Active Learning for Prediction of Prosodic Word Boundaries in Chinese TTS Using Maximum Entropy Markov Model

Ziping Zhao[1,2]
[1]Key Laboratory of Trustworthy Computing, Shanghai, East China Normal University
[2]College of Computer and Information Engineering, Tianjin Normal University, Tianjin, China
Email: gign_2001@yahoo.com.cn

Xirong Ma[1]
[1]College of Computer and Information Engineering, Tianjin Normal University, Tianjin, China
Email: maxirong2006@gmail.com

***Abstract*—For a Chinese speech synthesis system, hierarchical prosody structure generation is a key component. The prosodic word, which is the basic prosodic unit, plays an important role in the naturalness and intelligibility of Chinese Text-To-Speech system. However, obtaining human annotations of prosodic words to train a supervised system can be a laborious and costly effort. To overcome this, we explore active learning techniques with the goal to reduce the amount of human-annotated data needed to attain a given level of performance. In this paper Active Maximum Entropy Markov Model(AMEMM) is used to predict Chinese prosodic word boundaries in unrestricted Chinese text. Experiments show that for most of the cases considered, active selection strategies for labeling prosodic word boundaries are as good as or exceed the performance of random data selection.**

***Index Terms*—Prosodic Word, Text-to-Speech System (TTS), Active Learning, Maximum Entropy Markov Model**

I. INTRODUCTION

In Mandarin speech, the prosodic word is the basic rhythmic unit rather than lexical word. In real speech, prosodic words should be uttered continuously and closely without breaks, which play an important role in the naturalness and intelligibility of Chinese Text-to-Speech (TTS) system. Prediction of prosodic word from the text has become a key component in the prosodic analysis module of the TTS system. The Experiments show that using the prosodic word as the basic prosodic units improves naturalness over using lexical words[1]. Because the prosodic words greatly influence the rhythm of synthetic utterances, proper prediction of the prosodic word boundaries will directly affect the naturalness and correctness of TTS directly.

At present, from existing research in the field there have been some effective methods put forward. The existing methods for prosodic words prediction fall into two categories.

Earlier work usually adopted rule-based methods, beginning from Gee and Grosjean's work on performance structures [2]. Jianfen Cao [3,4] and Hongjun Wang [5] have also carried out the similar investigation for Chinese. The common idea of all these methods is to find some rules that could recreate the prosodic structure of a sentence from syntax, by way of a large number of experiments and empirical observation. The method is easily explicable and understandable, but it has its limitations. It poses strict demand for the system developer to summarize these rules. Moreover, it is hard to update and improve in practical applications, and the set of rules is domain specific, which hinders its general applicability [6].

With the rapid development of statistical machine learning, machine learning approaches have been more and more widely investigated for prosodic boundary prediction. Many different statistical methods have been tried, including Classification and Regression Tree (CART) used by Wang and Hirschberg [7], and Hidden Markov Model proposed by Paul and Alan [8]. Zhao has described methods for automatically predicting prosodic phrase by combining decision tree and TBL [9]. In Li's experiment, he attempted to predict prosody phrase break based on Maximum Entropy(ME) Model [10]. In [11], a statistical model based on word length, part of speech (POS) and current word was introduced for prosodic word tagging. In this model, each prosodic word boundary will not affect the next boundary. In [12], a SVM based method was proposed. An HMM based statistical method for prosodic word prediction was used in [13]. In [14], Zhao has described methods for automatically predicting prosodic word by combining MEMM [15] and TBL [16].

However, automatically predicting prosodic word boundaries with high precision and recall ratio requires a large amount of hand-annotated data, which is expensive to obtain. Meanwhile unlabeled data may be relatively easy to collect, but there have been few ways to use them. Active learning overcomes this problem by using large

doi:10.4304/jsw.8.12.3222-3228

amounts of unlabeled data together with the labeled data, to build better classifiers.

In this paper, we propose an alternative active learning strategy based on MEMM [15] for the prosodic word boundary prediction task. Without large-scale labeled data, the proposed method greatly reduces the training time and gets similar or better results when compared to the conventional supervised learning model.

Active learning has been studied in the context of many natural language processing (NLP) applications such as information extraction [17,18], text classification [19,20], word segmentation [21-23], named entity recognition(NER) [24] and chunking [25]. Active learning has also been applied to support-vector machines [26,27]. In the language processing framework, uncertainty-based methods have been used for automatic speech recognition [28]. To the best of our knowledge, active learning has not been used for prosodic word prediction.

The paper unfolds as follows: Section 2 describes MEMM; the principles and mathematical representation of MEMM are introduced. Section 3 presents the MEMM based method to predict the prosodic word boundary in detail. Section 4 gives the description of the active learning model. Section 5 gives the evaluations on each method. And the experiment results and discussion are made in Section 6. Section 7 presents the conclusion and the view of future work.

II. MAXIMUM ENTROPY MARKOV MODEL

The Hidden Markov model (HMM) is a powerful tool for predicting sequential data, and has achieved great success in the last decade.

However, HMM assumes that features are independent. As a generative model, HMM defines a joint distribution over label and observation sequences means that all possible observation sequences must be enumerated; as a result, richer features are not easily added. The second problem is that it sets the HMM parameters to maximize the likelihood of the observation sequence; however, it is inappropriately uses a generative joint model to solve a conditional problem in which the observations are given.

Compared with HMM, maximum entropy Markov model(MEMM) and other discriminative finite-state models can easily use more features. As an alternative to HMM, we offer MEMM to address two HMM problems. First, MEMM maximizes the conditional probability of the sequential data rather than the joint probability, as HMM does. As sequence labeling task is usually taken as a problem of conditional probability, MEMM is the more appropriate tool to use. Second, MEMM can exploit overlapping features by estimating the probability under the maximum entropy framework.

MEMM is a conditional probability model in which the HMM transition and observation functions are replaced by a single function $P(s_i \mid s_{i-1}, o)$ that provides the probability of the current state s_i given the previous state s_{i-1} and the current observation o. In contrast to HMM, in which the current observation only depends on the current state, in this model the current observation may also depend on the previous state, as shown in Figure 1[15].

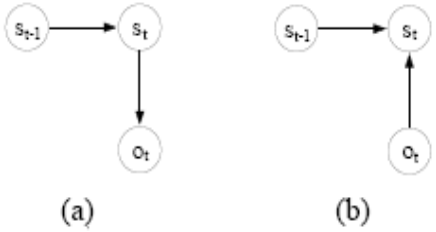

Figure 1. HMM(a) and MEMM(b)[15]

MEMM estimates the probability for $p(s_i \mid s_{i-1}, o)$ under the ME principle in order to utilize the overlapping features. The ME principle assumes that the trained model is consistent with certain constraints derived from the training data, and it makes the fewest assumptions about the data. To predicate the current state s, the context information of s is extracted from the training data and represented as the feature function [29]:

$$f(h,s) = \begin{cases} 1 & if\ h = h^* \ and\ s = s^* \\ 0 & otherwise \end{cases} \quad (1)$$

where h is the context information of s, and h^* (or s^*) is the concrete instance of h (or s).The following constraints are imposed so that the expectation for each feature in the learned model is consistent with its empirical value in the training corpus. More formally, the constraints can be expressed as [29]:

$$E_p(f) = E_{\tilde{p}}(f) \quad (2)$$

where $E_{\tilde{p}}(f)$ is the empirical expectation defined as

$$E_{\tilde{p}}(f) = \sum_{h,s} \tilde{p}(h,s) f(h,s) \quad (3)$$

$E_p(f)$ is the expectation model defined as

$$E_p(f) = \sum_{h,s} p(h,s) f(h,s) \quad (4)$$

$p(h,s)$ is further decomposed according to the multiplication rule:

$$p(h,s) = p(h) \times (s \mid h) \quad (5)$$

For efficiency, the following modification is usually made:

$$p(h,s) \approx \tilde{p}(h) \times p(s \mid h) \quad (6)$$

The expectation model is then reformulated as

$$E_p(f) = \sum_{h,s} \tilde{p}(h)\, p(s \mid h)\, f(h,s) \quad (7)$$

Under these constraints, the ME principle guarantees a learned model that is as uniform as possible. It can be obtained by maximizing the conditional entropy of the training data [29]:

$$H(p) = -\sum_{h,s} p(h)\, p(s \mid h) \log p(s \mid h) \quad (8)$$

So it defines each state-observation transition function to be a log-linear model:

$$P_{s'}(s \mid o) = \frac{1}{Z(h, s')} \exp(\sum_i \lambda_i f_i(o, s)) \quad (9)$$

In Formula(9), $Z(h, s')$ is a normalization factor. λ_i is the multiplier parameter with respect to each feature function which can be estimated by Generalized Iterative Scaling (GIS), Improved Iterative Scaling (IIS) [30] or L-BFGS[31] algorithms.

Finally the Viterbi dynamic programming algorithm is used to search for the best sequence of states.

III. MEMM BASED METHOD FOR PREDICTION OF PROSODIC WORD BOUNDARIES

A. Prosodic Words

Experiments show that Chinese utterance is structured in a prosodic hierarchy. As proposed by Cao [32], prosodic word(PW), prosodic phrase(PP) and intonation phrase(IP) are the three prosodic units, which are in a hierarchical relation, utilized in the prosodic scheme for our Mandarin speech synthesis system. An utterance can contain several IPs, an IP can contain several PPs, and a PP can contain several PWs respectively. It is shown that the prosodic word is more likely to be two syllables long and very few prosodic words will have more than 3 syllables. Figure 2 shows the prosodic structure of a Chinese sentence.

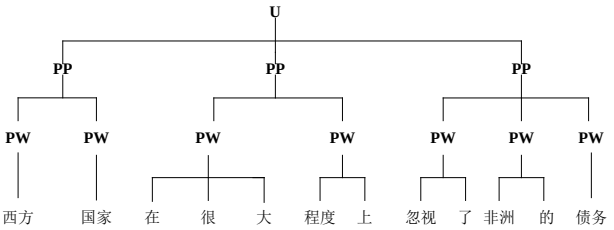

Figure 2. Prosodic structure tree (U for intonation phrase, PP for prosodic phrase, PW for prosodic word)

In the prosodic hierarchy tree, the lexicon word is the smallest unit. The task of building prosodic structure could be reduced to deciding the type for each syntactic word boundary, which is actually a classification problem.

For automatic prediction of prosodic word boundaries, the sentences in training corpus are labeled with follows:

XiFang/s/S GuoJia/n/S Zai/p/B Hen/d/I Da/a/E ChengDu/n/B Shang/m/E HuShi/v/B Le/u/E FeiZhou/ns/B De/u/E ZhaiWu/n/S。/w

(Western Countries have ignored African's debt to a large degree)

Here 'B'(Beginning) represents the beginning of a prosodic word, 'E'(End) is the end of a prosodic word, 'I'(Inside) represents the middle of a prosodic word. 'S'(Single) means that the prosodic word includes one lexicon word or the prosodic word is only a part of a lexicon word.

Therefore, the prosodic word prediction converts to a sequence data labeling problem. Due to the achievement of the MEMM, MEMM is adopted to model this labeling process.

B. Feature Selection in Prosodic Word Prediction

Like Li [9], a semi-automatic approach is used for feature selection in the paper. Features are obtained by two steps, the first of which is to establish feature templates, and the second is to extract features from training corpus according to the feature templates.

Feature templates are established manually from context information. For our specific application, most commonly used features include the part-of-speech(POS), the length in syllables and the word itself of the words surrounding the boundary. The neighbor words are restricted to two words before the boundary and one word after the boundary.

TABLE I.
ATOMIC FEATURES USED IN MEMM

Feature tag	Feature explanation
W-1	previous lexicon word
W0	current lexicon word
W+1	next lexicon word
P-1	Part-of-speech of the previous lexicon word
P0	Part-of-speech of the current lexicon word
P+1	Part-of-speech of the next lexicon word
WL-1	The length of the previous lexicon word, in Chinese characters
WL0	The length of the current lexicon word, in Chinese characters
WL+1	The length of the next lexicon word, in Chinese characters

The features used in the model are shown in Table 1.

Because prosodic word prediction is a complicated labeling problem, atomic features of words and POS tags are not sufficient to describe actual language phenomenon. Based on the atomic features, combined features are created to describe the context relationship. Some examples of combined templates are shown in Table 2.

TABLE II.
EXAMPLES OF COMBINED FEATURES USED IN MAXIMUM ENTROPY MARKOV MODEL

Combined features	P-1P0P+1,W-1W0W+1, W-1W0,W0W+1,W-1W+1,W-1P+1,P-1P+1,P-1W+1, P0P+1,W0P+1,W0P0,P-1W0,P0W+1

IV. ACTIVE LEARNING ALGORITHM

Using the supervised learning methods, large number of labeled data is required for training these models. Labeling-required training data is a time-consuming job. Reduction of the dependence on large amount labeled training data relies on great growth of learning ability. Active learning is a solution for the problem with scarce labeled data and rich unlabeled data in supervised learning.

In the active learning framework, the statistical learning model iteratively selects the instances on which it is going to be trained on. In the widely used pool-based approach, we start with a small labeled training set L and a large pool of unlabelled data U. In each round, a model is trained on L and it is used to select a batch n of instances from U which are considered to be informative. Then these selected samples are annotated by domain expert, added to L and the loop is repeated. The iteration becomes halt when the stopping criterion is met. The active data selection is expected to improve the system accuracy compared with the random data selection.

Generally an active learning approach consists of two independent parts: (f, q), where f denotes the learning machine, namely some model as mentioned before, and q denotes the query function ,i.e. selection function which implements the selection to select the most informative samples from U. So the most important part in active learning method is the query function q which mainly decides an algorithm performance of active learning.

A. Selection Algorithm

Unlabeled samples which a classifier labels with low confidence are considered to not have been learned well enough and considered good candidates for labeling in order to refine the classifier's expertise. We utilize the probabilistic confidence of the MEMM to assign the degree of uncertainty to an example. In the case of a probabilistic classifier, such as the MEMM employed in this work, confidence can be directly assessed via the posterior probability assigned to an observation by the following Equation.

$$x^* = \arg\max(1 - P(y_1^* \mid x, \lambda) \quad (10)$$

Let x be some observed input data sequence, such as a sequence of words in training data, where y_1^* is the most-likely label sequence obtained by the Viterbi decoder.

The entropy criterion can be approximated instead over the set N of N -best sequences, leading to the following expression:

$$\begin{aligned} x^* &= \arg\max \varphi(x) \\ &= \arg\max - \sum_y P(y \mid x; \lambda) \log P(y \mid x; \lambda) \quad (11) \\ &\approx \arg\max - \sum_{y \in N} P(y \mid x; \lambda) \log P(y \mid x; \lambda) \end{aligned}$$

But the function mentioned above has its limitation [33]. In our study, we prefer not only the most informative example in terms of uncertainty measure, but also the most representative example in terms of density measure. The density measure can be evaluated based on how many examples there are similar or near to it.

To address these issues, we propose a modified information density query strategy based on MEMM, which is formulated as (12).

$$DS(x^*) = \frac{1}{U} \sum_{u=1}^{U} \cos(x, x^u) \times (-\sum P(y \mid x; \lambda) \times \log P(y \mid x; \lambda)) \quad (12)$$

An example with larger value $DS(x^*)$ means the node has the larger uncertainty and is more useful information for the system. The density uncertainty measure is used to rank the unlabeled instances and select a certain number of unlabeled instances to update the training instance set for the next iteration.

V. EXPERIMENT

A. The Experiment Corpus

In our experiments, a speech corpus for training and testing are used. 11000 sentences are randomly selected from the People's Daily corpus read by a radiobroadcaster. The sentences with three-level prosodic boundaries are labeled manually by listening to the record speech.

To check consistency of annotation across different people, an exploratory experiment was carried out. Three annotators were first trained on the same 100 sentences. At this stage, they were required to discuss criteria for annotation so that they could achieve agreement on most of the annotations in the 100 sentences. Then they were asked to annotate a small subset of the corpus. All three annotators achieved agreement on 85%. That is to say pretty good consistency existed among the three annotators.

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 96.5%.

B. The Evaluation Criteria

The precision, recall ratio and F1-score are adopted as the evaluation criteria. The precision and recall are defined as: Pre= C_1 / C_2 , Rec= C_1 / C_3 . C_1 is the number of prosodic phrase boundaries correctly recognized, C_2 is the total number of prosodic phrase boundaries recognized, and C_3 represents the total number of real prosodic phrase boundaries in the test corpus.

The F1-score is calculated as: $F = 2 \times \mathrm{Pre} \times \mathrm{Rec} / (\mathrm{Pre} + \mathrm{Rec})$.

VI. RESULTS AND DISCUSSION

Two factors may influence the performance of active learning. One is the size of the initial labeled instance set,

and the other one is the number of classified unlabeled data selected for the next iteration.

A set of experiments are developed for active learning with different sizes of initial labeled instance sets. The testing size is the same. The stopping point is 100. During each iteration, the top 100 most informative samples were picked up, labeled by annotators and added into the training set. The experimental results of F1-score are shown in Table 3. The second column shows the performance of the MEMM using just the initial data. And the performance after active learning is shown in the third column in Table 3.

Then, we design the experiments of active learning with different numbers of selected unlabeled data for the next iteration. In the experiment, the size of initial training set L is 100.The experimental results of F1-score are shown in Table 4.

In order to show the effective of active learning, the results on the active learning method for MEMM are compared with the results when the instance is chosen randomly from the training corpus in Table 5. The results are also obtained based on MEMM when the instance is chosen randomly from the training corpus. It is easy to see that the active learning approach outperforms the random sampling, in spite of the result of active learning is lower than the result of the general MEMM with total 10,000 sentences.

Due to difference in the corpus and evaluation metric, these results may not be comparable in all respects. Yet from the statistics above, we could safely say that MEMM model combined with active learning method is more efficient to resolve prosodic word prediction problem.

A. Results

TABLE III.
THE RESULT OF DIFFERENT SIZES OF INITIAL LABELED DATA

Initial labeled data	Initial F1-score(%)	Final F1-score(%)
500	75.2	82.9
400	72.4	81.6
300	70.3	80.8
200	69.2	79.3
100	67.1	77.4

TABLE IV.
THE RESULT OF DIFFERENT NUMBERS OF SELECTED UNLABELED DATA FOR THE NEXT ITERATION

Numbers of unlabeled data for next iteration	F1-score(%)
10	72.1
20	73.2
30	74.4
50	75.8
100	77.2

TABLE V.
THE RESULTS OF ACTIVE-LEARNING, RANDOM SAMPLING

Training Data Size	Precision(%)	Recall(%)	F1-score(%)
Random 5000	91..5	84.2	87.7
Total 10,000	94.2	86.1	89.9
Active learning 5000	93.4	84.4	88.7

VII. CONCLUSION

In this paper, we introduce an active learning method to solve the task of prosodic word prediction. To the best of our knowledge, the presented work is the first to apply AMEMM to Chinese prosodic word prediction. Experiments show that the method can achieve comparable performance to the supervised learning models for prosodic word prediction.

Our future work is to incorporate more contextual information into the models. How to integrate that information into MEMM and further improve the performance of prediction in terms of the precision, recall and F1-score is one of the directions in the future.

ACKNOWLEDGMENT

The work described in this paper was substantially supported from the National Science Foundation of China (Grant No: 61103074), the Natural Science Foundation of Tianjin(Grant No: 11JCYBJC00600), the Open Project of Shanghai Key Laboratory of Trustworthy Computing of China under Grant No.53H10058 and the Technology Fund Planning Project of Higher Education, Tianjin(Grant No: 20110816,20110818).

REFERENCES

[1] Yao Qian, Min Chu, “Segmenting unrestricted Chinese text into prosodic words instead of lexicon words”, Proceedings of the 2001 International conference on acoustic, speech and signal processing, 2001, Salt Lake City, pp. 825-828.

[2] Gee J.P., Grosjean F.,"Performance structures: A psycholinguistic and Linguistic Appraisal", Cognitive Psychology, Vol. 15,1983, pp.411–458.
[3] Jianfen Cao, "Prediction of Prosodic Organization Based on Grammatical Information", Journal of Chinese Information Processing, Vol. 17.,2003, pp. 41–46.
[4] Jianfen Cao, Weibin Zhu, "Syntactic and Lexical Constraint in Prosodic Segmentation and Grouping", Proceedings of Speech Prosody 2002. 2002,France.
[5] Hongjun Wang, "Prosodic words and prosodic phrases in Chinese", Chinese Language, Vol. 6.2000, pp.525–536.
[6] Xiaonan Zhang, Jun Xu, Lianhong Cai, "Prosodic Boundary Prediction based on Maximum Entropy Model with Error-Driven Modification", Proceedings of ISCSLP 2006.
[7] Wang M., Hirschberg J," Predicting Intonational Boundaries Automatically from Text: The ATIS Domain" ,Proceedings of the DARPA Speech and Natural Language Workshop, 199l, pp. 378–383
[8] Paul Taylor, Alan. W. Black, "Assigning phrase breaks from part-of speech sequences", Computer Speech and Language, Vol. 12(4),1998, pp. 99–117.
[9] Z Sheng, T Jianhua, C Lianhong, "Learning rules for Chinese prosodic phrase prediction", International Conference on Computational Linguistics, Proceedings of the first SIGHAN workshop on Chinese language processing, Vol. 18,2002.
[10] Jianfeng Li, Guoping Hu and Renhua Wang, "Chinese prosody phrase prediction based on maximum entropy model", Proceedings of Interspeech 2004, Jeju Island, Korea, 2004, pp. 729-732.
[11] Z Sheng, T Jianhua, C Lianhong: ZHENG Min, and CAI Lianhong, "Statistical model based on probability frequency for Mandarin prosodic structure prediction", Journal of Tsinghua University, 2006, No. 46.pp.78-81
[12] Honghui DONG, Jianhua TAO, and Bo XU, "Prosodic Word Prediction Using the Lexical Information", Natural Language Processing and Knowledge Engineering, 2005, pp.189-193
[13] Dong Honghui, Tao Jianhua, "Prosodic Word Prediction using a Maximum Entropy Approach", Proceedings of ISCSLP2006.
[14] Ziping Zhao, Xirong Ma, "Prediction of Prosodic Word Boundaries in Chinese TTS Based on Maximum Entropy Markov Model and Transformation Based Learning", Proceedings of CIS2012,2012,pp.258-261.
[15] A. McCallum, D.Freitag and F.Pereira, "Maximum Entropy Markov Models for Information Extraction and Segmentation", Proceedings of ICML 2000, Standford, CA, USA, 2000, pp.591-598.
[16] E Brill, "Transformation-based error-driven learning and natural language processing: A case study in Part-of-Speech tagging", Computationa1 Linguistics, 21(4), 1995, pp.543-565.
[17] Cynthia A. Thompson, Mary Elaine Califf, and Raymond J. Mooney, "Active learning for natural language parsing and information extraction", Proceedings of the 16th International Conf. on Machine Learning, pp. 406-414.
[18] M. Tang, X. Luo, and S. Roukos, "Active learning for statistical natural language parsing", Proceedings of ACL 2002,pp.120-127.
[19] Lewis David D. and William A. Gale, "A sequential algorithm for training text classifiers", Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 3-12.
[20] Andrew McCallum and Kamal Nigam, "Employing EM and pool-based active learning for text classification", Proceedings of the 8th International Conference, 1998, pp.359-367.
[21] Chen Jinying, Andrew Schein, Lyle Ungar and Martha Palmer, "An empirical study of the behavior of active learning for word sense disambiguation", Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, pp. 120-127.
[22] Zhu Jingbo and Eduard Hovy, "Active learning for word sense disambiguation with methods for addressing the class imbalance problem", Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 783-790.
[23] M. Sassano, "An empirical study of active learning with support vector machines for Japanese word segmentation", Proceedings the ACL 2002,2002, pp .505-512.
[24] Shen Dan, Jie Zhang, Jian Su, Guodong Zhou and Chew-Lim Tan, "Multi-criteria-based active learning for named entity recognition", Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics.
[25] Ngai Grace and David Yarowsky, "Rule writing or annotation: cost-efficient resource usage for based noun phrase chunking", Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, pp. 117-125.
[26] G. Schohn and D. Cohn, "Less is more: Active learning with support vector machines", Proceedings of ICML 2000.
[27] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification", Journal of Machine Learning Research, vol. 2, 2001,pp. 45-66
[28] D. Hakkani-Tiir, G. Riccardi, and A. Gorin, "Active learning for automatic speech recognition", Proceedings of ICASSP 2002.
[29] Jinghui Xiao, Xiaolong Wang and Bingquan Liu, "The Study of a Nonstationary Maximum Entropy Markov Model and its Application on the POS-Tagging Task", Asian Language Information Processing, Vol.6, NO.2 ,2007, pp.1-28
[30] A. Berger, S. A. Della Pietra and V.J. Della Pietra, "A Maximum Entropy Approach to Natural Language Processing", Computational Linguistics, 22 (1), 1996, pp. 39-71.
[31] Hanna Wallach, "Efficient training of conditional random fields", Master's thesis, University of Edinburgh, 2002.
[32] J.Cao, "Rhythm of Spoken Chinese-Linguistic and Paralinguistic Evidences", Proceeding of ICSLP2000 Conference, 2000, pp. 357-360.
[33] Christopher T. Symons,etc, "Multi-Criterion Active Learning in Conditional Random Fields", Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence.

Ziping Zhao Director of Software Engineering in Tianjin Normal University and CCF member. He got his Doctor's Degree from Nankai University in 2008.

He started the teaching career in 2008 in Tianjin Normal University. In 2010, he studied in the Key Laboratory of Trustworthy Computing of East China Normal University as a visiting scholar. In 2010, he became the Director of Soft Engineering in Tianjin Normal University. His research fields are speech synthesis, machine learning and natural language processing.

He undertakes the project of National Natural Science Foundation, Natural Science Foundation of Tianjin and has published many papers in journals and conferences at home and abroad.

Xirong Ma Master Supervisor, Dean of computer and information engineering in Tianjin Normal University and CCF member. She got her Doctor's Degree from Nankai University in 2003.

She started the teaching career in 1984 in Tianjin Normal University, promoted as professor in 2002. From 2004 to 2007, she is engaged in the post-doctoral research at University of Science and Technology Beijing. In 2000, she became the Dean of college of computer and information engineering in Tianjin Normal University. Her research fields are affective computing, pattern recognition.

She undertakes many projects of National Natural Science Foundation, Provincial and Ministry Science Foundation and has published over 20 papers in journals and conferences at home and abroad.