

# Research on Framework of Speech Recognition Combining Text-Speech with Semantic Similarity

Wei Kang, Fangfang Shi, and Xianyi Cheng

**Abstract**—With the deepening of the speech recognition research, improving the accuracy of the general recognition engine is becoming more and more difficult. For the noise problem of Chinese speech recognition, in this paper we made a brief review and analysis about the current related articles of the semantic similarity and the text-speech similarity. We compared the advantages and disadvantages of various methods in speech recognition and pointed out the difficulties and the common problems existed in the research. At last, we give a speech recognition framework combining the semantic similarity and text-speech similarity. It provides a new way to improve the accuracy of speech recognition.

**Keywords**—Similarity, speech recognition, natural language processing, semanteme.

## I. INTRODUCTION

WITH the progress of the society and the rapid development of science and technology, The help of computer to people is more and more big, become indispensable good helper for people, but has always been the people communicate with computer through the keyboard, mouse and so on, this limits the communication between people and computer, more limits the consumers. For most people can use the computer even disabled people, let the computer understand human language, understand people's intentions, people began the study of speech recognition. Speech recognition main application scenarios is that speech input is one of a finite set of elements, let the system to identify the input correspondence to which element. At present, commonly used speech recognition method is: first to build a voice template, save the voice signal characteristics of each word or sentence; Then for each speech input, calculate degree of similarity the input speech signal characteristics of each voice signal characteristics with the template, put words or sentences which its biggest degree of similarity as user's expectations. Here used degree of similarity is the degree of similarity among the original speech signal characteristics. Literature<sup>[1]</sup> use a kind of new thought: first of all, using a general recognition engine give a primitive recognition result to input voice , the results may be difference with each

element of candidate set, using degree of similarity between text, calculated the degree of similarity initial recognition results with candidate set each element, put words or sentences which its biggest degree of similarity as user's expectations.

With the deepening of the speech recognition research, improve the accuracy of general recognition engine is becoming more and more difficult. Especially the particularity of Chinese language, only the literal string is used to calculate the text-speech similarity, improve the accuracy of speech recognition is very limited, in this paper, first briefly reviewed related speech similarity research and analyzed, compared the advantages and disadvantages of various methods, and points out that the difficulties of literal string similarity and existing problems in the research and discuss, and on this basis, we give method for the text-speech semantic similarity calculation based on natural language processing technology.

## II. RESEARCH ON SIMILARITY OF TEXT-SPEECH

### A. The method of similarity calculation no using rule

#### (1) The algorithm of edit distance

The basic idea of Edit distance [2] is that calculated need the least insert or delete or replace from the original string to the target string. The algorithm can be applied to speech recognition, spelling error correction, DNA analysis and process the retrieval [3]. Due to itself strict order, the identify the effect is bed in change some of the copied paper.

#### (2) The algorithm of the longest common subsequence

The algorithm of the longest common subsequence is for short LCS, it will a string delete several characters respectively, in the case of not changing the order of the remaining characters, the same sequence of characters which length is the longest.

The algorithm is more effective in ignore irrelevant elements, the most area which is applied is calculating the similarity between the DNA molecule<sup>[4]</sup>, but it can also be used for speech recognition, data compression, term extraction<sup>[5]</sup> <sup>[6]</sup> text retrieval, etc., can also be applied to the detection code similarity measure, because that the longest common subsequence is strict order by algorithm to get , it is discount greatly for those who changed the order of the statements in a

Wei Kang and Fangfang Shi is with the School of Electronic Engineering, Nantong University, Nantong 226019, China; e-mail: kangwei19871105@126.com).

Xianyi Cheng was with the School of Computer Science and technology, Nantong University, Nantong 226019, China (e-mail: xycheng@ntu.edu.cn).

code block, the expression of the operand or applications such as the order of the operator is plagiarism detection efficiency.

### (3) *The GST algorithm*

The GST algorithm is mainly find the largest sub-string by greed algorithm search two strings, the basic idea is to compare the text string with pattern strings of each element, if found different elements, text string restart comparison from the beginning, the longest match is concluded. Then the two strings which the longest match marked respectively, forming a cover, covered characters cannot be used to match again. All cover that greater than the minimum matching length form a maximum coverage eventually. The GST algorithm can effectively detect the change order article or program code, etc., at the same time also can deal with the problem of repeated sentences detection. Algorithm efficiency is low, its time complexity is bigger, the best time complexity is  $O(n^2)$ , the worst time complexity is  $O(n^3)$ <sup>[7]</sup>.

### (4) *The RKR-GST algorithm*

RKR - GST algorithm[8] is mainly aiming at the problem that GST efficiency is low, it introduced KP algorithm that the efficiency is high into the GST algorithm, so as to improve the GST algorithm.

The basic idea of KR algorithm is to calculate hash value of pattern that string's length is  $m$  according to a hash function, the text string that the length is  $t$  divided by length  $m$ , get sub-string which length is  $t-m+1$ , calculated hash value of the sub-string which the length is  $m$  using the same hash function, to get  $t-m+1$  hash value. The pattern with the same hash value can match the text sub-string, and pattern are not match text string with the different hash value.

The RKR-GST algorithm is difference with GST, it does not need to compare each element of a text string with pattern string, but when the pattern sub-string and text sub-string have the same hash value, need to compare them, so its efficiency is higher than GST algorithm. The RKR-GST algorithm and GST algorithm has the same idea, but in the best case for RKR-GST algorithm, the time complexity is  $O(n)$ , the worst case time complexity, RKR-GST and GST is as same, so RKR-GST efficiency is improved greatly, relative to GST algorithm.

### *B. Similarity algorithm based on phonetic rule*

The similarity algorithm based on phonetic rules mainly analyze differences of pinyin string from the perspective of syllables in Chinese characters. Fundamental ideas of the algorithm is that divide syllable into three parts, namely initials, finals and tones, because of value is limited to the initials, finals and tones, can be enumerated define all edit distance from one value to another value. Divide any two syllables into three parts above, according to the definition of the edit distance, find out each edit distance corresponding to syllable, the sum of three parts edit distance define as edit distance between two syllables. According to the edit distance between the two syllables can calculate the similarity. For example: if the definition similar pronunciation with the distance between the initials and finals differences as 1,

including retroflex, before and after nasal, etc.;The same initials and the different finals or the same finals and the different initials, and do not belong to the above situation, the distance is defined 2; The different initials and finals, the distance is defined 4; When tone is differences, the distance is defined 1. So for "lei1" and "nei4" can easily obtain the edit distance is 2, then can calculate the similarity of the two characters.

This algorithm has good correction effect for everyday speech recognition system or caused the sound of words confusing error by Chinese spell input tool homophone or nearly. But it has good processing effects only in the case of homophone or nearly sound word, and for other types of samples adaptability also cannot achieve the ideal effect.

## III. RESEARCH ON SEMANTIC SIMILARITY

### *A. Based on statistical*

Using statistical techniques to calculate semantic similarity between words is an unsupervised machine learning method, mainly includes:

#### *(1) The algorithm based on word co-occurrence analysis*

Concurrence refers to the vocabulary common in a document collection. With a word as the theme, you can find some frequently appeared together in collocation with the words, these words as its co-occurrence words together, the vocabulary describe the lexical semantic context or context. Word co-occurrence algorithm is only based on co-occurrence relationship to calculate the similarity between itself, namely the semantic similarity of two words with the two co-occurrence frequency increases with the increase in vocabulary. The premise of algorithm implementation is: assuming that often appear together words are synonyms or related words.

#### *(2) The algorithm based on the document vector space model*

This algorithm is mainly based on the document space model view all text in the corpus and word in the dictionary, respectively, as a text space and the space of a word. Two word semantic similarity define as the corresponding feature vector similarity between two words. If view a feature vector of vocabulary as a point in characteristic vector, then the degree of similarity between two words feature vector can be measured by the distance between two points in the feature space. In this algorithm, similarity calculation is mainly by vector space between the words, conversion for calculating the distance between two points in the vector space indirectly. The application of this method is more, such as latent semantic indexing method<sup>[9]</sup>.

#### *(3) The algorithm based on syntax analysis*

This algorithm using the syntax analyzer to analyze the dependence relation between words, on the basis of the syntax analysis, for example, Gregory realized similarity between noun by using weighted Jaccard coefficient calculation method, Dekang extracted the existence dependency triples,

and by using maximum likelihood estimation to calculate the information, then draws the semantic similarity between two words.

#### (4) Improvement method based on large-scale corpus

The traditional method based on corpus is prepared by using corpus to calculate the degree of similarity between words. It is not easy to update words and found no register words, so this kind of algorithm exist the problem of data sparseness, which may lead to estimate the probability by maximum likelihood estimation is unreliable, to correct the problem, the improved method based on the scale corpus was put forward. It mainly is to use a variety of different smoothing algorithm that descent the number of appear, so that does not appear can also take the probability of events rest, to make up for the deficiency of the traditional method based on corpus.

### B. The method based on knowledge

#### (1) The method based on ontology

The basic idea is to use ontology to calculate semantic distance between words, ontology is a kind of concept model, the basic unit of the semantic is concept, each concept was represented by the relationship between the concept and properties. Generally, Calculating semantic distance of word based on ontology refers to put the concepts to a tree or tree hierarchy, in the tree, between any two nodes must be only one path, so the length of the path can be a measure of the semantic distance of two concepts. Abroad, scholars mainly use the WordNet as ontology to research the calculation of semantic similarity between words. In China, scholars mainly

use Hownet[10] or "synonym word Lin"<sup>[11]</sup> to study the calculation of semantic similarity between words.

#### (2) the method based on the dictionary

The method based on dictionary refers to the whole dictionary mapping into a directed graph, each word is equivalent to a vertex in the dictionary, if the definition of a word or comment contains another word, then we draw a directed edge from the word to another word. For any known word, then, we all can construct a adjacent graph (sub-chart of dictionary chart), the adjacent figure includes all point to the word of known word and the word that known word point to word. With the information that the adjacent graph make an adjacent matrix, then we can use different calculation method to achieve the degree of similarity between two words.

## IV. THE FRAMEWORK OF SPEECH RECOGNITION

### A. The framework of speech recognition which mix text-speech and semantic similarity

It is inevitable to caused the interference by homophones or close the speech. Due to the similarity algorithm based on phonetic rules is to analyze the difference of phonetic in Chinese characters from the perspective of syllables, although similarity calculation methods without rules was proposed earlier, its development also is relatively mature, but this method is calculated mainly similarity in English string, and it ignores the particularity of Chinese characters, so choose the phonetic rules combined with semantic similarity method to eliminate the noise of the Chinese speech recognition has a positive role. The framework of speech recognition which text-speech and semantic similarity is shown in fig. 1.

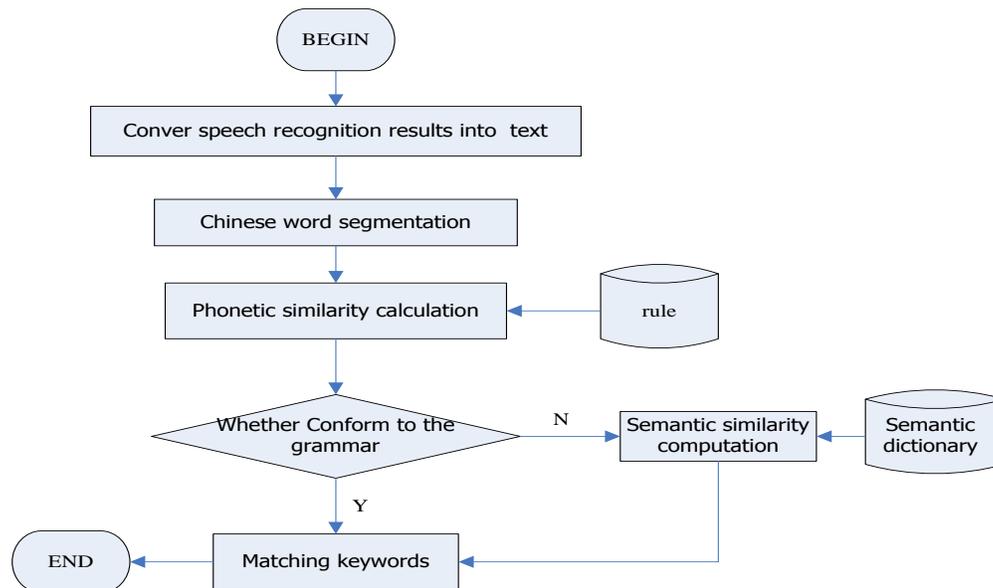


Fig. 1 The framework of speech recognition which text-speech and semantic similarity

### B. The principle of the framework

Because the algorithm of similarity based on phonetic rules error correction is simply rely on the rules judgment, and it is not necessarily conform to the rules of Chinese grammar for

the Chinese recognition results after initial error correction. In order to solve this problem, in this paper, we judgment the preliminary recognition result, and further calculate semantic similarity for recognition results to does not conform to the grammatical rules , and then output the final result. For the

method of semantic similarity, we choose the method based on semantic dictionary to calculate semantic similarity. Its performance is mainly depends on good and bad in the training corpus, it has big influence on sparse data and data noise interference, and build a domain-specific corpus, we have to cost a lot of manpower, but also to constantly update maintenance. Similarity calculation method based on ontology information can be obtained directly from resources, can accurately reflect the complicated relations between words, its calculation is simple, however, the concept of words describe detailed is not unified, so lead to semantic information possible error reflects the semantic difference between words. Based on dictionary method is simple and easy to implement, its development is more mature, although greatly influenced by human subjectivity, it also can reflect the objective reality performance for a particular field.

### C. The analysis of experiment results

As an application example of the text-speech similarity, we designed and developed a voice information query system for Nantong city bus. Through the system, the user through the voice input can query the "begin" have "end", system can give a direct train, also includes transfer bus (limited) at a time. When user input "begin" and "end" at the same time, it will show all trains through begin station.

Query examples: if the user's speech input begin "easy home bridge", recognition engine recognition results for "preferred bridge", when system with Nantong bus stops in

the collection by calculation text similarity calculation get listed top 10 stops, such as: easy, easy home home bridge destination, small stone bridge, friendship bridge, bridge, etc., more convenient for users to choose. If in combination with semantic similarity computation, give an answer only "easy home bridge". Similarly the user voice input end to stand "san-jiu welding", recognition engine recognition results is "san-jiu han street", combined with semantic similarity computation, the user select "san-jiu welding" station as the end. System shows that:

(1) In the sit easy home bridge [8 line] after arrive at [7 station] get off the bus [as a port outside west road ] the walk toward the south 150 meters.

(2) In sit easy home bridge [13 line] after arrive at [8 station] get off the bus [san-jiu welding], and so on. Table 1 shows the experimental contrast to various methods.

Arrival rate refers to the 100 different routes experiment, the system can give an answer the number of times; Optimization rate refers to the answer is the optimal route in the number of times; Stability refers to the 100 words in the same routes experiment, we give the same answer correctly.

From table I we can known that RKR-GST is not ideal, is affected by recognition engine; RKR GST+semantic method can meet the application requirements, but optimization rate is Not as good as Rule + dictionary method.

TABLE I  
CONTRAST TO VARIOUS METHODS

	RKR-GST	Rule+dictionary	RKR-GST+semantic
Arrival rate%	21.4	46.7	90.4
Optimization rate%	45.2	76.2	74.2
Stability%	34.8	78.7	85.3

## V. CONCLUSION

Use phonetic rules to solve the problem of noise to be caused by the homophones or near speech in Chinese, it is important to error correction in speech recognition, after the error correction which does not comply with the rules of grammar, to compute semantic similarity by dictionary, the final results mach the keyword. phonetic similarity calculation simplify the calculation of the latter complexity, and reduce the probability of the latter error, makes up for disadvantage that the former can't to be more precise semantic analyzing and keyword matching, combination rules of phonetic similarity with semantic similarity analysis, it provides a new way to improve the overall keyword matching.

### ACKNOWLEDGMENT

This work is supported by Natural Science Foundation of China National(No: 61340037, 61202006).

### REFERENCES

[1] Toktar E, Jamhour E, Maziero E. RSVP policy control using XACML[C]. In: Policies for Distributed Systems and Networks, POLICY 2004, Proceedings, Fifth IEEE International Workshop, 2004: 87-96.

[2] Yu Haiying, comparison algorithm of LCS and GST in String similarity measure[J]. Electronic science and technology, 2011, 24(2): 102-103.

[3] Cao Bin, Yi Jianwei and Chen Huixin. The process approach based on Levenshtein distance[J]. Computer integrated manufacturing system. 2012, 18(8): 1766-1773.

[4] Xu Zhihai; Yu Yang. LCS algorithm based on similarity of gene sequences. In: Proceedings - PACCS 2011: 2011 3rd Pacific-Asia Conference on Circuits, Communications and System. 2011.

[5] Pan Hong and Xu Chaohui. The research of application on LCS algorithm in term extraction [J]. Journal of intelligence, 2010, 29(5): 853-857.

[6] Tasi, Cheng-Shiun; Zang, Yong-Ming; Liu, Chien-Hung; Huang, Yueh-Min. Applying VSM and LCS to develop an integrated text retrieval mechanism. In: Expert Systems with Applications. 2012, 39(4): 3974-3982.

[7] Jiang; Zhemin. The study of plagiarism detection for program code. Jiang Hao. In: Advances in Computer Science and Education Applications - International Conference, CSE 2011, Pro. 2011: 128-133.

[8] Xiao Li and Xiao Jingzhong. the algorithm analysis Based on RKR - GST[J]. Journal of southwest university for nationalities (natural science edition). 2010, 36(5): 836-840.

[9] Li Wenchao and Yang Nini. The study of semantic similarity based on the ontology[J]. Science, technology and engineering. 2012, 12 (21): 5328-5330.

[10] Xue Huifang. Sentence similarity computing theory and application research[J]. Northwestern university professional degree master's thesis. 2011.

[11] Tian Jiule and Zhao Wei. The method of word similarity calculation based on synonym word Lin[J]. Journal of jilin university. 2010, 28(6): 602-608.

**Wei Kang**, male, was born in November 5, 1987. Education: Master Of Degree in Nantong University from sept, 2012 to now. Bachelor of degree in the University of electronic science and technology of zhongshan institute from sept, 2006 to june, 2010. Research interests: natural language processing and pattern recognition.

**Xianyi Cheng**, Men, was born in December 1956. Dr., Professor,. Long engaged in artificial intelligence, pattern recognition, image processing, teaching and research work in the field of natural language understanding. Participate in 4 items of national natural science fund project Host 1 key task projects focus on science and technology, province, provincial project 6 items. Successively in the International Journal of Advanced Robotic Systems ", "Journal of software, computer research and development," Journal of Chinese information, "China engineering science", "Journal of intelligent system such as magazine published over 100 papers, which are three SCI&EI collected more than 70 articles, published treatise 5.