

Hybrid Models for Chinese Named Entity Recognition

Lishuang Li, Tingting Mao, Degen Huang, Yuansheng Yang

Department of Computer Science and Engineering

Dalian University of Technology

116023 Dalian, China

{computer, huangdg, yangys}@dlut.edu.cn

maotingting1007@sohu.com

Abstract

This paper describes a hybrid model and the corresponding algorithm combining support vector machines (SVMs) with statistical methods to improve the performance of SVMs for the task of Chinese Named Entity Recognition (NER). In this algorithm, a threshold of the distance from the test sample to the hyperplane of SVMs in feature space is used to separate SVMs region and statistical method region. If the distance is greater than the given threshold, the test sample is classified using SVMs; otherwise, the statistical model is used. By integrating the advantages of two methods, the hybrid model achieves 93.18% F-measure for Chinese person names and 91.49% F-measure for Chinese location names.

1 Introduction

Named entity (NE) recognition is a fundamental step to many language processing tasks such as information extraction (IE), question answering (QA) and machine translation (MT). On its own, NE recognition can also provide users who are looking for person or location names with quick information. Palma and Day (1997) reported that person (PER), location (LOC) and organization (ORG) names are the most difficult sub-tasks as compared to other entities as defined in Message Understanding Conference (MUC). So we focus on the recognition of PER, LOC and ORG entities.

Recently, machine learning approaches are widely used in NER, including the hidden Markov model (Zhou and Su, 2000; Miller and Crystal, 1998), maximum entropy model (Borthwick, 1999), decision tree (Qin and Yuan,

2004), transformation-based learning (Black and Vasilakopoulos, 2002), boosting (Collins, 2002; Carreras et al., 2002), support vector machine (Takeuchi and Collier, 2002; Yu et al., 2004; Goh et al., 2003), memory-based learning (Sang, 2002). SVM has given high performance in various classification tasks (Joachims, 1998; Kudo and Matsumoto, 2001). Goh et al. (2003) presented a SVM-based chunker to extract Chinese unknown words. It obtained higher F-measure for person names and organization names.

Like other classifiers, the misclassified testing samples by SVM are mostly near the decision plane (i.e., the hyperplane of SVM in feature space). In order to increase the accuracy of SVM, we propose a hybrid model combining SVM with a statistical approach for Chinese NER, that is, in the region near the decision plane, statistical method is used to classify the samples instead of SVM, and in the region far away from the decision plane, SVM is used. In this way, the misclassification by SVM near the decision plane can be decreased significantly. A higher F-measure for Chinese NE recognition can be achieved.

In the following sections, we shall describe our approach in details.

2 Recognition of Chinese Named Entity Using SVM

Firstly, we segment and assign part-of-speech (POS) tags to words in the texts using a Chinese lexical analyzer. Secondly, we break segmented words into characters and assign each character its features. Lastly, a model based on SVM to identify Chinese named entities is set up by choosing a proper kernel function.

In the following, we will exemplify the person names and location names to illustrate the identification process.

2.1 Support Vector Machines

Support Vector Machines first introduced by Vapnik (1996) are learning systems that use a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical theory. SVMs are based on the principle of structural risk minimization. Viewing the data as points in a high-dimensional feature space, the goal is to fit a hyperplane between the positive and negative examples so as to maximize the distance between the data points and the hyperplane.

Given training examples:

$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}, x_i \in R^n, y_i \in \{-1, +1\}, \quad (1)$$

x_i is a feature vector (n dimension) of the i -th sample. y_i is the class (positive(+1) or negative(-1) class) label of the i -th sample. l is the number of the given training samples. SVMs find an “optimal” hyperplane: $(wx + b) = 0$ to separate the training data into two classes. The optimal hyperplane can be found by solving the following quadratic programming problem (we leave the details to Vapnik (1998)):

$$\begin{aligned} \max \quad & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i y_i \alpha_j y_j K(x_i, x_j) \\ \text{subject to} \quad & \sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq c, i = 1, 2, \dots, l. \end{aligned} \quad (2)$$

The function $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ is called kernel function, $\varphi(x)$ is the mapping from primary input space to feature space. Given a test example, its label y is decided by the following function:

$$f(x) = \text{sgn} \left[\sum_{x_i \in sv} \alpha_i y_i K(x_i, x) + b \right]. \quad (3)$$

Basically, SVMs are binary classifiers, and can be extended to multi-class classifiers in order to solve multi-class discrimination problems. There are two popular methods to extend a binary classification task to that of K classes: *one class vs. all others* and *pairwise*. Here, we employ the simple *pairwise* method. This idea is to build $K \times (K - 1) / 2$ classifiers considering all pairs of classes, and final decision is given by their voting.

2.2 Recognition of Chinese Person Names Based on SVM

We use a SVM-based chunker, YamCha (Kudo and Masumoto, 2001), to extract Chinese person names from the Chinese lexical analyzer.

1) Chinese Person Names Chunk Tags

We use the Inside/Outside representation for proper chunks:

- I Current token is inside of a chunk.
- O Current token is outside of any chunk.
- B Current token is the beginning of a chunk.

A chunk is considered as a Chinese person name in this case. Every character in the training set is given a tag classification of B, I or O, that is, $y_i \in \{B, I, O\}$. Here, the multi-class decision method pairwise is selected.

2) Features Extraction for Chinese Person Names

Since Chinese person names are identified from the segmented texts, the mistakes of word segmentation can result in error identification of person names. So we must break words into characters and extract features for every character. Table 1 summarizes types of features and their values.

Type of feature	Value
POS tag	n-B, v-I, p-S
Whether a character is a surname	Y or N
Character	surface form of the character itself
The frequency of a character in person names table	Y or N
Previous BIO tag	B-character, I-character, O-character

Table 1. Summary of Features and Their Values

The POS tag from the output of lexical analysis is subcategorized to include the position of the character in the word. The list of POS tags is shown in Table 2.

POS tag	Description of the position of the character in a word
<POS>-S	One-character word
<POS>-B	first character in a multi-character word
<POS>-I	intermediate character in a multi-character word
<POS>-E	last character in a multi-character word

Table 2. POS Tags in A Word

If the character is a surname, the value is assigned to Y, otherwise assigned to N.

The “character” is surface form of the character in the word.

We extract all person names in January 1998 of the People’s Daily to set up person names table and calculate the frequency of every charac-

ter (F) of person names table in the training corpus. The frequency of F is defined as

$$P(F) = \frac{\text{the number of } F \text{ as a character of person names}}{\text{the total number of } F}, \quad (4)$$

if $P(F)$ is greater than the given threshold, the value is assigned to Y, otherwise assigned to N.

We also use previous BIO-tags as features.

Whether a character is inside a person name or not, it depends on the context of the character. Therefore, we use contextual information of two previous and two successive characters of the current character as features.

Figure 1 shows an example of features extraction for the i -th character. When training, the features of the character “Min” contains all the features surrounded in the frames. If the same sentence is used as testing, the same features are used.

	i				
Position	-2	-1	0	+1	+2
Character	Jiang	Ze	Min	zhu	xi
POS tags	n-S	n-B	n-E	n-B	n-E
Whether the character is a surname	Y	N	N	N	Y
The frequency of a character in the person names table	Y	Y	Y	N	N
Previous BIO tags	B	I	I	O	O

Figure 1. An example of features extraction

3) Choosing Kernel Functions

Here, we choose polynomial kernel functions: $K(x, x_i) = [(x \cdot x_i) + 1]^d$ to build an optimal separating hyperplane.

2.3 Recognition of Chinese Location Names Based on SVM

The identification process of location names is the same as that of person names except for the features extraction. Table 3 summarizes types of features and their values of location names extraction.

Type of feature	Value
POS tag	n-B, v-I, p-S
Whether a character appears in location names characteristic table	Y or N
Character	surface form of the character itself
Previous BIO tag	B-character, I-character, O-character

Table 3. Summary of Features and Their Values

The location names characteristic table is set up in advance, and it includes the characters or words expressing the characteristics of location names such as “sheng (province)”, “shi (city)”, “xian (county)” etc. If the character is in the location names characteristic table, the value is assigned to Y, otherwise assigned to N.

3 Statistical Models

Many statistical models for NER have been presented (Zhang et al., 1992; Huang et al., 2003 etc). In this section, we proposed our statistical models for Chinese person names recognition and Chinese location names recognition.

3.1 Chinese Person Names

We define a function to evaluate the person name candidate PN . The evaluated function $TotalProbability(PN)$ is composed of two parts: the lexical probability $LP(PN)$ and contextual probability $CP(PN)$ based on POS tags.

$$TotalProbability(PN) = \alpha LP(PN) + (1 - \alpha) CP(PN), \quad (5)$$

where PN is the evaluated person name and α is the balance coefficient.

1) lexical probability $LP(PN)$

We establish the surname table ($SurName$) and the first name table ($FirstName$) from the students of year 1999 in a university (containing 9986 person names).

Suppose $PN = LF_1F_2$, where L is the surname of the evaluated person name PN , $F_i (i=1,2)$ is the i -th first name of the evaluated person name PN .

The probability of the surname $P_l(L)$ is defined as

$$P_l(L) = \frac{P_{l0}(L)}{\sum_{y \in SurName} P_{l0}(y)}, \quad (6)$$

where $P_{l0}(L) = \log_2(N(L) + 2)$, $N(L)$ is the number of L as the single or multiple surname of person names in the $SurName$.

The probability of the first name $P_f(F)$ is defined as

$$P_f(F) = \frac{P_{f0}(F)}{\sum_{y \in FirstName} P_{f0}(y)}, \quad (7)$$

where $P_{f0}(F) = \log_2(N(F) + 2)$, $N(F)$ is the number of F in the $FirstName$.

The lexical probability of the person name PN is defined as

$$LP(PN) = P_l(L) \times P_f(F_1) \quad \text{if}(PN = LF_1) \quad (8)$$

$$LP(PN) = P_l(L) \times C_b \times (P_f(F_1) + P_f(F_2)) \quad \text{if}(PN = LF_1F_2),$$

where C_b is the balance coefficient between the single name and the double name. Here, $C_b=0.844$ (Huang et al., 2001).

2) contextual probability based on POS tags $CP(PN)$

Chinese person names have characteristic contextual POS tags in real Chinese texts, for example, in the phrase “dui Zhangshuai shuo (say to Zhangshuai)”, the POS tag before the person name “Zhangshuai” is prepnoun and verb occurs after the person name. We define the bigram contextual probability $CP(PN)$ of the person name PN as the following equation:

$$CP(PN)=\frac{PersonPOS(< lpos, PN, rpos >)}{TotalPOS}, \quad (9)$$

where $lpos$ is the POS tag of the character before PN (called POS forward), $rpos$ is the POS tag of the character after PN (called POS backward), and $PersonPOS(< lpos, PN, rpos >)$ is the number of PN as a person name whose POS forward is $lpos$ and POS backward is $rpos$ in training corpus. $TotalPOS$ is the total number of the contextual POS tags of every person name in the whole training corpus.

3.2 Chinese Location Names

We also define a function to evaluate the location name candidate LN . The evaluated function $TotalProbability(LN)$ is composed of two parts: the lexical probability $LP(LN)$ and contextual probability $CP(LN)$ based on POS tags.

$$TotalProbability(LN) = \alpha LP(LN) + (1 - \alpha) CP(LN), \quad (10)$$

where LN is the evaluated location name and α is the balance coefficient.

1) lexical probability $LP(LN)$

Suppose $LN=F_0F^+S$, $F^+=F_1\dots F_n$, ($i=1, \dots, n$), where F_0 is the first character of the evaluated location name LN , F^+ is the middle characters of the evaluated location name LN , S is the last character of the evaluated location name LN .

The probability of the first character of the evaluated location name $P_h(F_0)$ is defined as

$$P_h(F_0) = \frac{P_{h_0}(F_0)}{P'_{h_0}(F_0)}, \quad (11)$$

where $P_{h_0}(F_0) = \log_2(C(F_0)+2)$, $C(F_0)$ is the number of F_0 as the first character of location names in the Chinese Location Names Record.

$P'_{h_0}(F_0) = \log_2(C'(F_0)+2)$, $C'(F_0)$ is the total number of F_0 in the Chinese Location Names Record.

The probability of the middle character of the evaluated location name $P_f(F^+)$ is defined as

$$P_f(F^+) = \sum_{i=1}^n \frac{P_f(F_i)}{P'_f(F_i)}, \quad (12)$$

where $P_f(F_i) = \log_2(C(F_i)+2)$, $C(F_i)$ is the number of F_i as the i -th middle character of location names in the Chinese Location Names Record.

$P'_f(F_i) = \log_2(C'(F_i)+2)$, $C'(F_i)$ is the total number of F_i in the Chinese Location Names Record.

The probability of the last character of the evaluated location name $P_l(S)$ is defined as

$$P_l(S) = \frac{P_l(S)}{P'_l(S)}, \quad (13)$$

where $P_l(S) = \log_2(C'(S)+2)$, $C(S)$ is the number of S as the last character of location names in the Chinese Location Names Record.

$P'_l(S) = \log_2(C'(S)+2)$, $C'(S)$ is the total number of S in the Chinese Location Names Record.

The lexical probability of the location name LN is defined as

$$LN = (P_h(F_0) + P_f(F^+) + P_l(S)) / Len(LN), \quad (14)$$

where $Len(LN)$ is the length of the evaluated location name LN .

2) contextual probability based on POS tags $CP(LN)$

Location names also have characteristic contextual POS tags in real Chinese texts, for example, in the phrase “zai Chongqing shi junxing (to be held in Chongqing)”, the POS tag before the location name “Chongqing” is prepnoun and verb occurs after the location name. We define the bigram contextual probability $CP(LN)$ of the location name LN similar to that of the person name PN in equation (9), where PN is replaced with LN .

4 Recognition of Chinese Named Entity Using Hybrid Model

Analyzing the classification results (obtained by sole SVMs described in section 2) between B and I, B and O, I and O respectively, we find that the error is mainly caused by the second classification. The samples which attribute to B class are misclassified to O class, which leads to B class vote's diminishing and the corresponding named entities are lost. Therefore the Recall is lower. In the meantime, the number of the misclassified samples whose function distances to the hyperplane of SVM in feature space are less than 1 can reach over 83% of the number of total misclassified samples. That means the misclassi-

fication of a classifier is occurred in the region of two overlapping classes. Considering this fact, we can expect to improve SVM using the following hybrid model.

The hybrid model includes the following procedure:

- 1) compute the distance from the test sample to the hyperplane of SVM in feature space.
- 2) compare the distance with given threshold.

The algorithm of hybrid model can be described as follows:

Suppose T is the testing set,

- (1) if $T \neq \Phi$, select $x \in T$, else stop;
- (2) compute $g(x) = \sum_{i=1}^l \alpha y_i K(x_i, x) + b$
- (3) if $|g(x)| > \varepsilon$, $\varepsilon \in [0,1]$ output $f(x) = \text{sgn}(g(x))$, else use the statistic models and output the returned results.
- (4) $T \leftarrow T - \{x\}$, repeat(1)

5 Experiments

Our experimental results are all based on the corpus of Peking University.

5.1 Extracting Chinese Person Names

We use 180 thousand characters corpus of year 1998 from the People’s Daily as the training corpus and extract other sentences (containing 1526 Chinese person names) as testing corpus to conduct an open test experiment. The results are obtained as follows based on different models.

1) Based on Sole SVM

An experiment is carried out to recognize Chinese person names based on sole SVM by the method as described in Section 2. The Recall, Precision and F-measure using different number of degree of polynomial kernel function are given in Table 4. The best result is obtained when $d=2$.

	Recall	Precision	F-measure
$d=1$	87.22%	94.26%	90.61%
$d=2$	87.16%	96.10%	91.41%
$d=3$	84.67%	95.14%	89.60%

Table 4. Results for Person Names Extraction Based on Sole SVM

2) Using Hybrid Model

As mentioned in section 4, the test samples which attribute to B class are misclassified to O class and therefore the Recall for person names extraction from sole SVM is lower. So we only deal with the test samples (B class and O class)

whose function distances to the hyperplane of SVM in feature space (i.e. $g(x)$) is between 0 and ε . We move class-boundary learned by SVM towards the O class, that is, the O class samples are considered as B class in that area. 93.64% of the Chinese person names in testing corpus are recalled when $\varepsilon=0.9$ (Here, ε also represents how much the boundary is moved). However, a number of non-person names are also identified as person names wrongly and the Precision is decreased correspondingly. Table 5 shows the Recall and Precision of person names extraction with different ε .

	Recall	Precision	F-measure
$\varepsilon=1$	93.05%	75.17%	83.16%
$\varepsilon=0.9$	93.64%	81.75%	87.29%
$\varepsilon=0.8$	93.51%	85.91%	89.55%
$\varepsilon=0.7$	93.05%	88.31%	90.62%
$\varepsilon=0.6$	92.39%	90.21%	91.29%
$\varepsilon=0.5$	91.81%	91.87%	91.84%
$\varepsilon=0.4$	91.02%	93.28%	92.13%
$\varepsilon=0.3$	90.56%	95.05%	92.75%
$\varepsilon=0.2$	90.03%	95.48%	92.68%
$\varepsilon=0.1$	88.66%	95.82%	92.10%

Table 5. Results for Person Names Extraction with Different ε

We use the evaluated function *TotalProbability(PN)* as described in section 3 to filter the wrongly recalled person names using SVM. We tune α in equation (5) to obtain the best results. The results based on the hybrid model with different α are listed in Table 6 (when $d=2$). We can observe that the result is best when $\alpha=0.4$. Table 7 shows the results based on the hybrid model with different ε when $\alpha=0.4$. We can observe that the Recall rises and the Precision drops on the whole when ε increases. The synthetic index F-measures are improved when ε is between 0.1 and 0.8 compared with sole SVM. The best result is obtained when $\varepsilon=0.3$. The Recall and the F-measure increases 3.27% and 1.77% respectively.

	Recall	Precision	F-measure
$\alpha=0.1$	90.37%	95.76%	92.99%
$\alpha=0.2$	90.37%	96.03%	93.11%
$\alpha=0.3$	90.43%	96.03%	93.15%
$\alpha=0.4$	90.43%	96.10%	93.18%
$\alpha=0.5$	90.63%	95.76%	93.13%
$\alpha=0.6$	90.43%	95.97%	93.12%

$\alpha=0.7$	90.43%	95.90%	93.09%
$\alpha=0.8$	90.43%	95.90%	93.09%
$\alpha=0.9$	90.37%	95.90%	93.05%

Table 6. Results for Person Names Extraction Based on The Hybrid Model with Different α

	Recall	Precision	F-measure
$\varepsilon=1$	92.53%	84.96%	88.58%
$\varepsilon=0.9$	93.05%	88.81%	90.88%
$\varepsilon=0.8$	92.86%	90.95%	91.89%
$\varepsilon=0.7$	92.46%	92.04%	92.25%
$\varepsilon=0.6$	91.93%	93.22%	92.58%
$\varepsilon=0.5$	91.48%	94.26%	92.85%
$\varepsilon=0.4$	90.76%	95.25%	92.95%
$\varepsilon=0.3$	90.43%	96.10%	93.18%
$\varepsilon=0.2$	90.04%	96.15%	92.99%
$\varepsilon=0.1$	88.73%	96.23%	92.32%

Table 7. Results for Person Names Extraction Based on The Hybrid Model ($\alpha=0.4$)

5.2 Extracting Chinese Location Names

We use 1.5M characters corpus of year 1998 from the People’s Daily as the training corpus and extract sentences of year 2000 from the People’s Daily (containing 2919 Chinese location names) as testing corpus to conduct an open test experiment. The results are obtained as follows based on different models.

1) Based on Sole SVM

The Recall, Precision and F-measure using different number of degree of polynomial kernel function are given in Table 8. The best result is obtained when $d=2$.

	Recall	Precision	F-measure
$d=1$	84.66%	91.95%	88.16%
$d=2$	86.69%	93.82%	90.12%
$d=3$	86.27%	94.23%	90.07%

Table 8. Results for Location Names Extraction Based on Sole SVM

2) Using Hybrid Model

The results for Chinese location names extraction based on the hybrid model are listed in Table 9 (when $d=2$; $\alpha=0.2$ in equation (10)). We can observe that the Recall rises and the Precision drops on the whole when ε increases. The synthetic index F-measures are improved when ε is between 0.1 and 0.7 compared with sole SVM. The best result is obtained when $\varepsilon=0.3$. The Recall increases 3.55%, the Precision decreases 1.05% and the F-measure increases 1.37%.

	Recall	Precision	F-measure
$\varepsilon=1$	90.75%	83.00%	86.71%
$\varepsilon=0.9$	90.85%	85.33%	88.01%
$\varepsilon=0.8$	91.42%	87.42%	89.37%
$\varepsilon=0.7$	91.65%	89.05%	90.33%
$\varepsilon=0.6$	91.75%	90.38%	91.06%
$\varepsilon=0.5$	91.32%	90.98%	91.15%
$\varepsilon=0.4$	90.66%	91.87%	91.26%
$\varepsilon=0.3$	90.24%	92.77%	91.49%
$\varepsilon=0.2$	89.10%	93.28%	91.15%
$\varepsilon=0.1$	87.83%	93.38%	90.52%

Table 9. Results for Location Names Extraction Based on The Hybrid Model ($\alpha=0.2$)

6 Comparison with other work

The same corpus was also tested using statistics-based approach to identify Chinese person names (Huang et al, 2001) and location names (Huang and Yue, 2003). In their systems, lexical reliability and contextual reliability were used to identify person names and location names calculated from statistical information drawn from a training corpus. The results of our models and the statistics-based methods (Huang 2001; Huang 2003) are shown in Table 10 for comparison. We can see that the Recall and F-measure in our method all increase a lot.

		Recall	Precision	F-measure
Person names	Our models	90.10%	96.15%	93.03%
	Huang (2001)	88.62%	92.37%	90.46%
Location names	Our models	90.24%	92.77%	91.49%
	Huang (2003)	86.86%	91.48%	89.11%

Table 10. Results of Our Method and Huang (2001; 2003) for Comparison

7 Conclusions and Future work

We recognize Chinese named entities using a hybrid model combining support vector machines with statistical methods. The model integrates the advantages of two methods and the experimental results show that it can achieve higher F-measure than the sole SVM and individual statistical approach.

Future work includes optimizing statistical models, for example, we can add the probability information of Chinese named entities in real texts to compute lexical probability, and we can

also use trigram models to compute contextual probability.

The hybrid model is expected to extend to foreign names in transliteration to obtain improved results by sole SVMs. The identification of transliterated names by SVMs has been completed (Li et al., 2004). The future work includes: set up statistical models for transliterated names and combine statistical models with SVMs to identify transliterated names.

References

- William J. Black and Argyrios Vasilakopoulos. 2002. Language Independent Named Entity Classification by Modified Transformation-based Learning and by Decision Tree Induction. *The 6th Conference on Natural Language Learning*, Taipei.
- Andrew Eliot Borthwick. 1999. A Maximum Entropy Approach to Named Entity Recognition. *PhD Dissertation*. New York University.
- Xavier Carreras, Lluís Marquez, and Lluís Padró. 2002. Named Entity Extraction Using AdaBoost. *The 6th Conference on Natural Language Learning*, Taipei.
- Michael Collins. 2002. Ranking Algorithms for Named-entity Extraction: Boosting and the Voted Perceptron. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-2002)*, Philadelphia, 489-496.
- Chooi-Ling Goh, Masayuki Asahara and Yuji Matsumoto. 2003. Chinese Unknown Word Identification Based on Morphological Analysis and Chunking. *The Companion Volume to the Proceedings of 41st Annual Meeting of the Association for Computational Linguistics (ACL-2003)*, Sapporo, 197-200.
- De-Gen Huang, Yuan-Sheng Yang, and Xing Wang. 2001. Identification of Chinese Names Based on Statistics. *Journal of Chinese Information Processing*, 15(2): 31-37.
- De-Gen Huang and Guang-Ling Yue. 2003. Identification of Chinese Place Names Based on Statistics. *Journal of Chinese Information Processing*, 17(2): 46-52.
- Thorsten Joachims. 1998. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *In Proceedings of the European Conference on Machine Learning*, 1398:137-142.
- Taku Kudo and Yuji Matsumoto. 2001. Chunking with Support Vector Machines. *In Proceedings of NAACL 2001*.
- Li-Shuang Li, Chun-Rong Chen, De-Gen Huang and Yuan-Sheng Yang. 2004. Identifying Pronunciation-Translated Names from Chinese Texts Based on Support Vector Machines. *Advances in Neural Networks-ISNN 2004, Lecture Notes in Computer Science*, Berlin Heidelberg, 3173: 983-988.
- Scott Miller and Michael Crystal. 1998. BBN: Description of the SIFT System as Used for MUC-7. *Proceedings of 7th Message Understanding Conference*, Washington.
- David D. Palmer. 1997. A Trainable Rule-Based Algorithm for Word Segmentation. *In Proc of 35th of ACL & 8th conf. of EACL*, 321-328.
- Wen Qin and Chun-Fa Yuan. 2004. Identification of Chinese Unknown Word Based on Decision Tree. *Journal of Chinese Information Processing*, 18(1): 14-19.
- Erik Tjong Kim Sang. 2002. Memory-based Named Entity Recognition. *The 6th Conference on Natural Language Learning*, Taipei.
- Koichi Takeuchi and Nigel Collier. 2002. Use of Support Vector Machines in Extended Named Entity Recognition. *The 6th Conference on Natural Language Learning*, Taipei.
- Vladimir N. Vapnik. 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, Berlin.
- Vladimir N. Vapnik. 1998. *Statistical Learning Theory*. John Wiley & Sons, New York.
- Ying Yu, Xiao-Long Wang, Bing-Quan Liu, and Hui Wang. 2004. Efficient SVM-based Recognition of Chinese Personal Names. *High Technology Letters*, 10(3): 15-18.
- Jun-Sheng Zhang, Shun-De Chen, Ying Zheng, Xian-Zhong Liu and Shu-Jin Ke. 1992. Large-Corpus-Based Methods for Chinese Personal Name. *Journal of Chinese Information Processing*, 6(3): 7-15.
- Guo-Dong Zhou and Jian Su. 2002. Named Entity Recognition Using an HMM-based Chunk Tagger. *Proceedings of the 40th Annual Meeting of the ACL*, Philadelphia, 473-480.