Grapheme-to-Phoneme Conversion in Mandarin Chinese Text-to-Speech System

Hongwei Ding and Oliver Jokisch
Laboratory of Acoustics and Speech Communication, Dresden University of Technology, Germany
Email: {hongwei.ding; oliver.jokisch}@ias.et.tu-dresden.de

Abstract

We present a lexicon-based model for segmenting Chinese text into dictionary entries and for providing pronunciations for these words. This approach adopts a matching algorithm combined with several heuristic rules to resolve the ambiguities. It can achieve total accuracy over 95%, which proved to be an effective solution to grapheme-to-phoneme conversion for Mandarin Chinese.

Introduction

The written Chinese texts are composed with strings of characters without blanks to delimit words. The first step towards word-based indexing is to break a sequence of characters into words. This process is called word segmentation. On the other hand, it is not possible to bypass the word-segmentation problem. The main reason is that many Chinese characters are homographs, whose pronunciation depends upon word affiliation.

The Problem of Word Segmentation

There are difficulties with the word identification process. First of all, almost all characters are free morphemes, which can be one-character words by themselves. They can also join other characters to form multi-character words. Second, compounding is the predominant word formation device in modern Chinese. It is difficult to tell whether a low-frequency compound is a word or phrase. Third, the same pool of characters is also used in constructing proper names, which brings difficulty in personal name identification [2].

Strategies in Word Segmentation

In order to cope with this problem, there exist some methods which can be classified into

(1) Purely statistical approaches [1];

(2) Heuristic rule-based methods [2];

(3) Statistical approaches which incorporate lexical knowledge [3].

Many statistical methods are based on a large pre-segmented text corpus for their analysis. The easiest and most effective one is the lexical based algorithm with supplementary rules. This is also adopted in our TTS system DRESS, but is modified to pass our system.

The paper first introduces our synthesis system. It then presents the solution of word identification and phonetic conversion. Finally, it points out the possibility for future research.

Synthesis System

The Mandarin Chinese Text-to-Speech system developed at TU Dresden is a syllable-based waveform concatenation synthesis. It consists of text analysis and acoustic synthesis. The acoustic synthesis is already accomplished with high naturalness. A syllable-based inventory takes the cross-syllable co-articulation into consideration [4]. A neural network is responsible for learning and modifying the duration and intonation [5]. Because of the unsolved problem of grapheme-to-phoneme conversion, the word boundaries had been inserted manually in the process of synthesis. This paper presents the solution of word segmentation, which makes the whole text-to-speech system to operate automatically.

Word Segmentation

The processing stage of word segmentation includes an algorithm of maximum matching with word lexicon, several ambiguity resolution rules, and some solutions to deal with time, numeral expressions and to identify personal names.

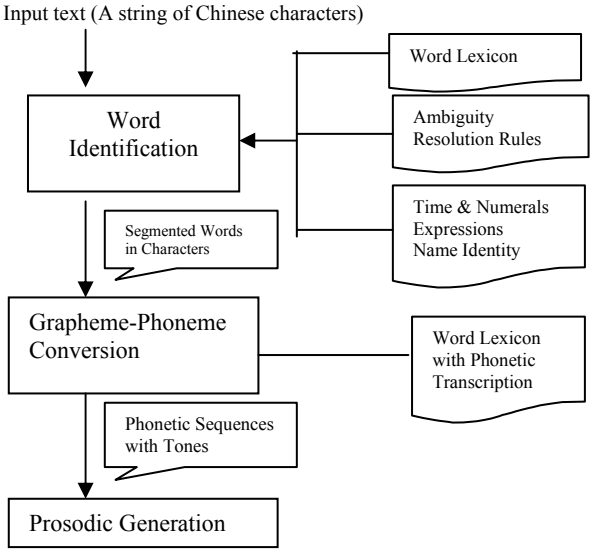

Figure 1: Grapheme-phoneme conversion

Maximum Path-Matching

The lexical-based word identification approach is matching, the basic strategy is to match the input characters string with a large set of entries stored in a pre-compiled lexicon to find all (or part of) possible segmentations. Another variant of maximal matching done in [2] says that the most plausible segmentation is the three-word chunks with maximal length. This algorithm is adopted in our system.

For example, suppose C1, C2, ... Cn represent characters in a string. We search the lexicon to find out the three words. If the following are possibilities of three-word segmentation:

C1 C2 C3C4
C2C2 C3 C4C5
C1C2 C3C4 C5C6

This matching algorithm will pick the third segmentation, which has three words with maximum length.

In this method, the lexicon plays the critical role in the result of word segmentation. The lexicon is a list of character strings of mainland Mandarin Chinese words (about 100,000) with their pronunciations in pinyin (transcriptions of Mandarin Chinese used in mainland China). In order to resolve the ambiguities, which can not be handled by the matching algorithm, a set of heuristic rules are applied in the analysis.

Ambiguity Resolution Rules

Some morphological rules [2] will be applied in our word segmentation. Two of them are introduced here.

Smallest Standard Deviation

The first one is that in Chinese, the word lengths are usually evenly distributed. For example, if the matching algorithm can not resolve the following ambiguities:

C1C2 C3C4 C5C6
C1C2C3 C4 C5C6

According to the smallest variance of word lengths, the first segmentation will be the choice.

Freedom of Monosyllabic Words

If the above rule can still not solve the ambiguity, the degree of morphemic freedom will be considered. The frequency of occurrence of a character can serve as an index of its degree of morphemic freedom. A high frequency character is more likely to be a one-character word, and vice versa. For example, if these two segmentations are possible:

C2 C2 C3C4
C1 C2C3 C4

Character C2 has a larger sum of frequency than Character C4, the first segmentation is preferred. Apart from these two main rules, some other morphological rules will also be used. The rate of ambiguity resolution of morphologic rules will be tested before they are adopted.

Personal Names and Numeral Expressions

Though the MMS can gain a very good performance. It was lacking a good name, numeral and time algorithm. In our system, the time and numeral expressions and name identifiers will be built in. Personal names, expressions of numerals and times will be considered as a word.

Results

In order to test the efficiency of these rules, the self-prepared database [5] will be used. The frequency of the one morpheme word is calculated, and saved in a list as a reference for the application of the second rule. The database consists of 63 sentences, the accuracy amounts to 96%.

Grapheme-to-phoneme Conversion

The segmented words will then be converted into phonetic strings, which are the pinyin transcription employed in our synthesis system. The same dictionary will be used, so that every word can find its corresponding pronunciation in the dictionary. After the grapheme-to-phoneme conversion, the results will be fed into the first step of acoustic synthesis module – the prosodic generation (see Figure 1), so that the whole text-to-speech system can run automatically. A naturally synthesized sentence can be generated from our system.

Discussion

This lexicon-based method is reported as one of the most effective ones in word segmentation. In our grapheme-phoneme conversion, a word list with a good coverage is compiled. With the above-mentioned supplementary modifications, our word segmentation achieves a very good performance.

But, on the other hand, no dictionary can be exhaustive, though the precision of word identification is comparably high. It covers only words that are in the dictionary. The algorithm can not identify some unknown words in texts.

In order to achieve further improvements in the future, a statistical model can also be integrated into the system, and the part of speech tagging is also necessary for an accurate word segmentation.

References

[1] Sproat, R. and Shih, C., A Statistical Method for Finding Word Boundaries in Chinese Text. In *Computer Processing of Chinese and Oriental Languages*, 4:336-351, March 1990

[2] Chen, K. and Liu, S., Word Identification for Mandarin Chinese Sentences. In *Proceeding of COLING*, pp. 23-28, August 1992

[3] Fan, C.-K. and Tsai, W.-H.., Automatic Word Identification in Chinese Sentences by the Relaxation Technique. In Computer Processing of Chinese and Oriental Languages, 4: 33-56, 1988

[4] Helbig, J. and Ding, H., A Syllable Based Mandarin Chinese Speech Synthesis Regarding Cross Syllable Coarticulation Effects. In Proceeding ICSP 1997, Seoul, Korea, pp. 173-176

[5] Jokisch, O. and Ding, H. et al., Learning Syllable Duration and Intonation of Mandarin Chinese. ICSLP 2002, Denver, pp. 1777-1780.