Learning Rules for Chinese Prosodic Phrase Prediction

[†]Zhao Sheng [‡]Tao Jianhua [§]Cai Lianhong
Department of Computer Science and Technology
Tsinghua University, Beijing, 100084, China
[†]szhao00@mails.tsinghua.edu.cn {[‡]jhtao, [§]clh-dcs}@tsinghua.edu.cn

Abstract

This paper describes a rule-learning approach towards Chinese prosodic phrase prediction for TTS systems. Firstly, we prepared a speech corpus having about 3000 sentences and manually labelled the sentences with two-level prosodic structure. Secondly, candidate features related to prosodic phrasing and the corresponding prosodic boundary labels are extracted from the corpus text to establish an example database. A series of comparative experiments is conducted to figure out the most effective features from the candidates. Lastly, two typical rule learning algorithms (C4.5 and TBL) are applied on the example database to induce prediction rules. The paper also suggests general evaluation parameters for prosodic phrase prediction. With these parameters, our methods are compared with RNN and bigram based statistical methods on the same corpus. The experiments show that the automatic rule-learning approach can achieve better prediction accuracy than the non-rule based methods and yet retain the advantage of the simplicity and understandability of rule systems. Thus it is justified as an effective alternative to prosodic phrase prediction.

1 Introduction

Prosodic phrase prediction or prosodic phrasing plays an important role in improving the naturalness and intelligence of TTS systems. Linguistic research shows that the utterance produced by human is structured in a hierarchy of prosodic units, including phonological phrase, intonation phrase and utterance. (Abney, 1995) But the output of text analysis of TTS systems is often a structure of syntactic units, such as words or phrases, which are not equivalent to the prosodic ones. Therefore the object of prosodic phrasing is to map the syntactic structure into its prosodic counterpart.

A lot of methods have been introduced to predict prosodic phrase in English text such as Classification and Regression Tree (Wang and Hirschberg, 1992), Hidden Markov Model (Paul and Alan, 1998). For Chinese prosodic phrasing, the traditional method is based on handcrafted rules. Recurrent Neural Network (Ying and Shi, 2001) as well as POS bigram and CART based methods (Yao and Min, 2001) is also experimented recently. Due to the difference in training corpus and evaluation methods between researchers, these results are generally less comparable.

In this paper, a rule-learning approach is proposed to predict prosodic phrase in unrestricted Chinese text. Rule-based systems are simple and easy to understand. But handcrafted rules are usually difficult to construct, maintain and evaluate. Thus two typical rule-learning algorithms (C4.5 induction and transformation-based learning) are employed to automatically induce prediction rules from examples instead of human. Generally speaking, automatic rule-learning has two obvious advantages over the previous methods:

1) Statistical methods like bigram or HMM usually need large training corpus to avoid sparse data problem while rule-learning doesn't have the restriction. In the case of prosodic phrase prediction, the corpus with prosodic labelling is often relatively small. Rule-learning is just suitable for this task.
2) CART, RNN or other neural network methods have good learning ability but the learned knowledge is represented as trees or network weights, which are not so much understandable as rules.

Once rules are learned from examples, they can be analyzed by human to check if they agree with the common linguistic knowledge. We can add prediction rules converted from our linguistic knowledge to the rule set, which is especially useful when the training corpus doesn't cover wide enough phenomena of prosodic phrasing. Furthermore, we can try to interpret and understand rules learned by

machine so as to enrich our linguistic knowledge. Hence rule-learning also helps us mine knowledge from examples.

Since features related to prosodic phrasing come from various linguistic sources, several comparative experiments are conducted to select the most effective features from the candidates. The paper also suggests general evaluation parameters for prosodic phrase prediction. With these parameters, our methods are compared with RNN and bigram based statistical methods on the same corpus. The experiments show that the automatic rule-learning approach can achieve better prediction accuracy than the non-rule based methods and yet retain the advantage of the simplicity and understandability of rule systems. The paper proceeds as follows. Section 2 introduces the rule-learning algorithms we used. Section 3 describes prosodic phrase prediction and its evaluation parameters. Section 4 discusses the feature selection and rule-learning experiments in detail. Section 5 reports the evaluation results of rule based and none-rule based methods. Section 6 presents the conclusion and the view of future work.

2 Rule Learning Algorithms

Research on machine learning has concentrated in the main on inducing rules from unordered set of examples. And knowledge represented in a collection of rules is understandable and effective way to realize some kind of intelligence. C4.5 (Quinlan, 1986) and transformation-based learning (Brill, 1995) are typical rule-learning algorithms that have been applied to various NLP tasks such as part-of-speech tagging and named entity extraction etc.

Both algorithms are supervised learning and can be used to induce rules from examples. But they also have difference from each other. Firstly the C4.5 rule induction is a completely automatic process. What we need to do is to extract appropriate features for our problem. As to transformation-based learning (henceforth TBL), transformation rule templates, which determine the effectiveness of the acquired rules, have to be designed manually before learning. Thus TBL can only be viewed as a semi-automatic method. Secondly the induction of C4.5 rules using a divide-and-conquer strategy is much faster than the greedy searching for TBL ones. In view of the above facts, C4.5 rules are induced from examples first in our experiments. And then the rules are used to guide the design of rule templates for TBL. See section 4.8 for detail.

3 Prosodic Phrase Prediction

3.1 The Methodology

Linguistic research has suggested that Chinese utterance is also structured in a prosodic hierarchy, in which there are mainly three levels of prosodic units: prosodic word, prosodic phrase and intonation phrase (Li and Lin, 2000).. Figure 1 shows the prosodic structure of a Chinese sentence. In the tree structure, the non-leaf nodes are prosodic units and the leaves are syntactic words. A prosodic phrase is composed of several prosodic words, each of which in turn consists of several syntactic words. Since intonation phrase is usually indicated by punctuation marks, we only need to consider the prediction of prosodic word and phrase.

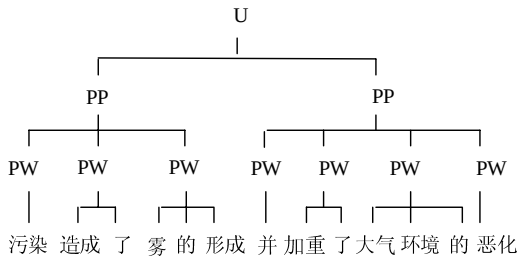

Figure 1: Two-level prosodic structure tree (U for intonation phrase, PP for prosodic phrase, PW for prosodic word)

Suppose we have a string of syntactic words i.e. $w_1, w_2, ... w_n$, the boundary between two neighbouring words is represented as $< w_i - w_{i+1} >$. There are total three types of boundaries labelled as B_0 (w_i, w_{i+1} are in the same prosodic word), B_1 (the words are in the same prosodic phrase, but not the same prosodic word), or B_2 (the words are in different prosodic phrases) respectively. Thus prosodic phrase prediction is to predict such boundary labels, which can be viewed as a classification task. We believe these labels are determined by the contextual linguistic information around the boundary. If we have a speech corpus with prosodic labelling, features related to prosodic phrasing can be extracted at each boundary and combined with the corresponding boundary labels to establish an example database. Then rule-learning

algorithms are executed on the database to induce rules for predicting boundary labels.

3.2 Evaluation Parameters

As a classification task, prosodic phrase prediction should be evaluated with consideration on all the classes. The rules induced from examples are applied on a test corpus to predict the label of each boundary. The predicted labels are compared with labels given by human, which are thought to be true, to get a confusion matrix as follows:

True labels	Predicted labels		
	B_0	B_1	B_2
B_0	C_{00}	C_{01}	C_{02}
B_1	C_{10}	C_{11}	C_{12}
B_2	C_{20}	C_{21}	C_{22}

Table 1: Confusion matrix

C_{ij}s are the counts of boundaries whose true label are B_i but predicted as B_j. From these counts, we can deduce the evaluation parameters for prosodic phrasing.

$$\mathrm{Re}\, c_i = C_{ii} / \sum_{j=0}^{2} C_{ij} \qquad (i = 0,1,2) \tag{1}$$

$$\mathrm{Pr}\, e_i = C_{ii} / \sum_{j=0}^{2} C_{ji} \qquad (i = 0,1,2) \tag{2}$$

$$F_i = 2 * \mathrm{Re}\, c_i * \mathrm{Pr}\, e_i / (\mathrm{Re}\, c_i + \mathrm{Pr}\, e_i)(i = 0,1,2) \tag{3}$$

$$Acc_1 = \sum_{i=0}^{2} C_{ii} / \sum_{j=0}^{2} \sum_{i=0}^{2} C_{ij} \tag{4}$$

$$Acc_2 = (\sum_{j=1}^{2} \sum_{i=1}^{2} C_{ij} + C_{00}) / \sum_{j=0}^{2} \sum_{i=0}^{2} C_{ij} \tag{5}$$

$\mathrm{Re}c_i$ defines the recall rate of boundary label B_i. $\mathrm{Pr}\, e_i$ defines the precision rate of B_i. F_i is a combination of recall and precision rate, suggested by (Rijsbergen, 1979). Acc_1 is the overall accuracy of all the labels. If we merge B_1 and B_2 into one label, which can be viewed as the prediction of prosodic word boundary, Acc_2 defines the overall accuracy of this case.

4 Experiments

4.1 The Corpus

In our experiments, the speech corpus of our TTS system is used for training and testing. The corpus has 3167 sentences, which are randomly selected from newspaper and read by a radiobroadcaster. We manually labelled the sentences with two-level prosodic structure by listening to the record speech. For example, the sentence in Figure 1 is labelled as "污染/B_1 造成/B_0 了/B_1 雾/B_0 的/B_0 形成/B_2 并/B_1 加重/B_0 了/B_1 大气/B_0 环境/B_0 的/B_1 恶化/B_2". Preliminary tests show that manually labelling can achieve a high consistency rate among human. Therefore it is reasonable to make the manually labelled results as the target of learning algorithms.

The sentences of the corpus are also processed with a text analyzer, where Chinese word segmentation and part-of-speech tagging are accomplished in one step using a statistical language model. The segmentation and tagging yields a gross accuracy rate over 94%. The output of the text analyzer is directly used as the training data of learning algorithms without correcting segmentation or tagging errors because we want to train classifiers with noisy data in the real situation.

Here are some statistical figures about the corpus. There are 56446 Chinese characters in the corpus, which constitute 37669 words. The number of prosodic word boundaries is 16194 and that of prosodic phrase ones is only 7231. The average length of syntactic word, prosodic word, prosodic phrase and sentence are 1.5, 2.4, 7.8 and 17.0 in character, respectively.

4.2 Candidate Features

Feature selection is crucial to the classification of prosodic boundary labels. Linguistic information around the word boundary is the main source of features. The features may come from different levels including syllable, word, phrase, sentence level. And the type of features may be phonetic, lexical, syntactic, semantic or pragmatic. Which features have most close relation with prosodic phrasing and how to represent them are still open research problems. In our approach, we decide to list all the possible features first and figure out the most effective ones by experiments. The features we currently consider are presented in the following.

4.2.1 Phonetic information

Chinese is well known as a monosyllabic, tonal language. And phonetic study shows sound will change in continuous speech because of context or prosodic structure. Retroflex, neutral tone and tone sandhi are important phonetic phenomena that cause sound variation. (Li and Lin, 2000). Thus phonetic information about phone and syllable is related to prosodic phasing. There are too many tonal syllables (about 1300) in Chinese to consider. Instead, the initials and finals of the syllables (total about 60) near a word boundary are taken into accounts, which are represented as *SYIF* in the following text.

Similarly the tones of the syllables, denoted by *TONE*, are also included as phonetic features.

4.2.2 Lexical information

Words in natural language have different occurrence frequency. And words that have high occurrence frequency may be especially important to prosodic phrasing (e.g. some functional words in Chinese, “的”, “和” etc). Therefore lexical word is treated as a candidate feature, represented as *WORD*.

4.2.3 Syntactic information

Syntactic information has close relation with prosodic structure. *POS*, which denotes part-of-speech of words, is a basic syntactic feature much easier to obtain with automatic POS taggers. And it has been widely adopted in previous researches. Since POS tag sets varies with taggers, we try to determine the best one for predicting prosodic phrase by experiments.

4.2.4 Other information

From the statistical figures of the corpus, both prosodic word and phrase have limitation in length. The length of syntactic word (*WLEN*), the length of the sentence in character (*SLENC*) and word (*SLENW*) are considered as length features. In HMM-based methods, the chain of boundary labels in a sentence is supposed to conform to Markov assumption. And according to experience, it is less possible for two boundaries with label B_2 to locate very close to each other. Thus the label of previous boundaries (*BTYPE*) and the distances from them to current position are also possible features.

4.3 Example Database

All of the possible features are extracted from the corpus at each boundary to establish an example database. Table 2 shows parts of the example entries of two word boundaries in Figure 1. Each row is a type of feature. The row name has a format of feature name plus a number. The number indicates which word the feature comes from. And the range of the number is limited by a window size. For example, *POS_0* denotes part-of-speech of the word just before the word boundary, *POS_-1* denotes that of the second word previous to the boundary and *POS_1* denotes that of the word just after the boundary. The rest may be deduced by analogy. *BTYPE_0* is the label of current boundary and also the target to be predicted.

Features \ Boundaries	<污染—造成>	<形成—并>
SYIF_0	an	eng
SYIF_1	z	b
TONE_0	3	2
TONE_1	4	4
WORD_0	污染	形成
WORD_1	造成	并
POS_0	vn	v
POS_1	v	c
POS_-1	w	u
WLEN_0	2	2
WLEN_1	2	1
BTYPE_0	B_1	B_2

Table 2: Example database entries

4.4 Feature Selection Experiments

Once the example database is established, we can begin to induce rules from it with rule learners. If all the features were used in one experiment, the feature space would get too large to learn rules quickly. Moreover we want to eliminate less significant features from the database. A series of comparative experiments is carried out to figure out the effective features. C4.5 learner is used to perform the learning task in the following experiments.

4.4.1 Baseline experiment (No.1)

Since *POS* features are widely used, a baseline experiment is performed with only two *POS* features that are *POS_0* and *POS_1*. The *POS* tag set has total 30 tags from the tagger.

4.4.2 POS window-size (No.2-9)

The window size determines the number of words whose features are considered. Suppose the window size is *L+R*, which means the features of *L* words left to the boundary and *R* words right to it are used. We design experiments with the combination of different value of *L* and *R* to find the best window of *POS* features. The features in the window are denoted by *POS{-L+1, R}* in a range form.

4.4.3 POS set (No.10-11)

Experiments are conducted on three *POS* sets, which are *BSET*, *LSET* and *CSET*. *BSET* is the basic *POS* set from the tagger. *LSET* is an enlarged version of *BSET*, which includes the most frequent 100 words as independent tags. *CSET* is built with clustering technique. Each *POS* in the *BSET* is represented as a 6-dimension vector, whose components are the probabilities of the boundary labels after and before that *POS*. Then these vectors are clustered into 10 groups. The window size used is 1+1.

4.4.4 *Other experiments (No.12-17)*

WORDLEN and *SLEN* are added into the baseline system to investigate the importance of length features in No.12 and 13. *SYIF*, *TONE* features of syllables around the boundary are considered in No.14. Previous boundary labels (*BTYPE_-1*, *BTYPE_-2*) are tested in the experiments No.15 and 16. *WORD* features are used in No.17 to find if there exist some words that have special prosodic effects.

No.	Features	POS tag set	F_0	F_1	F_2	Acc_1	Acc_2
1	*POS{0,1}*	*BSET*	0.69	0.72	0.76	0.72	0.79
2	*POS{0,0}*	*BSET*	0.57	0.53	0.14	0.50	0.64
3	*POS{-1,0}*	*BSET*	0.55	0.59	0.37	0.54	0.68
4	*POS{0,2}*	*BSET*	0.70	0.72	0.76	0.72	0.79
5	*POS{-1,1}*	*BSET*	0.71	0.71	0.76	0.72	0.79
6	*POS{-1,2}*	*BSET*	0.71	0.70	0.75	0.71	0.79
7	*POS{-2,1}*	*BSET*	0.71	0.70	0.75	0.71	0.79
8	*POS{-2,2}*	*BSET*	0.70	0.70	0.75	0.71	0.79
9	*POS{-3,3}*	*BSET*	0.71	0.70	0.74	0.71	0.79
10	*POS{0,1}*	*LSET*	0.72	0.74	0.77	0.74	0.81
11	*POS{0,1}*	*CSET*	0.67	0.67	0.73	0.68	0.75
12	*POS{0,1},WLEN{0,1}*	*BSET*	0.81	0.77	0.76	0.79	0.86
13	*POS{0,1},WLEN{0,1},SLEN*	*BSET*	0.82	0.76	0.74	0.78	0.87
14	*POS{0,1},TONE,SYIF*	*BSET*	0.71	0.72	0.75	0.72	0.79
15	*POS{0,1},BTYPE_-1*	*BSET*	0.75	0.74	0.76	0.75	0.82
16	*POS{0,1},BTYPE_{-1,-2}*	*BSET*	0.75	0.73	0.76	0.74	0.82
17	*POS{0,1},WORD{0,1}*	*BSET*	0.64	0.72	0.72	0.70	0.78

Table 3: Results of feature selection (F_0, F_1, F_2, Acc_1, Acc_2 are defined in section 3.2)

4.5 Feature selection results

The results of these experiments are listed in Table 3. From the evaluation figures in the table, we can draw the following conclusions on the effect of the features on prosodic phrase prediction:

1) Part-of-speech is a basic and useful feature. A window size of 2+1 is already enough. Larger window size will greatly lengthen the time of training but make no significant improvement on the accuracy rate.
2) The largest POS set *LSET* performs better than smaller ones like *BSET* and *CSET*. That's because small POS sets lead to small feature space, which may be not big enough to distinguish the training examples.
3) Length features are beneficial to prosodic phrase prediction.
4) Phonetic features are less useful than what we think before.
5) Former boundary information is also useful. When training, the former and latter boundary labels are both known, but when testing, exact former boundary labels do not exist. We can use the boundary labels that are already predicted to help make decision on current label. Although the error prediction of former labels may lead to error of current prediction, the result shows the accuracy rate is improved.
6) *WORD* feature is not appropriate to use, since the using of it greatly enlarges the feature space and needs more training examples.

4.6 C4.5 Experiments

According to the feature selection results, we know some features are effective to prosodic phrase prediction but some are not. And the solely using of effective features doesn't result in a high enough accuracy rate. In order to improve the prediction accuracy, we combine the effective features such as *WLEN{-1, 1}*, *BTYPE{-1}*, *SLEN* and *POS{-1,1}* in *LSET* tag set together to induce C4.5 rules.

4.7 Examples of C4.5 Rules

As mentioned above, rule systems have the advantage of simplicity and understandability. We examine the rules learned by C4.5 and find they certainly reflect the usage of prosodic structure in some sense. Here are some rules followed by example sentences with the current boundary labels in bold:

1) if POS_1 == 了 then BTYPE_0 = B_0
   我/B_0 去/B_1 **参观/B_0** 了/B_1 动物园/B_2
2) if POS_1 == 的 then BTYPE_0 = B_0
   学校/B_0 的/B_1 环境/B_1 不错/B_2

3) if POS_0 == 不 then BTYPE_0 = B_0
肚子/B_1 不/$\boldsymbol{B_0}$ 饿/B_2
4) if POS_0 == v && POS_1 == 于 then BTYPE_0 = B_0
他/B_1 生/$\boldsymbol{B_0}$ 于/$B_1$1998 年/B_2
5) if POS_1 == c && WLEN_0 > 2 then BTYPE_0 = B_2
游击队/$\boldsymbol{B_2}$ 并/B_1 没有/B_1 解散/B_2
她/ $\boldsymbol{B_1}$ 并/B_0 不/B_1 想/B_0 来/B_2
6) if POS_-1 == n && POS_0 == 是 && BTYPE_-1 == B_0 then BTYPE_0 = B_2
中国/B_0 是/$\boldsymbol{B_2}$ 伟大/B_0 的/B_0 国家/B_2

Rule 1, 2 and 3 shows the special prosodic effect of functional words such as "了", "的", "不", which tends to adhere to prosodic words in the sentences. Rule 4 exemplifies that the syntactic structure "Verb+于" usually acts as a prosodic word. Rule 5 concerns the conjunction word, the boundary before which would be B_2 (prosodic phrase boundary) if the previous word had a length above 2. The B_2 boundary is thought to accentuate the word before the conjunction. Rule 6 deals with the structure "Noun+是". We can see that these rules coincide with the experience of prosodic phrasing by human.

4.8 TBL Experiments

A general TBL toolkit (Grace and Radu, 2001) is used in our TBL experiments. The analysis on C4.5 rules casts lights on the design of the transformation rule templates of TBL. Since the same features as C4.5 learning are used in the rule templates, linguistic knowledge, which has been embodied by C4.5 rules, should also be captured by transformation rule templates. Suppose a C4.5 rule, "if (*POS_0* == n && *POS_1* == u) then *BTYPE_0* = B_0", has a high prediction accuracy, it is reasonable to make this rule as an instantiation of TBL rule templates. Table 4 lists some of the rule templates used in TBL experiments.

POS_0 POS_1 => BTYPE_0
POS_-1 POS_0 POS_1 => BTYPE_0
BTYPE _0 POS_0 POS_1 => BTYPE_0
BTYPE _0 POS_-1 POS_0 POS_1 => BTYPE_0
POS_0 POS_1 WLEN_0 WLEN_1=> BTYPE_0
WORD_0 POS_0 POS_1 => BTYPE_0
WORD_0 POS_-1 POS_0 POS_1 => BTYPE_0
BTYPE_0 WORD_0 POS_0 POS_1=>BTYPE_0
.....

Table 4: Rule templates for TBL

The left part of a rule template is a list of features, and the right is the target, *BTYPE_0*. For example, "*POS_0 POS_1 => BTYPE_0*", which is a short form of "if (*POS_0 == X* && *POS_1 == Y*) then *BTYPE_0 = Z*", means if current *POS* were *X* and the next *POS* were *Y*, the boundary label would be *Z*. *X, Y, Z* are template variables. Let *X*=n *Y*=u *Z*=B_0, the template is instantiated into the C4.5 rule above.

Due to the mechanism of TBL rules, there exist rule templates like "*BTYPE_0 POS_0 POS_1 => BTYPE_0*", in which the former *BTYPE_0* is the label before applying the rule and the latter is after applying it. That's actually what transformation means. When training, the initial boundary labels are all set to B_1. At each step, the algorithm tries all the possible values for template variables to find an instantiated rule that can achieve the best score. When testing, the initial boundary labels are set the same way, and then transformation rules are applied one by one.

5 Evaluation Results

To evaluate the generalization ability of the acquired rules, 5-fold cross validation tests are executed on the corpus for both C4.5 and TBL. We reimplemented the RNN algorithm and POS bigram statistical model to predict prosodic word boundary on the same corpus for comparison. Since our corpus is not large enough for HMM training and the CART method is also decision-tree based as C4.5, we didn't realize them in our experiments. The evaluation results are shown in Table 5.

Both the C4.5 rules and the TBL rules outperform the RNN algorithm and POS bigram method because the overall accuracy rates Acc_2 of the rule based methods are higher. TBL achieves comparable accuracy with C4.5 induction, which demonstrates that the design of transformation rule templates is successful.

Comparing Acc_1 and Acc_2 in Table 5, we discover that prosodic word boundaries can be more accurately predicted than prosodic phrase ones. It can be explained as follows. Prosodic word is the smallest prosodic unit in the prosodic hierarchy, which has more relation with the word level features such as POS, word length etc. Prosodic phrase is a larger prosodic unit less related to word level features, thus it cannot be predicted accurately using these features.

Tests	Rec_o	Pre_0	F_0	Rec_1	Pre_1	F_1	Rec_2	Pre_2	F_2	Acc_1	Acc_2
C4.5	0.914	0.837	0.874	0.814	0.822	0.818	0.712	0.829	0.766	0.829	0.904
TBL	0.849	0.884	0.866	0.782	0.848	0.814	0.851	0.613	0.713	0.818	0.895
bigram	0.653	0.746	0.696	0.874	0.816	0.844	N/A	N/A	N/A	N/A	0.793
RNN	0.764	0.803	0.783	0.883	0.857	0.870	N/A	N/A	N/A	N/A	0.837

Table 5: Evaluation results

6 Conclusion and Future Work

In this paper, we describe an effective approach to generate rules for Chinese prosodic phrase prediction. The main idea is to extract appropriate features from the linguistic information and to apply rule-learning algorithms to automatically induce rules for predicting prosodic boundary labels. C4.5 and TBL algorithms are experimented in our research. In order to find the most effective features, a series of feature selection experiments is conducted. The acquired rules achieve a best accuracy rate above 90% on test data and outperform the RNN and bigram based methods, which justifies rule-learning as an effective alternative to prosodic phrase prediction.

But the problem of prosodic phrase prediction is far from solved. The best accuracy rate got by machine is still much lower than that by human. In our future work, the study on this problem will go more deep and wide. Other machine learning methods will be experimented and compared with C4.5 and TBL. Features from deep syntactic, semantic or discourse information will be paid more attention to (Julia and Owen, 2001). And the speech corpus will be enlarged to cover more types of text and speaking styles.

Acknowledgements

Our work is sponsored by 863 Hi-Tech Research and Development Program of China (No: 2001AA114072). We also would like to thank the anonymous reviewers of the First SigHAN Workshop for their comments.

References

Abney Steven. (1995) Chunks and dependencies: bringing processing evidence to bear on syntax. Computational Linguistics and Foundations of Linguistic Theory, CSLI.

Eric Brill. (1995) Transformation-Based Error –Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging. Computational Linguistics 21(4):543- 565.

C.J. van Rijsbergen. (1979) Information Retrieval. Butterworths, London.

Grace Ngai and Radu Florian. (2001) Transformation-Based Learning in the Fast Lane. Proceedings of the 39th ACL Conference.

Julia Hirschberg, Owen Rambow. (2001) Learning Prosodic Features using a Tree Representation. Eruospeech2001.

Li Aijun, Lin Maocan. (2000) Speech corpus of Chinese discourse and the phonetic research. ICSLP2000.

Michelle Wang and Julia Hirschberg. (1992) Automatic classication of intonational phrase boundaries. Computer Speech and Language 6:175–196.

Paul Taylor and Alan W Black. (1998) Assigning phrase breaks from part-of-speech sequences. Computer Speech and Language v12.

Quinlan,J.R. (1986) Induction of decision trees. Machine Learning, 1(1):81-106.

Yao Qian, Min Chu, Hu Peng. (2001) Segmenting unrestricted chinese text into prosodic words instead of lexical words. ICASSP2001.

Zhiwei Ying and Xiaohua Shi. (2001) An RNN-based algorithm to detect prosodic phrase for Chinese TTS. ICASSP2001.