Tornado in Multilingual Opinion Analysis: A Transductive Learning Approach for Chinese Sentimental Polarity Recognition

Yu-Chieh Wu **Li-Wei Yang** **Jeng-Yan Shen** **Liang-Yu Chen** **Shih-Tung Wu**

Department of Research and Development
Tornado Technologies Co., Ltd
2F., No.6, Lane 11, Sec. 6, Mincyuan E. Rd., Neihu District
Taipei City 114, Taiwan

{yuchiehwu, willie_yang, yan, larrick_chen, walter_wu}@tornado.com.tw

Abstract

In this paper, we present our statistical-based opinion analysis system for NTCIR-MOAT track this year. Our method involves two different approaches: (1) the machine learning-based prototype system (on the basis of support vector machines (SVMs)) and (2) stochastic estimation of the character-level of words. The former were the real applications of state-of-the-art machine learning algorithms, while the latter comprises of ad-hoc opinioned word, phrase analysis. We submitted both two runs to NTCIR-MOAT in this year. The prototype system was first designed for traditional Chinese. We also directly port it to Simplified Chinese text with dictionary-based word translation. To make the model more robust, we present the idea of transdutive learning to our models. The main advantage of this approach is that it learns the hypothesis from labeled data meanwhile adapt to the large unlabeled data. Our method could not only be applied to SVM-based approaches, but also is applicable with the other non-machine learning algorithms.

The experimental results showed that our method (approach 1) can effectively identify the opinioned sentences in 0.661 and 0.611 f-measure rates under the lenient test. In terms of polarity judgment, our method achieves 0.284 and 0.294 in F-measure rates of the proposed two approaches, respectively. In the relevant sentence judgment track, our group achieved the best and the second best results among all other participants. Owing to the lack of labeled training data, we trust that our method could be further enhanced by feeding with more consistent and large annotated corpus.

Keywords: Opinion analysis, transductive learning, support vector machines

1. Introduction

In recent years, there has been a great volume of research reports in methods for sentiment analysis. NTCIR-MOAT (Seki et al., 2007) firstly gave the pilot study for the Asian opinion analysis, in particular Chinese and Japanese. The task of the MOAT this year focuses on five different sub-tasks, namely, opinioned sentence detection, sentence polarity recognition, source and target identification, and relevant sentence extraction. The dataset comprises of traditional and simplified Chinese, English, and Japanese. Participants received the sampled data and required to submit the results under the time constraints.

Extracting sentiment from text is a hard challenge with applications throughout natural language processing and machine learning techniques. Blogs, News, online discussion groups, BBS, and even more hybrids of the above sources could be the potential main roles for the sentiment analysis. A large set of past work on sentiment analysis has covered a wide range of tasks, including mining the comparative sentences (Jindal and Liu, 2006), identifying the opinion features (Hu and Liu, 2004), extracting the subjective expressions (Riloff and Wiebe, 2003), opinion source extraction (Choi et al., 2005; Choi et al., 2006).

In this paper, we focus on reporting the empirical results of the proposed two variant opinioned sentence analysis approaches on Chinese at NTCIR-7. The two methods integrate two well-known machine learning algorithms, i.e. SVM, and Naïve Bayes and different processing techniques. It is well-known that SVM is one of the state-of-the-art non-structural learning algorithms. Hence we merely make use of existing classification framework for the classification tasks. However, owing to the limit size of labeled data, we further adapt the typical supervised learning to semi-supervised prototypes that may benefit from large unlabeled data. A potential number of previous works (Joachims, 1999; Chapelle et al., 2008) also showed that the semi-supervised SVM significantly reduces testing errors than the original full supervised SVMs. In terms of the second analysis module, we integrate *N*-grams (a series of segmented Chinese words) into the Naïve Bayes classifier. On the basis of *N*-gram level of words, we design an ad-hoc approach for bagging *N*-grams and dictionary features into the Naïve Bayes classifier. In addition, we also present a transductive solution to adapt the trained Naïve Bayes classifier to be able to include the unlabeled data, especially testing examples. Preliminary experiment showed that transductive learning styles lead better empirical results in particular when the labeled data size is limited.

The remainder of this paper is organized as follows: Section 2 gives the overview of our sentiment analysis systems. In Section 3, the machine learning-based method is described. In Section 4, we present the transductive and non-learning-based approaches. Section 5, we report the experimental results of the two methods, SVM-based and statistical-based. In Section 6, we draw the conclusion and future works.

2. System Overview

Figure 1 illustrates the architecture of the designed Chinese opinioned analysis system. At the beginning, the Chinese text processing modules segments the input documents into sentence-level snippets by a set of pre-defined well-known Chinese punctuations, such as 。, ！, ？, ；, etc. Then, in the next stage, the word tokenizing and POS tagging module split the Chinese words from the sentence and give the part-of-speech (POS) labels for each word. These pre-processing steps give the important information for the downstream purposes. For the first run (Run 1 module), we apply the traditional text categorization method with semi-supervised learning algorithms, i.e., transductive support vector machines (TSVM), while the statistical analysis module performs the multi-pass solutions for determine the polarity of the given sentence. Both modules incorporate the polarity labels into learning. In other words, rather than determining the polarity after determining opinioned sentence, our methods directly recognize the sentence polarity.

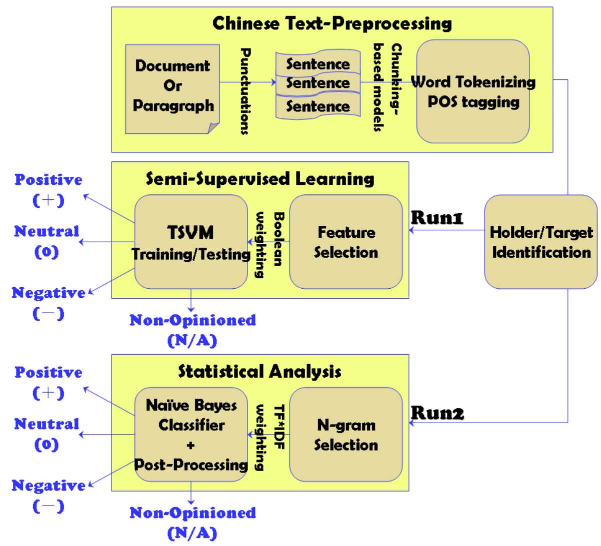

Figure 1: Overall system architecture

Table 1 summarizes the features of the submitted two runs for the Chinese (includes Traditional and Simplified versions) MOAT track. Other than the first and the last raw of the table, methods for holder/target identification and relevance judgment will be described in this Section.

Table 1: The system spec of our methods

Taiwan Tornado	Run 1	Run 2
Opinionated Sentence Detection	Word-based Transductive SVM with polynomial kernel	*N*-gram-based multinomial Naive Bayes
Holder and Target	Sequential labeling model	Heuristic search
Relevance	Language Model-based with two-stage smoothing method	TS42 with pseudo feedback
Polarity Judgment	Transductive SVM with polynomial kernel	*N*-gram-based multinomial Naïve Bayes + Post-processing

2.1. Chinese POS Tagging

Different from many western languages, there are no explicit boundaries between words in most far-eastern languages, such Chinese, Japanese, and Korean. Hence, almost the first step of Asian language processing technologies should resolve the word segmentation problems first (this is quite similar to the English word tokenization step).

There are two different ways for segmenting Chinese words, one is to apply a fixed-length word extraction (Wu et al., 2006; Wu et al., 2007), and the other is to employed a well-trained word segmentation tool (Levow, 2006). The fixed length word extraction approach defines the fixed word length, for example one (unigram), two (bigram), etc. Thus, a series of continuous tokens are grouped as a word. Usually, the overlapping bigram level of words was drastically and successfully employed in many Chinese information retrieval researches (Savoy, 2005; Min et al., 2005; Chen et al., 2005; Wu et al., 2006; Wu et al., 2007).

The second type of treating words is to adopt a well-trained word segmentation tool. In this paper, we had developed our own Chinese word segmentation and POS tagging modules based on the sequential labeling framework.

An example of different word segmentation strategy can be found in Table 2. In this paper, the Unigram and Bigram were selected for relevance judgment, while the auto-tagged words were used for opinionated sentence detection and classification.

Table 2: Samples of the segmented and POS-tagged Chinese sentences

Original sentence	911 攻擊對美國經濟的衝擊為何？
Unigram	911 攻 擊 對 美 國 經 濟 的 衝 為 何 ？
Bigram (overlapping)	911 攻擊 擊對 對美 美國 國經經濟 濟 的 的衝 衝擊 擊為 為何 ？
Word (auto-tagged)	911 攻擊 對 美國 經濟 的 衝擊為何 ？
POS tags	911(VE) 攻擊(VC) 對(P) 美國(Nc) 經濟 (Na) 的(DE) 衝擊(Na) 為(VG) 何 (Nes) ？(QUESTIONCATEGORY)

* the corresponding POS tags is the same as Academic Sinica's Treebank V3.1

2.2. Holder/Target Identification

In this paper, we employed the conventional sequential tagging models to recognize source (holder) and target in the given input strings. The step can also be viewed as recognizes named entity chunks in text. Figure 2 illustrates such idea.

As shown in Figure 2, the first column is the atomic Chinese character, while the last column indicates the source/target label of the character. For example, "無" with B-source means that this character is the first character of the opinion source. The second column in Figure 2 is the Chinese part-of-speech tag information. As reported by (Ng and Low, 2004; Wu, 2007), tagging Chinese words could be achieved by the sequential tagging models.

The learner we adopted in this paper is linear kernel support vector machines with L2-norm (DeCoste and Keerthi, 2005; Wu et al., 2007). Owing to the lake of sufficient labeled data, the official released sampled data is used to train our source/target detection models. In order to validate the suitable features, we randomly split 25% of the training data to validate. Once the optimal feature set is verified, the whole NTCIR-7 sampled data is feed into our model. Our Run 1 result was simply built based on the above steps.

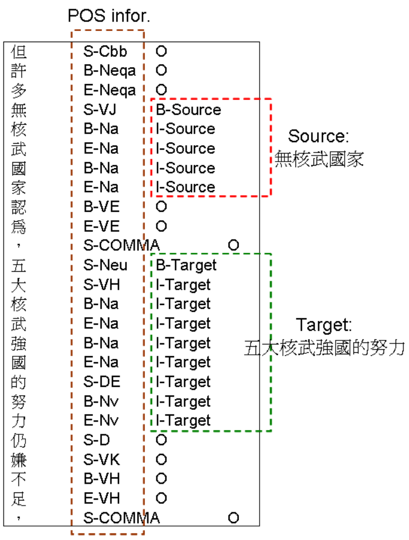

Figure 2: Example of holder/target detection

2.3. Relevance Judgment

The relevant sentence judgment can be viewed as retrieving related sentences to the given topic. Although the training-based methods often produce robust performance, sufficient labeled training data is required. On the contrary, the retrieval-based approaches which merely concern the input query produce very prominent results. Therefore, we adopt the retrieval-like strategy for relevance judgment task.

The method we used in this paper is quite simple. The query comprises the <DESC> description part of the given topic. Then the retriever returns a set of relevant sentences to the query.

For run 1, we adopted the state-of-the-art retrieval algorithms, i.e. language models as base implementation. The retrieval method of run 2 is an enhanced TFIDF models with regarding the word length and strength. We use the lemur[1] as our language model retriever and Tornado Search 5.0[2] to perform the second retrieval method. The two-stage Dirichlet smoothing strategy is used to smooth the language models (prior=10, lambda=0.4). In run 1, we only retrieve the highest 60% of the sentences as relevant for all topics. For run 2, we enlarge this ratio and perform various ratio of the retrieval rate.

3. Semi-Supervised Learning Module

3.1. Transductive Support Vector Machines

Assume we have a set of labeled training examples:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n),\ \ x_i \in \Re^D,\ y_i \in \{+1, -1\}$$

where x_i is a feature vector in D-dimension space of the i-th example, and y_i is the label of xi either positive or negative. Inductively training SVMs involves in minimize the following object (primal form, soft-margin) (Cortes and Vapnik, 1995):

$$\text{minimize}: W(\alpha) = \frac{1}{2}\vec{W}\cdot\vec{W} + C\sum_{i=1}^{n} Loss(\vec{W}x_i, y_i) \tag{1}$$

The loss function indicates the loss of training error. Usually, the hinge-loss is used. The factor C in (1) is a parameter that allows one to trade off training error and margin. A small value for C will increase the number of training errors.

To determine the class (+1 or -1) of an example x can be judged by computing the following equation.

$$y(x) = \operatorname{sign}\left(\left(\sum_{x_i \in SVs} \alpha_i y_i K(x, x_i)\right) + b\right) \tag{2}$$

α_i is the weight of training example x_i (α_i>0), and b denotes as a threshold. Here the xi should be the support vectors (SVs), and are representative of training examples. The kernel function K is the kernel mapping function, which might map from $\mathcal{R}^D$ to $\mathcal{R}^{D'}$ (usually D<<D'). The natural linear kernel simply uses the dot-product as (3).

$$K(x, x_i) = dot(x, x_i) \tag{3}$$

A polynomial kernel of degree d is given by (4).

$$K(x, x_i) = (1 + dot(x, x_i))^d \tag{4}$$

One can design or employ off-the-shelf kernel types for particular applications. In particular to the use of polynomial kernel-based SVM, it was shown to be the most successful kernels for many natural language processing (NLP) problems (Kudo and Matsumoto, 2001; Wu et al., 2007; Wu, 2007).

On the other hand, the TSVM has the following objective function.

$$\text{minimize}: W(\alpha) = \frac{1}{2}\vec{W}\cdot\vec{W} + C\sum_{i=1}^{n} Loss(\vec{W}x_i, y_i) + C^*\sum_{j=1}^{m} Loss(\vec{W}x_j, y_j)$$

Unfortunately, the above form is not convex and has no optimum solution. Also the label set of the unlabeled data y_j is obtained by first training the supervised SVM with the labeled data and get the non-real tags for the unlabeled parts. Usually, the ratio of the positive and

[1] http://www.lemurproject.org/
[2] Trial version is freely downloadable: http://reg.tornado.com.tw/ts50downloadform/form20080828.jsp

negative labels is a good criterion to project the labels on the unlabeled data. To make it more precise, we preserve only 95% of the positive training examples as truly positive, while the 5% with the remaining negative instances are treated as negative examples.

In this paper, we adopt the SVM^{Light} package (Joachims, 1999) for training the transductive SVM since it provide different kernel functions to optimize the transductive objective function. We also tried our L2-norm SVM (linear kernel) but the performance does not as good as L1-norm (polynomial kernel with d=1).

3.2. Preparing Training Data

Due to huge amount of inconsistent annotation among human labeler, the real training data is far difficult available. As indicated by opinion analysis pilot work (Seki et al., 2007), the agreement rate of the three human is almost 50%. In other words, the opinioned sentence will be accepted as opinioned sentence has the same opportunity as throwing two-sided coins. The reason is quite intuitive, i.e., the tagged source text is come from news articles (CIRB20, or CIRB40) which were written by a few certain journalists who were trained to make up detached news texts. Hence, the written words, phrases, sentences are relevant to some topics but rare opinioned.

Thus, to prepare training data, we adopt a dispirit strategy for the source/target identification module. That is the training data is come from annotator A. On the contrary, the collection of opinioned sentence detection and polarity judgment was derived from majority vote among the three annotators with the following orders, Pos > Neg > Neu > None.

For example, if a sentence is annotated as:

Pos/Man1 Pos/Man2 None/Man3

Then, the sentence is viewed as positive since the positive label receives the most annotation. Similarly, if the sentence is annotated as:

Pos/Man1 Neg/Man2 None/Man3

Then, the sentence is still viewed as positive even the positive, negative, and none got the same number of annotation. Such preference will enable us to bias to collect more training examples of positive label. The label order is specially designed for this purpose.

4. Statistical Analysis Module

The statistical approach is also performed for the purpose of comparison. For run1, we merely adopt the NTCIR-7 sampled data for training, while run2 additionally incorporate NTCIR-7 dataset to train. Each sentence in the training data set is re-assigned into one of the four categories: *N, Y-POS, Y-NEU, Y-NEG*, which stand for Non-opinionated, Opinionated-Positive, Opinionated-Neutral, and Opinionated-Negative, respectively.

In every category, each sentence is divided into bigrams. For a category C_k, the weight of the bigram B_i is calculated as followed:

$$W_{Bi} = \frac{1+\log F_{Bi}}{1+\log F_{Ck}} \bullet \log \frac{N_{all}}{N_{Bi}}$$

Where

F_{Bi} : the number of sentences containing B_i
F_{Ck} : the total number of sentences in C_k
N_{Bi} : the number of categories containing B_i
N_{all} : the total number of categories, which is 4.

Thus, given a new sentence S_j, the *similarity score* of S_j for a category C_k is calculated as followed:

$$S_{Sj,Ck} = \text{Average}(W_{Bi}) / T_{Ck}$$

Where $\text{Average}(W_{Bi})$ is the average of weight of every bigram occurred in both S_j and C_k ; T_{Ck} is the total number of unique bigram contained in C_k.

We label every new sentence into the category with the highest similarity score. For the relief to the problem of insufficient training data, the technique of *pseudo-feedback* is applied here. That is, we label every sentence in NTCIR-7 Opinion Analysis Task Traditional Chinese test data, and the sentences with similarity score higher than a threshold *t* are selected to enlarge the training data set. Then, every sentence in the test data set is labeled again.

5. Experiments

In this section, we firstly introduce the all settings used in this paper. Experimental results of the Chinese (traditional) run are presented in the following. NTCIR provide a series of evaluations for the system performance. We therefore directly report the actual accuracy on the participated tracks.

5.1. Settings

In this paper, we employed the SVMlight to perform polarity judgment task since it support different kernel functions. The feature set used in polarity judgment task is simply the bag-of-words (BOW). We had tried eliminated irrelevant words with chi-square scores, but the accuracy greatly decreases. By means of the union set of Chinese words (tokenized by our word segmentation tool), the optimal performance is obtained.

To prevent from over-training for our sequential labeling tagger, the rare features were eliminated from training. Here we filter the features that appear less than twice. The settings were mainly follows our previous works (Lee and Wu, 2007; Wu et al., 2007; Wu, 2007). On the other hand, for the settings of the SVM used in this thesis was based on the modified-finite Newton SVM (DeCoste and Keerthi, 2005). We replicate the original SVM-MFN algorithm by our own. Since the SVM is designed for classifying binary decision, to extend to multiclass, we apply the one-versus-all (OVA) type.

The feature set for the sequential labeling taggers are selected as a bottom-up feature search. First, a set of seed feature is used as the baseline performance. Then another set of candidate features are add at one time incrementally. If the added feature is useful (produce better performance), then it will be kept. Otherwise the algorithm goes next feature. To verify the system performance, 25% of the original training data is used.

5.2. Results of the Traditional Chinese Text

We submitted two runs: run1 (SVM-based) and run2 (statistical-based) approaches to NTCIR MOAT evaluation systems for all the sub-tasks, includes opinioned sentence detection, polarity judgment, source/target identification, and relevant sentence determination.

Table 3 lists the official results for the relevant sentence determination task. Our methods: TTRD run1, run2 clearly reached the best F-measure rates in comparison to the other groups. Even CHUK could surprisingly achieve 0.979 precision rate, the recall rate is substantially lower than us. In this experiment, we allow more relevant sentences for run2 to enhance the coverage of the run1. The trade-off between recall and precision is mainly dominated by the ratio of retrieval module. The larger the retriever extracts, the higher recall it obtains. Obviously, our group stands for the best trade-off in this task.

Second, the experimental result of the opinioned sentence detection is listed in Table 4. In this experiment, all groups achieve similar performance, i.e., 0.6~0.7. In this track, our methods reach moderate performance in comparison to the other groups. Compared to statistical-based approaches, the TSVM seem to be more effective in identifying opinioned sentence. Note that the training set for our TSVM is merely the sample data which provided by NTCIR. As noted by (Xu et al., 2007), development of additional training examples is very useful. We trust our method could be enhanced by feeding with larger labeled examples.

Table 3: Official results of the relevant sentence determination

Relevance Judgment	Opinionated (Lenient)			Opinionated (Strict)		
	P	R	F	P	R	F
CHUK	0.979	0.406	0.574	0.995	0.531	0.692
CityUHK	N/A	N/A	N/A	N/A	N/A	N/A
iclpku	0.953	0.563	0.708	0.994	0.677	0.805
NLCL	0.857	0.411	0.556	0.928	0.485	0.637
NTUCopeOpi	0.885	0.644	0.745	0.961	0.710	0.817
TTRD Run1	0.899	0.804	0.849	0.967	0.841	0.900
TTRD Run2	0.889	0.868	0.878	0.966	0.897	0.930

Table 4: Official results of the opinioned sentence detection

Opinioned Sentence	Opinionated (Lenient)			Opinionated (Strict)		
	P	R	F	P	R	F
CHUK	0.730	0.521	0.608	0.852	0.600	0.704
CityUHK	0.652	0.870	0.745	0.818	0.922	0.867
iclpku	0.701	0.628	0.663	0.857	0.700	0.770
NLCL	0.476	0.742	0.580	0.583	0.741	0.653
NTUCopeOpi	0.565	0.897	0.693	0.708	0.931	0.804
TTRD Run1	0.511	0.934	0.661	0.645	0.939	0.765
TTRD Run2	0.566	0.662	0.611	0.738	0.674	0.705
UniNe	0.543	0.927	0.685	0.692	0.938	0.796

Next, the opinioned sentence polarity judgment result is shown in Table 5. As described above, our method for the opinioned polarity judgment is one-pass, i.e., our TSVM directly predict the polarity label for an input sentence whether the sentence is non-opinioned. Owing to the lake of evaluated sentences, both our run1 and run2 result yield worse accuracy than the other groups. The main reason is that the abnormal high inconsistent annotation for the labeled dataset. Due to the low agreement rate, it is very difficult to acquire the real label for each training sentence. Table 6 and Table 7 show the source and target detection results of our methods.

Table 5: Official results of the sentence polarity judgment

Polarity Judgment	Polarity (Lenient)	# Evaluated	Polarity (Strict)	# Evaluated
CHUK	0.421	1137	0.493	754
CityUHK	0.397	1843	0.475	1135
Iclpku	0.319	1370	0.392	879
NLCL	0.338	1957	0.400	1169
NTUCopeOpi	0.248	2039	0.300	1180
TTRD Run1	0.284	1440	0.335	847
TTRD Run2	0.294	2022	0.353	1178
UniNe	0.421	1137	0.493	754

Table 6: Official results of the source detection

Holder detection	Holder (Lenient)			Holder (Strict)		
	P	R	F	P	R	F
CHUK	0.825	0.825	0.825	0.824	0.824	0.824
iclpku	0.599	0.599	0.599	0.582	0.582	0.582
NTUCopeOpi	0.503	0.503	0.503	0.482	0.482	0.482
TTRD Run1	0.565	0.565	0.565	0.550	0.550	0.550
TTRD Run2	0.595	0.595	0.595	0.584	0.584	0.584

Table 7: Official results of the target detection

Target detection	Target (Lenient)			Target (Strict)		
	P	R	F	P	R	F
CHUK	0.606	0.606	0.606	0.633	0.633	0.633
iclpku	0.021	0.021	0.021	0.023	0.023	0.023
TTRD Run1	0.036	0.036	0.036	0.037	0.037	0.037
TTRD Run2	0.107	0.107	0.107	0.136	0.136	0.136

5.3. Results of the Simplified Chinese Text

To port our method to Simplified Chinese, we need to convert not only the encoding style, but also the word collocations. In this paper, we adopt …… Table 8, Table 9, and Table 10 list the official results of our methods on Simplified Chinese text.

Table 8: Official results of the relevant sentence determination in Simplified Chinese

Relevance Judgment	Relevance (Lenient)			Relevance (Strict)		
	P	R	F	P	R	F
ICLPKU	0.978	0.656	0.785	0.985	0.674	0.800
NLCL	0.971	0.585	0.730	0.983	0.590	0.737
NLPR	0.951	1.000	0.975	0.963	1.000	0.981
NTU	0.966	0.769	0.856	0.975	0.786	0.870
TTRD Run1	0.951	0.698	0.805	0.963	0.701	0.811
TTRD Run2	0.968	0.736	0.836	0.976	0.749	0.847
ISCAS	0.970	0.929	0.949	0.983	0.937	0.959

6. Conclusion

In this paper, we present two variant transductive learning strategies for MOAT track at NTCIR-7. In addition to the word segmentation and POS tagging module, all of our methods were designed based on the sampled released data from NTCIR. Owing to the limit training data, our group could be improved by feeding with larger labeled data. But the experimental result gave surprising well performance in comparison to the other participants who includes far more training examples than us. This is the first time we join the MOAT track. Although many of the linguistic resources are not easily available, we try to establish the basic prototype with limited training data.

Table 9: Official results of the opinioned sentence detection in Simplified Chinese

Opinioned Sentence	Opinionated (Lenient)			Opinionated (Strict)		
	P	R	F	P	R	F
BUPT	0.604	0.399	0.481	0.631	0.442	0.520
ICLPKU	0.480	0.800	0.600	0.449	0.821	0.580
NEUNLP	0.472	0.712	0.568	0.436	0.734	0.547
NLCL	0.432	0.699	0.534	0.367	0.706	0.483
NLPR	0.582	0.775	0.665	0.610	0.892	0.724
NTU	0.594	0.609	0.601	0.631	0.752	0.686
TTRD Run1	0.412	0.964	0.577	0.348	0.970	0.512
TTRD Run2	0.446	0.756	0.561	0.396	0.755	0.519
ISCAS	0.465	0.744	0.572	0.427	0.812	0.560

Table 10: Official results of the sentence polarity judgment in Simplified Chinese (recall-based evaluation)

Polarity judgment	Opinionated (Lenient)			Opinionated (Strict)		
	P	R	F	P	R	F
ICLPKU	0.216	0.361	0.271	0.127	0.233	0.165
NTU	0.307	0.313	0.310	0.237	0.280	0.256
TTRD Run1	0.179	0.419	0.251	0.100	0.280	0.148
TTRD Run2	0.220	0.374	0.277	0.155	0.296	0.204

References

[1] O. Chapelle, V. Sindhwani and S.S. Keerthi. Optimization Techniques for Semi-Supervised Support Vector Machines. Journal of Machine Learning Research 9, pp. 203-233, 2008.

[2] J. Chen, R. Li, and F. Li. Chinese Information Retrieval Using Lemur. In Proc. of the NTCIR-6 Workshop Meeting, 2005.

[3] Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. Identifying Sources of Opinions with Conditional Random Fields. In Proc. of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 355-362, 2005.

[4] Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. Joint Extraction of Entities and Relations for Opinion Recognition. In Proc. of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 431-439, 2006.

[5] C. Cortes and V. Vapnik. Support-Vector Network. Machine Learning 20, 273–297, 1995.

[6] M. Hu and B. Liu. Mining Opinion Features in Customer Reviews. Proceedings of Nineteeth National Conference on Artificial Intelligence (AAAI), pp. 755-760, 2004.

[7] N. Jindal, and B. Liu. Identifying Comparative Sentences in Text Documents. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). pp. 244-251, 2006.

[8] T. Joachims. Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, MIT Press, pp. 169-184, 1999.

[9] T. Joachims, Transductive Inference for Text Classification using Support Vector Machines. International Conference on Machine Learning (ICML), pp. 200-209, 1999.

[10] S. Keerthi and D. DeCoste. A Modified Finite Newton Method for Fast Solution of Large Scale Linear SVMs. Journal of Machine Learning Research. 6: 341-361, 2005.

[11] T. Kudo and Y. Matsumoto. Chunking with Support Vector Machines. In Proceedings of the Meeting of North American Chapter of Association for Computational Linguistics (NAACL), pp. 192-199, 2001.

[12] Y.S. Lee and Y.C. Wu. A Robust Multilingual Portable Phrase Chunking System. Expert Systems with Applications, 33(3): 1-26, 2007.

[13] G.A. Levow. The Third International Chinese Language Processing Bakeoff: Word Segmentation and Named Entity Recognition. In Proc. of the 5th SIGHAN Workshop on Chinese Language Processing. pp. 108-117, 2006.

[14] J. Min, L. Sun, and J. Zhang. ISCAS in English-Chinese CLIR at NTCIR-5. In Proc. of the NTCIR-6 Workshop Meeting, 2005.

[15] H.T. Ng and J.K. Low. Chinese Part-of-speech Tagging One-at-a-time or All-at-once? Word-based or Character-based? In Proc. of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (EMNLP), 2004.

[16] E. Riloff and J. Wiebe. Learning Extraction Patterns for Subjective Expressions. In Proc. of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 105-112. 2003

[17] J. Savoy. Comparative Study on Monolingual and Multilingual Search Models for Use with Asian Languages. ACM trans. on Asian language information processing. 4(2): 163-189, 2005.

[18] Y. Seki, D.K. Evans, L.W. Ku, H.H. Chen, N. Kando, and C.Y. Lin. Overview of Opinion Analysis Pilot Task at NTCIR-6. In Proc. of the NTCIR-6 Workshop Meeting, pp. 265-278, 2007.

[19] Y.C. Wu, J.C. Yang, and Q.X. Lin. Description of the NCU Chinese word segmentation and named entity recognition system for SIGHAN Bakeoff 2006. In Proc. of the 5th SIGHAN Workshop on Chinese Language Processing. pp. 209-212, 2006.

[20] Y.C. Wu, J.C. Yang, and Y.S. Lee. Description of the NCU Chinese Word Segmentation and Named Entity Recognition System for SIGHAN Bakeoff 2007. In Proceedings of the 6th SIGHAN Workshop on Chinese Language Processing. pp. 161-166, 2008.

[21] Y.C. Wu, J.C. Yang, and Y.S. Lee. An Approximate Approach for Training Polynomial Kernel SVMs in Linear Time. In Proc. of the Annual Meetings of the Association for Computational Linguistics (ACL), pp. 65-68, 2007.

[22] Y.C. Wu. Incorporating Top-level and Bottom-level Information for Chinese Word Dependency Analysis. PhD thesis, National Central University, 2007.

[23] R. Xu, K.F. Wong, and Y. Xia. Opinmine-Opinion Analysis System by CUHK for NTCIR-6 Pilot Task. In Proc. of the NTCIR-6 Workshop Meeting, pp. 350-357, 2007.

[24] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. In Proc. of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR). pp. 334-342, 2001.