Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Research on Implicit Intent Recognition Method Based on Prompt Learning

Shuhua Liu
Northeast Normal University
Lanting Li
Northeast Normal University
Ming Fang
Northeast Normal University
Chih-Cheng Hung
Kennesaw State University
Shihao Yang (yangsh861@nenu.edu.cn)
Northeast Normal University

Research Article

Keywords: Implicit intent, prompt learning, deep semantics, CIID

Posted Date: July 29th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1891913/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Research on Implicit Intent Recognition Method Based on Prompt Learning

Shuhua Liu[1], Lanting Li[1], Ming Fang[1], Chih-Cheng Hung[2] and Shihao Yang[1*]

[1*]Department of Information Science and Technology, Northeast Normal University, Changchun, 130000, China.

[2]Center for Machine Vision and Security Research, Kennesaw State University, Marietta, 75566, USA.

*Corresponding author(s). E-mail(s): yangsh861@nenu.edu.cn;
Contributing authors: liush129@nenu.edu.cn; lilt967@nenu.edu.cn;
fangm000@nenu.edu.cn; chung1@kennesaw.edu;

Abstract

As one of the core modules of the dialogue system, intent recognition plays an important role in human-computer interaction. Most of the existing intent recognition research is limited to simple, direct, and explicit intent recognitions. However, the natural human-computer interactions are flexible and diverse, and the expressions are often the euphemistic implicit intentions. Therefore, the implicit intent recognition brings new research challenges in this field. This paper pioneers a Chinese Implicit Intent Dataset CIID, which covers 7 common intents from different fields, and the data is the text containing the user's implicit intent. Based on this corpus, it is the first time prompt learning is employed for implicit intent recognition and by constructing a suitable prompt template, the model can get "relevant hints" to dig out the true intention of the user. Finally, this paper evaluates a range of classification models on CIID dataset. Experimental results show that the recognition rate of the proposed model is 97.6%, and achieves the state-of-the-art recognition accuracy. Furthermore, since it is difficult to collect the user's implicit intention data, this paper also explores the performance of these classification models on the CIID dataset with few-shot settings, and the experimental results show when the training data is reduced to 4.7%, the recognition rate of the proposed model can still keep 92.4%, which is significantly higher than other baseline models, the results further prove this proposed method is advanced and robust.

Keywords: Implicit intent, prompt learning, deep semantics, CIID

1 Introduction

Human's intention can be divided into explicit intention and implicit intention according to the way of expression[5]. Explicit intent means that the user clearly states his needs by speech or text. Implicit intent refers to the fact that the user does not express his requirements directly, and it is necessary to infer the user's true intention by analyzing their utterances[16, 23]. For example, the intent for booking hotel, the explicit intent expression is "book a hotel near the train station for one night", and the implicit intent expression may be "I missed the train and have to find a place to live nearby". It can be seen from

these examples that the expressions of implicit intent is more in line with people's daily speaking habits than the mechanical and monotonous expressions of explicit intent. However, due to the difficulty of implicit intent recognition and the lack of implicit intent corpus, most of the existing intent recognition methods focus on simple and clear utterances with explicit intents. With the increasing application of artificial intelligence in people's daily life, it is urgently required that the computer is able to infer human implicit intents besides it can understand simple and direct service instructions. By improving the intelligence level of human-computer interaction, the intelligent system provides people with more considerate services. Therefore, this paper focuses on the implicit intention recognition in the field of human-computer dialogue.

Human's intentions are usually subjective and the intentions expressed by the same sentence vary from person to person, therefore, it is very difficult to realize the generalized intention recognition. Existing intent recognition manually divides the dataset into several limited intent categories, and then converts the intent recognition into a text classification issue. The most commonly datasets for explicit intent recognition are the air travel dataset ATIS[11] and SNIPS[6] collected from the Snips personal voice assistant. In addition, there are intent recognition dataset of the banking BANKING77[3], the Chinese intent recognition dataset CAIS[20], etc. Compared with explicit intent recognition, researches on implicit intent recognition are few and there is no public dataset. As a result of that, this paper proposes a Chinese Implicit Intent Dataset CIID. Social media such as Weibo and Facebook provide a platform for users to express their thoughts, wishes and demands[21]. With the popularity of these platforms, more and more users post their status information on these platforms, many of which express their intentions explicitly or implicitly. Therefore, inspired by the literature[15], this paper collects data from Sina Weibo which is the biggest Chinese social media platform, and then conducts filtering, cleaning and annotating these data. The implicit intent dataset is constructed with seven categories, that is, ordering food, checking weather, shopping, traveling, booking movie ticket, setting reminder and booking hotel.

Implicit intent data usually contains deep-level semantic information, it is difficult for general classification models to capture this information so as to accurately identify the true intent of users. Therefore, this paper proposes an implicit intent recognition method based on prompt learning to mine users' implicit intents. Prompt refers to processing the input text and refactoring the task according to a specific template in order to more fully utilize the pre-trained language model for feature extraction. On the one hand, this method can give the pre-trained model a prompt and make it recall the learned knowledge in the pre-training stage so as to maximize the potential of the pre-trained model. On the other hand, by adding a suitable natural language template, the method provides "intent hints" to help the model infer the user's true intent. The proposed method is verified on the implicit intent dataset CIID and the experimental results show that the method achieves the state-of-the-art recognition results.

The main contributions of this paper are listed as follows:

(1) A cross-domain Chinese implicit intent dataset CIID is constructed, which contains 7 common service intents and a total of 5,042 annotated utterances. Few previous works involve the implicit intent recognition, however, it is of great significance for a more intelligent and natural human-computer interaction. Therefore, the proposed dataset CIID can promote the development of implicit intent recognition.

(2) An implicit intent recognition method based on prompt learning is proposed. This method enables the model to fully mine the implicit semantics in the utterance by constructing a prompt template. It is the first attempt of prompt learning applied to implicit intent recognition. The method is also evaluated in few-shot settings. The experimental results demonstrate the proposed method is advanced.

(3) Comprehensive experiments are carried out on a range of benchmark classifiers. Compared with the general pre-trained classification model BERT, the recognition accuracy of our proposed model is improved by 1.6%, and the F1 score is improved by 1.7%, achieving the state-of-the-art experimental results.

The rest of this paper is organized as follows: Section 2 discusses the related work of explicit

intent recognition and implicit intent recognition respectively; Section 3 introduces the proposed implicit intent recognition dataset CIID and its construction method; Section 4 describes the proposed implicit intent recognition method in details; Section 5 gives experimental results and analysis; Section 6 is conclusions and future work.

2 Related works

Intention recognition as the basis of human-computer interaction has been widely studied. Early intention recognition relies on rules written by the domain experts. Ramanand et al.[24] proposed a method based on rules and graphs to obtain the intention templates for consumption intentions, which achieved a good classification result. Their method is easy to understand, but requires different experts to manually formulate feature rules for different fields, and the recognition results heavily rely on the expert rules, so it has certain limitations. The researchers then turned their focus to statistical learning methods. Peng et al.[22] proposed a naive Bayesian approach for intent recognition. The support vector machine model proposed by Haffner et al.[10] has a great performance in the field of classification and is widely used in the task of intent recognition. However, the semantic features based on statistical learning methods are usually determined manually and empirically, and do not consider the contextual information of conversational utterances, so the deep semantics cannot be accurately identified.

The current mainstream of intent recognition methods are based on deep learning. Kim et al.[14] first tried to use CNN for text classification tasks, proposed a textCNN model, and achieved an ideal results. Ravuri et al.[25] proposed to use RNN and LSTM respectively to achieve intent recognition, and verified it on the ATIS dataset. The results show that the recognition performance of LSTM is superior to RNN. Balodis et al.[1] proposed an intent detection system based on FastText word embeddings and neural network classifiers, and achieved the state-of-the-art results on three datasets. With the pre-trained language models such as BERT[7] are proposed, the performance of NLP tasks has been greatly improved. Chen et al.[4] proposed a BERT-based model to joint intent classification and slot filling task. Compared to the attention-based recurrent neural network models, the model achieves significant improvements in intent classification accuracy, slot filling F1-score, and sentence-level semantic framing accuracy on multiple public benchmarks.

Most of the existing works focus on explicit intent research, on the contrast, few researches are involved for implicit intent recognition. Fu et al.[9] made a preliminary exploration in this field and proposed a method for automatic recognition of implicit consumption intention in social media. They regarded implicit consumption intention recognition as a multi-label classification problem which comprehensively used user's attention behavior, intention attention behavior, intention forwarding behavior and other features. Li et al.[15] used an attention-based encoder-decoder model for implicit intent detection and first constructed a parallel corpus of implicit-explicit intent for training the model, and then recognize the converted explicit intent. Jia et al.[12] used transfer learning to integrate the data of Jing Dong Q&A platform and a small amount of labeled Weibo data to construct the training set, and proposed a bidirectional long short-term memory(BiLSTM) neural network based on attention mechanism to identify users' implicit consumption intentions.

The current research on intent recognition is dominated by data-driven deep learning methods, thus, the lack of implicit intent datasets is one of the main reasons which restrict the development of this field. Most of currently public datasets are for explicit intent recognition, here list part of the datasets and their attribute information in Table 1. The earliest proposed ATIS[11] is a commonly used intent recognition dataset and the data comes from the corpus in the field of air traveling which contains 5,371 utterances and 21 intents. The Banking-77[3] is also a single-domain intent recognition dataset, derived from online queries of the bank customers, including 13,083 utterances and 77 fine-grained intents. Both SNIPS[6] and CAIS[20] datasets are cross-domain intent recognition datasets, containing 13,784 and 10,013 utterances respectively. Above datasets only involve the explicit intent information, which is monotonous and straightforward. However, human's daily communication is flexible, complex, and often contains many

overtones. Therefore, improving robots or computers the ability of understand implicit intention is an essential and critical step for implementation human-like intelligence. Therefore, this paper proposes a Chinese implicit intent dataset CIID, which contains 7 common user intents across different domains, with a total of 5042 utterances.

For the proposed CIID dataset, we adopt an advanced prompt-based method to identify the implicit intents in the utterances. We firstly construct prompt templates for the input which can convert the original classification task into a masked language modeling task, then use the pre-trained language model to predict the masked token in the template, and finally map it to the true label. For implicit intent recognition tasks, appropriate prompts can provide semantic hints to the model, assisting the model to infer the user's true intent. For example, an utterance " I missed the train and have to find a place to live nearby", after constructing the prompt the sentence could be " I missed the train and have to find a place to live nearby, so I want to ____", which can help the model to recognize the user's implicit intent —book a hotel.

3 CIID dataset

In order to enable robots to understand the deep semantics of the users and develop the research of intent recognition, this paper proposes a Chinese Implicit Intention Dataset CIID. To the best of our knowledge, no such dataset has been proposed in previous studies. The overview of this dataset and the construction process will be briefly described.

3.1 Overview

The CIID dataset contains a total of 5042 utterances, covering 7 service intents that commonly used in human-computer interaction, including ordering food, checking weather, shopping, traveling, booking movie ticket, setting reminder and booking hotel. The distribution of each category is shown in Fig.1. Among them, shopping intent accounts for the highest proportion, the reason is because users are more inclined to share their shopping intentions or product-related information in social media, on the contrary, hotel reservations may involve the user privacy, so fewer related intention utterances are collected.

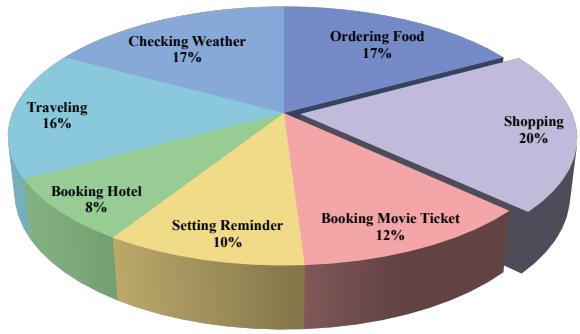


Fig. 1 Data distribution in CIID, showing the proportion of each category in the dataset

All data of CIID are textual information containing users' implicit intentions, and some examples are showed in Table 2. Implicit intent text does not contain any explicit intent instruction while expresses the corresponding intent. For example, for the corpus in Table 2, "I'm so hungry, but I don't want to cook." Although there is no expression related to ordering foods, actually, it is the user's true intention.

3.2 Dataset construction

The construction of the CIID dataset includes two stages. In the first stage, a small-scale seed dataset is manually constructed. In the second stage, implicit intention utterance templates are abstracted from the obtained seed dataset, and social media data is crawled based on these templates to expand the dataset. The specific construction process of CIID is shown in Fig.2.

3.2.1 Construct the seed dataset

In order to promote the intelligence in human-computer interaction, the dataset selects 7 common user services of the intelligent assistants as intent labels. We first recruit a number of volunteers and let them express their service requirements in a variety of ways, but these utterances cannot contain the keywords directly related to the intention. For example, for checking weather intent, participants could not directly ask "what is the weather?" and even could not say the

Table 1 Intent recognition datasets

Dataset	Size	Data Source	Intents	Explicit/Implicit(E/I)
ATIS[11]	5,871	Airline Travel Information Systems	21	E
SNIPS[6]	14,484	Snips	7	E
Banking-77[3]	13,083	Customer service queries	77	E
Chatbot Corpus[2]	206	Telegram chatbot	2	E
Ask Ubuntu[2]	162	StackExchange platforms	5	E
Web Apps[2]	89	StackExchange platforms	8	E
CAIS[20]	10,013	Chinese Artificial Intelligence Speakers	–	E
CIID(Ours)	5,042	Weibo	7	I

Table 2 Part of the utterances in CIID

Utterances	Intents
I'm so hungry, but I don't want to cook.	Ordering Food
KFC is really tasty. I haven't eaten it for a long time...	Ordering Food
It has been raining for several days, when will the sun come out?	Checking Weather
I want to wear a skirt tomorrow, will it be cold?	Checking Weather
It's almost New Year's Eve, it's time to start stocking up.	Shopping
I need a new cup with a straw! !	Shopping
The world is so big and I want to see it.	Traveling
I've been waiting for this movie for a long time, finally released!	Booking Movie Ticket
My poor memory...I forgot to sign in for a few days.	Setting Reminder
I have missed the assignment submission time for the third time.	Setting Reminder
I have to stay outside for two nights for the postgraduate entrance examination next week.	Booking Hotel
Is there any places to live around Disneyland?	Booking Hotel

keyword "weather". Finally, we collected more than 300 utterances from all the participants, and obtained a small-scale seed dataset after further deduplication and annotation operations.

3.2.2 Dataset augmentation

Due to the limited scale of manually collected dataset, this paper fully utilizes the rich data resources from the social media to expand the seed dataset. Compared with other data resources such as movie scripts and personal voice assistants, Weibo not only contains richer user demand status, but also is easier to obtain. Therefore, this paper conducts corpus from the popular social media platform—Sina Weibo [1] to obtain more posts containing implicit intentions.

Specifically, we abstract a number of implicit intention utterance templates which can represent the utterances based on the seed dataset, in order to ensure the richness of the data, there are more than 20 implicit intention utterance templates of each category. Then, Weibo posts in the

[1] https://weibo.com

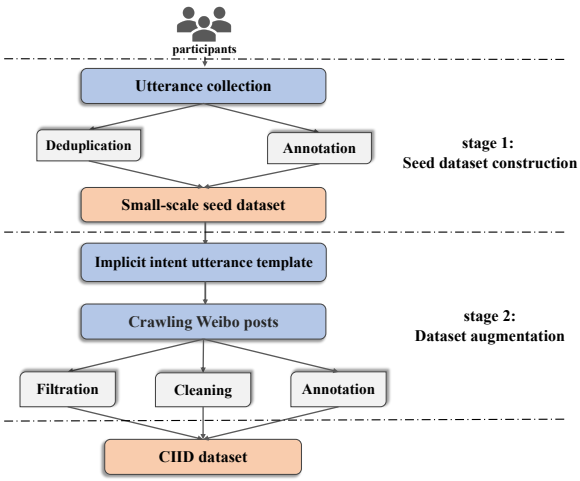


Fig. 2 The process of dataset construction

past three years is crawled by the scrapy crawler according to the generated templates, and a total of about 20,000 posts are obtained, we filter out the posts that really contain implicit intents to expand the dataset. Fig.3 shows an example of the data augmentation process, " I forgot to clock in again this morning " is an utterance in the seed dataset, and the intent category is setting reminder. First, we analyze the keywords in the sentence that determine the intention, and then construct the implicit intention utterance template "____forgot to ____again____" according to the obtained template. Finally, we crawl the Weibo posts which have the similar semantics to the seed dataset. Since the crawled posts are very noisy and contain lots of redundant information such as external links and expressions, we further filter out the posts which are semantically inconsistent with the seed dataset. For example, a searched post is " I forgot what I dreamt about this morning again! ", although it conforms to the pattern in the template in Chinese, but its semantics are not the same as the seed data. After filtering, we continue to clean the data to remove other redundant information such as expression signals and other nformation that may involve user privacy such as someone's name. Finally, a total of 5042 implicit intent annotated utterances are obtained, as shown in Fig.1.

4 Intent recognition method based on prompt learning

4.1 Prompt learning

Prompt learning is the new paradigm of natural language processing, which can prompt the pre-trained model to recall the learned knowledge in the pre-training stage by constructing a suitable prompt, so as to unify the downstream tasks with the pre-training tasks[18]. The literature[26] is an earlier work on prompt learning and proposes a PET method which is a semi-supervised training process. PET reformulates the input example into a cloze style sentence to help the language model understand the given task. Jiang et al.[13] used a paraphrase-based method to automatically generate high-quality and diverse prompt templates, while using an ensemble approach to combine answers from different prompt templates. The literature[19] improved the prompt template, and proposed to use some pseudo-prompts, such as the [unused] token in the vocabulary, to replace the manually-defined prompts, so that the prompt tokens can automatically learn and update during training stage. This paper extends prompt learning to the field of implicit intent recognition for the first time. We infer that a suitable prompt can not only give hints to the pre-trained model, but also provide semantic hints to the entire recognition model, so as to help the model understand the implicit intent better.

4.2 Implicit intent recognition method

This paper proposes an implicit intent recognition method based on prompt learning to understand the implicit intents better. The architecture is shown in Fig.4. Firstly, we construct different prompt templates $prompt_1 \dots prompt_n$ for the input, and these templates contain the masked tokens that need to be predicted. At the same time, the mapping between the predicted word and the label is established. In the following step, a pre-trained Chinese BERT is adopted to predict words at the masked positions and maps them to the actual labels. Finally, the $prompt_i (1 \leq i \leq n)$ with the best performance is selected for the intent category prediction.

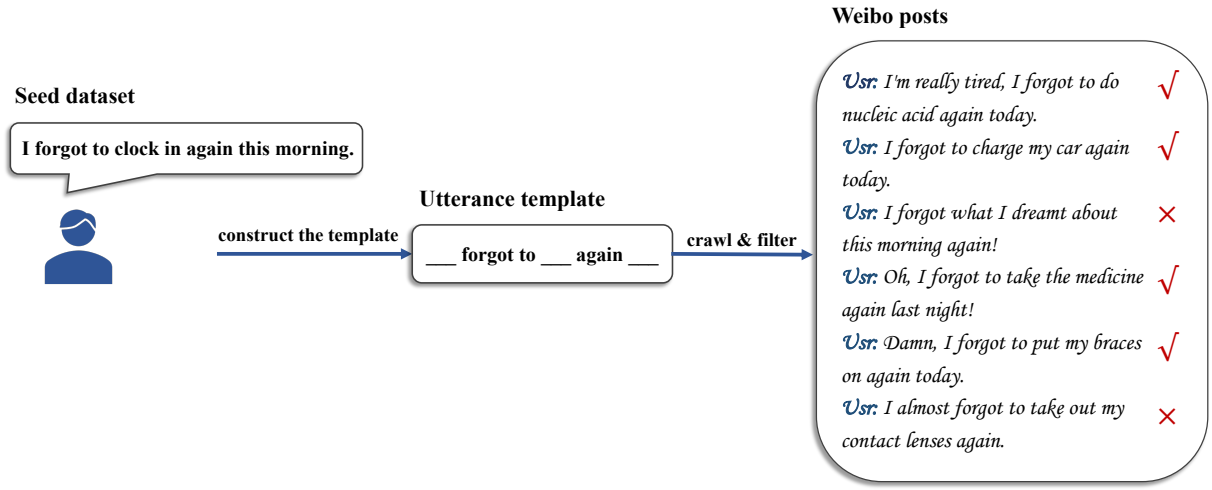


Fig. 3 Example of the data augmentation process

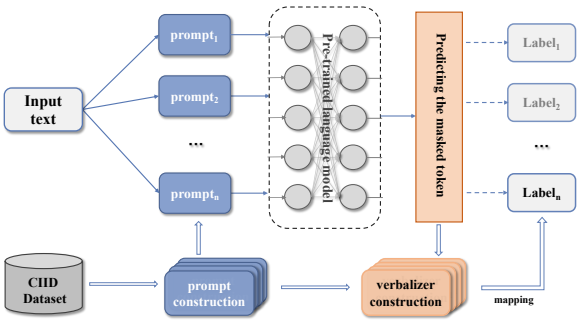


Fig. 4 The architecture of the proposed model

4.2.1 Construction of prompt and verbalizer

In order to explore the influence of giving a certain explicit prompt to the model for identifying the implicit intent, this paper uses artificially created prompts and verbalizer to build the model, which is the most natural and direct way. Specifically, for the input text X, the prompt is defined as function P, the input after constructing the prompt is:

$$X' = P(X) \tag{1}$$

Where X' is a sentence that contains X and the masked tokens. We denote the class label as $L = (l_1, l_{2,\ldots,} l_k)$,and denote the predicted vocabulary as $V = (v_1, v_{2,\ldots,} v_m)$, the verbalizer represents the mapping of $L \rightarrow V$. Since the selection of prompt templates is very important to the performance of the whole model, this paper designs different explicit intent prompt templates for the input based on human knowledge and experience.

Table 3 shows the four prompts constructed in this paper and a verbalizer example corresponding to prompt P_4. For different prompt templates, the predicted words in the verbalizer will be changed.

4.2.2 Prediction and answer mapping

By constructing the prompt, we obtain the reconstructed text X' containing the masked tokens, the intent classification task is transformed from assigning a meaningless label for the input to selecting the most suitable word for the masked positions. Therefore, BERT is employed to predict the masked tokens in X'. Since the Chinese BERT predicts character-by-character, this paper predicts the two masked tokens m_1, m_2 respectively and obtain the predicted score distribution s_1, s_2.

$$s_1 = \underset{w \in v_1}{\text{MLM}}\left(w \mid m_1\right) \tag{2}$$

$$s_2 = \underset{w \in v_2}{\text{MLM}}\left(w \mid m_2\right) \tag{3}$$

$\text{MLM}\left(w \mid m\right)$ represents the unnormalized score assigned to the character w at the mask position m, v_1 is the set of the first characters in all predicted words, and v_2 is the set of the second characters.

Then, the scores of all predicted words in verbalizer are calculated by the obtained character score distributions. Here, the word score is calculated by simply multiplying scores of the two characters. After calculating the score distribution s of the predicted words, this paper normalizes

Table 3 Model settings

Prompt	P_1(X)=“ X & so I want to [mask] ”
	P_2(X)=“ X & the intent of this sentence is [mask] ”
	P_3(X)=“ X & this sentence expresses the intent of the [mask] field ”
	P_4(X)=“ X & the topic of this sentence is about [mask] ”
Verbalizer	V(“checking weather”)=[“weather” ,“meteorology”]
	V(“shopping”)=[“commodity” ,“purchase”]
	V(“ordering food”)=[“eating” ,“catering”]
	V(“traveling”)=[“tourism” ,“travel”]
	V(“booking hotel”)=[“hotel” ,“accommodation”]
	V(“booking movie ticket”)=[“movie” ,“theater”]
	V(“setting reminder”)=[“remind” ,“schedule”]
Language Model	BERT-base-chinese

it and obtains the probability distribution p of all predicted words. Then, the probabilities of the predicted words belonging to the same category are averaged, and the obtained probability distribution $\bar{p}$ represents the corresponding label distribution. The prediction process is shown in Fig.5.

$$s = s_1 * s_2 \tag{4}$$

$$p = \text{ normalize } (s) \tag{5}$$

$$\bar{p} = \underset{l(v_i)=l(v_j)}{\text{average}} \left(p_{v_i}, p_{v_j}\right) \tag{6}$$

The probability distribution $\bar{p}$ of the class labels is obtained, and the maximum probability is selected as the final predicted class $\hat{l}$ of the model.

$$\hat{l} = \text{argmax}(\bar{p}) \tag{7}$$

4.2.3 Training loss

In the training phase, the probability distribution $p\left(v \mid X'\right)$ of all predicted words is predicted by the pre-trained model, and then the mean probability of predicted words belonging to the same category is calculated as the corresponding label probability distribution $\bar{p}(l \mid X)$. The loss function is defined as:

$$L(y, \bar{p}) = -\sum_{i=1}^{k} y_i \log\left(\bar{p}_i\right) \tag{8}$$

where k is the number of labels and y represents the ground truth.

5 Experiments

To validate the constructed CIID dataset and proposed implicit intent recognition method based on prompt learning in this paper, comprehensive experiments are carried out on a range of classification models. In this section, the experimental setup, experimental results and analysis are presented.

5.1 Baselines

(1) Non-pretrained models. This paper uses non-pretrained models such as BiLSTM [17], FastText [1], TextCNN [14] as baseline models. First, the input sentences are encoded by pre-trained word vectors, then features are extracted through the neural network layers, and finally intent classification is performed through a softmax function.

(2) Pre-trained model. The pre-trained model BERT[7] is adopted as the baseline model.

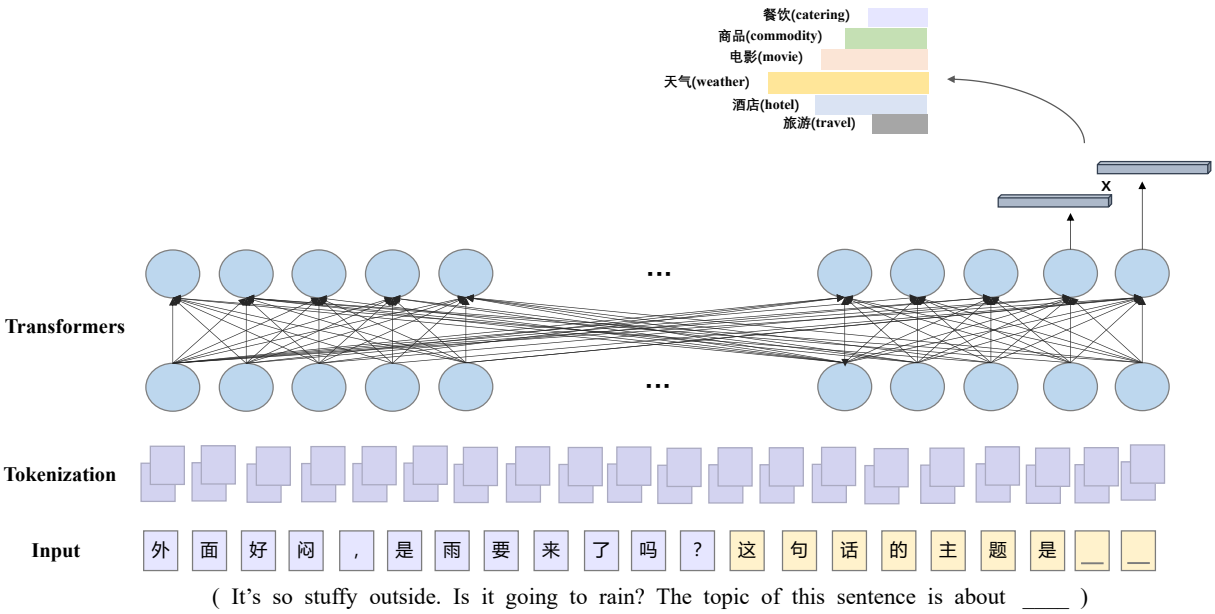


Fig. 5 The process of prediction.Firstly, the pre-trained language model predicts the character scores in the masked positions;Then multiply the character scores to get the word score;Finally, the word probability is calculated for category prediction

After fine-tuning the BERT model on the proposed dataset CIID, the semantic representation vector [cls] is extracted, and then the intent classification is performed after the linear transformation.

5.2 Experimental settings

The hardware platform is Intel® Core™ i9-10,900 with 2.8 GHz, NVIDIA GeForce RTX 3090, and Ubuntu 18.04.3 OS. The proposed implicit intent recognition model is built based on the OpenPrompt[8] toolkit which is provided by the Natural Language Processing Laboratory of Tsinghua University, and the rest of models are implemented based on the Pytorch framework. This paper divides the training set, validation set and test set into 4242, 300, 500 and the detailed statistics are shown in Table 4. Among all the models evaluated on the CIID dataset, the learning rate of the non-pretrained model, pre-trained model and our proposed model is set to 1e-3, 5e-5, 3e-5, and the epoches are 100, 10, 10 respectively. To avoid overfitting, all of the models adopt an early stopping strategy. The evaluation indicators are F1 score and Accuracy.

Table 4 Statistics of the CIID dataset

		Utterance	
Split	Samples	Avg.length	$\|V\|$
Train	4242	18.23	2435
Val	300	18.31	1038
Test	500	18.33	1300
Total	5042	18.24	2539

$|V|$ represents the vocabulary size of the dataset.

5.3 Experimental results and analysis

Table 5 shows the experimental results with different prompt templates for classification on the proposed CIID dataset. It can be seen that different prompts affect the recognition accuracy of the model. But no matter which prompt template is used (P_1, P_2, P_3 or P_4), the recognition rate of the corresponding model is higher than that of other five classification models in Table 6. Among

these prompts, the model P4 obtains the best performance, so this study uses the results of P_4 to compare with other models below.

Table 5 Performance of different prompt templates

Prompt	Accuracy	F1
P_1	0.970	0.969
P_2	0.968	0.969
P_3	0.970	0.968
P_4	0.976	0.975

Table 6 shows the experimental results of all the evaluated models on the CIID dataset, where Precision, Recall and F1 refer to the macro average. As can be seen from Table 6, the F1 score and accuracy of our proposed model are 0.975 and 0.976 respectively, and achieve the state-of-the-art results. From the results we can conclude that: (1) BERT model has certain advantages over other non-pretrained models such as BiLSTM[17], FastText[1] and TextCNN[14] because it owns the general language representation obtained by pre-training based on the large-scale corpus. (2) Compared with the BERT[7] model, our prompt-based model can more fully utilize the knowledge learned in the pre-training stage, furthermore, the prompts also play an important role in understanding text semantics.

Fig.6 analyzes the F1 scores of the baseline models BiLSTM+Att[27], BERT[7] and the proposed prompt-based model on each category. The BiLSTM+Att model adds the attention layer on the basis of the BiLSTM model which can assign weights for the words to enhance the performance and interpretability of the model. However, the BiLSTM+Att model has the lowest F1 score in each category in Fig.6, which shows that only relying on optimizing the neural network cannot fully capture the implicit semantics in the sentence. The BERT model is pre-trained on a large-scale corpus and acquires the general knowledge, which is helpful for understanding implicit semantics. Therefore, F1 scores of all categories are higher than BiLSTM+Att model. Compared with above models, our proposed prompt-based model achieves the highest F1 score in almost all categories, which shows our method can not only understand the implicit intent correctly but also be applicable to different fields.

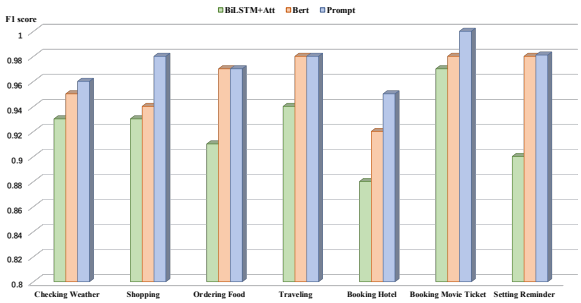


Fig. 6 F1 score of each category in different models

5.4 Ablation experiments for few-shot settings

The lack of data sources has always been an important problem for intention recognition, especially for implicit intention recognition. Because the acquisition and annotation of the implicit intent data is very difficult, it urgently requires the model shows good recognition performance even when the training set is small. Therefore, this paper explores the performance of the proposed model under different-scale training sets. The experimental results are shown in Table 7.

The experimental results in Table 7 show that when the training set is reduced to about 47% of the original, the accuracy of our model is 96.4%, and when the training set is reduced to 4.7% of the original, the accuracy of our model can still reach 92.4%. The results prove that our proposed model presents strong robustness and good recognition accuracy even when the training data is few-shot.

To make the conclusions more convincing, this paper conducts comparative experiments on different models to explore the influence of the training scale, the results are shown in Fig.7. When the training samples are reduced to 1000, the recognition accuracy of all evaluated models decreases slightly. However, when the training samples are reduced to 500, the recognition accuracy of the proposed model is 94%, while that of other models have dropped significantly. With the continuous reduction of the training samples, the recognition accuracy of the traditional neural network models

Table 6 Performance comparison between prompt-based model and other baselines

Classifier	Intent Classification			
	Precision	Recall	F1	Accuracy
BiLSTM[17]	0.909	0.899	0.903	0.914
BiLSTM + Attention[27]	0.929	0.921	0.924	0.928
FastText[1]	0.947	0.942	0.943	0.946
TextCNN[14]	0.955	0.952	0.953	0.954
BERT[7]	0.958	0.959	0.958	0.960
Prompt-based Model(Ours)	0.973	0.978	0.975	0.976

Table 7 Performances of prompt-based model under different training scales

	Training Set	Proportion	Accuracy	F1
Prompt-based Model	4242	100%	0.976	0.975
	2000	47.1%	0.964	0.967
	1000	23.6%	0.952	0.953
	500	11.8%	0.940	0.937
	200	4.7%	0.924	0.924
	100	2.4%	0.902	0.890
	50	1.2%	0.866	0.846
	20	0.5%	0.794	0.765
	10	0.2%	0.776	0.747

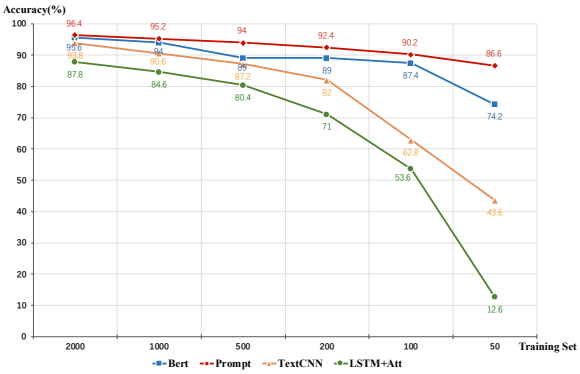


Fig. 7 Influence of training scale on different models

drop sharply. Compared with the neural network model, BERT[16] has a smaller accuracy decrease owing to its prior knowledge. The prompt-based model can obtain stable recognition accuracy even if the training set is extremely small, which proves that the proposed model has the highest recognition accuracy and also has a strong few-shot learning ability.

6 Conclusions and future works

This paper elaborates the research on implicit intention recognition. Firstly, a new implicit intention dataset CIID is built, which contains 7 common and representative intents in human-computer interaction, with a total of 5042 utterances. Secondly, a prompt-based approach is proposed to identify the implicit intents in the CIID

dataset, and is compared with a series of benchmark models. The experimental results show that the proposed model achieves the best performance currently since constructing a suitable prompt can enhance the inferring ability to users' true intents. Finally, we evaluate the models in few-shot settings, and it turns out that the prompt-based model performs well even when the training samples are insufficient.

In the future, greater variety of intents will be built according to the daily requirements in human-computer interaction, and covering more aspects of our lives as many as possible. In addition, since human intentions are often subjective and vary from person to person, we will further explore the fusion of multimodal information and combine the user portraits to achieve a better recognition performance.

Acknowledgments. This work was supported by the project of Changchun Bureau of Science and Technology [grant number 21ZY31] and the project of Jilin province development and reform commission [grant number 2022C047-5].The authors gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

Statements and Declarations

- Funding

  This research was funded by the project of Changchun Bureau of Science and Technology under the Grant 21ZY31.
- Competing interests

  The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- Ethics approval

  Written informed consent for publication of this paper was obtained from the Northeast Normal University and all authors.
- Authors' contributions

  This study was completed by the co-authors. Shuhua Liu conceived the research. The major experiments and analyses were undertaken by Lanting Li and Shihao Yang. Ming Fang was responsible for data processing and drawing figures. Chih-Cheng Hung edited and reviewed the paper. All authors have read and approved the final manuscript.

References

[1] Balodis K, Deksne D (2019) Fasttext-based intent detection for inflected languages. Information 10(5):161

[2] Braun D, Mendez AH, Matthes F, et al (2017) Evaluating natural language understanding services for conversational question answering systems. In: Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue, pp 174–185

[3] Casanueva I, Temčinas T, Gerz D, et al (2020) Efficient intent detection with dual sentence encoders. arXiv preprint arXiv:200304807

[4] Chen Q, Zhuo Z, Wang W (2019) Bert for joint intent classification and slot filling. ArXiv abs/1902.10909

[5] Chen Z, Liu B, Hsu M, et al (2013) Identifying intention posts in discussion forums. In: Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: human language technologies, pp 1041–1050

[6] Coucke A, Saade A, Ball A, et al (2018) Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. arXiv preprint arXiv:180510190

[7] Devlin J, Chang MW, Lee K, et al (2019) BERT: Pre-training of deep bidirectional

transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota, pp 4171–4186, https://doi.org/10.18653/v1/N19-1423, URL https://aclanthology.org/N19-1423

[8] Ding N, Hu S, Zhao W, et al (2021) Openprompt: An open-source framework for prompt-learning. arXiv preprint arXiv:211101998

[9] Fu B, Liu T (2016) Implicit user consumption intent recognition in social media. Journal of Software 27(11):2843–2854

[10] Haffner P, Tur G, Wright JH (2003) Optimizing svms for complex call classification. In: 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03)., IEEE, pp I–I

[11] Hemphill CT, Godfrey JJ, Doddington GR (1990) The atis spoken language systems pilot corpus. In: Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990

[12] Jia Y, Han D, Lin H, et al (2020) Consumption intent recognition algorithms for weibo users. Beijing Da Xue Xue Bao 56(1):68–74

[13] Jiang Z, Xu FF, Araki J, et al (2020) How can we know what language models know? Transactions of the Association for Computational Linguistics 8:423–438. https://doi.org/10.1162/tacl_a_00324, URL https://aclanthology.org/2020.tacl-1.28

[14] Kim Y (2014) Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Doha, Qatar, pp 1746–1751, https://doi.org/10.3115/v1/D14-1181, URL https://aclanthology.org/D14-1181

[15] Li C, Du Y, Wang S (2017) Mining implicit intention using attention-based rnn encoder-decoder model. In: International conference on intelligent computing, Springer, pp 413–424

[16] Liu J, Li Y, Lin M (2019) Review of intent detection methods in the human-machine dialogue system. In: Journal of Physics: Conference Series, IOP Publishing, p 012059

[17] Liu P, Qiu X, Huang X (2016) Recurrent neural network for text classification with multi-task learning. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. AAAI Press, IJCAI'16, p 2873–2879

[18] Liu P, Yuan W, Fu J, et al (2021) Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. arXiv preprint arXiv:210713586

[19] Liu X, Zheng Y, Du Z, et al (2021) Gpt understands, too. arXiv preprint arXiv:210310385

[20] Liu Y, Meng F, Zhang J, et al (2019) Cm-net: A novel collaborative memory network for spoken language understanding. arXiv preprint arXiv:190906937

[21] Park DH, Fang Y, Liu M, et al (2016) Mobile app retrieval for social media users via inference of implicit intent in social media text. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, pp 959–968, https://doi.org/10.1145/2983323.2983843

[22] Peng F, Schuurmans D, Wang S (2004) Augmenting naive bayes classifiers with statistical language models. Information Retrieval 7(3):317–345

[23] Qian Y, Ding X, Liu T, et al (2017) Identification method of user's travel consumption intention in chatting robot. SCIENTIA SINICA Informationis 47(8):997

[24] Ramanand J, Bhavsar K, Pedanekar N (2010) Wishful thinking-finding suggestions and'buy'wishes from product reviews. In: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, pp 54–61

[25] Ravuri SV, Stolcke A (2015) Recurrent neural network and lstm models for lexical utterance classification. In: INTERSPEECH, https://doi.org/10.21437/interspeech.2015-42

[26] Schick T, Schütze H (2021) Exploiting cloze-questions for few-shot text classification and natural language inference. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. Association for Computational Linguistics, Online, pp 255–269, https://doi.org/10.18653/v1/2021.eacl-main.20, URL https://aclanthology.org/2021.eacl-main.20

[27] Zhou P, Shi W, Tian J, et al (2016) Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers), pp 207–212, https://doi.org/10.18653/v1/p16-2034