing a MC-BLSTM-MSCNN-based dialog state tracking model by adopting a combined modeling mode of current state

input and context statements of a human-computer dialog state system so as to extract the characteristics of a slot value pair; the dialog state tracking model includes: a semantic decoding module and a context coding module;

the semantic decoding module is used for directly interacting the vector representation  $r$  and the candidate slot value pair representation  $c$  input by the user and judging whether the user explicitly expresses the intention matched with the current candidate pair or not according to the vector representation  $r$  and the candidate slot value pair  $c$ ;

the context coding module for acquiring specific slot  $t$  inquired by system to user  $q$  And the system asks the user to confirm the particular slot value pair  $(t_s, t_v)$  And according to said  $t_q$  And  $(t_s, t_v)$  Calculating a measure of similarity, wherein calculating the measure of similarity comprises: system output behavior  $(t_q, t_s, t_v)$  Candidate slot value pair  $(c_s, c_v)$  And a user input sentence representation  $r$ ; system output behavior  $(t_q, t_s, t_v)$  Candidate slot value pair  $(c_s, c_v)$  And user input statement representation  $r$  is a parameter that will be used to calculate the similarity measure  $d$ ;

and the third modeling module is used for constructing a Bi-LSTM matching model based on the domain-outside recovery mechanism of the shift attention mechanism, so that the identified user intention and the user slot value are input into the shift network for weight distribution of the attention mechanism, and the coding of the conversation state and the matching of conversation control are realized.

2. The closed domain intelligent human-machine dialog system of claim 1 wherein the multi-feature fusion depth intent recognition model comprises at least: a multi-channel embedded layer, a high-speed channel layer and a multi-granularity convolutional layer.

3. The closed domain intelligent human-machine dialog system of claim 1 wherein the slot name and value vector space representation of the candidate slot-value pairs is represented by  $c_s$  And  $c_v$  Given, mapping the slot names of the candidate slot-value pairs and the tuples of the vector-space representation of values into a single vector  $c$  of the same dimension as the vector representation  $r$

of the user input, wherein the two vector representations interact to learn a similarity measure for distinguishing the interaction between the user input sentence and the slot-value pair that it expresses or does not express, the similarity measure being specifically as follows:

wherein the content of the first and second substances, representing element-wise vector multiplication, which reduces the rich feature set in  $d$  to a single scalar, allows downstream networks to better utilize their parameters by learning the nonlinear interaction between the feature sets in  $r$  and  $c$ .

4. Closed domain intelligent human-machine dialog system according to claim 1, characterized in that the system outputs a behavior  $(t)_q, t_s, t_v$  Candidate slot value pair  $(c)_s, c_v$  And the user input sentence representation  $r$  has the following relationship:

$$m_r = (c_s \cdot t_q) r$$

$$m_c = (c_s \cdot t_s)(c_v \cdot t_v) r$$

where, represents the dot product, the computed similarity terms act as a gated gating mechanism that is represented only by the utterance when the system queries the current candidate bin or bin value pair.

5. The closed-domain intelligent human-computer interaction system of claim 4, wherein the context coding module is further configured to input the obtained vector representation  $m$  and the vector representation  $d$  of semantic similarity into a Softmax layer for classification.

6. The closed domain intelligent human-machine dialog system of claim 1 wherein the Bi-LSTM matching model comprises:

the attention module based on displacement is used for matching the characteristics of the slot value pairs with the current intention of the user, and the obtained matching degree is used as the importance basis of the slot value pairs, so that different attention weights of different slot value pairs under different intentions are realized;

the user speech feature extraction module is used for extracting user speech features according to the long-term and short-term memory network;

the standard answer classification module is used for splicing the slot value pair characteristics and the user utterance characteristics, inputting the slot value pair characteristics and the user utterance characteristics into a bidirectional long-short term memory network together, and finally performing softmax classification;

and the session control module supports an out-of-domain recovery mechanism and is used for adding data of the open domain data set into the closed domain data set to request recovery from the outside of the user domain.

7. The closed domain intelligent human-computer dialog system of claim 6 wherein the shift-based attention module is configured to map the feature set to a single output, wherein the input and the output are matrices formed by concatenating feature vectors, and the mapping is implemented by the following formula:

wherein,  $X$  represents the input groove value pair information, and the calculation formula of  $\lambda$  in the above formula is:

where  $W$  is the vector of the user's current intent and  $a$  is a learnable scalar.

8. The closed domain intelligent human-machine dialog system of claim 6 wherein the user utterance feature extraction module extracts user utterance features according to a long-short term memory network, comprising:

the method comprises the following steps of adopting two LSTM networks, respectively inputting current input sentences of a user into the two networks, simultaneously analyzing the sentences of the user from forward and reverse time sequences, obtaining the output of each hidden node at each time point, connecting the output of each hidden node at each time point of the two LSTM networks through averaging, and inputting the output into the next-stage network, wherein the specific formula is as follows:

wherein the content of the first and second substances, the hidden node outputs representing the positive direction at time  $t$ , and (4) representing each output of the hidden nodes in the reverse direction at the time  $t$ , averaging the outputs of the nodes at the same time to form a two-dimensional tensor, and inputting the two-dimensional tensor into a subsequent network.

9. The closed domain intelligent human-machine dialog system of claim 6 wherein the last layer of the Bi-LSTM matching model is a softmax full connectivity layer, as follows:

the layer predicts an intention category  $y$  corresponding to the  $S$  statement by taking the output  $h$  of the hidden layer as input, and the classifier performs probability analysis on the matching degree of the current statement of the user and each intention category and outputs the category with the highest probability as a final prediction category;

meanwhile, a loss function, namely, the loss function, corresponding to softmax is adopted, and the loss function, namely, the loss function, is as follows:

wherein  $t \in R^m$  is the correct class, denoted by one-hot,  $y \in R^m$  is the probability of each type that the softmax classifier predicts,  $m$  is the number of classification types, and  $\sigma$  is the hyperparameter set at the time of L2 regularization.

## Description

Intelligent man-machine conversation system of closed domain

Technical Field

The invention relates to the technical field of man-machine conversation, in particular to an intelligent man-machine conversation system of a closed domain.

Background

Most of the man-machine conversation systems in the industry at present adopt a mode of manually combing business logic and writing rules for conversation matching, higher cognitive

intelligence is not achieved, academic circles are mainly concentrated on certain implementation parts of the man-machine conversation systems, such as conversation state tracking, natural language understanding and the like, a complete solution for fully considering business scenes is not provided, and the problems of low fluency and single field of the man-machine conversation systems cannot be easily solved in the prior art. There are three difficulties in the interactive dialog system.

(1) Short text and grammatical deficits present challenges to the intent of understanding accuracy.

In a man-machine interaction scene, most of user inputs are spoken words in a spoken language form, and the system has the characteristics of short text, grammar deficiency and the like, so that the difficulty of user intention recognition in a dialogue system is increased.

(2) The complexity of multiple rounds of human-computer conversation is challenging to the degree of intelligence and mobility of the conversation model.

The dialogue system is different from a general question-and-answer system, and needs to track the state in the dialogue process, so that the slot value pair information in the user input needs to be analyzed.

(3) The man-machine conversation system has the challenges brought by all AI in multiple links.

The intelligent service man-machine conversation system needs to realize a complete AI system comprising a plurality of processes, and designs a very detailed solution which has low coupling degree, ensures high availability and high performance of each module and can form a complete data processing flow.

The purpose recognition in the current man-machine dialogue system can be used as a text semantic classification problem, and the traditional purpose recognition uses a method in machine learning, such as a Support Vector Machine (SVM), a Bayesian network and the like. The dialog state tracking technology is not well implemented and

applied; the dialog management mostly adopts the finite state machine and rule matching, and generates corresponding dialog actions according to different states or rules. For example, in chinese patent application No. 201410781154.9, an intention recognition method of a Support Vector Machine (SVM) is used. The system classifies and numbers collected data by using numbers, determines a characteristic value and an intention type, then performs dimensionality reduction on the collected data by using a Principal Component Analysis (PCA) method, selects a proper kernel function to map a characteristic vector to a high-dimensional space so as to separate the originally inseparable data, trains parameters by using a pre-classified Support Vector Machine (SVM) model and performs offline verification, and finally identifies the intention by using the real-time collected data. In the patent application No. 201511001276.2, a finite state machine is used, first obtaining text converted from speech input by a user; performing semantic recognition on the text to obtain the intention of the user; matching the user intention with a jump condition; and jumping to a corresponding proxy module according to a jumping condition matched with the intention of the user and a finite state automaton so as to execute the function of the proxy module and obtain an execution result.

However, the conventional dialog system is intended to ignore short text property and grammar deficiency property in a man-machine dialog scene, so that the recognition accuracy is not high; dialog state tracking is not well applied, resulting in the semantic information of the context being ignored; the dialog control ignores the control and recovery of the dialog jumping outside the domain and does not implement different feedback for the context, failing to respond to the dynamic changes of the dialog.

#### Disclosure of Invention

The present invention is directed to solving at least one of the above problems.

Therefore, an object of the present invention is to provide a closed-domain intelligent human-machine dialog system, which has higher intention



recognition accuracy, dialog state tracking accuracy and dialog control stability, and thus improves the cognitive intelligence capability of the human-machine system.

In order to achieve the above object, an embodiment of the present invention provides an intelligent human-machine interaction system of a closed domain, including: the first modeling module is used for constructing a multi-feature fusion depth intention recognition model based on a bidirectional long-time memory network BiLSTM and a convolutional neural network CNN so as to extract and fuse features of short texts input by a user with different granularities and dimensions; the second modeling module is used for constructing a MC-BLSTM-MSCNN-based dialog state tracking model by adopting a combined modeling mode of current state input and context statements of a human-computer dialog state system so as to extract the characteristics of a slot value pair; and the third modeling module is used for constructing a Bi-LSTM matching model based on the domain-outside recovery mechanism of the shift attention mechanism, so that the identified user intention and the user slot value are input into the shift network for weight distribution of the attention mechanism, and the coding of the conversation state and the matching of conversation control are realized.

In addition, the intelligent human-computer interaction system for closing the domain according to the above embodiment of the present invention may further have the following additional technical features:

in some examples, the multi-feature fusion depth intent recognition model includes at least: a multi-channel embedded layer, a high-speed channel layer and a multi-granularity convolutional layer.

In some examples, the dialog state tracking model includes: the semantic decoding module is used for directly interacting the vector representation  $r$  and the candidate slot value pair representation  $c$  input by the user and judging whether the user explicitly expresses the intention matched with the current candidate pair or not according to the vector representation  $r$  and the candidate slot value

pair  $c$ ; the context coding module is used for obtaining a system request parameter  $t_q$  and a word vector  $(t_s, t_v)$  of a confirmation action, and calculating the similarity measurement according to the  $t_q$  and the word vector  $(t_s, t_v)$ , and comprises the following steps: system output behavior  $(t_q, t_s, t_v)$ , candidate slot value pairs  $(c_s, c_v)$ , and user input statement representation  $r$ .

In some examples, the slot names and vector space representations of values of the candidate slot-value pairs are given by  $c_s$  and  $c_v$ , the tuple is mapped to a single vector  $c$  of the same dimension as the vector representation  $r$  of the user input, wherein the two vector representations interact to learn a similarity measure for distinguishing interaction between the user input statement and the slot-value pairs that it expresses or does not express, the similarity measure being specified as follows:

wherein the content of the first and second substances, representing element-wise vector multiplication, which reduces the rich feature set in  $d$  to a single scalar, allows downstream networks to better utilize their parameters by learning the nonlinear interaction between the feature sets in  $r$  and  $c$ .

In some examples, the following relationship exists between the system output behavior  $(t_q, t_s, t_v)$ , the candidate slot value pair  $(c_s, c_v)$ , and the user input statement representation  $r$ :

$$m_r = (c_s \cdot t_q) r$$

$$m_c = (c_s \cdot t_s)(c_v \cdot t_v) r$$

where, represents the dot product, the computed similarity terms act as a gated gating mechanism that is represented only by the utterance when the system queries the current candidate bin or bin value pair.

In some examples, the context encoding module is further configured to input the obtained vector representation  $m$  and the vector representation  $d$  of semantic similarity into a Softmax layer for classification.

In some examples, the Bi-LSTM matching model includes: the attention module based on displacement is used for matching the characteristics of the slot value pairs with the current intention of the user, and the obtained matching degree is used as the importance basis of the slot value pairs, so that different attention weights of different slot value pairs under different intentions are realized; the user speech feature extraction module is used for extracting user speech features according to the long-term and short-term memory network; the standard answer classification module is used for splicing the slot value pair characteristics and the user utterance characteristics, inputting the slot value pair characteristics and the user utterance characteristics into a bidirectional long-short term memory network together, and finally performing softmax classification; and the session control module supports an out-of-domain recovery mechanism and is used for adding data of the open domain data set into the closed domain data set to request recovery from the outside of the user domain.

In some examples, the shift-based attention module is configured to map the feature set to a single output, where the input and the output are matrices formed by splicing feature vectors, and the mapping is implemented by the following formula:

wherein,  $X$  represents the input groove value pair information, and the calculation formula of  $\lambda$  in the above formula is:

where  $W$  is the vector of the user's current intent and  $a$  is a learnable scalar.

In some examples, the user utterance feature extraction module extracts user utterance features according to a long-short term memory network, including: the method comprises the following steps of adopting two LSTM networks, respectively inputting current input sentences of a user into the two networks, simultaneously analyzing the sentences of the user from forward and reverse time sequences, obtaining the output of each hidden node at each time point, connecting the output of each hidden node at each time point of

the two LSTM networks through averaging, and inputting the output into the next-stage network, wherein the specific formula is as follows:

wherein the content of the first and second substances, the hidden node outputs representing the positive direction at time  $t$ , and (4) representing each output of the hidden nodes in the reverse direction at the time  $t$ , averaging the outputs of the nodes at the same time to form a two-dimensional tensor, and inputting the two-dimensional tensor into a subsequent network.

In some examples, the last layer of the Bi-LSTM matching model is a softmax fully-connected layer, as follows:

the layer predicts an intention category  $y$  corresponding to the  $S$  statement by taking the output  $h$  of the hidden layer as input, and the classifier performs probability analysis on the matching degree of the current statement of the user and each intention category and outputs the category with the highest probability as a final prediction category;

meanwhile, a loss function, namely, the loss function, corresponding to softmax is adopted, and the loss function, namely, the loss function, is as follows:

wherein  $t \in R^m$  is the correct class, denoted by one-hot,  $y \in R^m$  is softmax classification. The probability of each type that the machine predicts,  $m$  is the number of classification types, and  $\sigma$  is the hyperparameter set at the time of L2 regularization.

According to the intelligent man-machine conversation system of the closed domain, the accurate intention recognition technology based on short text user input of multi-feature fusion, the conversation state tracking technology based on context joint modeling, and the conversation control technology based on state input supporting shifting attention and an out-of-domain recovery mechanism are adopted to improve the man-machine conversation system, so that the man-machine conversation system has higher intention recognition accuracy, higher correctness of conversation state tracking and higher stability of

conversation control, and the cognitive intelligence capability of the man-machine system is further improved.

Additional aspects and advantages of the invention will be set forth in part in the description which follows and, in part, will be obvious from the description, or may be learned by practice of the invention.

#### Drawings

The above and/or additional aspects and advantages of the present invention will become apparent and readily appreciated from the following description of the embodiments, taken in conjunction with the accompanying drawings of which:

FIG. 1 is a block diagram of a structure of a closed domain intelligent human-machine dialog system according to an embodiment of the invention;

FIG. 2 is a schematic diagram of a multi-feature fusion depth intent recognition model according to one embodiment of the invention;

FIG. 3 is a diagram of a MC-BLSTM-MSCNN based dialog state tracking model according to an embodiment of the invention;

FIG. 4 is a diagram of a Bi-LSTM matching model for an out-of-domain restoration mechanism based on a shift attention mechanism according to an embodiment of the present invention;

FIG. 5 is a diagram of an example of an expansion of a out-of-domain recovery data set, according to an embodiment of the present invention.

#### Detailed Description

Reference will now be made in detail to embodiments of the present invention, examples of which are illustrated in the accompanying drawings, wherein like or similar reference numerals refer to the same or similar elements or elements having the same or similar function throughout. The embodiments described below with reference to the accompanying drawings are illustrative only for the purpose of explaining the

present invention, and are not to be construed as limiting the present invention.

In the description of the present invention, it is to be understood that the terms "center", "longitudinal", "lateral", "up", "down", "front", "back", "left", "right", "vertical", "horizontal", "top", "bottom", "inner", "outer", and the like, indicate orientations or positional relationships based on those shown in the drawings, and are used only for convenience in describing the present invention and for simplicity in description, and do not indicate or imply that the referenced devices or elements must have a particular orientation, be constructed and operated in a particular orientation, and thus, are not to be construed as limiting the present invention. Furthermore, the terms "first" and "second" are used for descriptive purposes only and are not to be construed as indicating or implying relative importance.

In the description of the present invention, it should be noted that, unless otherwise explicitly specified or limited, the terms "mounted," "connected," and "connected" are to be construed broadly, e.g., as meaning either a fixed connection, a removable connection, or an integral connection; can be mechanically or electrically connected; they may be connected directly or indirectly through intervening media, or they may be interconnected between two elements. The specific meanings of the above terms in the present invention can be understood in specific cases to those skilled in the art.

An intelligent human-machine dialog system of a closed domain according to an embodiment of the present invention is described below with reference to the accompanying drawings.

FIG. 1 is a block diagram of a structure of a smart human-machine dialog system that encloses a domain, according to one embodiment of the invention. As shown in fig. 1, the intelligent human-machine dialog system 100 of the closed domain includes: a first modeling module 110, a second modeling module 120, and a third modeling module 130.

The first modeling module 110 is configured to construct a multi-feature fusion depth intention

recognition model based on the bidirectional long-time and short-time memory network BiLSTM and the convolutional neural network CNN, so as to extract and fuse features of short texts input by a user with different granularities and dimensions.

Wherein the multi-feature fusion depth intent recognition model comprises at least: multi-channel embedded layers, high-speed channel layers, and multi-grain convolutional layers, such as shown in fig. 2.

Specifically, the embodiment of the invention aims at the requirement characteristics of task word length finiteness, text non-standardization and the like in a task-oriented man-machine conversation scene, improves the three aspects of multi-channel, multi-granularity and high-speed channel direct connection, constructs a multi-fusion depth intention recognition model based on the BiLSTM and the CNN, extracts and fuses the characteristics of short texts input by users with different granularities and dimensions, and thus improves the accuracy of user intention recognition.

The multi-channel embedded layer, the high-speed channel layer, and the multi-grain convolutional layer are described below with reference to fig. 2.

Multi-channel embedding layer: the traditional neural network adopts a single-channel word embedding layer as input, and is divided into a trainable embedding layer and an untrained embedding layer: the first trainable embedded layer is modified in the model training process to better express the semantics suitable for the scene, however, quantitative analysis on the modification cannot be performed at present, the situation of large semantic deviation caused by excessive modification may be encountered, and most values of a single embedded layer are lost after the single embedded layer is randomly discarded, so that subsequent calculation is influenced; the second untrained embedding layer directly uses the trained word vector as input, and the embedding layer cannot be automatically optimized according to input data in the model training process, so that the second untrained embedding layer is not necessarily completely suitable for the task scene, and similarly, a single embedding layer can lose

more information after passing through a random discarding layer to influence subsequent calculation. Based on this, the embodiment of the invention uses the dual-channel embedding layer as an input, and balances the original semantic information and the dynamically modified semantic information by setting whether the embedding layer can be trained or not, and meanwhile, the dual-channel input can enhance the information intensity and has higher optimization gradient in subsequent operations such as pooling convolution and the like.

High-speed channel layer: more random discarding layers and pooling layers are added in the neural network for model performance, and features are abstracted layer by layer. On one hand, the requirements of feature extraction, data dimension reduction, overfitting prevention and the like are met, and on the other hand, the problem that part of information in original data is lost is also brought. In the case of short question text in this scenario, it becomes important to fully utilize the original information. Based on this, the embodiment of the invention adds the high-speed channel for direct connection, directly fuses and outputs the original embedded layer information and the BLSTM output to the convolutional neural network layer, fully utilizes the original information and reduces the information loss under the condition of not influencing the output of the embedded layer and preventing over-fitting.

Multi-granularity convolutional layers: most of traditional models adopt single-granularity convolution kernels, and only feature extraction is carried out on sentence information on single granularity, so that the problem that feature extraction input by a user is insufficient exists. In consideration of the short text property of the scene, the short text input by the user needs to be more fully mined to obtain more accurate semantic information. Based on this, the embodiment of the present invention performs multi-range feature extraction on the question in a multi-granularity convolution kernel manner.

The second modeling module 120 is configured to construct a MC-BLSTM-MSCNN-based dialog state tracking model by using a combined modeling



manner of current state input and context statements of the human-computer dialog state system, so as to extract slot value pair features. Specifically, aiming at the problem that the traditional slot position analysis based on the labeling sequence strongly depends on a large-scale labeling data set, and the current slot value pair extraction method ignores the semantic relation between context sentences, the embodiment of the invention adopts a mode of jointly modeling the current state input and the context sentences of the man-machine conversation state system to construct a slot value pair feature extraction model based on MC-BLSTM-MSCNN, thereby improving the accuracy of identifying the specific target of the user in a certain specific field in the man-machine conversation system.

Specifically, in one embodiment of the present invention, as shown in connection with FIG. 3, the dialog state tracking model includes a semantic decoding module and a context encoding module.

The semantic decoding module is used for directly interacting the vector representation  $r$  and the candidate slot value pair representation  $c$  input by the user and judging whether the user explicitly expresses the intention matched with the current candidate pair or not according to the vector representation  $r$  and the candidate slot value pair  $c$ . For example, whether "i want to go to hangzhou" matches "with" destination hangzhou "requires the use of pre-trained high quality word vectors.

Further, the slot names of candidate slot-value pairs and the vector-space representation of the values are given by  $cs$  and  $cv$ , the tuple is mapped to a single vector  $c$  of the same dimension as the vector representation  $r$  of the user input, wherein the two vector representations interact to learn a similarity measure for distinguishing the interaction between the user input statement and the slot-value pairs that it expresses or does not express, the similarity measure being specified as follows:

wherein the content of the first and second substances, representing the element-wise vector multiplication, can be seen as a more intuitive

similarity measure, which reduces the rich feature set in  $d$  to a single scalar, allowing downstream networks to better utilize their parameters by learning the nonlinear interaction between the feature sets in  $r$  and  $c$ .

Further, since the semantic decoding module is not enough to extract intentions in a human-machine conversation, the context of the conversation must be considered for dialog state tracking in order to better understand the statements. While all previous system output and user input statements are important, the last system output statement is most relevant. The system output typically exists in one of two:

the system requests: the system asks the user for the value of a particular slot  $t_q$ . For example, the system asks "what price range you want?" then the user gives any answer, but the model must infer the reference price range, not other slots (e.g., area or food).

And (3) system confirmation: the system asks the user to confirm whether a particular slot-value pair  $(t_s, t_v)$  is part of its required constraints. For example, the system asks "how to eat Beijing duck," the user answers yes or no, and the model must be aware of the behavior of the system to correctly update the state.

Based on this, the context coding module is configured to obtain the word vector  $(t_s, t_v)$  of the system request parameter  $t_q$  and the confirmation action (zero vector, if not), and calculate the following similarity measure according to  $t_q$  and  $(t_s, t_v)$ , including: system output behavior  $(t_q, t_s, t_v)$ , candidate slot value pairs  $(c_s, c_v)$ , and user input statement representation  $r$ . Wherein the following relationship exists among the system output behavior  $(t_q, t_s, t_v)$ , the candidate slot value pair  $(c_s, c_v)$  and the user input statement representation  $r$ :

$$m_r = (c_s \cdot t_q) r$$

$$m_c = (c_s \cdot t_s)(c_v \cdot t_v) r$$

where,  $\cdot$  represents the dot product, the computed similarity terms act as a gated gating mechanism that is represented only by the utterance when the

system queries the current candidate bin or bin value pair. This type of interaction is necessary to confirm the system action: if the system is to be validated by the user, the user may not mention any slot values, but only respond positively or negatively, meaning that the model must take into account the interaction between the output statements, candidate slot value pairs, and output slot value pairs of the system. If (and only if) the latter two are the same, the model will take into account the positive or negative polarity of the user utterance when making subsequent binary decisions.

And finally, the context coding module is also used for uniformly inputting the obtained vector representation  $m$  and the vector representation  $d$  of the semantic similarity into a Softmax layer for classification.

The third modeling module 130 is used for constructing a Bi-LSTM matching model based on the domain-outside restoration mechanism of the shift attention mechanism, so as to input the identified user intention and the user slot value into the shift network for weight distribution of the attention mechanism, and realize the coding of the conversation state and the matching of conversation control. Specifically, aiming at the problem of difficult tracking of the dialog state in the dialog system, the embodiment of the invention adopts the idea of the shift attention mechanism, constructs an accurate Bi-LSTM matching network based on the shift attention mechanism, inputs the identified user intention and the user slot value into the shift network for weight distribution of the attention mechanism, and realizes accurate coding of the dialog state and accurate matching of the dialog control. Meanwhile, aiming at the problem of poor robustness in a dialogue system, the embodiment of the invention reconstructs the training data set by introducing the open domain data, thereby realizing the mechanism of recovery outside the domain and improving the robustness of the system.

Specifically, in one embodiment of the present invention, as shown in FIG. 4, the Bi-LSTM matching model includes: a shift-based attention module, a user utterance feature extraction

module, a standard answer classification module, and a dialog control module.

The attention module based on the shift is used for matching the characteristics of the slot value pair with the current intention of the user, taking the obtained matching degree as the importance basis of the slot value pair and realizing that different slot value pairs have different attention weights when different intentions.

More specifically, the shift-based attention module is configured to map a series of feature sets to a single output, where the input and the output are matrices formed by splicing feature vectors, and the mapping is specifically implemented by the following formula:

wherein,  $X$  represents the input groove value pair information, and the calculation formula of  $\lambda$  in the above formula is:

where  $W$  is the vector of the user's current intent and  $a$  is a learnable scalar. The output of the module is averaged as the final bin pair feature information, taking into account that the number of bin pairs identified in each input dialog may be different. The module performs attention mechanism calculation based on shift once for each slot value pair, different attention points exist in each calculation, and output results of all attention mechanisms are connected to serve as final slot value pair characteristic information.

And the user speech feature extraction module is used for extracting the user speech features according to the long-term and short-term memory network.

In particular, long-short term memory network (LSTM), as a model widely used in natural language processing, achieves a good effect in various tasks, and is implemented by the following formula:

$$f_t = \sigma(W_f [h_{t-1}, X_t] + b_f)$$

$$i_t = \sigma(W_i [h_{t-1}, X_t] + b_i)$$

$$o_t = \sigma(W_o [h_{t-1}, X_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

considering that the memory of the long-short term memory network to natural language is accumulated in time series, the feature of the intermediate state output of the model is based on the previous text feature, and then the utterance input by the user is mostly short text information. Based on this, in order to fully mine the information input by the user, in the embodiment of the present invention, two LSTM networks are used, the current input sentence of the user is respectively input into the two networks, the user sentence is simultaneously analyzed from the forward and reverse time sequences, in order to obtain more sufficient text information, the output of each hidden node at each time point is obtained, and the hidden node outputs of the two LSTM networks at each time point are connected by averaging and input into the next-stage network, specifically, the following formula:

wherein the content of the first and second substances, the hidden node outputs representing the positive direction at time  $t$ , and (4) representing each output of the hidden nodes in the reverse direction at the time  $t$ , averaging the outputs of the nodes at the same time to form a two-dimensional tensor, and inputting the two-dimensional tensor into a subsequent network.

And the standard answer classification module is used for splicing the obtained slot value pair characteristics and the user speech characteristics, inputting the spliced slot value pair characteristics and the user speech characteristics into a bidirectional long-short term memory network together, and finally performing softmax classification.

Wherein, the last layer of the Bi-LSTM matching model is a softmax full-connection layer, and the formula is as follows:

the layer predicts the intention category  $y$  corresponding to the  $S$  statement by taking the output  $h$  of the hidden layer as input, and the classifier performs probability analysis on the matching degree of the current statement of the user and each intention category and outputs the

category with the highest probability as a final prediction category.

Meanwhile, the loss function, namely, the loss function, is used for realizing the following relationship by adopting the loss function, namely, the loss function:

wherein  $t \in R^m$  is the correct class, denoted by one-hot,  $y \in R^m$  is the probability of each type that the softmax classifier predicts,  $m$  is the number of classification types, and  $\sigma$  is the hyperparameter set at the time of L2 regularization.

And the session control module supports an out-of-domain recovery mechanism and is used for adding data of the open domain data set into the closed domain data set to request recovery from the outside of the user domain. In particular, simple supervised learning is not sufficient to learn a stable continuous decision strategy. Because the model can only output expert conclusions and cannot perform self-error recovery and user out-of-domain request recovery, in the embodiment of the invention, the robustness of the model is increased by adding the open domain data set data to the closed domain data set. For example, as shown in FIG. 5, an example of out-of-domain recovery data set augmentation is illustrated.

For example, the following steps are carried out: an original data set is provided containing  $n$  sessions  $D = [d_0, d_1, \dots, d_N, d_N]$ , where  $d_N$  is a multi-turn closed domain session with  $|d_N|$  branches that can be turned. Then take a chat data set  $D_c = [q_0, r_0, (q_M, r_M), \dots, q_M, r_M]$ . Where  $q_m, r_m$  are questions and answers, a new data set  $D$  is created by repeating the following steps:

1. randomly selecting a section of dialogue  $d_n$  in the  $D$ ;
2. randomly selecting a turn  $t_i = a_i, u_i$  at  $d_n$ ;
3. randomly selecting a pair of questions and answers  $(q_m, r_m)$  at  $D_c$ ;
4. recombination  $t_i = a_i, q_m$ ];
5. added to the data set.

In summary, according to the intelligent human-computer dialogue system with a closed domain in the embodiment of the present invention, the human-computer dialogue system is improved by using the accurate intent recognition technology based on the short text user input with multi-feature fusion, the dialogue state tracking technology based on context joint modeling, and the dialogue control technology based on the state input with shifted attention and the out-of-domain recovery mechanism, so that the human-computer system has higher intent recognition accuracy, the correctness of dialogue state tracking and the stability of dialogue control, and further the cognitive intelligence capability of the human-computer system is improved.

In the description herein, references to the description of the term "one embodiment," "some embodiments," "an example," "a specific example," or "some examples," etc., mean that a particular feature, structure, material, or characteristic described in connection with the embodiment or example is included in at least one embodiment or example of the invention. In this specification, the schematic representations of the terms used above do not necessarily refer to the same embodiment or example. Furthermore, the particular features, structures, materials, or characteristics described may be combined in any suitable manner in any one or more embodiments or examples.

While embodiments of the invention have been shown and described, it will be understood by those of ordinary skill in the art that: various changes, modifications, substitutions and alterations can be made to the embodiments without departing from the principles and spirit of the invention, the scope of which is defined by the claims and their equivalents.

#### Patent Citations (4)

Publication number	Priority date	Publication date
<a href="#">CN106156003A</a> *	2016-06-30	2016-11-23

CN106776578A *	2017-01-03	2017-05-31
CN107169035A *	2017-04-19	2017-09-15
CN107239445A *	2017-05-27	2017-10-10
Family To Family Citations		

\* Cited by examiner, † Cited by third party

### Non-Patent Citations (3)

Title
Natural Language Generation for Spoken Dialogue System using RNN Encoder-Decoder Networks;Van-Khanh and Le-Minh Nguyen; 《CoNLL 2017, the 21st Conference on Computational Natural Language Learning》 ;20170812;第1页至第6页 *
Revisiting the Effectiveness of Off-the-shelf Temporal Modeling Approaches for Large-scale Video Classification;Yunlong Bian et al.; 《https://arxiv.org/abs/1708.03805》 ;20170812;第2页 *
Text Classification Improved by Integrating Bidirectional LSTM with Two-dimensional Max Pooling;Peng Zhou et al.; 《COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers》 ;20161121;第5页 *

\* Cited by examiner, † Cited by third party



## Cited By (22)

Publication number	Priority date	Publication date
Family To Family Citations		
<a href="#">CN109241255B</a> *	2018-08-20	2021-05-18
<a href="#">CN109165284B</a> *	2018-08-22	2020-06-16
<a href="#">CN109241265B</a> *	2018-09-17	2022-06-03
<a href="#">CN109299267B</a> *	2018-10-16	2022-04-01
<a href="#">CN109635079A</a> *	2018-10-25	2019-04-16
<a href="#">CN109741751A</a> *	2018-12-11	2019-05-10
<a href="#">CN109785833A</a> *	2019-01-02	2019-05-21

---

CN109767758B *	2019-01-11	2021-06-08
----------------	------------	------------

---

CN109885668A *	2019-01-25	2019-06-14
----------------	------------	------------

---

WO2020211006A1 *	2019-04-17	2020-10-22
------------------	------------	------------

---

CN110334340B *	2019-05-06	2021-08-03
----------------	------------	------------

---

CN110276072B *	2019-06-10	2021-07-23
----------------	------------	------------

---

CN110209791B *	2019-06-12	2021-03-26
----------------	------------	------------

---

CN110309170B *	2019-07-02	2021-04-13
----------------	------------	------------

---

---

CN110413752B \*

2019-07-22

2021-11-16

---

CN110689878B \*

2019-10-11

2020-07-28

---

CN110853626B \*

2019-10-21

2021-04-20

---

CN110765270B \*

2019-11-04

2022-07-01

---

CN111353029B \*

2020-02-22

2020-09-22

---

CN111737458A \*

2020-05-21

2020-10-02

---

CN111833872B *	2020-07-08	2021-04-30
CN112182191A *	2020-10-16	2021-01-05

\* Cited by examiner, † Cited by third party, ‡ Family to family citation

## Similar Documents

Publication	Publication Date	Title
<a href="#">CN108415923B</a>	2020-12-11	Intelligent man-machine conversation system of closed domain
<a href="#">CN108182295B</a>	2021-09-10	Enterprise knowledge graph attribute extraction method and system
<a href="#">US9875440B1</a>	2018-01-23	Intelligent control with hierarchical stacked neural networks
<a href="#">CN108319686B</a>	2021-07-30	Antagonism cross-media retrieval method based on limited text space
<a href="#">US20200097820A1</a>	2020-03-26	Method and apparatus for classifying class, to which sentence belongs, using deep neural network
<a href="#">CN108664589B</a>	2022-03-15	Text information extraction method, device,

		system and medium based on domain self-adaptation
<a href="#">Seo et al.</a>	2017	Neural speed reading via skim-rnn
<a href="#">Irsoy et al.</a>	2013	Bidirectional recursive neural networks for token-level labeling with structure
<a href="#">Li et al.</a>	2018	Image sentiment prediction based on textual descriptions with adjective noun pairs
<a href="#">Ni et al.</a>	2021	Recent advances in deep learning based dialogue systems: A systematic survey
<a href="#">Liu et al.</a>	2020	A new method of emotional analysis based on CNN-BiLSTM hybrid neural network
<a href="#">WO2021052137A1</a>	2021-03-25	Emotion vector generation method and apparatus
<a href="#">CN112131372B</a>	2021-02-02	Knowledge-driven conversation strategy network optimization method, system and device
<a href="#">Yuan et al.</a>	2015	Twitter sentiment analysis with recursive neural networks
<a href="#">CN112507039A</a>	2021-03-16	Text understanding method based on external knowledge embedding

<a href="#">Lin et al.</a>	2019	SpikeCD: a parameter-insensitive spiking neural network with clustering degeneracy strategy
<a href="#">Sur</a>	2021	CRUR: coupled-recurrent unit for unification, conceptualization and context capture for language representation-a generalization of bi directional LSTM
<a href="#">CN110532558A</a>	2019-12-03	A kind of more intension recognizing methods and system based on the parsing of sentence structure deep layer
<a href="#">Jo et al.</a>	2018	Time series analysis of clickstream logs from online courses
<a href="#">Vrigkas et al.</a>	2016	Exploiting privileged information for facial expression recognition
<a href="#">Song</a>	2019	Distilling knowledge from user information for document level sentiment classification
<a href="#">Bai et al.</a>	2021	Exploiting more associations between slots for multi-domain dialog state tracking
<a href="#">Chandra et al.</a>	2021	Utilizing Gated Recurrent Units to Retain Long Term Dependencies

		with Recurrent Neural Network in Text Classification
<a href="#">Ge et al.</a>	2019	The Application of Deep Learning in Automated Essay Evaluation
<a href="#">M'Charrak</a>	2018	Deep learning for natural language processing (nlp) using variational autoencoders (vae)

## Priority And Related Applications

### Priority Applications (1)

Application	Priority date	Filing date	Title
<a href="#">CN201710973047.XA</a>	2017-10-18	2017-10-18	Intel man mac conv syst: clos: dom

### Applications Claiming Priority (1)

Application	Filing date	Title
<a href="#">CN201710973047.XA</a>	2017-10-18	Intelligent man-machine conversation system of closed domain

### Legal Events

Date	Code	Title
2018-08-17	PB01	Publication
2018-08-17	PB01	Publication
2018-09-11	SE01	Entry into force of request for substai
2018-09-11	SE01	Entry into force of request for substai



















2020-12-11 GR01 Patent grant

2020-12-11 GR01 Patent grant

## Concepts

machine-extracted

[Download](#) Filter table 

Name	Image	Sections	Co
 recovery		claims,abstract,description	17
 memory		claims,abstract,description	12
 fusion		claims,abstract,description	10
 bidirectional		claims,abstract,description	7
 neural		claims,abstract,description	7
 extraction		claims,description	12
 interaction		claims,description	12
 behavior		claims,description	8
 calculation method		claims,description	7
 corresponding		claims,description	7
 substance		claims,description	6
 granularity		claims,description	5
 analytical method		claims,description	4
 extract		claims,description	3
 long-term memory		claims,description	3
 short-term memory		claims,description	3
 displacement reaction		claims,description	2
 sand		claims	1



■ cognitive

abstract,description

5

Show all  
concepts  
from the  
description  
section

---

[About](#)

[Send Feedback](#)

[Public Datasets](#)

[Terms](#)

[Privacy Policy](#)