Single Character Chinese Named Entity Recognition

Xiaodan Zhu, Mu Li, Jianfeng Gao and Chang-Ning Huang
Microsoft Research, Asia
Beijing 100080, China
xdzhu@msrchina.research.microsoft.com
{t-muli,jfgao,cnhuang}@microsoft.com

Abstract

Single character named entity (SCNE) is a name entity (NE) composed of one Chinese character, such as *"中" (zhong1, China)* and *"俄"（e2,Russia）*. SCNE is very common in written Chinese text. However, due to the lack of in-depth research, SCNE is a major source of errors in named entity recognition (NER). This paper formulates the SCNE recognition within the source-channel model framework. Our experiments show very encouraging results: an F-score of 81.01% for single character location name recognition, and an F-score of 68.02% for single character person name recognition. An alternative view of the SCNE recognition problem is to formulate it as a classification task. We construct two classifiers based on maximum entropy model (ME) and vector space model (VSM), respectively. We compare all proposed approaches, showing that the source-channel model performs the best in most cases.

1 Introduction

The research of named entity recognition (NER) becomes very popular in recent years due to its wide applications and the Message Understanding Conference (MUC) which provides a standard test-bed for NER evaluation. Recent research on English NER includes (Collins, 2002; Isozaki, 2002; Zhou, 2002; etc.). Chinese NER research includes (Liu, 2001; Zheng, 2000; Yu, 1998; Chen, 1998; Shen, 1995; Sun, 1994; Zhang, 1992 etc.)

In Chinese NEs, there is a special kind of NE, called single character named entity (SCNE), on which there is little in-depth research. SCNE is a NE composed of only one Chinese character, such as the location name *"中" (zhong1,China)* and *"俄" (e2,Russia)* in the phrase *"中俄贸易" (zhong1-e2-mao4-yi4, trade between China and Russia)*. SCNE is very common in written Chinese text. For instance, SCNE accounts for 8.17% of all NE tokens according to our statistics on a 10MB corpus. However, due to the lack of research, SCNE is a major source of errors in NER. Among three state-of-the-art systems we have, the best F-scores of single character location (SCL) and single character person (SCP) are 43.63% and 43.48% respectively. This paper formulates the SCNE recognition within the source-channel model framework. Our results show very encouraging performance. We achieve an F-score of 81.01% for SCL recognition and an F-score of 68.02% for SCP recognition. An alternative view of the SCNE recognition problem is to formulate it as a classification task. For example, *"中"* is a SCNE in *"中俄贸易"*, but not in *"北京四中"(bei3-jing1-si4-zhong1, Beijing No.4 High School)*. We then construct two classifiers respectively based on two statistical models: maximum entropy model (ME) and vector space model (VSM). We compare these two classifiers with the source-channel model, showing that the source-channel model is slightly better. We then compare the source-channel model with other three state-of-the-art NER systems.

The remainder of this paper is structured as follows: Section 2 introduces the task of SCNE recognition and related work. Section 3 and 4 propose the source-channel model and two classifiers for SCNE recognition, respectively. Section 5 presents experimental results and error analysis. Section 6 gives conclusion.

2 SCNE Recognition and Related Work

We consider three types of SCNE in this paper: single character location name (SCL), person name (SCP), and organization name (SCO). Below are examples:

1. **SCL**: *"中"*and *"俄"* in *"中俄贸易"*
2. **SCP**: *"周" (zhou1, Zhou)* in *"周总理" (zhou1-zong3-li3,Premier Zhou),*
3. **SCO**: *"国" (guo2, Kuomingtang Party) and "共" (gong4, Communist Party) in "国共合作" (Guo2-gong4-he2-zuo4, Cooperation between Kuomingtang Part and Communist Party)*

SCNE is very common in written Chinese text. As shown in Table 1, SCNE accounts for 8.17% of all NE tokens on the 10MB corpus. Especially, 14.65% of location names are SCLs. However, due to the lack of research, SCNE is a major source of errors in NER. In our experiments described below, we focus on SCL and SCP, while SCO is not considered because of its small number in the data.

	# SCNE	# NE	#SCNE / #NE
PN	5,892	129,317	4.56%
LN	32,483	221,713	14.65%
ON	356	122,779	0.29%
Total	38,731	473,809	8.17%

Table 1. Proportion of SCNE in NE

To our knowledge, most NER systems do not report SCNE recognition results separately. Some systems (e.g. Liu, 2001) even do not include SCNE in recognition task. SCNE recognition is achieved using the same technologies as for NER, which can be roughly classified into rule-based methods and statistical-based methods, while most of state-of-the-art systems use hybrid approaches.

Wang (1999) and Chen (1998) used linguistic rules to detect NE with the help of the statistics from dictionary. Ji(2001), Zheng (2000), Shen(1995) and Sun(1994) used statistics from dictionaries and large corpus to generate PN or LN candidates, and used linguistic rules to filter the result, and Yu (1998) used language model to filter. Liu (2001) applied statistical and linguistic knowledge alternatively in a seven-step procedure. Unfortunately, most of these results are incomparable due to the different test sets used, except the results of Chen (1998) and Yu (1998). They took part in Multilingual Entity Task (MET-2) on Chinese, held together with MUC-7. Between them, Yu (1998)'s results are slightly better. However, these two comparable systems did not report their results on SCNE separately. To evaluate our results, we compare with three state-of-the-art system we have. These systems include: MSWS, PBWS and LCWS. The former two are developed by Microsoft® and the last one comes from by Beijing Language University.

3 SCNE Recognition Using an Improved Source-Channel Model

3.1 Improved Source-Channel Model[1]

We first conduct SCNE recognition within a framework of improved source-channel models, which is applied to Chinese word segmentation. We define Chinese words as one of the following four types: (1) entries in a lexicon, (2) morphologically derived words, (3) named entity (NE), and (4) factoid. Examples are

1. lexicon word: *朋友 (peng2-you3, friend).*
2. morph-derived word: *高高兴兴 (gao1-gao1-xing4-xing4* , *happily*)
3. named entity: *微软公司(wei1-ruan3-gong1-si1, Microsoft Corporation)*
4. factoid[2]: *1月9日(yi1-yue4-jiu3-ri4, Jan 9th)*

Chinese NER is achieved within the framework. To make our later discussion on SCNE clear, we introduce the model briefly.

We are given Chinese sentence S, which is a character string. For all possible word segmentations W, we will choose the one which achieves the highest conditional probability W* = $\text{argmax}_w P(W|S)$. According to Bayes' law and dropping the constant denominator, we acquire the following equation:

[1] This follows the description of (Gao, 2003).

[2] We define ten types of factoid: date, time (TIME), percentage, money, number (NUM), measure, e-mail, phone number, and WWW.

$$W^* = \arg\max_W P(W)P(S \mid W) \quad (1)$$

Following our Chinese word definition, we define word class C as follows: (1) each lexicon word is defined as a class; (2) each morphologically derived word is defined as a class; (3) each type of named entities is defined as a class, e.g. all person names belong to a class PN, and (4) each type of factoids is defined as a class, e.g. all time expressions belong to a class TIME. We therefore convert the word segmentation W into a word class sequence C. Eq. 1 can then be rewritten as:

$$C^* = \arg\max_C P(C)P(S \mid C). \quad (2)$$

Eq. 2 is the basic form of the source-channel models for Chinese word segmentation. The models assume that a Chinese sentence S is generated as follows: First, a person chooses a sequence of concepts (i.e., word classes C) to output, according to the probability distribution $P(C)$; then the person attempts to express each concept by choosing a sequence of characters, according to the probability distribution $P(S|C)$.

We use different types of channel models for different types of Chinese words. This brings several advantages. First, different linguistic constraints can be easily added to corresponding channel models (see Figure 1). These constraints can be dynamic linguistic knowledge acquired through statistics or intuitive rules compiled by linguists. Second, this framework is data-driven, which makes it easy to adapt to other languages. We have three channel models for PN, LN and ON respectively. (see Figure 1)

However, although Eq. 2 suggests that channel model probability and source model probability can be combined through simple multiplication, in practice some weighting is desirable. There are two reasons. First, some channel models are poorly estimated, owing to the sub-optimal assumptions we make for simplicity and the insufficiency of the training corpus. Combining the channel model probability with poorly estimated source model probabilities according to Eq. 2 would give the context model too little weight. Second, as seen in Figure 1, the channel models of different word classes are constructed in different ways (e.g. name entity models are *n*-gram models trained on corpora, and factoid models are compiled using linguistic knowledge). Therefore, the quantities of channel model probabilities are likely to have vastly different dynamic ranges among different word classes. One way to balance these probability quantities is to add several *channel model weight* CW, each for one word class, to adjust the channel model probability $P(S|C)$ to $P(S|C)^{CW}$. In our experiments, these weights are determined empirically on a development set.

Given the source-channel models, the procedure of word segmentation involves two steps: first, given an input string S, all word candidates are generated (and stored in a lattice). Each candidate is tagged with its class and the probability $P(S'|C)$, where S' is any substring of S. Second, Viterbi search is used to select (from the lattice) the most probable word segmentation (i.e. word class sequence C^*) according to Eq. 2.

Word class	Channel model	Linguistic Constraints
Lexicon word (LW)	$P(S\|LW)$=1 if S forms a lexicon entry, 0 otherwise.	Word lexicon
Morphologically derived word (MW)	$P(S\|MW)$=1 if S forms a morph lexicon entry, 0 otherwise.	Morph-lexicon
Person name (PN)	Character bigram	family name list, Chinese PN patterns
Location name (LN)	Character bigram	LN keyword list, LN lexicon, LN abbr. list
Organization name (ON)	Word class bigram	ON keyword list, ON abbr. List
Factoid (FT)	$P(S\|G)$=1 if S can be parsed using a factoid grammar G, 0 otherwise	Factoid rules (presented by FSTs).

Figure 1. Channel models (Gao, 2003)

3.2 Improved Model for SCNE Recognition

Although our results show that the source-channel models achieve the state-of-the-art word segmentation performance, they cannot handle SCNE very well. Error analysis shows that 11.6% person name errors come from SCP, and 47.7% location names come from SCL. There are two reasons accounting for it: First, SCNE is generated in a different way from that of multi-character NE. Second, the context of SCNE is different from other NE. For example, SCNE usually appears one after another such as *"中俄贸易"*. But this is not the case for multi-character NE.

To solve the first problem, we add two new channel models to Figure 1, that is, define each type of SCNE (i.e. SCL and SCP) as a individual class (i.e. NE_SCL and NE_SCP) with its channel probability $P(S_j \mid NE_SCL)$, and $P(S_j \mid NE_SCP)$. $P(S_j \mid NE_SCL)$ is calculated by Eq. 3.

$$P(S_j \mid NE_SCL) = \frac{|SCL(S_j)|}{\sum_{i=1}^{n} |SCL(S_i)|} \quad (3)$$

Here, S_j is a character in SCL list which is extracted from training corpus. $|SCL(S_j)|$ is the number of tokens S_j , which are labeled as SCL in training corpus. n is the size of SCL list, which includes 177 SCL. Similarly, $P(S_j \mid NE_SCP)$ is calculated by Eq. 4, and the SCP list includes 151 SCP.

$$P(S_j \mid NE_SCP) = \frac{|SCP(S_j)|}{\sum_{i=1}^{n} |SCP(S_i)|} \quad (4)$$

We also use two *CW* to balance their channel probabilities with other NE's.

To solve the second problem, we trained a new source model $P(C)$ on the re-annotated training corpus, where all SCNE are tagged by SCL or SCP. For example, *"周"* in *"周总理"*is tagged as SCP instead of PN, and *"中"* in *"中俄"* is tagged as SCL in stead of LN.

4 Character-based Classifiers

In this section, SCNE recognition is formulated as a binary classification problem. Our motivations are two folds. First, most NER systems do not use source-channel model, so our method described in the previous section cannot be applied. However, if we define SCNE as a binary classification problem, it would be possible to build a separate recognizer which can be used together with any NER systems. Second, we are interested in comparing the performance of source-channel models with that of other methods.

For each Chinese character, a classifier is built to estimate how likely an occurrence of this Chinese character in a text is a SCNE. Some examples of these Chinese character as well as their probabilities of being a SCNE is shown in Table 2.

SCL	#SCL/#All Tokens	SCP	#SCP/#All Tokens
沪	0.60465	冯	0.24342
俄	0.31612	阮	0.23810
蜀	0.23810	崔	0.13913
华	0.17516	乾	0.12500
港	0.12841	裴	0.11111
…	…	…	…
深	0.00454	易	0.00033
宁	0.00436	米	0.00030
加	0.00421	施	0.00030
孟	0.00261	常	0.00020
希	0.00228	安	0.00017

Table 2. The probability of a character as SCNE

We can see that the probabilities of being a SCNE of many characters are very small. Thus, SCNE recognition is an 'unbalanced' classification problem. That is, in most cases, it is safer to assume that a character is not a SCNE.

We construct two classifiers respectively based on two statistical models: maximum entropy model (ME) and vector space model (VSM). Local context characters (i.e. left or right characters within a window) are used as features.

4.1 Maximum Entropy

ME provides a good framework to integrate various features from different knowledge sources. Each feature is typically represented as a binary constraint *f*. All features are then combined using a log-linear model shown in Eq. 5.

$$P_\lambda(y \mid x) = \frac{1}{Z_\lambda(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right) \quad (5)$$

where λ_i is a weight of the feature f_i , and $Z(x)$ is a normalization factor.

Weights (λ) are estimated using the maximum entropy principle: to satisfy constraints on observed data and assume a uniform distribution (with the maximum entropy) on unseen data. The training algorithm we used is the improved iterative scaling (IIS) described in (Berger et al, 1996)[3]. The context features include six characters: three on the left of the SCNE, and three on the right. Given the context features, the ME classifier would estimate the probability of the candidate being a SCNE. In our example, we treat candidates with the probability larger than 0.5 as SCNEs. To get the precision-recall curve, we can vary the probability threshold from 0.1 to 0.9.

4.2 Vector Space Model

VSM is another model we used to detect SCNE. Similar to ME, we use six surrounding characters as the features, as shown in Figure 2.

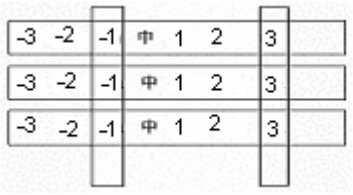

Figure 2. Context window

In this approach, we apply the standard *tf-idf* weighting technique with one minor adaptation: the same character appearing in different positions within the context window is considered as different terms. For example, character *Cj* appearing at position *i*, $i \in \{-3,-2,-1,1,2,3\}$, is regarded as term *Cji,*. Term weighting of *Cji* is acquired with Eq.6.

$$Wei(C_{ji}) = TF(C_{ji}) * IDF(C_{ji}) * PW_i \quad (6)$$

With this adaptation, we can apply an additional weighting coefficient *PWi* to different position, so as to reflect the importance of different positions. *PWi* is determined in a heuristic way as shown in Table.3 with the underlying principle that the closer the context character is to the SCNE candidate, the larger *PWi* is.

[3] Thank Joshua Goodman for providing the ME toolkit.

Pos	-3	-2	-1	1	2	3
PWi	1	4	7	7	4	1

Table 3. Weights assigned to deferent positions

A precision/recall curve can be obtained by multiplying a factor to one of the two cosine distances we get, before comparing them.

5 Experiment results

5.1 Evaluation Methodology

To achieve a reliable evaluation, we developed an annotated test set. First, we discuss a standard of Chinese NE and SCNE. Most previous researches define their own standards; hence results of different systems are not comparable. Recently, two widely accepted standards were developed. They are (1) MET-2 (Multilingual Entity Task)[4] for Chinese and Japanese NE, and (2) IEER-99'[5] for Chinese NE. IEER-99 is a slightly modified version of MET-2. Our NE/SCNE standard is based on these two well-known standards.

Second, we manually annotated a 10MB training corpus and a 1MB test corpus. The texts are randomly selected from People's Daily, including articles from 10 subjects and 5 writing styles. This test set is much larger than MET-2 test data (which is about 106 KB), and contains more SCNE for evaluation.

The evaluation metrics we used include precision (P), recall (R), and F-score. F-score is calculated as $\frac{(\beta^2+1.0)*P*R}{(\beta^2*P)+R}$, while β=1 in our experiments.

5.2 Results of Source-Channel Models

We show the SCNE recognition results using the source-channel models described in Section 3. Two versions of NE models are used. M1 is the original model described in Section 3.1. M2 is the one adapted for SCNE, shown in Section 3.2. The results in Table 4 show that obvious improvement can be achieved on SCL and SCP after adapting source-channel models for SCNE. As shown in Table 5, the improvement of SCL and SCP has significant impact on performance of LN and PN

[4] http://www.itl.nist.gov/iaui/894.02/related_projects/muc/
[5] http://www.nist.gov/speech/tests/ie-er/er_99/er_99.htm

recognition. We can see that the increase of F-score of LN is 5.13%, and PN is 0.92% absolutely.

	SCL			SCP		
	P%	R%	F	P%	R%	F
M1	83.77	15.25	25.80	*	*	*
M2	84.38	77.90	81.01	76.14	61.47	68.02

* No SCP is detected

Table 4. Improvement of SCL and SCP recognition

	LN			PN		
	P%	R%	F	P%	R%	F
M1	88.60	71.35	79.04	83.23	76.17	79.54
M2	88.20	80.49	84.17	83.51	77.63	80.46

Table 5. Improvement of LN and PN recognition

5.3 Results of Different Methods

In Figures 3 and 4, we compare the results of the source-channel models with two classifiers described in Section 4: ME and VSM.

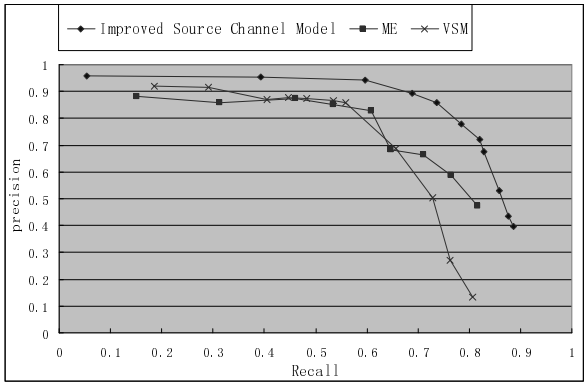

Figure 3. Result of different methods on SCL

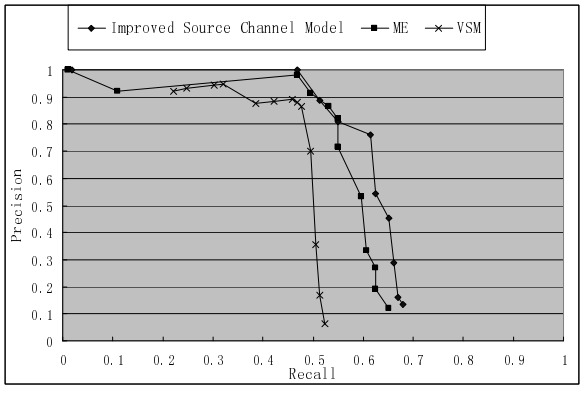

Figure 4. Result of different methods on SCP

We can see that source-channel model achieves the best result. This can be interpreted as follows. The source-channel models use more information than the other two methods. The feature set of ME or VSM classifiers includes only six surrounding characters while the source-channel models use much rich global and local information as shown in Figure 1. Based on our analysis, we believe that even enlarging the window size of the local context, the performance gain of these classifiers is very limited, because most error tags cannot be correctly classified using local context features. We can then say with confidence that the source-channel models can achieve comparable results with ME and VSM even if they used more local context.

5.4 Comparison with Other State-of-The-Art Systems

The section compares the performance of the source-channel models M2, with three state-of-the-art systems: **MSWS, LCWS** and **PBWS.**

1. The **MSWS** system is one of the best available products. It is released by Microsoft® (as a set of Windows APIs). **MSWS** first conducts the word breaking using MM (augmented by heuristic rules for disambiguation), then conducts factoid detection and NER using rules.
2. The **LCWS** system is one of the best research systems in mainland China. It is released by Beijing Language University. The system works similarly to **MSWS**, but has a larger dictionary containing more PNs and LNs.
3. The **PBWS** system is a rule-based Chinese parser, which can also output the NER results. It explores high-level linguistic knowledge such as syntactic structure for Chinese word segmentation and NER.

To compare the results across different systems, we have to consider the problem that they might have different tagging format or spec. For example, the LCWS system tags the two-character string *"中俄"* as a location name, and tags *"周总理"* other than *"周"* as a person name. We then manually convert all tagging results of these three systems according to our spec. The results are shown in Table 6.

	SCL			SCP		
	P%	R%	F	R%	P%	F
MSWS	*	*	*	71.42	31.25	43.48
LCWS	100.0	21.43	35.30	66.67	25.00	36.36
PBWS	92.31	28.57	43.63	100.0	18.75	31.58
M2	84.38	77.90	81.01	76.14	61.47	68.02

* No SCL is detected

Table 6. Comparison with other systems

We can see that our system (M2) achieves the best results in both SCL and SCP recognition. PBWS has the second best result in recognizing SCL (43.63%), and MSWS in SCP (43.48%). However, they achieved the worst result on SCP and SCL, respectively.

5.5 Error Analysis

Through human checking, we list the typical errors as follows:

1. Ambiguity between NE: *"美中贸易全国委员会"(mei3-zhong1-mao4-yi4-quan2-guo2-wei3-yuan2-hui4, National Committee on United States-China Relations)* is a ON, but *"美", "中"* are usually wrongly recognized as SCL. On the contrary, *"日中友好七团体"(ri4- zhong1-you3-hao3-qi1-tuan2-ti3, seven Janpan-China friendship organizations)* is tagged as ON falsely. So *"日"(ri4,Japan), "中"* are missed.
2. SCNE list acquire from training data cannot covers some cases in test data: *"中卡足球赛"(zhong1-ka3-zu2-qiu2-sai4, China-Qatar soccer match), "卡"(ka3, Qatar)* here is stand for *"卡塔尔"(ka3-ta1-er3, Qatar),* which is out of SCL list.
3. Other errors: *"中巴驶出大门"(zhong1-ba1-shi3-chu1-da4-men2, the middle bus drives out from the gate),"中"(zhong1, middle), "巴" (ba1, bus)*are recognized falsely as SCL. Because*"中"* and *"巴"* can also stand for *China* and *Pakistan."巴"* can even stand for other countries such as *"巴西"*（*ba1-xi1, Brazil*）.

Errors in (1) account for about 40% of all errors. SCNE is usually a part of multi-character NE, such as *"美", "中" in "美中贸易全国委员会"*.Viterbi search has to make a decision: recognizing the multi-character NE, or recognizing SCNE. Current features we used seem not powful enough to resovle this ambiguity well. Errors in (3) come from another kind of ambiguities such as ambiguity between SCNE and normal lexicon words. They are partly caused by noises in training data, because SCNE are very likely to be neglected by annotators, which makes training data more sparse. Both errors in (1) and (3) are not easy to handle.

Our immediate work is to cope with errors in (2), which account for about 8.9% of all errors. We can obtain additional SCNE entries from resources such as abbreviation dictionaries. However, the procedure to select SENE entries should be careful, because the SCNE characters we do not cover currently might be rare to act as SCNE, and difficult to recall. Besides, unsupervised methods can be applied to the task, considering insufficiency of the training data of the task.

6 Conclusion

Although SCNE is very common in written Chinese text, due to the lack of in-depth research, SCNE is a major source of errors in NER. This paper formulates the SCNE recognition within the source-channel model framework. Our experiments show very encouraging results: an F-score of 81.01% for single character location name recognition, an F-score of 68.02% for single character person name recognition. An alternative view of the SCNE recognition problem is to formulate it as a classification task. We construct two classifiers respectively based on maximum entropy model (ME) and vector space model (VSM), respectively. We compare all proposed approaches, showing that the source-channel model performs the best in most cases.

References

A. Berger, S. Della Pietra and V. Della Pietra, 1996. A maximum entropy approach to natural language processing. *Computational Linguistics* 22(1):39-71.

Andrew Borthwick, John Sterling, Eugene Agichtein, and Ralph Grishman. 1998. NYU: Description of the MENE Named Entity System as Used in MUC-7. *Proceedings of the Seventh Message Understanding Conference (MUC-7)*

Hsin-Hsi Chen, Yung-Wei Ding, Shih-Chung Tsai and Guo-Wei Bian, 1998. Description of the NTU System used for MET-2, *Proceedings of the Seventh Message Understanding Conference (MUC-7).*

Michael Collins, 2002. Ranking Algorithm for Named-Entity Extraction: Boosting and Voted Perceptron. *Proceedings of 40th Annual Meeting of Association for Computational Linguistics.*

Jianfeng Gao, Mu Li and Chang-Ning Huang, 2003. Improved Source-Channel Models for Chinese Word Segmentation. *Proceedings of 41th Annual Meeting of Association for Computational Linguistics*.

Hideki Isozaki and Hideto Kazawa, 2002. Efficient Support Vector Classifier for Named Entity Recognition, *Proceedings of 19th International Conference on Computational Linguistics.*

Heng Ji, Zhensheng Luo, 2001. Inverse Name Frequency Model and Rule Based Chinese Name Identification. (In Chinese) *Natural Language Understanding and Machine Translation*. Tsinghua University Press. pp. 123 -128.

Kaiying Liu, 2001. Research on Chinese Proper Noun and Internet Words Recognition. (In Chinese) *Proceedings of Conference of the 20th Anniversary of CIPSC*. Tsinghua University Press. pp. 7-13.

Song Rou, Benjamin K T'sou, 2001. Primary Study on Chinese Proper Noun. (In Chinese) *Proceedings of Conference of the 20th Anniversary of CIPSC.* Tsinghua University Press,14-19.

Dayang Shen, Maosong Sun, 1995. Chinese Location Name Recognition. (In Chinese) *Development and Applications of Computational Linguistics.* Tsinghua University Press.

Jian Sun, Jianfeng Gao, Lei Zhang, Ming Zhou, and Changning Huang, 2002. Chinese named entity identification using class-based language model. *COLING 2002*. Taipei, Taiwan, August 24-25, 2002.

Maosong Sun, Changning Huang, Haiyan Gao, Jie Fang, 1994. Chinese Person Name Recognition. (In Chinese) *Journal of Chinese Information Processing.*

Xing Wang, Degen Huang, Yuansheng Yang, 1999. Identifying Chinese Names Based on Combination of Statistics and Rules. (In Chinese) *proceedings of JSCL-99*. Tsinghua University Press. pp. 155 -161

Shihong Yu, Shuanhu Bai and Paul Wu. Description of the Kent Ridge Digital Labs System Used for MUC-7. *Proceedings of the Seventh Message Understanding Conference (MUC-7)*, 1998.

Junsheng Zhang, 1992. Chinese Person Name Recognition. (In Chinese) *Journal of Chinese Information Processing*. 9(2) .

Jiahen Zheng, Xin Li, Hongye Tan, 2000. The Researh of Chinese Names Recognition Based on Corpus,. (In Chinese) *Journal of Chinese Information Processing.* 14(1): 7-12

Guodong Zhou and Jian Su, 2002. Named Entity Recognition using an HMM-based Chunk Tagger. *Proceedings of 40th Annual Meeting of Association for Computational Linguistics*.