

## A Semantic Recognition Method of Radiotelephony Communication for Apron Controller Based on Deep Learning

Yujie Qiao<sup>1+</sup>, Lin Zuo<sup>2</sup>, Shiyu Huang<sup>1</sup> and Zhiyuan Shen<sup>1</sup>

<sup>1</sup> College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China.

<sup>2</sup> Air Traffic Management Bureau East China Regional, Civil Aviation Administration of China, Shanghai, China

**Abstract.** With the rapid development of China's civil aviation industry, the security and efficiency of airport operation is necessary to improve. Apron control transfer has become a trend at the China airport with 10 million throughputs. However, there is lack of specific research on semantic recognition of apron controller's ground air calls. This paper firstly establishes the framework of apron control semantic recognition. Then it analyzes the structure and content characteristics of apron control ground air call. The feature template of key instructions is extracted based on data from Urumqi airport. Finally a deep learning method is used to achieve semantic recognition of apron controller under special circumstances. The simulation results verify the effectiveness of the model.

**Keywords:** Apron control; Radiotelephony communication; Deep Learning; Semantic Recognition

### 1. Introduction

In the 1950s, IBM launched the natural language processing based on the rule-based leading stage and translated the natural language by computer. In this stage, many algorithms that can promote the development of natural language processing are put forward successively, but the main disadvantage is that the algorithm cannot accurately identify the semantic information contained in the language. In the 1970s, based on Since the 21st century, due to the enhancement of computing power, researchers have begun to analyse corpus and put forward many new corpora. Then deep network [1] and deep learning [2] have been gradually applied to semantic recognition. The field brings revolutionary influence, which greatly improves the recognition rate and reliability of semantic recognition [3].

Due to the late start and slow development of computer research in China, the research on natural language processing, including semantic recognition, is lagging behind compared with foreign countries. However, inspired by foreign countries, domestic researchers have made great breakthroughs in natural language processing, such as Chinese word segmentation, machine translation, document retrieval, semantic comparison and so on. For example, Chen Zhigang of Tianjin University has completed the research of text classification system by using vector space model method [4]; Liu Zhenlu of Northeast University has analysed the LDA method, clustering the Chinese text classification corpus of Chinese Academy of Sciences by using LDA and the accuracy has reached more than 90% [5], Zhao Peng has proposed the concept of sketch semantic web by using the whole image as the network input Line semantic recognition [6]; Zhang Qiqi carries out experimental analysis on the semantic analysis model based on the improved convolutional neural network, and designs a robot voice interaction system based on arm [7].

At the same time, Baidu and iFLYTEK are emerging in the field of artificial intelligence and speech recognition. As an international authoritative standard in the field of semantic recognition, the error rate of

---

+ Corresponding author. Tel.: + 16651692966.  
E-mail address: Qiaoyujie12138@126.com.

speech recognition (WER) can reflect the reliability of recognition engine. The WER of human speech is about 4%, and the commercial recognition software is within 15%, which is the threshold of reliability. As a voice service technology provider for more than 60% of the world's top 100 enterprises, nuance voice company's WER is up to 10%; relying on in-depth learning technology, Google takes the lead in achieving 8% recognition error rate; iFLYTEK, the leader of domestic voice recognition, also reached the reliability threshold of 15% in 2014, especially in conference speech, to within 5%, In the field of Chinese dialects, it is the first to enter the practical threshold.

## 2. Framework of Semantic Recognition

Semantic recognition system is essentially a pattern recognition system [8], which consists of three basic modules: feature extraction, pattern matching and reference pattern library. Its basic structure is shown in Figure 1:

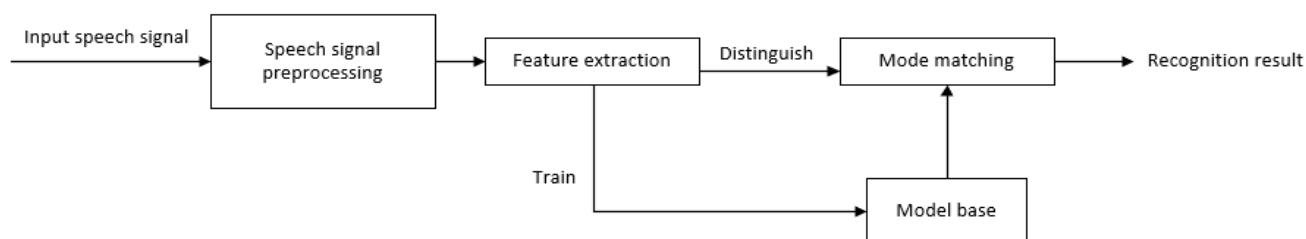


Figure 1: Principle structure of semantic recognition.

It can be seen from the above figure that the most important component of semantic recognition is feature extraction and pattern matching. [9] In the field of semantic recognition, deep learning is mainly used in the training of feature extraction in the principle structure of Figure 1. The advantage of deep learning is that it can use non (semi) supervised feature learning and hierarchical features to extract more efficient algorithms and automatically extract more complex and effective features from massive data, so as to replace the traditional manual feature acquisition.

It can be seen from Figure 1 that the optimal result is directly related to the selected features, the quality of the voice model itself, the accuracy and fit of the template.

## 3. Method

### 3.1. Construction of Speech Database from Urumqi Airport

The corpus used in this paper is based on the example of Urumqi Diwopu Airport, 500 relatively standard and standard voice instructions extracted from the apron control simulation practice training and proofread by the first line controller. The recorded text after proofreading constitutes the text source of this corpus.

The plane structure of Diwopu airport in Urumqi is relatively simple, so it is relatively easy to collect the land air communication instructions from the airport floor controllers. The airport has one runway (runway 07 / 25), two parallel taxiways (a, t), seven fast breakaway taxiways (A1, A2, A3, A4, F, K, I), 7 connecting taxiways (A5, A6, a7, A8, A9, A10, a11) and 5 apron taxiways (T2, T3, T4, Ka, KB), with a total of 96 parking spaces.

After proofreading 500 common apron controller voice instruction texts constitute the standard database of this paper's semantic recognition test. Among them, there are 55 release permit type instructions, 60 start-up type instructions, 275 taxi path type instructions, 98 instructions for adjusting aircraft status (such as changing speed, waiting in place, etc.), 62 control handover type instructions, and 50 other (such as communication inspection, location verification, etc.).

The above speech texts constitute the standard corpus of this paper. The corpus of the above 500 instructions has 317 non repetitive words after word segmentation, so the size of the word list of the apron control land air communication is 317.

### 3.2. Feature Extraction

Through the general analysis of the particularity and natural language characteristics of air land call instructions, based on the rules of air land call, dependency grammar and valence grammar, we designed a structured template for control instructions, which helps the computer system to better realize the semantic recognition function of controller instructions.

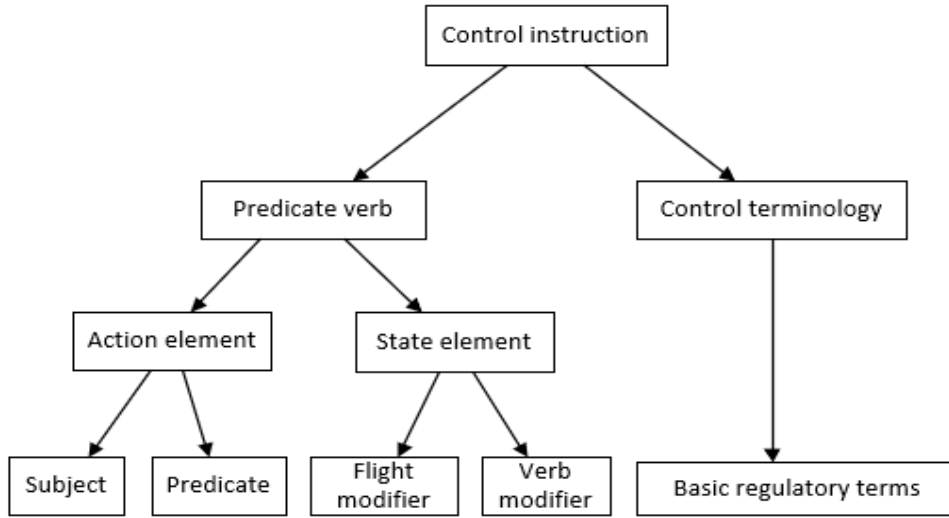


Figure 2: Structure template of control instruction

As shown in Figure 2, the structural template of control instruction includes the following six parts: predicate verb, subject, object, flight modifier, verb modifier and basic regulatory terms. Among them, flight modifiers refer to the words used to describe the characteristics of flight status, and action modifiers refer to the adverb phrases used to describe actions.

In the feature parameter extraction process, we usually extract the Mel frequency spectrum coefficient (MFCC) of the input speech signal as the feature sequence. The conversion relationship between the frequency parameter of MFCC and the linear frequency is as follows:

$$f_{mel} = 2595 \log_{10} \left( 1 + \frac{f}{700} \right) \quad (1)$$

The principle of the whole speech decoding process is formally summarized as follows:

$$W^* = \arg \max_w P(W|X) \quad (2)$$

The formula can be interpreted as:  $X$  is the given input sound sequence,  $W$  is the optimal output word sequence after analysis and extraction, which can be converted into:

$$W^* = \arg \max_w \frac{P(X|W)P(W)}{P(X)} \quad (3)$$

Among them,  $P(W)$  is the language model, which describes the prior probability of a word sequence  $W$ ;  $P(W|X)$  is the acoustic model, which describes the conditional probability of a given word sequence  $W$  generating a sound sequence  $X$ ;  $P(X)$  is the probability of acquiring a sound sequence  $X$ , which can be ignored.

## 4. Semantic Recognition and Matching of Control Special Situation Instructions

### 4.1. Collection of Test Set

Common special conditions in apron operation include scratch between aircraft and aircraft or support vehicle, oil leakage of aircraft or vehicle, road surface pollution caused by foreign object invasion (FOD), aircraft using wrong taxiway, etc.

This paper takes Urumqi airport as an example to simulate five typical special situation scenarios, and collects 10 voice instructions collected in the special situation simulation process as the test set. We collect the

special voice instructions, input them to MATLAB as the test set voice for recognition, and extract the parameters to lay the foundation for comparison with the standard corpus in the later stage.

## 4.2. Key Words of Apron Control

Before comparing the special situation instructions in the test database with the instructions in the standard corpus, we need to classify the key words contained in the instructions, realize the correspondence of each category between the recognition results in the test database and the keywords in the standard database, carry out targeted matching detection, and evaluate the recognition results more accurately.

Through the classification of 317 thesaurus from the standard corpus, 280 effective keywords are extracted and divided into the following five categories, as shown in Table 1.

Table 1: Word classification of land air communication instructions for apron control

Type	Example
Speed related	A. contains figures. 20 kilometers per hour
	B. Without numbers accelerate, decelerate, fast
Location related	C. static stand 13, runway holding point, runway 07, in place, ahead
	D. dynamics A-A9-T、 A10-T
Action related	E. State action taxi, wait, avoid, launch, contact
	F. Embellishment action stop, pay attention, speed up
Object related	G. call sign MU4468、 CZ9952
	H. other tower, guided vehicle, aircraft
Environment related	I. Terminology ground wind, dew point, corrected sea pressure

## 4.3. Results and Analysis

By comparing with the information in the actual voice instruction, check the correctness and completeness of each category of words, and get the evaluation format as (the number of correctly recognized words in this category / the number of words actually contained in the instruction). The results are shown in Table 2.

Table 2: Accuracy of key instruction words in recognition results

Type Number	A	B	C	D	E	F	G	H	I
1	0/0	0/0	1/1	0/0	1/1	1/2	1/1	1/1	0/0
2	0/0	0/0	2/2	0/0	1/1	1/2	1/1	2/2	0/0
3	0/0	0/0	2/3	1/1	0/0	1/1	1/1	0/0	0/0
4	0/0	0/0	1/1	0/0	1/2	1/1	1/1	0/1	0/0
5	1/1	1/1	2/2	0/0	0/0	0/0	1/1	0/1	0/0
6	0/0	1/1	1/2	0/0	2/2	0/0	1/1	1/1	0/0
7	1/1	1/1	3/3	0/0	1/1	0/0	1/1	1/2	0/0
8	0/0	0/0	3/3	0/1	1/1	0/0	0/1	0/0	0/0
9	0/0	0/0	2/2	0/0	2/2	2/2	1/1	0/1	0/0
10	1/1	1/1	1/1	0/0	0/0	1/1	1/1	1/1	0/0

In the table, we make statistics on the membership relationship of each word identified from the test set instructions, and get the recognition result  $f_m(x_n)$  of the nth type words in the m-th instruction. The average value is used to represent the membership result of the whole instruction word recognition. The recognition rate of the special situation instruction 1 is 90%.

Calculate the mean value of all test sets,  $f_1(x), f_2(x) \dots f_{10}(x)$ , so as to obtain the membership degree of key words between the air and land call instructions of the whole test set and the standard set. The results are as follows:

$$f(x) = \frac{\sum_{m=10}^{m=1} f_m(x)}{m} = 0.8334 \quad (4)$$

Therefore, the recognition rate of all special situation instructions in the test set is 83.34%, which is the reliability result of semantic recognition in this experiment based on MATLAB.

## 5. Conclusion

In recent years, with the rapid development of China's civil aviation industry and the arrival of the era of artificial intelligence, more and more new technologies will be used in the operation of civil aviation industry. Based on the background of apron control transfer, this paper analyzes the method and feasibility of semantic recognition for the ground air communication of apron controller.

## 6. Acknowledgements

The authors would like to thank the financial support by the Air Traffic Management Bureau of Civil Aviation Administration of China for the Research on Monitoring and Management of Fatigue State in Air Traffic Controller Program.

## 7. References

- [1] Bin X, Ningning T. DIOD: Fast and Efficient Weakly Semi-Supervised Deep Complex ISAR Object Detection[J]. IEEE Transactions on Cybernetics, 2018:1-13.
- [2] Bin X, Ningning T. Real-world ISAR Object Recognition using Deep Multimodal Relation Learning[J]. IEEE Transactions on Cybernetics, 2019:1-12.
- [3] Liu Cheng. Study on handover scope and responsibility division of apron control [J]. Science and technology innovation guide, 2017, (31) 14.
- [4] Tang Yinggang. Research on new equipment for apron control [J]. Machinery and technology, 2018, (7) 299.
- [5] Zhao Peng, Feng Chencheng, Han Li, Ji Xia. Sketch recognition method integrating deep learning and semantic tree [J]. Pattern recognition and artificial intelligence, 2019,32 (04): 361-368.
- [6] Zhang Qiqi. The realization of robot speech interaction and semantic recognition [D]. Jinan University, 2017.
- [7] Chen Xie, Liu Xunying, Wang Yongqiang. Efficient Training and Evaluation of Recurrent Neural Network Language Models for Automatic Speech Recognition[J]. IEEE Computational Intelligence Magazine, 2016, 24(11): 1-1.
- [8] Bengio Y. Learning Deep Architectures for AI[J]. Foundations & Trends® in Machine Learning, 2009, 2(1):1-127.
- [9] Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(5786):504-507.
- [10] Sutskever, I. Vinyals, O. & Le. Q. V. Sequence to sequence learning with neural networks[J]. In Proc. Advances in Neural Information Processing Systems,2014,27:3104-3112.