Comparison of Three Method for Entering Tone Recognition

Ying Liu
Department of Chinese Language and Literature, Tsinghua University, BeiJing,China,100084
yingliu@tsinghua.edu.cn

Abstract

The entering tones are recognized using rule-based method, maximum entropy model and support vector machine. For those syllables which are affirmatively entering tone syllables, or undoubtedly non-entering tone syllables, rules are made use of to recognize. But for the ambiguity syllables, statistics-based method is utilized to disambiguate. Rules can reduce the recognition problem's scale, and the statistical method can improve the accuracy of entering tone recognition. We compare the results of rule-based method, maximum entropy model and support vector machine. the precision using rules is the lowest and the precision using maximum entropy model and support vector machine are much higher than rule-based method. From the precision, the accuracy, F-measure and efficiency, maximum entropy model is better than support vector machine.

Key words: Entering Tone, Rule, Maximum Entropy Model, Support Vector Machine, F-measure

1. Introduction

Entering-tones are the tones which corresponding Chinese characters are ancient entering-tone characters. Entering tones, one of the four tones in classical Chinese pronunciation, still retain in certain dialects, and disappear for mandarin. For example, "yi1", "shi2", "bai3" and "liu4" stand for four different tone types for mandarin. If the four syllables stand for "一", "十", "百", and "六" respectively, the four syllables are entering tones. Because "一", "十", "百", and "六" are ancient entering tone characters. "一", "十", "百", and "六" are pronounced entering tone for YangZhou dialect and NanNing dialect now. In order to make dialect conversion system's effect better, entering tones should be recognized.[1][2][3][4][5][6][7]

Section 2 gives rules to recognize entering tones. Section 3 introduces maximum entropy model. Section 4 gives a simple introduction to Support Vector Machine. Section 5 presents how to recognize entering tone. Section 6 analyzes experimental results, and offers the systematic comparison of three methods. Section 7 draws a conclusion.

2. Rule-based method

There are 1334 syllables with tone for all Chinese characters of GB2312. 78 syllables are entering tone syllables and 905 syllables are non-entering tone syllables of 1334 syllables. Other 351 syllables are ambiguity syllables. The entering tone syllables and non-entering tone syllables are recognized using rule-based method. [1][2][3][4][5][6][7]

Whether a syllable is an entering tone is related to the vowel and initial consonant of the syllable. We summarized the rules of syllables as table1. [1][2][3][4][5][6][7]

Table 1 rules

	rule	example
1	The syllable in Mandarin which is Yin rhyme and the rising tone and which consonant is unaspirated stop and affricate is an entering tone word.	"鼻", "直"
2	The syllable in Mandarin which vowel is "üe" is an entering tone except "que2" and "xue1" .	"yue", "nüe", "lüe", "jue", "que1", "que4", "xue2",

This work is sponsored by Independent scientific research plan from the Ministry of Education in China, Construction of Social Pragmatic Information network based on large scale of corpus. (20111081010)

		“xue3”
3	The syllable in Mandarin which vowel is “a” and initial consonant is “f” ,“z” or “c” is an entering tone.	“fa”, “za” ,“ca ”
4	The syllable in Mandarin which vowel is “uo” and initial consonant is “k”, “zh” ,“ch” “sh” or “r” is an entering tone.	“kuo”, “zhuo” , “chuo” ,“ shuo”, “ruo”
5	The syllable in Mandarin which vowel is “ie” and initial consonant is “b”, “p” ,“t” or “n” is an entering tone.	“bie”, “pie” , “tie” ,“ni e”
6	The syllable in Mandarin which vowel is “ei” and initial consonant is “g”, “h” or “z” is an entering tone.	“gei”, “hei” , “zei”
7	The syllable in Mandarin which ends in nasal consonant “n” or “ng” except ancient entering tone words is a non-entering tone.	
8	The syllable in Mandarin which initial consonant is “m”, “n” “l” or “r” and the tone is level tone, the rising tone, or the falling-rising tone except “ma1”, “mo1”, “mo2”, “mo3”, “ru3”, “la1”, “nie1”, “lei1”, and “luo1” is a non-entering tone.	
9	The syllable which vowel is “i” and initial consonant is “z” ,“c” or “s” is a non-entering tone.	
10	The syllable which vowel is “uei”, “uai” or “er” is a non-entering tone.	
11	The syllable which initial consonant is “zh” ,“ch”,“sh” or “r” except “尺” and “辱” and which tone is the falling-rising tone is not a entering tone word.	
12	The syllable which vowel is “ou” and initial consonant is not “zh” ,“ch”, “sh” or “r” is not a entering tone word.	

3 Maximum Entropy Model

Maximum entropy model (MEM) is used for disambiguating for the ambiguity syllables. Maximum entropy model is equation (1). [8][9][10][11]

$$P(y \mid x) = Z_\lambda(x) \exp(\sum_i \lambda_i f_i(x, y))$$
$$Z_\lambda(x) = \frac{1}{\sum_y \exp(\sum_i \lambda_i f_i(x, y))} \quad (1)$$

x stands for a pinyin sequence, y stands for a pinyin sequence with entering tone tags. For example, for “社会主义建设”， x=“she4 hui4 zhu3 yi4 jian4 she4”, y=“she4 hui4 zhu3 yi4 jian4 she4/r”

“/r” is entering tone tag, and it expresses that the syllable is an entering tone. So “she4” is an entering tone and others are not entering tones.

λ_i in Eq. (1) expresses the weight of f_i. The $Z_\lambda(x)$ expresses normalization factor.

4. Support Vector Machine

Support Vector Machine, or SVM was first introduced in 1992. It is used in chunking recognition, POS tagging and word segmentation. [12]

Given a pinyin sequence $S=p_1p_2p_3\ldots p_n$, each pinyin syllable p_i will have a tag: if the Chinese character of the syllable is an ancient entering tone character, then $f(p_i)=1$, and if not, $f(p_i)=-1$. Thus, the problem of entering tone syllable recognition turns to a binary classification one. Each pinyin falls into a class, either 1 or -1, i.e. either entering tone or non-entering tone.

For example: “计算语言学实验室”,

x=“ji4 suan4 yu3 yan2 xue2 shi2 yan4 shi4”,

y=“ji4/-1 suan4/-1 yu3/-1 yan2/-1 xue2/1 shi2/1 yan4/-1 shi4/1”

“xue2”(学), “shi2”(实) and “shi4”(室) in Chinese character, are of entering tones, so it should be tagged “+1”, and the others are not entering tones and should be tagged “-1”.

For a binary classification problem, given training set $S=\{(x_1,y_1),(x_2,y_2),\ldots(x_m,y_m)\}$, and

$x_i \in X=R^n$, $y_i \in Y=\{+1,-1\}$, i=1,2,…, SVM can learn a function $f(x)$, and then the decision function $f(x)$ can classify new input vector x.

For linear classifier, SVM would train the set, and compute the best hyperplane $g(x)=w^T x+b$, which makes the decision boundary as far away from the data of both classes as possible.

The algorithm[12][13][14][15][16][17]:

(1) Given training set

$S=\{(x_1,y_1),(x_2,y_2),\ldots(x_n,y_n)\}$,

$y_i \in \{1,-1\}$, $1<=i<=n$

(2) $K(x_i,x_j)=(x_i \bullet x_j+1)^d$ (2)

(3)Maximize

$$W(\alpha)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i,x_j) \quad (3)$$

and satisfies

$$\sum_{i=1}^{n} y_i\alpha_i = 0, \alpha_i \geq 0, i \in \{1,2,3,....,\ n\} \quad (4)$$

, Lagrange coefficient $\alpha=(\alpha_1,\ldots,\alpha_n)$ is obtained.

(4) The vectors x_i which satisfies (4) is a support vector.

(5) $$b=\frac{1}{2}\left[\omega_0 \cdot x^*(+1)+\omega_0 \cdot x^*(-1)\right] \quad (5)$$

here $x^*(+1)$ is a support vector of entering tones, $x^*(-1)$ is a support vector of non-entering tones.

$$\|\omega_0\|^2=2W(\alpha_0)=\sum \alpha_i^0 \alpha_j^0 y_i y_j (x_i \cdot x_j) \quad (6)$$

(6) Make prediction for the category of x' according to

$$f(x')=\sum_{SupportVector} y_i\alpha_i K(x_i,x')+b \quad (7)$$

$\alpha=(\alpha_1,\ldots,\alpha_n)$ is from (3) and b is from (5).

5. The Recognition Process

The process is as follows:

(1)The newspaper corpora are transcribed into pinyin sequences using Crystal TTS tool[1],

(2) The pinyin sequences are proofread.

(3) We utilized ancient entering tone Chinese character list to recognize the entering tones. The syllables which aligned Chinese characters are ancient entering tone characters are entering tones.

(4) Entering tones or non-entering tones are recognized according the rules in section 2.

(5) For other ambiguity syllables, maximum entropy model or support vector machine is respectively to recognize the entering tones.

6. Experimental results

Experimental corpus is People's Daily newspaper corpus of 1998. The corpus covers 1,608,121 pinyin syllables. There are 1570 kinds of syllables in the corpus. The entering tone list includes 120 pinyin syllables. There are 254 syllables in the ambiguous list. 36% of the corpus syllables are in the ambiguous list. 4.48% pinyin syllables are in the entering tone list, and they are tagged "entering tone" simply. 90% of the corpus is used as the training set and 10% is used as the testing set.

A series of features should be considered for each pinyin syllable. Two kinds of features are taken in consideration. One is the ambiguous syllable, and the other is its context syllables.

Precision P, accuracy A, recall R and F-measure F are used to evaluate experimental results of statistical method.

$$P=\frac{|D \cap B|}{|B|}\times 100\ \% \quad (8)$$

$$R=\frac{|D \cap B|}{|D|}\times 100\% \quad (9)$$

$$F=\frac{2PR}{P+R}\times 100\% \quad (10)$$

$$A=\frac{|(D \cap B)\cup(C-D\cup B)|}{|C|}\times 100\% \quad (11)$$

Suppose set C contains all testing syllables of statistical method, set D contains all entering tone syllables, and set B contains all entering tone syllables that are predicted by maximum entropy model or support vector machine.

First, rules are used for recognizing entering tones.

A=(104191/161637)*100%=64.46%

The accuracy using rules is 64.46%.

[1] Programmed by Cai LianHong et al, Department of Computer Science and Technology, Tsinghua University

Table 2 Results of different position and same span

		1	2	3
SVM	A	88.82%	90.55%	89.12%
	P	88.37%	88.81%	87.85%
	R	85.54%	89.57%	87.01%
	F	86.93%	89.19%	87.43%
MEM	A	89.80%	91.45%	89.70%
	P	86.79%	89.27%	87.21%
	R	89.45%	90.90%	88.88%
	F	88.10%	90.08%	88.04%

MEM is maximum entropy model, and SVM is support vector machine in table 2 and table 3.

Table 2 gives the results of same context span and different target syllable position. The context span is 3 and target syllable position is respectively 1, 2 or 3. If target syllable position is 1, the first syllable of context window is the target syllable. If target syllable position is 2, the target syllable is set in the center of the context window.

According to table 2, if target syllable position is 2, the result is the best of all for SVM and MEM respectively. If the target syllable position is 1, the result is the worst for SVM and MEM respectively. Thus, if the context span is set, it is better to make the center of context window as the target syllable.

According to table 2, if target syllable position is 1, the accuracy, the recall rate and F of MEM are higher than those of SVM. if target syllable position is 2, the accuracy, the precision, the recall rate and F of MEM are higher than those of SVM.

Table 3 The results of different span

		3	5	7
SVM	A	90.55%	91.49%	91.61%
	P	88.81%	90.55%	90.75%
	R	89.57%	89.80%	89.87%
	F	89.19%	90.17%	90.31%
MEM	A	91.45%	91.60%	91.61%
	P	89.27%	89.58%	89.43%
	R	90.90%	90.96%	91.11%
	F	90.08%	90.26%	90.26%

Table 3 gives the results of different context span. The target syllable is set in the center of the context window and the context span is respectively 3, 5 or 7. The larger the context span is, the higher the accuracy, the precision, the recall rate and F are for maximum entropy model and support vector machine respectively. When the context span is 7 the precision is 0.01% more than that when the context span is 5 and F keeps unchangeable for maximum entropy model. But the precision is improved 0.12% and F is improved 0.14% when the context span is improved for Support Vector Machine.

But when the context span is 3, the the accuracy, the precision, the recall rate and F of MEM are higher than those of SVM.. When the context span is 5 or 7, the accuracy and the recall rate of MEM are higher than those of SVM.

In all, the precision using rules is the lowest and the precision using maximum entropy model and support vector machine are much higher than rule-based method. The efficiency of maximum entropy model is higher than the efficiency of support vector machine. From the precision and efficiency, maximum entropy model is better than support vector machine.

7 Conclusion

Rules , maximum entropy model and support vector machine all help improving the accuracy, the precision, the recall rate and F of entering tone recognition. The context span and the target syllable position play an important role in entering tone recognition for maximum entropy model and support vector machine. In order to improve the accuracy, the precision, the recall rate and F further, we should take into account more rules, more large scale training corpus and more features.

References

[1] Huang Borong, and Liao Xudong, Modern Chinese (The third edition) (in Chinese), Higher education press, Beijing, 2004.

[2] Linguistics staff, Chinese Dialect Glossary,(the second edition) (in Chinese). Chinese press,Beijing,2003.

[3] Li Rong,(editor in chief), Qian CengYi. JiNan Dialect Dictionary (in Chinese). Jiangsu education press, Nanjing.1997.

[4] Li Ming, the Model and Conversion of Dialect Conversion System. Master dissertation (in Chinese). Tsinghua university, Beijing. 2008.

[5] Yang YanDong, the Study and Realization of ChaoShan Dialect Text to Speech (in Chinese). Master dissertation. ShanTou university, ShanTou. 2005.

[6] Jiao Dongmei, and Ya Mei, "Recognition Methods of Entering Tone Characters" (in Chinese), Journal of Shandong College of Education, Vol. 43, No. 3, 1994, pp. 46-51, 91.

[7] Zhang Zirong, and Chu Min, "A Statistical Approach for Grapheme to Phoneme Conversion in Chinese" (in Chinese), Journal of Chinese Information Processing, Vol. 16, No. 3, 2002, pp. 39-45.

[8] Darroch, J. N, D. Ratcliff. "General Iterative Scaling for Log-linear Models", The Annals of Mathematical Statistics, Vol. 43, pp. 1470-1480.

[9] Berger, Adam L, Stephen A. Della Pietra, Vincent J. Della Pietra, "A Maximum Entropy Approach to Natural Language Processing", Computational Linguistic, Vol.22, No.1,1996,pp.39-71.

[10] Ratnaparkhi, Maximum Entropy Models for Natural Language Ambiguity Resolution,1998.

[11]Adam Berger, Stephen Della Pietra and Vincent Della Pietra, "A maximum entropy approach to natural language processing", Computational Linguistics, Vol. 22, No.1, 1996.

[12] Vapnik, V. N. The Nature of Statistical Learning Theory. Berlin: Springer-Verlag, 1995.

[13] Kecman V., Learning and Soft Computing, Support Vector machines, Neural Networks and Fuzzy Logic Models, The MIT Press, Cambridge, MA, 2001.

[14] L.P. Wang (Ed.), Support Vector Machines: Theory and Application, Springer, Berlin, 2005.

[15] Lee, J., Lee, D., "An improved cluster labeling method for support vector clustering", IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 2005: pp 461-464.

[16] L.P. Wang, Fu X.J., Data Mining with Computational Intelligence, Springer, Berlin, 2005.

[17] Ramirez, L., Durdle, N.G., Raso, V.J., Hill, D.L., "A support vector machines classifier to assess the severity of idiopathic scoliosis from surface topography", IEEE Transactions on Information Technology in Biomedicine, Vol. 10, 2006: pp.84-91