Text-to-pinyin conversion based on contextual knowledge and D-tree for Mandarin

Sen Zhang, Yves Laprie

TEXT-TO-PINYIN CONVERSION BASED ON CONTEXTUAL KNOWLEDGE AND D-TREE FOR MANDARIN

Zhang Sen, Yves Laprie
Speech Group, INRIA-LORIA B.P.101 54602 Villers les Nancy, France
Chinese Academy of Sciences, Beijing, China
zhangsen@yahoo.com

ABSTRACT

In this paper we discuss the technologies for text-to-Pinyin conversion based on context knowledge and D-tree for Mandarin Chinese. (1). The local and long-distance context knowledge were considered and exploited. (2). The context knowledge was represented by Regular Expressions. (3). D-trees were built based on the Regular Expressions. (4). The experimental results show that the approach using context knowledge and D-tree is an efficient solution to implement Chinese text-to-Pinyin conversion.

1. INTRODUCTION

Text-to-pinyin (TTP) conversion is an indispensable and important part in Chinese Text-to-Speech (TTS) systems. In Chinese TTS systems, the input text will be firstly converted into proper pinyin sequence, then the corresponding speech waveform will be generated based on these pinyin sequences. Pinyin is a transcription system which translates Chinese characters according to their pronunciations into Roman characters. Since many Chinese characters have multiple pronunciations, it is difficult to predict the proper pronunciation of a Chinese character in some cases. So, TTP is a challenging problem in both research and development of high-performance Chinese TTS.

Many researchers have already realized to use the context knowledge to deal with the Chinese TTP problem [1]. However, the difficulty is what kind of context knowledge should be exploited in Chinese TTP, what representation form of the context knowledge is efficient for Chinese TTP? In this paper, we try to answer these questions.

Table-search approach is a basic and essential schema for Chinese TTP conversion. Since it is simple and efficient, it is still widely used in state-of-the-art Chinese TTS systems. Usually, the table consists of more than 40,000 Chinese words and their pinyin (pronunciations), each word may have two or more Chinese characters. This schema used local context knowledge by collecting the frequently used words and phrases, which is referred to as “predicting the pronunciation of a character by a word”. In general, if the words and phrases in the table were carefully selected, 95% or better TTP conversion accuracy can be achieved. To deal with the left 5% error cases, long-distance context knowledge and linguistic knowledge should be further exploited. In Chinese TTP conversion, the errors often take place in high-frequency characters which usually play important role in semantic understanding.

Why don’t we use some schema which succeeded in English TTP conversion? The first reason is that Chinese has a special linguistic structure which is different from that of English. The second is that the pronunciation variation problem of Chinese characters is more difficult than that of English. Why don’t we use statistical LM, such as bi-grams or tri-grams, in Chinese TTP conversion? The reason is

that the context knowledge in LM can be also found in a large-size Chinese word-base, and the structure of the word-base is more flexible and more efficient than LM for Chinese TTP conversion.

As an efficient context knowledge representation method, d-tree has been used in continuous speech recognition to deal with the pronunciation variability in spontaneous speech [2]. In our research, we used d-tree to represent Chinese context knowledge which describes the conditions by which the pronunciation of a Chinese character can be predicted.

The outline of this paper: (1). We address some problems existed in Chinese TTP conversion and discuss the approaches already proposed to deal with these problems. (2). We discuss the local and long-distance context knowledge representation issues by Regular Expressions and d-tree. (3). We propose a schema based table-search and d-trees for Chinese TTP conversion. (4). We present some experimental results which showed that the approach using context knowledge and d-tree is an efficient solution to implement Chinese text-to-Pinyin conversion.

2. PROBLEMS IN CHINESE TTP

Let us begin with the problems which must be taken into account in Chinese TTP conversion.

2.1 How Many Characters?

Characters are fundamental elements in Chinese. But, how many of them? No one can tell the exact number of Chinese characters. Some modern Chinese dictionaries collected more than 20,000 characters. Statistical results based on large Chinese corpus show that a few thousands of Chinese characters can cover 99.99% characters used in Chinese media. Based on these studies, GB2312 was designed for Chinese information processing, in which 6763 most frequently used Chinese characters are included.

GB2312 specified how many Chinese characters should be encoded and how to encode them, but it didn't specify the pronunciation of each character. The mapping from the character-set into the code-set can be illustrated as in figure 2.1.

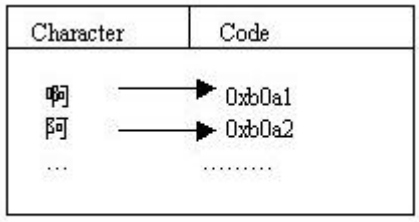

Figure 2.1 Structure of GB2312

2.2 Pronunciation and Syllable

Each Chinese character may have one or multiple pronunciations, usually each pronunciation only consists of one syllable, each syllable can be further viewed as two parts: INITIAL and FINAL. Tone is an important feature attached to syllable. Two syllables with different tones are treated as different syllables in this paper. The total number of phonologically allowed syllables in Chinese is about 1300. In other words, every about 5.2 characters share one syllable. The relation between characters and their pronunciations can be depicted as following figure 2.2.

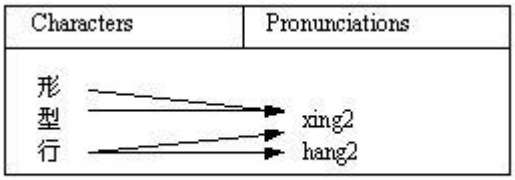

Figure 2.2 Mappings: chars to pronuns

Pinyin is a transcription system which translates the pronunciations of Chinese characters into Roman characters. Pinyin is an easy way to represent Chinese pronunciations, so it is widely used in both research and development of Chinese information processing. For example, "xing2" is a pinyin, "xing" is called the base syllable which has INITIAL "x"

and FINAL “ing”, “2” is the tone of this syllable.

2.3 Typical Problems

As we mentioned, one Chinese character can map to multiple syllables. Such character is called homograph. Almost 800 such characters exist in Chinese language. Many hard problems are caused by the homographs. Even exploiting context, the proper pronunciation of a homograph can not be easily predicted. In the following examples, the pronunciations of the characters in italic font can not be correctly determined even in a word. This shows that there are some cases that can not be solved by table-search method based on a large-sized word-table in Chinese TTP conversion.

Words	Pinyin
重 用	zhong4yong4 | chong2yong4
同 *行*	tong2xing2 | tong2hang2

Figure 2.3 Pronuns. in words

In general, the meaning and the part-of-speech (POS) of a character depend on its pronunciation, so do a lot of Chinese words. So, errors in TTP conversion usually lead to misunderstanding of a word or sentence in TTS system. Apart from the above examples in which the pronunciation of a character can not be determined by only using the word including the character, even worse cases exist in Chinese. The following example shows the pronunciation of a Chinese character can not be determined by considering a whole sentence. The meaning of the sentence will completely different (almost opposite) with the two possible pronunciations of a character in it.

他 *还* (*huan2* | *hai2*) 欠款 100 元。

Figure 2.4 pronun. in a sentence

For the sentence in the figure 2.4, the sentence structure analysis doesn’t give any hint of the correct pronunciation. In fact, even human beings don’t know the correct pronunciation only by this isolated sentence. These examples show that only considering the context information in the scope of a word or a sentence is still not enough for solving the Chinese TTP conversion problems.

We can give some factors which often cause some problems in Chinese TTP conversion.

(1). Characters which are homographs and can be used as family names. In Chinese, there are over 10 such characters (shown in figure 2.5), but they are frequently used. Such characters often cause word segment problem and then leads to TTP conversion problem.

乐 仇 单 曾 盛 车 石 应 解 查 ……

Figure 2.5 homographs as family names

(2). Characters which are homographs and have two or more POSs (i.e., N and V). In Chinese, there are over 100 such characters (shown in figure 2.6). In fact, most of the homographs have two or more POSs. Usually, such characters are among the most-frequently-used ones.

背 打 数 行 还 会 中 的 称 量 处……

Figure 2.6 homographs having more POSs

(3). Tone-sandhi which makes some tones vary in some special context, but the base syllables usually do not change. Tone-sandhi patterns can be represented by a few rules as follows, these rules should be applied within semantic block (generated by word segment and sentence structure analysis which may lead to some errors).

Tone 3 + ..+ tone 3 + tone x (x≠3) =>
Tone 2 + ..+ tone 2 + tone 3 + tone x

Figure 2.7 tone-sandhi rules

Apart from the above listed factors, other factors may also influence Chinese TTP conversion, i.e., word segment, etc. The accuracy of state-of-the-art TTP conversion has been over 99%.

3. CONVERSION APPROACHES

Let's first consider the approaches which have been used for English text-to-phoneme (TTP) conversion, then discuss which approaches should be proper for Chinese TTP conversion.

In general, the approach for English TTP conversion is based on a set of letter-to-phoneme (LTP) rules and a dictionary for exceptional words and special names, etc. For example, more than 4530 rules have been used for the pronunciation of letter 'a'. Some of these rules [3] are as follows:

[a] => [ax] / [nst] _ [bles]
[a] => [ae] / [ ] _ [naco]
[a] => [ax] / [hex] _ [gon]

Obviously, these rules exploit the left and right context to predict the pronunciation of a letter. However, these rules can not deal with all the problems in English TTP conversion. So, some exceptional words or special names will be collected in a dictionary to explicitly lis their pronunciations for table-search.

The approaches used in Chinese and English TTP conversion share some commons, but some differences exist. In Chinese TTP conversion, a large-size word-base is first used, then some rules will be applied to some exceptional cases. Usually, the large word-base contains local context knowledge, and can deal with most cases in TTP conversion for Mandarin Chinese. The contents of the word-base can be as follows.

Word	Pinyin
行动	xing2 dong4
银行	yin2 hang2
……	… …

Figure 3.1 contents of word-base

In fact, the contents of Chinese word-base are essentially identical to the rules used in English, though they are represented in different forms, i.e., the examples in figure 3.1 can be expressed in the following form:

```
[行] => [xing2]/[]_[动]
[行] => [hang2]/[银]_[]
```

The word-base is very efficient to predict the pronunciation of the Chinese homographs by using table-search. So, it is still used in state-of-the-art TTS system for TTP conversion. However, word-base only contains local context knowledge, the pronunciation of some Chinese homographs requires long-distance context knowledge to be predicted.

Now, we will discuss how to represent the long-distance context knowledge and how organize it in a proper structure for Chinese TTP conversion. To facilitate the TTP conversion, we adopt the Regular Expression (RE) and use the similar operators (symbols) to represent the context knowledge which haven't been included in the static word-base. In the following figure 3.2, we give an example for a Chinese monograph.

```
[行] => [xing2]/[]_[]
[行] => [xing2]/[]_[了|吗|呀|啊]
[行] => [hang2]/[(\.*)]_[(\d*)[树](\.*)]
[行] => [hang2]/(\.*)[在|懂|那|哪](\.*)_[(\.*)]
[行] => [xing2]/[(\.*)[一](\.*)]_[(\d*)[人](\.*)]
```

Figure 3.2 Long-distance context knowledge represent.

As we mentioned, there are about 800 monographs in Chinese, but only a small portion of them need long-distance context knowledge to predict their pronunciations. By checking the TTP conversion results, we selected 50 monographs out of 800 and built rules by using long-distance context knowledge as in figure 3.2. In these rules, we didn't use POS though it is an important feature. If POS was used, text parsing before TTP conversion should be performed. However, the parsing results for Chinese language is often too ambiguous since the grammar of Chinese is too free.

Based on the rules illustrated in figure 3.2, we built d-trees for each of the 50 homographs. The d-tree of a homograph is the combination of all the rules of that homograph. Figure 3.3 is a d-tree for the rules in figure 3.2.

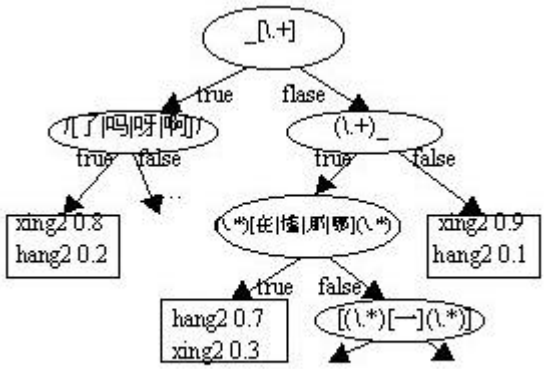

Figure 3.3: d-tree of monograph '行'

These 50 d-trees was trained by a Chinese text corpus with their Pinyin transcriptions. This corpus was developed by PKU in China [5], it includes about 1 million Chinese characters. In this corpus, only the pronunciations of homographs in pinyin were given. The following figure 3.4 shows a piece of the corpus. The POS tagging (after each "/") in this corpus was not used.

```
这{zhe4}/r  当然/d  只是/v  个{ge4}/q  黄金/n
的{de5}/u  幻梦/n  。/w
```

Figure 3.4 Piece of Chinese corpus

The probabilities at each leaf node of the d-trees (as in figure 3.3) could be yielded by the training process. By using d-trees, the pronunciation of some monographs can be scored.

The implementation of TTP conversion for Mandarin is illustrated in figure 3.5. First, the input text should be segmented into word sequence, then check if the input character is monograph, if it has an entry in the word-base and if it has a d-tree, the default pronunciation will be assigned to the isolated single character.

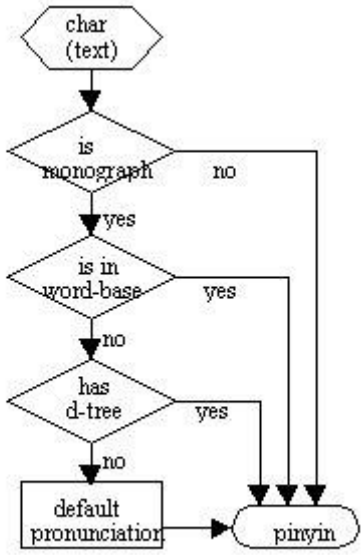

Figure 3.5 TTP conversion for Mandarin

4. EVALUATION

We selected a test set of 1,800 Chinese characters, in which 100 are monographs. The word-base consists of 50,000 words and their pronunciations, and 50 d-trees were build for monographs. The TTP conversion results were listed as follows.

Approaches	TTP Accuracy
Word-base	99.28%
Word-base + d-trees	99.61%

5. CONCLUSION

We presented an approach for Mandarin TTP conversion by using static word-base and the context knowledge which is represented by REs and d-trees. The experimental results seem quite good, but some errors still exist, i.e., neutral tone (tone 5) were often not correctly processed.

REFERENCES

[1] Lin-shan LEE, Voice dictation of Mandarin Chinese, IEEE Signal processing magazine, July, 1997

[2] Michael Riley, et al, Stochastic pronunciation modelling from hand-labelled phonetic corpora, Speech Communication 29, 1999

[3] http://www.nist.gov/speech/tools/addttp4-11tarZ.htm
Text-to-phone rules of English and software.

[4] C. Shih and R. Sproat, Issues in text-to-speech conversion for Mandarin, int. J. Comp. Ling. Chinese Language Processing, vol.1, no.1,1996

[5] A Text corpus with 1-million Chinese characters with segmentation, POS and pinyin transcription, Beijing University, China, 2000