Dealing With Polyphone in Text-to-Speech System Using How-Net

LV Xin, LIU Zhan-yi, ZHAO Tie-jun, WANG Li

Machine Translation Laboratory, Harbin Institute of Technology
{lxfancy, zhanyi, tjzhao, wangli}@mtlab.hit.edu.cn
Box 321, Harbin Institute of Technology, No.92 West DaZhi Str., Nangang Dist., Harbin, Heilongjiang, P.R.China, 150001

Abstract

Dealing with polyphones is an important part of Chinese text-to-speech system. Because the pronunciation of a Chinese character is directly related to the meaning of it, an algorithm based on semantic calculation using How-Net is introduced to determine the pronunciations of the polyphones in new words, which hasn't appeared in the polyphone list, the polyphone knowledge base or the corpus. The experiment results prove it can do good performance.

1. Introduction

Building a Chinese text-to-speech (TTS) system involves three major steps. In the first, text is converted to syllables, the symbols representing in a rough way the categories of Chinese Mandarin speech sounds. A second stage involves questions of prosody, i.e., the intonation and pausing; and the third stage is the back end, the component responsible for the production of the sounds from the specifications provided by the first two components.

It is known that some of Chinese characters have several pronunciations. This kind of Chinese characters is called polyphone. Each pronunciation represents different meaning. That is to say that if the sentence is transferred into the wrong syllables, the speech synthesized by the system will cause the listener misunderstand the meaning of the sentence. And the number of polyphones in the common characters is very large. For example, in the first ten common characters there are six polyphones "的,一,了, 不,和,大"[1]. Therefore dealing with polyphones properly in TTS not only can guarantee the correctness of the output speech but also will make the speech more understandable and natural.

Since the number of common polyphones is only a small part of all Chinese characters, listing all polyphones and theirs pronunciations in different contexts one by one has been used in TTS. But this method not only takes much time and labor but also is not easy to update in time. If a new word comes out, the system would confuse its pronunciation for no according list items. A new word is an input word, which hasn't appeared in the polyphone list, the polyphone knowledge base or the corpus. On the other hand a method based on rules has been used before. It needs the builders of the rule database to have linguistic experiences and almost all methods based on rules have the difficulties such as building rule modules, avoiding conflicting between rules and the covering rate of rules. Later Corpus-based techniques are widely used in speech processing and they often have good performance while ignoring the true complexities of language, based on the fact that complex linguistic phenomena can often be indirectly observed through simple superficies. So some researchers began to use the statistical method in dealing with polyphones, especially in transferring syllables to texts. In TTS system, this method is often used for dealing with polyphones in new words. However, this method heavily depends on the corpus. Sun Y.-Q. combined the last two methods to deal with polyphones in TTS. Though good performance has been achieved, it still can't keep away from the limitation of linguistic experiences and corpus. In addition, there are few documents written about converting text to syllables and papers about dealing with polyphones in new words have not been seen.

Some large-scale and computable semantic knowledge bases such as WordNet, MindNet and FrameNet were built abroad and have been applied in research on natural language processing and used in some real systems [2]. Recently in China more and more attention is paid on semantics in Chinese natural language processing and some semantic resources have been built such as How-Net. Since the pronunciations of polyphones are directly related to their meanings, using semantic knowledge to deal with polyphones, especially new words of polyphones, must be an effective method.

The paper unfolds as follows. Section 2 gives an overview of How-Net. Section 3 gives the definition of semantic similarity degree. Section 4 describes the algorithm to dealing with the polyphones when they appear in new words. In Section 5 some experiment results are listed and analyzed.

2. An Overview of How-Net

How-Net is a bilingual general knowledge-base describing relations between concepts and relations between the attributes of concepts [3]. Knowledge dictionary is the basic file of How-Net. In this file the concept and description of each word form a record. Each record in either language of Chinese and English has four items. Each item has two parts, between them there is a "=" as separator. In the left of "=" is the data item name and in the right is the value of the data item. The format is as follow:

W_X=word
E_X=some examples of this word
G_X=the part of speech of this word
DEF=the definition of this word

"X" indicates the classification of language, so if "X" is replaced by "C" it means this record is used for Chinese

and “E” means English. Here our task is dealing with polyphones in Chinese so what we need is only Chinese part. “DEF” indicates the definition of each word and the value of “DEF” is made of several sememes and the semantic relations between them and their key word. Sememe is the most basic unit. After reviewing and analyzing over 6,000 Chinese characters, in How-Net more than 800 sememes were drawn out and some semantic relations such as agent, patient, relevant and so on were listed. All these relations are represented by the prefixes inserted in front of sememes such as “*”, “@”, “$” and so on. So these relations were called prefix semantic relations of sememes and the according symbols are called prefix semantic relation symbols [4]. The philosophy of How-Net is “All things around the world (including substantial and immaterial ones) are always moving and changing in particular time and space. They often change from one state to another state and are shown by their attribute values.”[5] All these sememes and their relations can indicate definitions of all Chinese words.

“E_X” is used to introduce some examples for the word with several semantic records. These examples can be used to discriminate the pronunciations of each word.

There is an example:

W_C=扒

G_C=V

E_C=~猪肉，鸡爪~豆腐，~鸡

DEF=烹调

In the same line as “E_C” the symbol “~” indicates the word of the item “W_C”. That is to say that “扒猪肉”, “鸡爪扒豆腐” and “扒鸡” are the examples of this record. From “DEF” we can know the meaning of “扒” in “扒鸡” is “烹调”, which means “cook”.

Besides the semantic knowledge base, How-Net provides sememe classification trees, which organize all the sememes and their relations. In How-Net there are several classification trees respectively indicating “Event”, “Entity”, “Attribute”, etc. There is an example that is part of the classification tree of “Entity”:

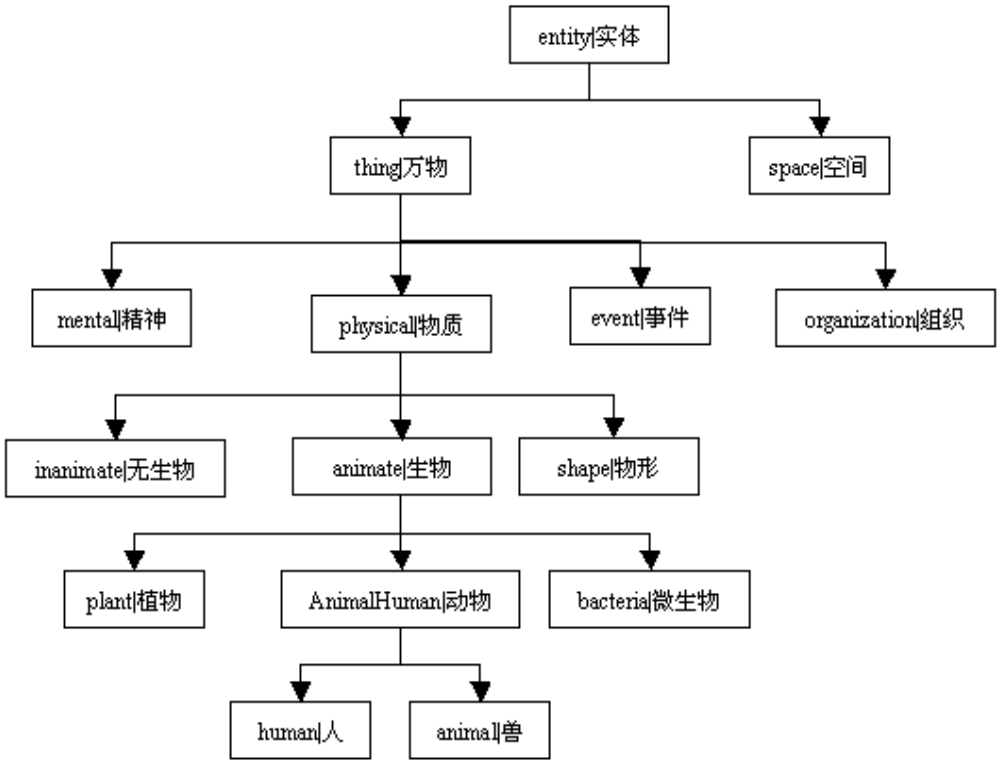

Figure 1. part of the structure of “Entity” classification tree

In the tree “entity” is the highest concept and its child nodes, “thing” and “space”, are the lower concepts. The rest may be deduced by analogy. “human” is a concept that belongs to “entity” but is five levels lower than “entity”.

3. Semantic Similarity Degree

From the classification tree in the last section we can see that the distance between the sememes in the tree can be calculated so we can use it to calculate the semantic similarity degree between two sememes. Then in the definition (DEF) of each word there are sememes, so we can calculate the semantic similarity degree between two words.

3.1. Semantic Similarity Degree Between Two Sememes

Definition 1: semantic distance between sememe a and sememe b is

DISTANCE-ATOM(a,b)=the shortest distance between a and b in the semantic classification tree

Definition 2: semantic similarity degree between sememe a and sememe b is

SIM-ATOM(a, b)=

$$(1-\frac{DISTANCE-ATOM(a,b)}{TreeHeight\times 2})\times 100 \quad (1)$$

3.2. Semantic Similarity Degree Between Two Words

Definition 3: Relation(I) is the set of all prefix semantic relations in DEF of word I.
Definition 4: Item-Relation-Atom(I, K) is the set of sememes which have the prefix semantic relation K in DEF of word I.
Definition 5: semantic similarity degree between word I and word J is

SIM-WORD-ITEM(I, J)=

$$\sum_{k\in \mathrm{Re}lation(I)} \underset{\substack{a\in Item-\mathrm{Re}lation-Atom(I,k)\\ b\in Item-\mathrm{Re}lation-Atom(J,k)}}{Max} SIM-ATOM(a,b) \quad (2)$$

These five definitions above are first introduced by Yang X.-F. (2000).[4]

4. Algorithm of Dealing with Polyphones in New Words

4.1. The Preprocessing of How-Net

In the file of How-Net there is no information about pronunciations of words, so we have to insert pronunciations in the records after the items of English have been deleted. We add an item named "Y_C", which is used to introduce the pronunciation syllables of the word with this definition item. For example:

W_C=扒
G_C=V
E_C=~猪肉，鸡爪~豆腐，~鸡
Y_C=pa2
DEF=烹调

"pa" is the spelling form of the syllable and "2" indicates the tone value of this character. We use the numbers from 1 to 5 to indicate the tone values of characters.

There are over 60,000 Chinese records in How-Net. It is impossible and unnecessary to add the pronunciations of all the records. What we have to do is only picking out the records of the single characters that are polyphones and add item "Y_C" in them. The rest records are not changed.

4.2. Description of The Algorithm

When a new word comes out, we use the following steps to determine its pronunciation:

- Look up in the item "E_C" among the records of the new word's polyphone. If the word is in "E_C" and the value of "G_C" is the same as the part of speech of this word, the pronunciation of the polyphone in the word is the syllable in "Y_C" of this record. Go to next step if it is not in "E_C".
- Look up in all the records. If the word is in How-Net, find out all its records in which "G_C" is the same as the part of speech of this word. Calculate the semantic similarity degrees between these records we have gotten and the records of the polyphone in the word. The pronunciation of the polyphone in the word is the syllable in "Y_C" of the record that has the highest semantic similarity degree with the word. Go to next step if the word is not in How-Net.
- Break up this word and calculate the semantic similarity degrees between the rest of the word and the records of the polyphone in the word. Find the pronunciation belonging to the record with the highest semantic similarity degree and the degree is above a certain threshold value. If the degree is lower than this threshold, it means that the pronunciation of this word can't be determined according to How-Net and it is left for dealing by other methods.

4.3. Example Explanation of The Algorithm

Here we give some examples to explain the algorithm of dealing with polyphones in new words.
Example 1: the new word is "行为" and in it "行" is a polyphone. There is a record of "行":

W_C=行
G_C=N
E_C=品~,~为
Y_C=xing2
DEF=属性,举止,人

We find "行为" is an example of this record, so the character "行" is pronounced "xing2" here.
Example 2: the new word is "银行" and in the records of "行" we can not find correlative "E_C" items. However there is a record of "银行" in How-Net:

W_C=银行
G_C=N
E_C=
Y_C=
DEF=场所,@留存,@取回,@借出,钱财,商

There are several records of "行" of which the parts of speech are "N" (noun):

W_C=行
G_C=N
E_C=
Y_C=hang2
DEF=属性,辈分,人

W_C=行
G_C=N
E_C=~业
Y_C=hang2
DEF=事务

W_C=行
G_C=N
E_C=品~
Y_C=xing2
DEF=属性,举止,人

W_C=行
G_C=N
E_C=商~
Y_C=hang2
DEF=场所,商

W_C=行
G_C=N
E_C=
Y_C=hang2

DEF=场所,@留存,@取回,@借出,钱财,商

So we calculate the semantic similarity degrees between “银行” and these records of “行”. The semantic similarity degree between “银行” and the last record of “行” is the highest. So the character “行” is pronounced “hang2” in “银行”.

Example 3: the new word is “行辈”. We can not determine the pronunciation of “行” during the first two steps of the algorithm. The rest of “行辈” except “行” is the character “辈”. The records of “辈” in How-Net with the same part of speech N are below:

W_C=辈
G_C=N
E_C=~分，晚~，长~，前~，后~，长他一~
Y_C=
DEF=属性,辈分,人

W_C=辈
G_C=N
E_C=无能之~，平庸之~，碌碌无为之~，尔~，吾~，我~数人
Y_C=
DEF=人,众

We calculate the semantic similarity between these records of “辈” and the records of “行” in Example 2. Among the similarity degrees we have got the one between the first “辈” and the first “行” is the highest. So the character “行” in “行辈” is pronounced “hang2”.

5. Experiment and Analysis

In order to test the ability of our algorithm to deal with polyphones in new words, we must do some experiments.

In our corpus for automatic detection of prosodic phrases in TTS, there are over 10,000 sentences and the pronunciation syllables aligned to the texts. And the polyphones appear about 40,000 times in over 6,000 different words. So we pick out all the words in which there are polyphones and the corresponding polyphones’ pronunciations. Since we have not used the polyphone list or the training corpus for dealing with polyphones before using this algorithm, we regard all these words as new words. Then these words are used as experiment datum to calculate the accuracy ratio.

We use the follow definitions to access the accuracy ratio of dealing with polyphones in new words:

Definition 6: accuracy ratio is the ratio of the number of new words in which the pronunciation of the polyphone is determined *properly* and the number of new words with the polyphone.

Firstly, we pick out all words with the character “行” and the character “薄” and then use our algorithm to determine the pronunciations. The experiment result is shown in Table 1.

Table 1: The experiment result of determine the pronunciations of “行” and “薄” in new words

Polyphone	Accuracy ratio
行	77.4%
薄	81.8%

Secondly, we find out 370 common polyphones in modern Chinese articles and sort them by alphabet. Then we pick out the first forty and their corresponding new words in our corpus to do another experiment. All these forty polyphones are listed in Appendix. Then the totally average accuracy ratio is 84%.

From the experiment results above we can see our method to deal with polyphones in new words using How-Net has done good performance and will raise the accuracy ratio of the whole process in converting text to syllables. And it has the following characters:

- The problem of data sparseness is solved in a large degree.
- This method lessens the laborious hand tagging of training corpus.
- The algorithm is easy to be applied.

Certainly in some situations this method can’t do well. For example, if the new word is somebody’s name, the characters have no real meanings, so the semantic analysis can’t help to determine their pronunciations. But this problem can be solved by counting the pronunciation frequency of the characters that are often used in people’s names. On the other hand, if we add context information into the calculating of semantic similarity degrees, perhaps we can get better performance. So in our future work we will try to do this.

6. Reference

[1] Sun, Y.-F., Zhang, K., Wang, X.-L., and Xu, Z.-M., *Research on Polyphone Based on Rules and Statistics*, NCMMSC5, Harbin, 1998, pages 323-327

[2] Zhou, Q., and Feng, S.-Y., *Build a Relation Network Representation for How-net*, 2000 International Conference on Multilingual Information Processing, Urumqi, 2000, pages 139-145.

[3] Gan, K.-W., and Wong, P.-W., *Annotating Information Structures in Chinese Texts Using How-Net*, ACL2000, Hong Kong, 2000 .

[4] Yang, X.-F., Li, T.-Q., and Hong, Q.-Y., *Example-Based Chinese Syntactic Structure Disambiguation*, http://www.keenage.com, 2000.

[5] Dong, Z.-D., and Dong, Q., *How-Net*, http://www.keenage.com, 1999.

[6] Yang, E.-H., Zhang, G.-Q., and Zhang, Y.-K., *The Research of Word Sense Disambiguation Method Based on Co-occurrence Frequency of Hownet*, http://www.keenage.com, 1998.

7. Appendix

Table 2: The forty polyphones used in our experiment

阿	腌	挨	艾	拗
扒	耙	把	百	柏
伯	般	膀	蚌	磅
炮	剥	薄	堡	暴
瀑	刨	背	臂	奔
绷	泌	秘	裨	辟
扁	便	瘪	别	槟
并	屏	泊	簸	卜