Chinese Part-of-speech Tagging Based on Fusion Model

Guang-Lu Sun[1] Fei Lang[2] Pei-Li Qiao[1] Zhi-Ming Xu[3]

[1]School of Computer Science & Technology, Harbin University of Science & Technology, Harbin, China {bati_sun@hit.edu.cn}
[2]Department of Foreign Languages Teaching, Harbin Science and Technology, Harbin
[3] School of Computer Science & Technology, Harbin Institute of Technology, China

Abstract

This paper proposes a new part-of-speech tagging algorithm based on the fusion model which combines Maximum Entropy model and Error Correction model. According to the analysis of the two models, the fusion tagging model is utilized with the profits of conditional probability model and rule based model. The selection of features and rule templates in the fusion model is discussed. Experimental results show that the new model achieves impressive accuracy in terms of the F-score: 93.73%.

Keywords: Maximum entropy, Part-of-speech tagging, Error correction, Fusion model

1. Introduction

In the field of computational linguistics, the final goal is to make computers process and structurize human language, even understand it. Therefore, a series of processing procedures are worked up from the bottom to the top, including tokenization, segmentation, Part-Of-Speech (POS), name entity recognition, text chunking, full parsing, semantic parsing, etc. Based on good behavior of these procedures, intelligent computer systems can be built, such as information retrieval, question and answering, information extraction and machine translation.

Most previous works applied different kinds of machine learning algorithms to POS tagging. Two factors that determine the tag of a word are its lexical probability and its contextual probability. Some approaches have been adopted, which can mainly fall into rule-based approaches, such as Transformation-Based method [1], and statistical approaches, such as Decision Tree [2], Hidden Markov model (HMM) [3], Maximum Entropy Model [4] and Support Vector Machines (SVM) [5].

A recent trend in the POS tagging task is to train several classifiers on the task and to combine their results to produce a final result for improving the tagging performance [6]. In the present work, Maximum Entropy (ME) model is firstly applied to the task. ME tagging model has been confirmed to get fairly high performance in the past works. But it is difficult to correct the errors in ME tagging model. For further improving the accuracy of tagging, Error Correction algorithm is then utilized to correct part of errors which are brought from ME tagging model. This paper focuses on POS tagging with the spec and corpus of Chinese People’s Daily Newspaper.

Section II describes in detail the ME POS tagging model. Section III introduce the Error Correction model and presents the fusion system combining two models Section IV presents experimental results of our system. Finally, we draw some conclusions.

Proceedings of the 11th Joint Conference on Information Sciences (2008)
Published by Atlantis Press

2. Maximum entropy model

The ME model is an effective machine learning model which is proposed to solve the classification problem [7]. One of the main advantages of using the ME model is the ability to incorporate various features into the conditional probability framework.

Given the histories H, the goal of the ME model is to find the optimal tag sequence $T = t_1, t_2, \dots , t_k$..

Let f_j denote the features of the ME model. f_j is defined as follows:

$$f_j(t,h) = \begin{cases} 1 & \textit{if} \quad h = h^* \quad \textit{and} \quad t = t^* \\ 0 & \textit{otherwise} \end{cases} \tag{1}$$

where t^* is a certain tag, and h^* is a certain instance of context.

The conditional entropy of $P(t/h)$ is defined as:

$$H(p) = -\sum_{t \in T, h \in H} \tilde{p}(h)\, p(t/h) \log p(t/h) \tag{2}$$

By maximizing the conditional entropy subject to the constraints, we can estimate $P(t/h)$ based on the maximum entropy theory. The model's distribution $P(t/h)$ can be inferred by means of Lagrange transformation:

$$p(t \mid h) = \frac{1}{Z(h)} \exp\left(\sum_j \lambda_j f_j(t,h) \right) \tag{3}$$

$$Z(h) = \sum_t \exp\left(\sum_j \lambda_j f_j(t,h) \right) \tag{4}$$

where $Z(h)$ is the normalization constant. λ_i is the multiplier parameter with respect to each feature function.

Given a set of features and training data, the improved iterative scaling algorithm can be used to find the optimal parameters λ_i .

3. Part-of-speech tagging based on the fusion system

3.1. Error correction model

The formalism of Transformation based learning (TBL) is first introduced by Eric Brill in 1992. The transformational rules which correct the error tags to the right ones are stored and used in turn for the purpose of template correction learning.

In Brill's TBL model, the base model is a heuristic probability which has low accuracy. For improving the performance of the TBL model, the ME model is used to replace the heuristic model.

The ME model is a supervised learning model which needs the training corpus. But it is the same as the training corpus used in TBL. The close test procedure results in few error tags exist. Therefor, the N-fold partitioning method is proposed to solve the problem. The one fold training corpus is tagged by the ME model which is trained by other N-1 fold training corpora. Through the cross-validation, all the training corpora used in TBL are rebuilt.

3.2. System description

POS tagging can be seen as the sequence analysis and labeling task. This type of task is often described as models which are from input sequences to sequences of labels. Given a word sequence $W = w_1, w_2, \dots , w_k$, where k is the number of words in the sentence, the result of POS tagging is assumed to be a sequence, in which the words are tagged with POS tags as follows:

... $[w_i\ w_{i+1}$ ... $w_{i+m}]$ $[w_{i+m+1}$ ... $w_{i+m+h}]$ ...
... $[P_i\ P_{i+1}$ ... $P_{i+m}]$ $[P_{i+m+1}$ ... $P_{i+m+h}]$ ...

where P_j corresponds to the POS tag which is used to indicate the type of part-of-speech.

With the formalization of the tagging task and two learning models described above, the tagging system is built as follows. The ME model produces the initial tag for each position. Ten times cross-validation is firstly applied to train the ME model. Then each part of corpus is tagged using the ME model which is trained by other nine parts of training corpora. The tagged corpora are used in the Error Correction model. In the test phase, the predictions of the tagging system on new text are determined by beginning with the ME model and then applying each correction rule of Error Correction model in turn.

The ME model highly depends on feature templates. The histories of the current position are sources for feature collection. We utilized the lexical information of the current word, the left and right context consisting of two words as atomic features. In addition, the affix information of the current word and the POS tag of the previous word are atomic features. Tab. 1 shows the features templates.

Table 1: Feature template based on the lexical information

Feature type	Features
Atomic features	W_i, W_{i-1}, W_{i-2}, W_{i+1}, W_{i+2}, P_i, P_{i-1}, P_{i-2}, P_{i+1}, P_{i+2}, S_{i-1}, PF_i, AF_i
Combined features	$W_{i-1}W_i$, W_iW_{i+1}, $W_{i-1}W_{i+1}$, $P_{i-1}P_i$, $P_{i-2}P_{i-1}$, P_iP_{i+1}, $P_{i-1}P_{i+1}$, $P_{i-1}P_iP_{i+1}$, $P_{i-2}P_{i-1}P_i$, $P_iP_{i+1}P_{i+2}$, W_iP_{i+1}, W_iP_{i+2}, P_iW_{i-1}, $W_{i-2}P_{i-1}P_i$, $P_iW_{i+1}P_{i+1}$, $P_{i-1}W_iP_i$, $S_{i-1}P_iP_{i+1}$, $S_{i-1}P_i$, $S_{i-1}P_{i-1}P_i$, P_iW_{i+1},

The heuristic that low frequency features are not reliable is used to cut off the features that occurred less than three times. Through feature selection, more reliable features could be used.

The rule templates which are formed from conjunctions of atomic features (except the affix information) in Table 1 match to particular combinations of features in the histories of the current word W_i. Tab. 2 shows the patterns of rule templates. 120 types of rule templates are built using the patterns.

Table 2: The patterns of rule template

Word patterns	W_i, W_{i-1}, W_{i-2}, W_{i+1}, W_{i+2}, $W_{i-1}W_i$, W_iW_{i+1}, $W_{i-2}W_{i-1}$, $W_{i+1}W_{i+2}$
Tag patterns	T_i, T_{i-1}, T_{i-2}, T_{i+1}, T_{i+2}, $T_{i-1}T_i$, T_iT_{i+1}, $T_{i-2}T_{i-1}$, $T_{i+1}T_{i+2}$, C_{i-1},C_{i-2}

4. Evaluations

At first, the corpus and measurement are introduced briefly. Then we present the performance of POS tagging.

4.1. Corpus and measurement

The corpus which is used in our system comes from the Chinese People's Daily Newspaper in 1998. It was manually annotated by Institute of Computational Linguistic of Peking University. The annotation tasks include word segmentation, POS tagging and Named Entity information. The training corpus is the corpus of 98.01 including about 124 thousand sentences and 1121 thousand words. The test corpus is the corpus of 9806 including about 137 thousand sentences and 1244 thousand words.

The performance is measured with three rates: precision (P), recall (R) and F-score (F), which are equal in POS tagging task.

4.2. Experimental results

In this experiment, we compare the performances of different POS tagging models. Tagging results are listed in Tab. 3. The POS tag that has maximum occurrence probability for each word is used to tag its corresponding word token. By this method, we have got the baseline result that is listed in the first row of Tab. 3. The results based on the ME model are listed in the second row of Tab. 3. Based on the fusion model, the result is listed in the third row of Tab. 3. All the results are obtained in open tests.

Table 3: POS tagging performance achieved by applying different systems

Model	F (%)
Baseline	*64.57*
ME Model	*92.41*
The Fusion Model	*93.73*

The accuracy of the fusion model is better than that of the ME model. The overall improvement is 1.32% in F-score. The fusion POS tagging model utilizes sufficient context information that can describe actual language phenomenon effectively. The correction rules make the relation of context words and tags much tight. For the fusion model can be seen as the combination of ME model and Error Correction model, the fusion model performs better than the ME model in combined characteristics of different models. Experimental results show that fusion POS tagging model is more efficient to resolve the POS tagging problem.

5. Conclusions

We propose a new algorithm of POS tagging based on the fusion model combining the ME model and the Error Correction model. The fusion model combines the conditional probability model and rule based model harmoniously. In open tests, the new tagging model obtained the F-score of 93.93% which is better than the ME model for POS tagging. The improvement is 1.32%.

Acknowledgement
This work is supported by National Natural Science Foundation 60673037, the High Technology Research and Development Program of China grant 2007AA01Z172, the Heilongjiang provincial Science Foundation F2007-06, Heilongjiang provincial special fund to support information industrialization.

6. References

[1] E. Brill. Transformation-based error-driven learning and natural language processing: A case study in part of speech tagging. *Computational Linguistics*, 21(1995)4, 543-565.

[2] David, M.M.. Statistical Decision-Tree Models for Parsing. *In Proceeding of the 33rd Annual Meeting of the ACL*, (1995), 276-283.

[3] Rabiner, L. R.. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *In Proceedings of the IEEE*, 77(2), 1989, pp. 257-285.

[4] Ratnaparkhi, A.. A Maximum Entropy Model for Part-Of-Speech Tagging. In *Proceedings of EMNLP'1996*, New Brunswick, New Jersey, USA, 1996, pp. 133-142.

[5] Jesús, G., Lluís, M. SVMTool: A General POS Tagger Generator Based on Support Vector Machines. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*. 2004.

[6] Jiang wei. Statistical Chinese Lexical Analysis and Its Reinforcement Learning Mechanism. *P.H D dissert of Harbin Institute of Technology*. 2007.

[7] Berger, A., 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.