Chinese Intention Recognition Based on Domain Dictionary and Hard Attention

Yanli Liu, Liancheng Xu

School of Information Science and Engineering Shandong Normal University Jinan China

Keywords: intention recognition, hard attention, short text classification, domain dictionary

Abstract: This paper proposes a dictionary classification method that incorporates hard attention mechanism in the classification of intention domain, an important link in human-computer dialogue.Based on the common method of classification of sentences, through the establishment of intent on the classification of the data of artificial domain dictionary, and then manually collected sentence to clean, cut and remove the stop operation, then iterative matching and dictionary before joining attention mechanism, will be more important information in the sentence first match, through the test on real data sets, this article selects the traditional machine learning method and the emerging neural network as a comparison method, verified: in this paper, on the accuracy and speed is better than the other.

1. Introduction

Man-machine dialogue (also known as dialogue system) is one of the most challenging problems in natural language understanding tasks. The origin of man-machine dialogue can be traced back to Turing test proposed by Turing in the 1960s [1].The purpose of human-computer dialogue is to enable the computer to understand the natural language communication between human and machine by human control and intervention.Making computers capable of understanding and using natural languages has become one of the main goals of the development of the fifth generation computer.

The research on man-machine dialogue is mainly divided into two aspects: voice dialogue and written dialogue.From the perspective of its ontology composition and business logic, we can divide it into domain task-oriented and open information interaction. Domain task-oriented takes specific task operations as the interactive target, and common flight, weather query and so on all belong to domain task-oriented interaction.The relatively non-domain-specific interactive tasks, such as chatting, are all open interactive tasks [2].According to the traditional view, human-computer interaction is mainly composed of input, understanding and output. Intention domain analysis is the main function of the input stage, and ensuring the real intention of the interaction problem is the premise for the correct human-computer interaction. Therefore, the classification of intention domain as the starting point of

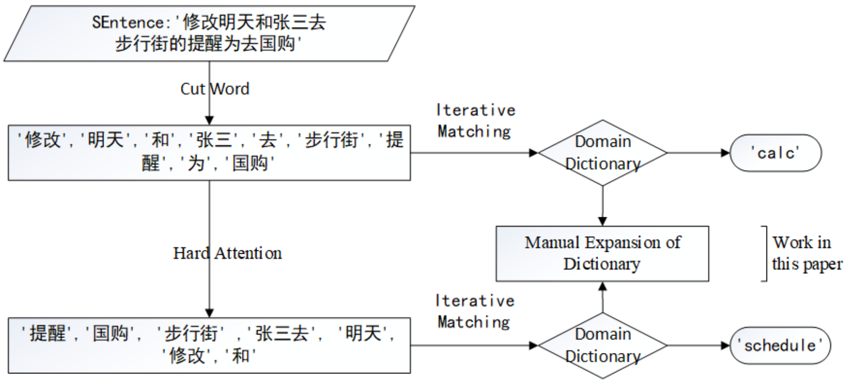

Figure 1. A simple example to explain the work of this article

Human-computer dialogue is particularly important.As a key link in human-computer dialogue, intention analysis is mainly to realize the domain recognition of user's voice or text input. Through the semantic analysis of input text information, the real intention analysis of interactive input can be realized, and then subsequent understanding and output can be made.

At present, intention domain classification is mostly applied in the single - round task dialogue of information robot.Most of the information robots use the interactive mode of passive questions and answers. These information robots get answers by connecting to search engines. Typically, such robots like Cortana equipped with Microsoft Windows operating system interact with people through connecting to bing search engine.Siri of iPhone is an information robot with both task-oriented and open information interaction. In addition to daily chatting, Siri can also manipulate and control devices through various instructions of people.

It can be seen that the research on the classification of intention domain plays an important role in the field of artificial intelligence. However, it is a pity that there are very few researches on the classification of intention domain at present. One of the important reasons is the lack of experimental data[3].The data used in this paper is the data set published by smp2017-ecdt , which is from the user log of smp2017-ecdt, and is the text form converted by the voice input received by the information robot.All the data is divided into 31 fields,They are app, bus, calc, chat, cinemas, contacts, cookbook, datetime, email, epg, flight, health, lottery, map, match, message, music, news, novel, poetry, radio, riddle, schedule, stock, telephone, train, translation, tvchannel, video, weather, and website.

In this paper, the commonly used methods in sentence classification are used to establish the dictionary of intention field. Then the sentences are cleaned and word segmentation is carried out. Through the attention mechanism, different weights are added to the sentences after word segmentation.

2. Related work

Most of the traditional intention classification methods adopt machine learning algorithm, rely on the principle of statistics, and use the manually annotated data set to conduct supervised training to obtain the classifier, so as to achieve the purpose of classification of new data [4].Naive bayes [5], support vector machine, maximum entropy model and logistic regression are the main representatives of machine learning classification methods. Compared with traditional machine learning methods, graph sorting method can achieve better results [6].In certain fields, such as consumer intention [9], customer service system [10] and film [11], they are widely used.Compared with sentences in other similar tasks, and the data of sentences classified in the intent-domain of human-computer dialogue task are sparser. Because many sentence structures of colloquial input are not complete, the accuracy of machine learning algorithm in the intent-domain classification task is not high.

With the development of neural networks, it has become a common method to classify sentences in intention domain.In literature [10], an LSTM model that is more accurate than the RNN model was obtained through training on multiple data sets. In literature [11], Microsoft staff made a comparative analysis of common CNN, LSTM, RNN and GRU models.Thus it is easy to see we use neural network to solve the intention domain classification task of the data in the scale and the quality is the key to complete the task, deep learning in training data set of annotation requires a lot of manpower material resources, Internet companies for the protection of the user and consideration on the development of their choice to strict protection of data, it's difficult for us to get in touch with the high quality set of training corpus.

Keyword matching is an excellent method to solve classification tasks in the field of intention. Its accuracy rate is higher than other machine learning methods. It also has excellent performance in computing speed and is widely used in various evaluation tasks.

In this paper, a new model is established by using the keyword matching method commonly used in sentence classification and attention mechanism, which solves the problem of multiple keywords matching that is difficult to be solved in traditional keyword matching methods.

3. Methods

Compared with other data, the data analyzed in this paper is shorter, and its sparsity and irregularity are more obvious.There are more prescriptive and colloquial data.Considering the characteristics of the data in this paper, we adopted the keyword matching method to study the task. Before keyword matching, we added the attention mechanism to increase the weight of the information keywords that we think are more important in the sentence.Experimental results show that our method is more accurate and faster than the traditional machine learning method, neural network method and keyword matching method, especially in some specific areas can even achieve 100% accuracy.

This chapter is composed of three parts. The first part: this paper analyzes the key words in each field according to the data characteristics, and establishes the intention domain dictionary.The second part is the manual expansion of dictionaries in different fields.In the third part, attention mechanism is added according to the key words in dictionaries in different fields.

3.1 Domain Dictionary Constructions

According to SMP2017-ECDT training set and development set, intention domain dictionary was established manually in this paper. A total of 31 domain dictionaries were constructed. Considering that the data is converted from artificial voice to text, in order to avoid the influence of recognition error on the experimental results, we modify some data before word segmentation. Dropping stop words when processing data is a common way to reduce memory and improve search efficiency. People can artificially remove functional words and lexical words in human language by importing stop word list. After removing stop words, the text can effectively improve the density of keywords and get faster speed in keyword matching. At the beginning of the experiment we in order to achieve better experimental result, then chose two kinds of different ways to cut the word document preprocessing: jieba participle support three kinds of word segmentation model, the search engine model is on the basis of accurate model for long words segmentation again, increase the rate of recall, apply to the search engine, word segmentation, so we choose this mode jieba participles; Considering that most of the short texts are ambiguous, we also choose LTP word segmentation method of Harbin Institute of Technology. LTP word segmentation framework based on machine learning can well solve ambiguity problems. By comparison, we found that the word segmentation tool of LTP of Harbin Institute of Technology was better than the commonly used word segmentation tool of jieba. Specifically, the word "fei" appears as a single word in the stammering segmentation of "kedaxunfei". In this way, when dictionary matching, the corpus containing "kedaxunfei" will be classified into the flight field with "fei" as the key word, resulting in classification errors. In order to construct dictionaries of different fields, we will count the word frequency of the cut words according to the field. Word frequency can measure the importance of a word to a domain file set in a corpus. Now there are a large number of word frequency statistical tools available on the Internet. We input the text of 31 fields into the statistical word frequency tool one by one, and select the entity words with word frequency more than 5% in each field as the dictionary words in this field. As shown in the figure, we have counted the word frequency in the field of telephone, and found that the occurrence frequency of such substantive words as "dadianhua" and "da" is quite high, so we chose these two words as dictionaries.

Table 1. Telephone Field Word Frequency Statistics

Num	Word	Count	Frequency %
1	Take a phone	12	13.33333
2	to	12	13.33333
3	call	4	4.44444
4	phone	3	3.33333
5	please	1	1.11111

3.2 Manual Expansion of Dictionary

Through the word frequency statistics in the previous chapter, we find that in some fields, there are no words with high word frequency, but many nouns belonging to a certain type appear. Through analysis, we found that the key words in Chinese text in these fields are these nouns, and the information robot relies on these nouns to perform search functions rather than perform certain operations. For example, in the field of health, we can find that the decisive words are various organs and common diseases by analyzing text features. Information robot interacts with people by searching such words. This requires us to build a dictionary in accordance with the characteristics of the dictionary association with characteristic words. Therefore, we decided to carry out manual expansion of dictionaries in certain fields, and collect the noun data in this field from the website of specific fields through the method of website crawler to form the dictionary of this field.

3.3 Hard Attention

In practical application, we find that the ambiguity of words has a great influence on the experimental results. For example, the word "and" is a conjunction in the sentence with a high frequency, but it is also an important dictionary word in the field of "calc". .For this reason, we added the Attention mechanism according to the gerund group features of the text, and the experiment proved that the Attention mechanism could significantly improve the accuracy of classification in the intention field. Attention is divided into hard Attention and soft Attention: soft Attention is to get the coding state through the deterministic score calculation, while hard Attention is a random process, which will sample a part of the hidden state of the input according to the probability to calculate. When the attention mechanism is not added, the input will ignore the importance of the input word with the increase or decrease of the input sequence. Therefore, this paper adopts the sampling method of Multi-bennon distribution to add hard attention to the input. First of all, we will input sentence structure:

$$c = \{c_1, c_2, \ldots, c_n\} \tag{1}$$

And then we add Attention for the input of each set of words (Attention), so that we get a new form of input:

$$V_c = \sum_{i=1}^{\|C\|} a_i * c_i * s_i \tag{2}$$

For each of the phrases c_i in the text, we will produce a can decide whether to produce the correct position of the words of weight a_i, generally speaking, the probability that phrases c_i are selected as input decoder were a_i . s_i stands for the position variable of the position information of the i word, which satisfies the Multi-bennon distribution of the a_i as the parameter:

$$p\left(s_i = 1 \mid s_{j<i}, a\right) = a_i \tag{3}$$

For example, the sentence “Modify the reminder of going to the pedestrian street tomorrow and Zhang San to go to the country “as shown in Figure 1. participle and add attention mechanism after we get new input form for ['tixing',guogou", 'buxingjie' ,'Zhangsan', 'mingtian', 'xiugai','he'], such a sentence "xiugaimingtiantixinghezhangsanquguogoubuxingjie" will be the right classification to the "schedule", instead of the wrong classification to the "calc".

4. Intention recognition based on domain dictionary and hard attention

We repeated iterative matching of the candidate strings according to the 30 domain dictionaries. When the candidate strings were successfully matched, we wrote the sentence where the candidate strings were located into the corresponding TXT. If the candidate strings were iteratively matched to the end and failed to match, we wrote this sentence into the "chat" domain.

Dictionary is an iterative matching process, and we need to frequently determine whether candidate words are in the dictionary, so the optimization of search algorithm is particularly important in the use of the dictionary.By comparing global traversal algorithm with partition search algorithm, we finally selected HashSet as the search algorithm in this paper.

At the same time, we have also selected the genertor in python that replaces memory resource consumption with computational resource consumption, and its feature of returning iterable objects for convenient cyclic use is also very consistent with the method used in this paper, namely, the calculation method of repeated iterative matching.

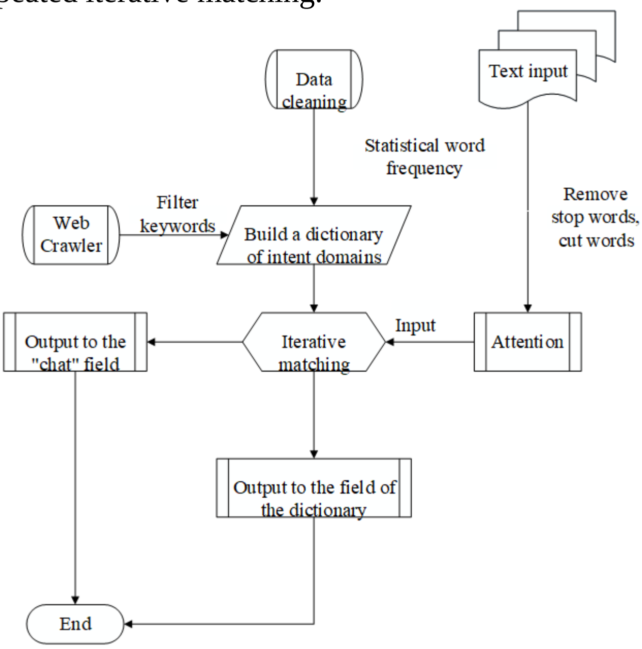

Figure 2. This article work flow chart

5. Experimental

5.1 Datasets

In this paper, the intention domain classification dictionary is established manually by studying the training set and the development set in the task 1 of the Chinese human-computer dialogue evaluation unit of the 7th national social media processing conference.The test set used in the human-computer conversation evaluation at the 6th national social media processing conference the previous year was used as the test data for this experiment.

5.2 Analysis of results

Contrast experimental naive bayes and support vector machine SVM are traditional machine learning algorithms. They seek common ground between sentences based on the principle of statistics [10].Naive bayes algorithm has a better effect in the areas of spam classification, but the email such as the length of the text length compared with the text of this article selected vary too much, and this article is a classification system of as many as 31, naive bayesian algorithm in the case of insufficient data have been trained classifier performance effect is not good.Support vector machine (SVM) algorithm has some advantages in small sample classification, but it is limited in the field of human-machine dialogue intention recognition by features such as large corpus information and less repeated language.We collated texts from 31 fields and stored them in another text as labels from 1 to 31. We compared the two experimental results with the correct answers of SMP2017-ECDT.The results are shown in the figure below: we found that the two experimental results are very poor, and because the number 4 data type, namely "chat" field data is more than other data, both classifiers classify most texts into the "chat" field, so the accuracy of both experiments is that "chat" type data accounts for 8.448% of all data.

In addition, CNN network, which is representative in sentence classification, was selected to classify the text of this paper [14]. The original text of this paper is a binary classification problem, which was expanded to 31 categories. The accuracy rate of the neural network established in this paper can reach more than 97% in the binary classification problem.After the expansion of the categories, due to the lack of text training data, the results obtained in this paper differ greatly from the binary classification accuracy, which is only 73.62%.

From this, we can infer that in the early stage of intention domain classification with less data, we can adopt the dictionary iterative matching algorithm with attention mechanism proposed in this paper to classify intention domain.

We selected the neural network to classify the text in the representative CNN network in the sentence classification [6]. The original text is a two-category problem. We extend it to 31 categories. The neural network established in the article is the correct rate in the second classification problem can reach over 97%. After the expansion of the article, due to the lack of text training data, the results obtained are quite different from the accuracy of the two classifications, only 73%.

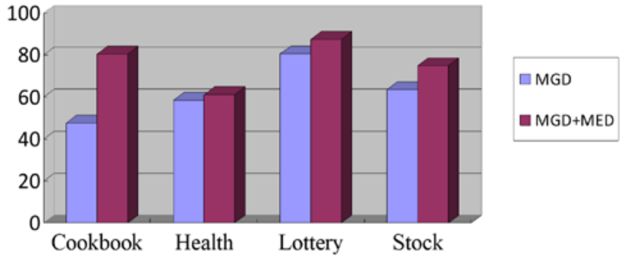

Figure 3. Comparison of the correct rate before and after the dictionary of artificial expansion in the field of CookBook and other fields. MGD and MGD-MED respectively represents the artificial expansion domain dictionary before and after

Table 2. Comparison of Results Representative Areas

Model	Average Correct Rate %	Correct Rate in Representative Areas %			
		App	Chat	Flight	Health
SVM	8.45	0.01	8.44	0.01	0.01
CNN	73.62	11.2	8.50	31.4	15.9
MGD	66.05	37.1	15.7	100	61.1
MGD-MED	65.37	37.1	30.51	100	77.6
MGD-MED+HA	77.07	56.5	46.1	100	87.3

From the chart we can see the following points

1) The dictionary iteratively matches can reach 100% correct rate in some areas (such as 'flights').

2) The method of artificially expanding the dictionary has significantly improved the classification of most fields.

3) Hard Attention input method plays a certain role in each field, especially in certain areas, the correct rate can be improved significantly.

The method proposed in this paper has a high accuracy rate in most areas of the classification direction of the data shortage intent, especially in the field of task-based vertical domain interaction, such as bus, cinemas, flight, message, train and weather. Close to a neural network model that relies on training of large numbers of data samples. The reason we can find out that because of the text or language interaction in these vertical fields, we can find more obvious keywords, such as "flight" and "airport" in flight, which are composed of keywords with higher frequency. The dictionary can achieve very high accuracy in the late iterative matching process.

Keywords in the field of interaction domain that rely on information search are noun tables in the domain to which the keyword belongs. Building a dictionary in such a field depends more on the expansion of keywords in the dictionary. For example, in the fields of cookbook, health, lottery, match, novel, etc., we have significantly improved the accuracy of the experimental results after crawling more relevant keywords through the web crawler technology, but such a method is difficult to apply to the actual system. Human maintenance is required.

In the open field of information interaction, the proposed method does not perform well. This is because the open type feature makes no keywords in the field match. In this method, the uncategorized sentences are simply put into "chat". "The field has not effectively improved the correct rate of the classification.

6. Conclusion

In this paper, the dictionary matching method commonly used for sentiment classification is applied to the intent domain classification task and artificially expands the dictionary according to the characteristics of each field, and the weight of the input words is changed by the hard attention to adapt to the Chinese grammar. Due to the sparsity of this type of task data and the characteristics of less tag data, our approach can be better in terms of speed and accuracy than other methods [15]. However, we still find that in some areas such as "movie" and "fiction" because the ambiguity accuracy of words can not be improved, of course, this is related to the user's habits. Therefore, the development prospects of user intent domain classification should be combined with user portraits in order to achieve higher accuracy and practicality.

Acknowledgments

National Natural Science Foundation of China: 61373148, Science Foundation of Ministry of Education of China: 14YJC860042, Shandong Provincial Social Science Planning Project: 18CXWJ01/18BJYJ04/17CHLJ33/17CHLJ30/17CHLJ18 /16CXWJ01

References

[1] Turing, Alan M. "Computing Machinery and Intelligence." *Creative Computing* 6.1 (1980): 44-53.

[2] Pickering C A, Burnham K J, Richardson M J. A review of automotive human machine interface technologies and techniques to reduce driver distraction [C] //2nd Institution of Engineering and Technology international conference on system safety. 2007: 223-228.

[3] Zhang W N, Chen Z, Che W, et al. The First Evaluation of Chinese Human-Computer Dialogue Technology [J]. 2017.

[4] Holtgraves T. Automatic intention recognition in conversation processing [J]. Journal of Memory and Language, 2008, 58 (3): 627-645.

[5] Dos Santos C, Gatti M. Deep convolutional neural networks for sentiment analysis of short texts [C] //Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 2014: 69-78.

[6] Wang X, Liu Y, Chengjie S U N, et al. Predicting polarities of tweets by composing word embeddings with long short-term memory [C] //Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015, 1: 1343-1353.

[7] Shan S, Zhao F, Wei Y, et al. Disaster management 2.0: A real-time disaster damage assessment model based on mobile social media data—A case study of Weibo (Chinese Twitter) [J]. Safety science, 2019, 115: 393-413.

[8] Báo cáo khoa học: Baselines and Bigrams: Simple, Good Sentiment and Topic Classification [J].

[9] Dong W, Zhao J, Hu R, et al. Differentiation of Chinese robusta coffees according to species, using a combined electronic nose and tongue, with the aid of chemometrics [J]. Food chemistry, 2017, 229: 743-751.

[10] Edlund J, Gustafson J, Heldner M, et al. Towards human-like spoken dialogue systems [J]. Speech communication, 2008, 50(8-9): 630-645.

[11] Lê Cao K A, Rossouw D, Robert-Granié C, et al. A sparse PLS for variable selection when integrating omics data [J]. Statistical applications in genetics and molecular biology, 2008, 7 (1).

[12] Zhang L, Lin L, Liang X, et al. Is faster r-cnn doing well for pedestrian detection? [C] //European conference on computer vision. Springer, Cham, 2016: 443-457. Li Y, Tian Y, Duan L Y, et al. Sequence multi-labeling: A unified video annotation scheme with spatial and temporal context [J]. IEEE Transactions on Multimedia, 2010, 12 (8): 814-828

[13] Kalchbrenner, Nal, Edward Grefenstette, and Phil Blunsom. "A convolutional neural network for modelling sentences." arXiv preprint arXiv: 1404.2188 (2014).

[14] Parra-López E, Bulchand-Gidumal J, Gutiérrez-Taño D, et al. Intentions to use social media in organizing and taking vacation trips [J]. Computers in Human Behavior, 2011, 27 (2): 640-654.