

A New Rule-based Method of Automatic Phonetic Notation on Polyphones

ZHENG Min, CAI Lianhong

(Department of Computer Science and Technology of Tsinghua University, Beijing, 100084)

Email: yi yi@media.cs.tsinghua.edu.cn, clh-dcs@tsinghua.edu.cn

Abstract: In this paper, a new rule-based method of automatic phonetic notation on the 220 polyphones whose appearing frequency exceeds 99% is proposed. Firstly, all the polyphones in a sentence are classified beforehand. Secondly, rules are extracted based on eight features. Lastly, automatic phonetic notation is applied according to the rules. The paper puts forward a new concept of prosodic functional part of speech in order to improve the numerous and complicated grammatical information. The examples and results of this method are shown at the end of this paper. Compared with other approaches dealt with polyphones, the results show that this new method improves the accuracy of phonetic notation on polyphones.

Keyword: grapheme-phoneme conversion; polyphone; prosodic word; prosodic functional part of speech; feature extracting;

1. Introduction

Grapheme-Phoneme Conversion is a method that transforms text forms into phonetic forms, such as phoneme forms for English and phonetic notation for Chinese. There already many researches are mainly based on reading and spelling rules^[1] which are summarized by authorities or based on corresponding relations between alphabetic sequences and phonetic symbol sequences which are obtained by data-driving approaches^[2] on alphabetic writing system. Then the system confirms whether the phonetic notation of a new word exists in the lexicon or not, that is to say, the system solves an Out-of-vocabulary/OOV problem. But this problem doesn't exist in Chinese. Generally, every Chinese character has a fixed phonetic notation, but some special characters have two or more phonetic notations. The statistic result from 《词海》, which is 2641 polyphones^[3] out of 16339 Chinese characters, shows the considerable proportion hold by polyphones in Chinese. So ascertaining the pronunciations of the polyphones is a basic and important problem in Chinese grapheme-phoneme conversion.

2. A rule-based method on polyphones

2.1 Classify the polyphones beforehand

Although the number of polyphones is large, the appearing frequency is widely discrepant. The statistic results in our experiment show: the accumulative frequency of the former 100 polyphones exceeds 93%, and the accumulative frequency of the former 180 polyphones exceeds 97%. The distributing chart of polyphones' accumulative frequency is shown in the chart 3.1. In order to improve the accuracy of the grapheme-phoneme conversion more, we classify the 700 polyphones in the GB_2312 Chinese-character system into three kinds in our text-to-speech system

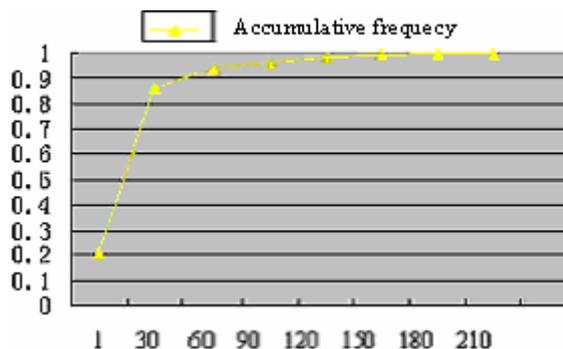


Chart 2.1 Accumulative frequency of polyphones

- More than 200 polyphones are rare, which scarcely can be seen in usual articles. We select default pronunciations to these rare polyphones. Such as the “读” is pronounced “dou4” when the meaning is “停顿” in the ancient.
- Some polyphones can not appear alone, the polyphonic situation only occurs when they compose a word with two or three characters. Such as “恶” is pronounced “wu4” only in some polyphonic words like “憎恶” or “可恶”. “恶” can only be pronounced “e4” when it appears alone.
- More than 160 polyphones often appear alone, two or more pronunciations and the corresponding part of speeches of which must be both kept in the lexicon. Such as the lemma of “好” must be kept

as “\hao3x\hao4f”, “发” must be kept as “\fa1d\fa4m\fa11” (x, f, d, m, l represent adjective, adverb, verb, noun and measure word respectively)

Excluding the first kind of polyphones, we mainly deal with the second and the third kinds about 220 polyphones which are often used in articles.

2.2 Concept of prosodic functional part of speech

The word-syncoating and word-tagging in a sentence usually have been done before grapheme-phoneme conversion in the text-to-speech system. But it often adds many complicated tags besides the 26 part of speeches, (Such as some classified tags for proper noun: nr represents name, ns represents location, nt represents incorporation or department, nz represents other proper nouns etc.), which have no use of ascertaining pronunciations of polyphones.

Such as the following sentence:

“我们系倒新来了几名留学生。”

Eight grammatical words can be extracted:

ca(1)=“我们”, ca(2)=“系”, ca(3)=“倒”, ca(4)=“新”,
ca(5)=“来了”, ca(6)=“几”, ca(7)=“名”.....

The sixth grammatical word “几” is pronounced “ji3” because the next grammatical word is a measure word. But the second and the third grammatical words are monomial polyphones, whose pronunciations can not be ascertained only by the neighboring grammatical words or the part of speeches.

In this paper, a new concept called “prosodic functional part of speech” is proposed. When pronunciations of polyphones can not be ascertained by usual grammatical words or part of speeches, we can turn to their prosodic functional part of speeches. The concept of prosodic functional part of speech is syncoating a sentence into several prosodic chunks and giving a prosodic part of speech to each chunk.

Such as the former sentence can be syncoated into five chunks :

ca(1)= “我们系”, ca(2)= “倒”, ca(3)= “新来了”,
ca(4)= “几名”, ca(5)= “留学生”

“系” and “我们” are combined into a nominal prosodic chunk which avoids the ambiguity of pronunciation because “我们系” is apparently a noun and “系” is pronounced “ji4” only in

a verb. So we can confirm that the pronunciation of “系” is “xi4”. We can also ascertain that the pronunciation of the polyphone “倒” is “dao4” because it is pronounced “dao3” only in a verb.

When we deal with the polyphones in the experiment, the part of speeches are simplified into 18 kinds shown in the following diagram 3.1:

Diagram2.1 Abbreviated part-of-speech diagram

	Part of speech		Part of speech
pn	Decorated word before noun	n	Noun
pv	Decorated word before verb	v	Verb
pm	Decorated word before numeral	m	Numeral
bn	Decorated word after noun	d	Adverb
bv	Decorated word after verb	q	Measure word
bm	Decorated word after numeral	s	Location word
nr	Surname word	p	Preposition
c	Connected word	u	Auxiliary word
t	Time word	r	Pronoun

We suppose that Chinese contains three kinds of prosodic chunks shown in the diagram 3.2. As every prosodic chunk is usually composed by a central word and some decorated words before or after it, each kind of prosodic chunk adds three kinds of prosodic functional part of speeches. The difference between prosodic and grammatical part of speech is that the former one can be a part of speech both for a single grammatical word and for a prosodic word. In fact, the pronunciation of the polyphone is closely related to the part of speech of the prosodic chunk.

Diagram2.2 Prosodic functional part-of-speech diagram

Prosodic chunk	Former decorated word	Central word	Next decorated word
Nominal prosodic chunk	Decorated words with “的”: 快乐的, 白的..... Ajective: 美丽、坏..... Indicative pronoun: 这个、由、从..... Others: 十分,很,国内,人均...	noun (includes foreign symbols etc) pronoun, surname word	Prefixal word: 叔叔、工程师..... Words after institutions: 大学,院,局,..... Others: 附近、左右、得很、的、了、的话、来说、等等、之下.....
		pn	bn

Verbal prosodic chunk	Capable word:愿意,可以,能,..... Time, location adverb:在我国... Adverb: 十分... Onomatopoeic word: 砰,铿,.... Decorated words with “地”: 积极地、快速地..... Others:逐字、深入、让大家,.....	Verb (includes 是、有)	Decorated words with “得”: (叫)得响,(吃)得快... Some auxiliary words:的,了... Directional words:(跑)去,走(向),爬(上).. Others:着、不起、(吃)不了、不已、过、以
	pv	v	bv
Numeral prosodic chunk	Indicative pronoun: 这个、由、从... Others::约、占、正、负.....	Numeral	Measure words: 斤、年、吨公里、块钱、件事..... Others:的、正...
	pm	m	bm

2.3 Feature extracting

For convenient, we suppose that the pronunciation of the polyphone only depends on the related information of the words or phrases in the same clause. This suppose is adapted to many polyphones and we extract eight features to establish rules based on the grammatical and prosodic functional information.

1. The grammatical and prosodic functional part of speeches of the words before or after the polyphone or the place that is located:

This feature can handle both the single polyphone and the polyphonic word. Such as the polyphone “长”:

“这棵树长了三厘米”, pronounced zhang3
“衣服太长了” pronounced chang2

Usually, “长” is pronounced “zhang3” when it is a verb and pronounced “chang2” when it is an adjective. It is an example that the pronunciation is ascertained depending on the part of speech of grammatical word.

Other kinds of polyphones such as “种”:

“他十分愿意种地。” pronounced zhong4
“我学会了一种新的算法。” pronounced zhong3

In the first sentence, we can ascertain the prosodic functional part of speech of the polyphone “种” is a verb because both the adverb “十分” and the capable

word “愿意” are decorated words before verb, then we can confirm the pronunciation of the “种” is “zhong4”. In the second sentence, the “种” is pronounced “zhong3” because “一” is a numeral which helps us know that “种” is a decorated word after numeral.

2. Special words before or after polyphones.

These considered grammatical words are usually beside the polyphone or a short distance away. Such as the following polyphone “和”:

“曲高和寡”、“和词” pronounced he4
“豆沙里竟然和了点泥” pronounced huo4
“我打麻将和三圈” pronounced hu2
“我和他是好朋友” pronounced he2

The pronunciation of the polyphone “和” is greatly related to several special neighboring grammatical words, such as when the sentence contains some grammatical words like “诗”, “词”, “曲”, the pronunciation is “he4” and when the sentence contains some grammatical words like “药”, “泥”, “面”, the pronunciation is “huo4” etc.

3. Special characters before or after polyphones;

This feature provides compensations to the former one when the grammatical syncopating is wrong.

Such as: 他的/确有/事” (false syncopating)

The pronunciation of “的” will not be falsely ascertained because of the rules: when “确” is after “的” the polyphone “的” is mostly pronounced “di2”.

4. The length of the words(phrases) before or after the polyphone or the place that is located:

Such as the polyphone “壳”:

“鸡蛋壳”、“外壳” pronounced ke2
“地壳” pronounced qiao4

As the pronunciation “qiao4” of “壳” only appears in the word “地壳”, we can ascertain that “壳” is pronounced “ke2” in a single grammatical word.

5. The position of polyphones in the grammatical word (beginning, end or middle);

Such as the polyphone “省”:

“省会”、“省略” pronounced sheng3
“自省”、“发人深省” pronounced xing3

The polyphone “省” can only be pronounced as “sheng3” when it is at the beginning of a grammatical word, the pronunciation “xing3” appear when it is in the middle or at the end of a grammatical word.

6. The position of the polyphone in the sentence (beginning, end or middle);

Some part of speeches can't appear in some special positions of a sentence, such as surname word can not appear in the end, auxiliary word and measure word can't appear at the beginning of a sentence.

7. The grammatical word and the part of speech which is at the end of the sentence;

For example, if the ending grammatical word in a sentence which contains the polyphone “为” is a numeral, “为” is usually pronounced as “wei2”:

“利润下降为去年的 20%” pronounced wei2

8. The punctuation which is at the end of the sentence that contains the polyphones;

This feature is used to handle the ambiguity arose by tone. Such as the polyphone “啊”:

“多美的花啊!” pronounced a1

“啊? 他病了!” pronounced a2

2.4 The example rules of handling polyphones

Such as the polyphone “为”:
Pronounced wei2:
1) Next punctuation is colon: “处理原则为:”
2) ...为+n: “以法律为准绳 以鲁迅为代表的”
3) 为+所+v: “为人民所喜爱”
4) 为+...+m: “亏损额为 1 万元”
Pronounced wei4:
1) “为” is at the beginning of a sentence: “为简便计”
2) 为+n+而+v: “为人民而献身”
3) ...为+v+n: “这都是为控制人口”
4) 为+n+v: “为我帮忙”

3. Experimental results

The main testing datum are three standard phonetic corpora labeled by a special software and cross-checked by some professionals. So it can be the standard datum to test the accuracy. We make two kinds of tests as follows:

Diagram 3.1 Comparative results between containing and not containing prosodic functional part of speech

	Accuracy without prosodic functional part of speech	Accuracy with prosodic functional part of speech
--	---	--

corpus I	0.955	0.995
corpus II	0.959	0.993
corpus III	0.937	0.983

Diagram 3.2 Comparative results between statistic method and rule-based method

	Accuracy in the statistic method	Accuracy in the rule-based method
corpus I	0.978	0.995
corpus II	0.979	0.993
corpus III	0.967	0.983

4. Conclusion

This paper proposes a new concept of prosodic functional part of speech and a new rule-based method on polyphones. As the results shown, the accuracy is increasingly improved, but errors still exist because the problems like atonic pronunciation is not roundly dealt with which also need to be improved in the future.

References

1. M.Davay and A.J.Vitale, *Algorithms for grapheme-phoneme translation for English and French: Application for database searches and speech synthesis*, Computational Linguistics, 1997,23:495-523
2. K.Torkolla, *An efficient way to learn English grapheme-to-phoneme rules automatically*, Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Minneapolis, 1993,2:199-202
3. Ying Zhiwei, Cai Peiqi, Chen Qihui, *Research of Chinese word segmentation in TTS system*, Computer Applications, 2000,2,20(2):8-11
4. Zhang Zirong, Chu Min, *A statistic approach for grapheme-to-phoneme conversion in Chinese*, Journal of Chinese information processing, 2002,16(3):39-45
5. H.Li and K.Yamanishi, *Text Classification Using ESC based Stochastic Decision Lists*, Proceedings of 8th International Conference on Information and Knowledge Management (CIKM'99), Kansas City, MO, USA, 1999,122-130