

The Contribution of Mutual Information in the Intonational Phrase Prediction in Chinese Text

Guo-Ping HU Ben-Feng CHEN Ming FAN Ren-Hua WANG

University of Science and Technology of China, Hefei, 230027
Email: applecor@mail.ustc.edu.cn

Abstract

The contribution of Mutual Information (MI) in the Intonational Phrase (IP) prediction is analyzed and verified in this paper. The basic idea of employing MI in IP prediction is that people is likely to pause between the less correlated words, where the MI value is low. The paper presents a decision tree based predictor which adopts POS as the main feature firstly as the baseline, and then the paper analyzes the correlation between MI and IP. The approach which only bases on the MI to predict the IP boundary is demonstrated in this paper, and three methods combining the MI and POS in the predictor is presented too. In the MI based approach, a considerable performance (F-Score: 64.2%) is achieved, and 3.4% promotion from the baseline is achieved after combining MI and POS in our experiment. All our work indicates that MI is an effective feature in the prosodic phrase boundaries prediction in Chinese text, and combining the MI and POS in the predictor is valuable.

Keywords: Intonational Phrase, Mutual Information, Prosody Boundary, Part-Of-Speech

1. Introduction

The essential goal of text analysis in a Text-to-Speech system is to extract abundant prosodic information from text in order to improve the naturalness and intelligibility of the speech output (Chu and Lu 1996). And intonational phrase (IP) boundary is an important part of the prosodic information that should be extracted from text for speech synthesis module including accent assignment, duration control and pause insertion.

There are many experiments which prove that prosody structures are not always identical to syntactic structures (Gee and Grosjean.1983) (Selkirk1984). So in recent studies, stochastic models have been used more frequently (Wang and Hirschberg 1991) (Hirschberg and Prieto 1996) (Lee and Oh 1999) and achieve very reasonable results.

These researches on prosodic phrase prediction are mostly focus on developing a more effective model based on part-of-speech (POS) feature. However, some researches have demonstrated that it is more serious to find new features but the training algorithm in IP boundary detection (Zhao, Tao and Cai 2002).

On the other hand, a word carries rich information related to phonetic feature. And mutual information (MI) is a word-based feature and is successfully applied in various areas on Natural Language Processing. Based on the hypothesis that people like to pause between the less correlated words, where the MI value is lower, we believe that MI is useful in the IP prediction and so carry out following experiments and this paper.

The rest of this paper is organized as following. Decision tree method based on POS feature is first introduced in section 2, and then try to predict the IP with the mutual information alone in section 3. In section 4, the combined model for POS and MI were presented, and the results will be presented in section 5. And we draw our conclusion and do some discussion in section 6.

2. Predict IP based on Part-of-Speech and Decision Tree Method

Stochastic models are the traditional approach in IP prediction. We describe this approach in our experiments first.

2.1. The Base Model

We employ the C4.5 decision tree algorithm as the base model of the predictor. The predictor considers the circumstance feature to decide each of the word boundaries should be IP boundary or not. And the analyzed circumstance features include:

1. POS of the surround words, three before the word boundary and two after.
2. Syllable numbers of surrounding words, three before

the word boundary and two after.

3. Word unigram(Frequency of word), three before the word boundary and two after.
4. The position of the word boundary in the sentence.

For the POS is the most significant feature in the feature set, we just let POS denote the above four features in the rest of the paper.

2.2. Smoothing

The basic model doesn't take into account the phenomenon that the IP should not be too short and should not be too long. So smoothing technique is needed for the base model.

So the reliability of the decision of the C45 is treated as the probability of the word boundary to be IP boundary, we called this target cost. Another function called link cost is also employed which derived from the distribution of the length of IP in trigram. DFS algorithm was implemented for the smoothing technology. And this smoothing is proved be we calculate the probability of one sentence

3. Predict Intonational Phrase based on Mutual Information

3.1. Original Mutual Information Definition

Mutual Information was reported in (Fano 1961) as a measure of the interdependence of two signals in a message. This bigram mutual information is a function of the probabilities of two events:

$$MI(x, y) = \log_2 \frac{P(x, y)}{P(x)P(y)}$$

In our application, we consider the events x, y as words in sequence in a sentence. Hence P(x), P(y) is the frequency of the word x, y in the statistical corpus. And MI (x, y) represents the correlation between word x and y.

3.2. Preprocessing MI value

When we just check the MI value of the some sentences, we find that the MI(x, y) value is quite low in the some conditions where should not be IP boundary, such as:

- 1) word x is numeral and word y is a quantifier;
- 2) word y is “了, 着, 的”;
- 3) word x is “该”, and word y is noun;

Because the numeral word has not been clustered into one non-terminal symbol, and the “了, 着, 的” have too high frequency which make the MI value low, so we adopt one rule-based preprocessing, where the rule just describes the condition like above. We replace the MI value with a large value (here it is 4) if a rule is matched. We totally collect 2903 rules from one corpus annotated with IP boundaries, and about 29.3% MI values are replaced, that means the MI value will take 71.7% of the IP prediction task

Following is an example sentence:

全市(1.73)国有工业(2.45)企业(0.30) || 实现(0.27)了(-0.32) 销售收入(1.95) || 七点九亿(1.63)元

(The national industries of the whole city gain 79 million Yuan as revenue in sale.)

The numbers in the brackets is the MI value of the two words beside, and the || denoted the IP boundaries.

The result of preprocessing is showed as following:

全市(1.73)国有工业(2.45)企业(0.30) || 实现(4) 了(-0.32) 销售收入(1.95) || 七点九亿(4)元

3.3. Correlation between MI and IP

In this study, polynomial fitness is applied to test the correlation of the mutual information and IP. And the fit target is set to 1 if the word boundary is an IP boundary, and 0 to the non IP boundaries. We consider the MI value after each Chinese character as input argument for fitting. If the character is the last character of one word, the MI is copied from the MI of the word boundary; else the MI of the character is set as one high value (such as 4.00). Following is the MI of each character in the sample sentence according above definition.

全(4.00) 市(1.73) 国(4.00) 有(4.00) 工(4.00) 业(2.45) 企(4.00) 业(0.30) 实(4.00) 现(4) 了(-0.32) 销(4.00) 售(4.00) 收(4.00) 入(1.95) 七(4.00) 点(4.00) 九(4.00) 亿(4) 元

Table 1 shows the fitting result. From the result, we can see that the MI has some significant relationship with IP boundary, and also the preprocessing contributes great in the fitting. And also the consideration of surround MI value is also useful.

3.4. Predict IP based on Mutual Information Only

Based on the above discussion, it is easy to predict the IP based on the MI. We just treat the fitting value as the possibility of the word boundary to be an IP boundary.

Note that the smoothing algorithm is also employed to get more rational IP prediction.

Table 1: polynomial fitting result from MI to IP boundary

MI value	Considered MI value	Multiple R	MS Residual
Original MI value	Left: 0, Mid: 1, Right: 0	0.195258	0.496413
	Left:7, Mid:1, Right:7	0.399134	0.434098
Preprocessed MI value	Left: 0, Mid: 1, Right: 0	0.501728	0.386319
	Left:7, Mid:1, Right:7	0.563018	0.352704

Multiple R means the multiple correlation coefficients of the IP boundary and the considered MI value

MS Residual means the fit error in average.

4. Combines POS and MI to predict IP

Both POS and MI are proved to be useful information in the IP prediction, and how to combine them to get a more powerful predictor is a problem now.

There are many approaches for this combination, we now just present the two ways which we have tried.

4.1. Take MI as a feature in Decision Tree

One easy way is to add the MI value into the Decision Tree model as one input features, we consider 5 MI value around the word boundary, left:2, mid:1, and right:2.

4.2. Integrate MI into target cost

The Second approach is the weighted-sum. At each word boundary, we set the probability ($P_{Combined}$) of being IP boundary as:

$$P_{Combined} = P_{POS}^w * P_{MI}^{1-w}$$

And then send the $P_{Combined}$ as the target cost to the smoothing stage, and get the prediction result with the DFS algorithm too.

4.3. Distinguish the competent condition of POS and MI

Integrating MI into target cost is more effective combining method, with which we can distinguish the condition where the POS can predict the IP boundary with high precision, and where the MI is more competent. Simply we can just describe the condition as the bigram of Part-Of-Speech. Following are some statistical result:

Table 2: The competent condition of POS and MI

POS before word boundary	POS after word boundary	Occurred Count	POS Accuracy	MI Accuracy
Verb	Verb	4766	0.854214	0.873689
Noun	Noun	12023	0.928987	0.939782
Noun	Verb	4698	0.876391	0.744785
Verb	Noun	4369	0.814112	0.726024
Noun	Adverb	2193	0.923196	0.786138
Auxiliary word	Noun	3250	0.877756	0.866769
All kind	All kind	101730	0.929356	0.884921

From table 2, POS is proved to more powerful in the IP prediction in most of the condition and averagely. But MI is also proved to get higher accuracy as the condition of verb-verb or noun-noun. And so we could modify the weight of the $P_{Combined}$ to let the competent method bear more weight in the prediction, just as following formula:

$$P_{Combined} = \begin{cases} P_{POS}^{w1} * P_{MI}^{1-w1} & \text{If in POS's more competent condition} \\ P_{POS}^{w2} * P_{MI}^{1-w2} & \text{In other condition} \end{cases}$$

$w1, w2$ are two argument need optimization, and $w1$ should larger than $w2$.

5. Experiments and Results

In this section, we will demonstrate the performance of our prediction model based on POS and MI

5.1. Corpus

Our Corpus for mutual information statistics is collected from the Chinese newspaper People's Daily, and the size of the corpus is 372M. The language model is automatically generated by the CMU-Cambridge Statistical Language Model Toolkit (Clarkson 1997). Good-Turing and back-off smoothing is applied for estimation in our model (Church and Gale 1991) (Katz, 1987).

To test our algorithm, 15580 sentences are selected from the People's Daily, and each sentence is labeled with IP boundaries manually by listening to their utterances and reading the text transcriptions. The sentence length, counted in Chinese characters, varies from 7 to 20. The distribution of sentence length is shown in figure 1.

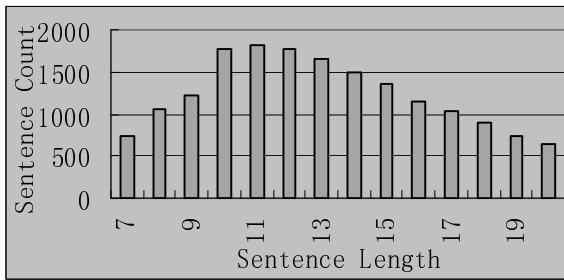


Figure 1. Distribution of sentence length in testing corpus

5.2. Evaluation criteria

There are various measures for evaluating the performance of prosodic phrasing parser. Here, what we really care about is the performance of predicting the phrase boundaries. We evaluate the performance of our model by following features: Precision (P), Recall Rate(R) and f-score (F), which are defined as:

$$P = \frac{N1}{N2}, R = \frac{N1}{N3}, F = \frac{2PR}{P+R}$$

Here N1 represents the correct predicted boundary count, N2 is the total predicted boundary count, and the N3 means the count of total boundary occur in the annotated testing corpus.

Since human can speak the sample sentence in different ways, there is not only one result for one sentence. So manually labeled phrase boundaries are used as a reference for evaluating the results obtained with automatic method. The most confident evaluation is the average score given by the person, such as the MOS evaluation in the speech naturalness. So the human evaluation is also performed in our comparison. There are 5 grades that can be assigned to each sentence: 5-excellent, 4-good, 3-acceptable, 2-bad and 1-worst.

5.3. Results

We split the 15580 sentences into two parts, 14580 sentences as the training set, and 1000 sentences as the testing set. All the training work (including decision tree training, smoothing argument estimation and fit parameter estimation are all done in the training set, and the w, w1, w2 are also optimized in the training set by enumerate each 0.05 interval from 0 to 1). Table 3 presents the experiments result of different MI-based model tested in the test set.

From the result, we can see that the MI-based model achieves considerable performance. Furthermore, when the two models are combined, a rather significant improvement would be obtained in both f-score and human evaluation. It must be noted that, here we only combine them in a simple

way- just add up their predicted result with weight. We believe that higher performance can be achieved with more effective approaches.

Table 3: Comparative Result of MI added into Decision Tree

Feature Set	Precision	Recall	F-Score	Human Evaluation
POS + Decision Tree Without smoothing	73.5%	49.5%	59.1%	3.10
POS + Decision Tree	64.6%	70.4%	67.4%	3.96
MI only	56.0%	75.2%	64.2%	3.84
POS and add MI in Decision Tree	64.3%	70.4%	67.2%	3.94
weighted-sum of POS and MI (w = 0.3)	65.8%	75.6%	70.4%	4.08
Distinguish condition of POS and MI (w1= 0.35, w2 = 0.25)	67.6%	74.3%	70.8%	4.10

6. Conclusion and Discussion

Now we can draw our conclusion: mutual information is proved to be an effective feature for automatic prediction of intonational phrase boundaries in Chinese text, and it can help the traditional stochastic model to obtain higher performance. The combined model of POS and MI can achieve the highest performance in our experiments.

However, we only implemented both the algorithm and the combination of MI with POS in a simple way. A more complex model may be developed to get better result or to predict other phonetic features from text, such as a hierarchical model, using n-gram mutual information or combining mutual information with syntactic parsing result. Thus, further research may be done to investigate into the relationship between mutual information and prosodic structure. A proper training and iterating algorithm may reveal more interesting information for prediction model. We hope our study may give some benefits to the other researches on automatic prediction of prosodic phrase, which remains an open task in the Chinese speech synthesis.

References

- [1]. Chu, M and Lu, S.N "A Text-to-Speech System with High Intelligibility and High Naturalness for

- Chinese ", *Chinese Journal of Acoustics*, Vol.15, No.1, 1996, pp.81-9
- [2]. Gee,j.P. and F.Grosjean. 1983. "Performance structures: a psycholinguistic and linguistic appraisal". *Cognitive Psychology*, 15,411 – 458
- [3]. Selkirk,E.O. 1984. "Phonology and syntax: the relation between sound and structure". Cambridge, MA: MIT Press.
- [4]. Shimei Pan, IBM Research Report: Exploring Features from Natural Language Generation for Prosody Modeling, *Computer Speech and Language*, vol. 16(3-4):457-90, July 2002
- [5]. Wang, M.Q and Hirschberg,j., "Predicting Intonational phrasing from text", *Proceeding of ACL*,1991,pp.285-292
- [6]. Hirschberg,j. and Prieto,P., "Traning Intonational phrasing rules automatically for English and Spanish text-to speech", *Speech Communication*, Vol,18,1996,pp.281-290.
- [7]. Lee,S and Oh, Y.H., "Tree-based modeling of prosodic phrasing and segmental duration for Korean TTS systems", *Speech Communication*, Vol.28,1999,pp.283-300
- [8]. Guo-Ping HU, Ben-Feng CHEN, "Developing Chinese TAK for Computer directly", *ISCSLP2002*
- [9]. Ben-Feng CHEN, Guo-Ping HU, "Large Lexicon Construction for TTS System", *ISCSLP2002*
- [10]. Veilens,N.M, Ostendorf M.,Price, P.J. and Shattuch-hufnagel, S., "Markov Modeling of Prosodic Phrase structure", *Preceeding of the 1990 International Conference on Acoustics, Speech and Signal Processing*, Vol.2, 1990,pp. 777-780
- [11]. Taylor,P. and Black, A.W., "Assigning phrasing breaks from part-of-Speech sequences", *Computer speech and language*, Vol.12, 1998,pp.99-117
- [12]. Ostendorf,M. and Veilleux,N., "A hierarchical stochastic model for automatic predctions of prosodic boundary locations", *Computatinal Linguistics*, Vol.20, No.1, 1994,pp.27-54 Sanders, E., Tarlor, P. "Using statistical methods to prdict phrase boundaries for speech synthesis". *Proc. EUROSPEECH*. 1995
- [13]. Shen,X. and Xu,B., "A CART based hierarchical stochastic model for prosodic phrasing in Chinese", *Proceeding of the 2nd International Symposium on Chinese Language Processing*, 2000, Beijing.
- [14]. M,Chu, Y,Qian, "Locating Boundaries for Prosodic Constituents in Unrestricted Mandarin Texts", *Computational Linguistics and Chinese Language Processing*, Vol.6, No.1, February 2001,pp. 61-82
- [15]. F. Jelinek (1990). "Self-organized language modeling for speech recognition", in *Reading in Speech Recognition*, Alex Waibel and Kai-Fu Lee (eds.), Morgan-Kaufmann, San Mateo, CA, pp.450-506
- [16]. Feng Chien. "PAT-Tree-Based adaptive keyword extraction for Chinese information retrieval", *SIGIR* (1997)
- [17]. Chuch,K., Hanks, P. 1989. "Word Association Norms, Mutual Informaiton, and Lexicography". In *Proceeding of the 27th ACL*.
- [18]. David, M. Magerman and Mitchel P.Macus. "Parsing a natural language using mutual information statistics". *Proceeding of AAAI-90*, 984-989. American Association for Artificial Intelligence.
- [19]. K T Lua, "Experiments on the Use of Bigram Mutual Information In Chinese Natural Language Processing", *International Coference on Computer Processing of Oriental Languages (ICCPOL)*, Nov,1995, Hawaii
- [20]. Fano,R. 1961. *Transmission of Information*. NewYork,New York: MIT Press
- [21]. Wu Ming-Wen and Su Keh-Yih. "Corpus-based automatic compound extraction with mutual information and relative frequency count", *Proceedings of R. O. C. Computational Linguistics Conference VI, Taiwan, ROCLING-VI*, 1993. 207-216.
- [22]. Kim-Teng Lua and Kok-Wee Gan. "An Application of Information Theory in Chinese Word Segmentation". *Computer Processing of Chinese & Oriental Languages*, vol.8,no.1 (June,1994): 115-124
- [23]. Calzolori,N.(1990),"Acquisition of lexical information from a large Textual Italian Corpus." *Proc. of COLING-90*, Vol2,54 – 59.
- [24]. Clakson,P., and Rosenfeld, R. "Statistical Language Modeling using the CMU-Cambridge Toolkit", *Proceedings of Eurospeech*, Vol 5,pp. 2707-2710, 1997
- [25]. K.Church and W.Gale. "A comparison of the enhanced Good-Turing and deleted estimation methods of estimating probabilities of English bigrams". *Computer Speech and Language*, 5:19-54, 1991
- [26]. S.Katz. "Estimation probabilities from sparse data for the language model component of a speech recognizer". *IEEE Transactions on Acoustics,Speech and Signal Processing*. ASSP-35: 400-401, 1987
- [27]. E.Fitzpartrick and J.Bchenko. "Parsing for prosody: What a Text-to-Speech System Needs from Syntax". *IEEE*,1989
- [28]. Shimei Pan and Kathleen Mckeown, "Learning Intonation Rules fro Concept to Speech Generation", *COLING-ACL98*, 1998.

- [29]. Wang,H.J., Chinese non-linear phonology (In Chinese),Peiking Univerisity Press, Beijing, 1999,pp. 229-281