Polyphonic Word Disambiguation with Machine Learning Approaches

Jinke LIU[1,2], Weiguang QU[1,2], Xuri TANG[3], Yizhe ZHANG[1,2], Yuxia Sun[1,2]
1.School of Computer Science, Nanjing Normal University, Nanjing,210046, China
2. The Research Center of Information Security and Confidentiality Technology of Jiangsu Province, Nanjing, 210097, China
3. School of Chinese Language and Literature, Nanjing Normal University, Nanjing, 210097, China
e-mail: lyliujinke@163.com, wgqu_nj@163.com

Abstract—**Five different classification models, namely RFR_SUM, CRFs, Maximum Entropy, SVM and Semantic Similarity Model, are employed for polyphonic disambiguation. Based on observation of the experiment outcome of these models, an additional ensemble method based on majority voting is proposed. The ensemble method obtains an average precision of 96.78%, which is much better than the results obtained in previous literatures.**

Keywords-*Polyphone disambiguation; Ensemble model; RFR_SUM; CRFs; Maximum Entropy; SVM; Semantic Similarity*

I. INTRODUCTION

Polyphony is one of the crucial problems in Chinese TTS systems which transform a sequence of characters into a sequence of Chinese Pinyin. In Chinese, polyphony is common. In some worst cases, one character may have up to five different types of pronunciation. For instance, the character “和” may be spoken in one of the following sound: hé、hè、hú、huó and huò[1]. According to [5], among the 10 most frequent characters, 6 of them are polyphones: “的”, “一”, “了”, “不”, “和”, “大”. Thus correct determination of how a character is read can improve TTS performance to a great extent.

Modern Chinese Dictionary[2] collects 1036 polyphonic characters and 580 polyphonic words. However, not all of them are frequently used. About 180 characters and 70 words take 95% and 97% of cumulative frequencies respectively in actual language use [1]. Among these 180 characters and 70 words, only 41 characters and 22 words are needed to extract because use frequency of their high frequency pronunciation is lower than 95%. To put it in another way, if the polyphonic ambiguity of these 41 characters and 22 words are successfully solved, the problem of polyphone ambiguity in Chinese should be basically solved. The present study is to approach these polyphones with machine-learning approach.

The choice of pronunciation of polyphones is determined by language convention and semantic content. There are currently two paradigms to approach the ambiguity: rule-based paradigm and statistics-based paradigm. Recent years have witnessed a growing number of researches on polyphone disambiguation with statistical machine learning. [1] proposes to use the ESC-based stochastic decision list to learn pronunciation rules for polyphones. In [2], polyphones are divided into two categories and disambiguate on POS level and semantic level separately. [3] presents a rule-based method of polyphone disambiguation, integrated with SVM-based weight estimation and [4] makes use of maximum entropy model to solve polyphone ambiguity.

This paper proposes an ensemble-learning approach for polyphonic disambiguation. The approach experiments with five machine learning models in polyphonic disambiguation and ensembles these five models with majority-voting to determine the final pronunciation of polyphones. The rest of the paper is organized as follows. Section II gives an overview of five models. In Section III, experiments with the five models and the ensemble learning are described in detail. IV compares the results obtained by ensemble model with related documents, followed by conclusions and plans for future work in Section V.

II. MACHINE LEARNING MODELS

A. *RFR_SUM Model*

Qu (2008) defines the concept of relative frequency ratio (RFR) and proposes the RFR_SUM model which disambiguates with context before and after the ambiguous word [6]. RFR of a word is the frequency ratio associated with relative position to the ambiguous word and is calculated between local frequency and global frequency. In RFR_SUM, the context is categorized into pre-context, the context before the word in question, and post-context, the context after the word in question. Thus the context of the word W_i in question can be characterized by the following formula:

$$SUM_m = \sum_{i=-l}^{-k} f_{m,left}(W_i) + \sum_{i=l}^{k} f_{m,right}(W_i)$$

Disambiguation can thus be done by comparing individual SUMs in different occurrences.

[1] These are Chinese Pinyin, annotated with tones.

[2] Institute of Linguistics of Chinese Academy of Social Sciences. Revised edition 3, 1996.7.

978-0-7695-4281-2/10 $26.00 © 2010 IEEE
DOI 10.1109/ICGEC.2010.67

B. *Conditional Random Fields*

Conditional Random Fields, presented firstly by Laferty (2001), is a conditional probability model used for tagging and partitioning sequence data. The model is an undirected graph that can calculate the conditional probability of output node based on the conditions of given input node. The tool kit adopted in our experiment is the CRF++ (version 0.50)[3] created by TakuKudo.

C. *Maximum Entropy Model*

Maximum entropy model is a method based on maximum entropy theory, in which the category with maximum entropy is selected as the optimum. The model has been applied in various fields of NLP, such as word segmentation, POS tagging and semantic disambiguation. In our experiment, the toolkit developed by Zhang Le is used.[4]

D. *Support Vector Machine*

Recent years have witnessed Support Vector Machine (SVM) as a prevailing machine learning tool applied in various fields. SVM is proposed by Vapnik in 1995 for pattern recognition, which seeks to find a hyper plane which has the largest distance to the nearest training data points, called support vectors, and thus best divides the two categories. We adopt libSVM implemented by Doctor Lin Chih-Jen of Taiwan University for experiment[5].

E. *Semantic Similarity Model*

Semantic similarity model calculates semantic similarity between sentences and then employs K nearest neighbor classifier to decide which category the polyphone should fall into. The tool for word similarity calculation is based on HowNet and the algorithm is proposed in [7] and [8]. Given two sentences SEN1 and SEN2, the procedure of disambiguation can be briefly described as below:

a. Given the polyphone word W and its position i and j in SEN1 and SEN2, and a window size N, four word set can be obtained: frontsen1= W_{i-1}^{i-N} , backsen1= W_{i+1}^{i+N} , frontsent2=W_{j-1}^{j-N} and backsen2=W_{j+1}^{j+N} .

b. Obtain front context semantic similarity FrontSim(frontsent1, frontsent2) and back context semantic similarity BackSim (backsent1, backsent2).

c. Obtain sentence semantic similarity：
SenSim = FrontSim(frontsent1, frontsent2)＋BackSim(backsent1, backsent2).

d. Apply K nearest neighbor classifier to classify W.

III. EXPERIMENTS AND ANALYSIS

A. *Experiment data and evaluation*

The experiment uses the 1998 People's Daily Corpus compiled by Peking University for experiment which contains half a year's newspaper of People's Daily in 1998. The corpus has a total of 13 million words. More than 40 polyphonic characters and 20 polyphonic words are selected from related literature for study. For each polyphone, 1600 sample sentences are retrieved from the corpus, 75% of which are used as training set and 25% of which as test set.

Due to space limitation, only 18 are selected in the paper for illustration. Table I gives the detailed experiment data, where Low stands for low-frequency, High for high-frequency, and B_L for baseline. Note that the average baseline is 74.23%. The results are evaluated by P, defined as follows:

$$P = \frac{\text{number of correct sentence}}{\text{number of overall sentence}} \qquad (1)$$

TABLE I. SAMPLE SENTENCES STATISTICAL

Word	Low	High	B_L
背	202	332	62.20
长	300	1396	82.31
重	232	693	74.92
得	512	735	69.65
干	90	1319	93.61
种	513	2208	81.15
倒	161	194	54.65
曾	412	2414	85.42
还	580	2765	82.66
只	754	2913	79.43
处	277	947	77.37
担	56	169	75.11
为	684	744	52.10
藏	113	121	51.71
合计	12	134	91.78
孙子	44	109	71.24
朝阳	85	138	61.88
地方	1281	1623	55.88

B. *RFR_SUM experiment results*

In RFR_SUM model, the relative frequency ratio of word, the relative frequency ratio of POS, and the relative frequency ratio of the combination of POS and word are independently experimented within a window of 5 for disambiguation. Table II gives the average precisions in the three experiments.

TABLE II. RESULTS OF RFR_SUM

Feature	Word	POS	Word&POS
P	87.75	92.44	94.65

C. *CRFs experiment results*

The advantage of CRFs is that new features can be added freely. Four feature templates and the results obtained

[3] Accessible at http://crfpp.sourceforge.net
[4] Accessible at http://homepages.inf.ed.ac.uk/s0450736/ME_toolkit.html
[5] Accessible at http://www.csie.ntu.edu.tw/~cjlin/libsvm

in the experiment are given in Table III. As is seen in Table III, Template 1 and Template 2 use word form and POS respectively. Template 4 uses the combination of word and POS in addition to word form and POS. However, the average precision drops 0.69% when Template 3 is used.

TABLE III. RESULTS OF CRFs

Template 1	Template 2	Template 3	Template 4
U1:%x[-2, 0]	U1:%x[-2, 1]	U1:%x[-2, 0]	U1:%x[-2,0]
U2:%x[-1, 0]	U2:%x[-1, 1]	U2:%x[-1, 0]	U2:%x[-1,0]
U3:%x[0,0]	U3:%x[1,1]	U3:%x[0,0]	U3:%x[0,0]
U4:%x[1,0]	U4:%x[2,1]	U4:%x[1,0]	U4:%x[1,0]
U5:%x[2,0]		U5:%x[2,0]	U5:%x[2,0]
		U6:%x[-2, 1]	U6:%x[-2,1]
		U7:%x[-1, 1]	U7:%x[-1,1]
		U8:%x[1,1]	U8:%x[1,1]
		U9:%x[2,1]	U9:%x[2,1]
			U10:%x[-2,0]/%x[-1,0]
			U11:%x[1,0]/%x[2,0]
			U12:%x[-2, -1]/%x[-1,-1]
			U13:%x[1, -1]/%x[2,-1]
92.46	94.14	95.27	94.58

D. *SVM experiment results*

In the experiment of SVM, the relative frequency ratio of word and POS in a window size of 4 is used as vector features. Table IV gives the results obtained through different kernel functions.

TABLE IV. RESULTS OF SVM

K_F	Linear	Polynomial	RBF	Sigmoid
P	90.33	90.47	91.86	89.94

In table IV, it can be seen that different kernel functions have an effect on the disambiguation result, but the difference is not significant. All kernel functions have a precision close to 90% and RBF kernel function has the best precision, reaching 91.04%.

E. *Semantic similarity model experiment results*

As the semantic similarity model uses K-nearest neighbor classifier for disambiguation, it is obvious that the parameter K will affect the final outcome. Table V lists out the experiment results for different Ks.

TABLE V. K VALUES AND EXPERIMENT RESULTS

K	3	4	5	6
P	90.56	91.23	91.05	90.37

F. *Ensemble learning*

Table VI lists out the best experiment results of 18 polyphones obtained with the five models discussed above, in which AVG stands for average precision of 60 polyphones obtained in the models.

TABLE VI. RESULTS OF FIVE MODELS AND MAJORITY VOTING

Word	S_S	SVM	M_E	RFR_ SUM	CRF	M_V
背	79.59	63.27	69.39	93.88	83.67	91.84
长	89.31	85.89	90.32	92.74	92.94	91.53
重	89.89	95.88	91.01	94.76	97.00	97.0
得	90.42	88.12	89.78	90.23	87.33	91.67
干	93.49	77.22	95.86	81.66	94.67	95.27
种	95.32	96.93	95.05	97.33	96.93	98.40
倒	79.63	85.53	80.26	97.76	89.47	94.74
曾	93.33	99.68	94.81	98.10	99.79	99.79
还	98.35	81.87	98.13	98.13	99.45	99.34
只	95.31	98.90	93.92	98.31	99.40	99.00
处	91.94	93.62	94.97	96.64	98.32	98.32
担	82.86	94.29	95.71	97.14	94.29	94.29
为	83.99	85.92	85.04	81.82	89.44	90.32
藏	75.33	76.67	78.33	93.33	90.00	96.67
合计	92.42	96.97	95.45	98.48	93.94	98.48
孙子	94.12	97.06	95.59	95.59	97.06	97.06
朝阳	91.04	91.04	85.07	94.03	97.01	98.51
地方	92.47	93.49	90.43	94.26	96.68	97.45
ARG	91.23	91.86	92.64	94.65	95.27	96.78

As can be seen in Table VI, results of five models (S_S, SVM, M_E, RFR_SUM and CRF) are diverse and complementary. For some polyphonic words, their precisions are probably very low in some models, but very high in some other models. For instance, precision of '背' is only 63.27% in SVM model, but goes up to 93.88% in RFR_SUM model.

The nature of insatiability in individual model and complementation among different models lead us to consider and adopt ensemble method for the final disambiguation. The principle adopted in the ensemble method is majority voting, namely the pronunciation which receives the most votes in the five models is chosen as the final pronunciation. The experiment result (M_V) given in Table VI obtains an average precision of 96.78%, showing that the ensemble effect is better than any single model, because it well ensembles advantages of every model and effectively eliminates the instability of individual model.

IV. COMPARISON

In order to evaluate the effect of ensemble method, the experiment results reported in [1] and [4] are selected to contrast the results obtained in ensemble method in the present study. The contrast is given in Figure 1 and Figure 2, where ARG is the average precision.

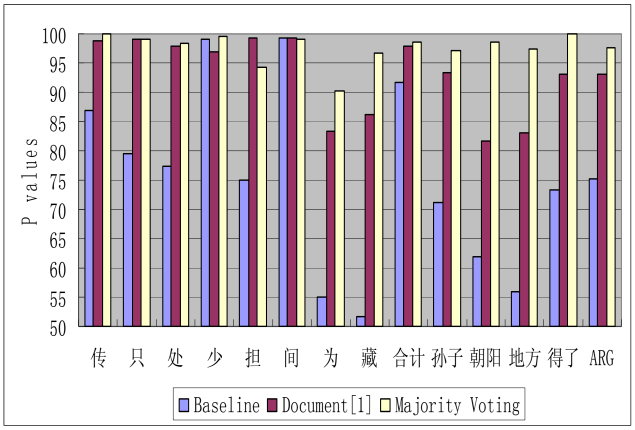

Figure 1. Comparison with [1]

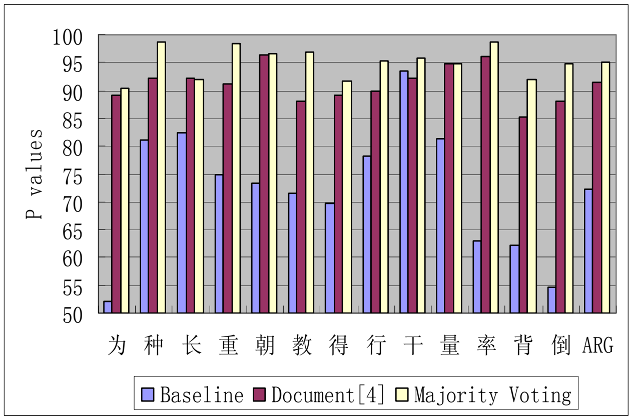

Figure 2. Comparison with [4]

Figure 1 and Figure 2 show that the average precision of ensemble method is 4.56% and 3.69% better than [1] and [4] respectively. Low precision polyphones such as '藏', '朝阳' and '地方' in [1] and [4] gain very high precision in majority voting method. High precision polyphones still keep high precision in ensemble method experiment. Obviously, ensemble model absorbs the merits of every model to improve low precision and keep high precision.

V. CONCLUSION

The present study has applied five different classification models for polyphonic disambiguation, namely RFR_SUM, CRFs, Maximum Entropy, SVM and Semantic Similarity Model. The experiments show that CRF and RFR_SUM perform better than the other three models. An ensemble method based on majority voting with the five models is also employed and reaches an average precision of 96.78% which is better than the results of [1] and [4], which shows that the ensemble model has more advantage than any individual model.

For the future work, we plan to conduct more experiments to obtain a better ensemble method, to combine RFR_SUM and Semantic Similarity Model so as to improve the instability of RFR_SUM in conditions of sparse data. And more knowledge will be introduced into the system[9].

ACKNOWLEDGMENT

This work is supported by Chinese National Fund of Natural Science under Grant 60773173 and 61073119.

REFERENCES

[1] Zi-Rong Zhang, Min Chu. A Statistical Approach for Grapheme-to-Phoneme Conversion in Chinese. Journal of Chinese Information Processing, 2002.

[2] Ming Fan, Guo-ping Hu. Multi-level Polyphone Disambiguation for Mandarin Grapheme-Phoneme Conversion. Computer Engineering and Applications, 2006.

[3] Guo-ping Hu, Zhi-gang Chen, Ren-hua Wang. A Rule-Based Approach with SVM-Based Weight Estimation for Phoneme Disambiguation of Polyphone. Proceedings of the 20th International Conference on Computer Processing of Oriental Languages, Shenyang, 2003.

[4] Fang-zhou Liu, Qin Shi. Maximum Entropy Based Homograph Disambiguation. Proceedings of the 9th National Conference on Human-Machine Language Communications, 2007.

[5] Yu-Qi Sun, Kai Zhang. Polyphone Study of Combination Based on rule and statistical. Proceedings of the 5th National Conference on Human-Machine Language Communications,1998.

[6] Wei-guang Qu. Disambiguation Study on Modern Chinese Word-Level Ambiguity. Beijing: Science Press, 2008.

[7] Qun Liu, Su-Jian Li. Word Similarity Computing Based on How-Net. The Third Symposium on Chinese Lexical Semantics, Taipei, 2002.

[8] Zheng-Dong Dong, Qiang Dong. How-Net. http://www.keenage.com.

[9]Yao Liu, Zhifang Sui, Qingliang Zhao,Yongwei Hu. Research on Construction of Medical Ontology. International Journal of Knowledge and Language Processing.2010,1(1):19-35.