Journal of Computational Information Systems3:5(2007) 1-6
Available at http://www.JofCI.org

A Study on Deception Detection Based on Classification for Chinese Text

Hu ZHANG† , Shan-de WEI, Hong-ye TAN, Jia-heng ZHENG

School of Computer &Information Technology, Shanxi University, Taiyuan 030006, China

Abstract

Deception detection on Chinese text is vital to the safety of people's life, the survival of enterprises and the stability of the country. The expansion of the Internet has significantly increased the amount of textual communication received and stored by individuals and organizations. Inundated with massive amounts of textual information transmitted through Computer-mediated Communication (CMC), people remain largely unsuccessful and inefficient in detecting those messages that may be deceptive. Proposing an automated deception detection method that could help people flag the possible deceptive messages in CMC is desirable, but first it is necessary to construct the deceptive and non-deceptive Chinese text corpora, which have not been published so far. Further, according to the corpora analysis results we put forward a novel classification-based method on deception detection for Chinese text. Our method, along with feature selection and parameter adjustment, respectively conduct deception detection experiments by using the Bayes classifier and SVM classifier. The experimental results show the precision rate, the recall rate and the F-value may achieve 79.6%, 75% and 0.77 separately by using SVM classifier in open test.

Keywords: Deception Detection; CMC; Classification; Chinese Text

1. Introduction

Deception has been studied widely in many fields of social science, and it is defined as the active transmission of messages and information to create a false conclusion [1-3]. With the development of computer technology and network technology, CMC(Computer-mediated Communication) has changed and improved our everyday life, bringing with it new venues for deception. How to detect deception from amounts of electronic text is very important to the safety of people's life, the survival of enterprises and the stability of the country. Therefore, the research of deception detection on Chinese Text-based CMC is of great significance and application value. CMC can be classified into text, audio, audio/video, and multi-media based formats. Text-based CMC uses only written forms without audio and video signals. The majority of information transferred through Internet is in the written forms, such as E-mail, Web text etc., so we focus on the research of deception detection in Chinese Text-based CMC (CTCMC) in this paper.

Up to now, there is little attention paid to deception detection on abroad, and the research is still on the initial stage on the whole. At present, the researches mainly focus on the two aspects: the theory research and experiment research of deception detection.

The theory research of deception was launched earlier, and some theories have become the theoretical foundation of deception detection and are used to build experiment hypothesis. They are listed as follows: ①Media richness theory (MRT) [4,5], developed by Daft R. and Lengel R. (1986), suggests that the communication media vary in capacity to transmit and process the "rich" information. The media's richness is the function of time needed for communicating sides to enable understanding or overcome different

† Corresponding author.
Email addresses: zhanghu@sxu.edu.cn (H. Zhang).

June, 2007

perspectives; ②Social presence theory(SPT) [4,6] was developed by Short et al. (1976) and focuses on the degree to which communicating parties in mediated environment sense or perceive one another in terms of being a “real person.”; ③Channel expansion theory(CET) [4,7] that was developed by Carlson J. and Zmud R.(1999), suggests that media richness may be dependent on the experience that a particular individual has had with a particular channel; ④Interpersonal deception theory(IDT) [4,8] that was proposed by Buller D. and Burgoon J. (1996) , studies deception as a communication activity, i.e. “how social interaction alters deception and how deception alters social interaction.”.

At present there are 8 experimental researches that provide experimental data to verify theory hypothesis on deception detection. Each research investigated the deception and its’ detection from the different angles, but the referred specific issues, the hypothesis, the used methodology and the concepts of the researches are different. Among them the 4 studies of Blair etc. (2005), George J. and Marett K. (2004), George etc. (2004), Marett K. and George J.(2005) investigated the effect on the occurrence of deception and detection in different communication forms [9-11]. The conclusions are listed as follows: ①Suspicious receivers have a higher accuracy of deception detection, and the more suspicious receivers will accept the less deceptive information; ②People have facticity bias when they detect deception, and think truthful information is more than deceptive information; ③In CMC, it is easier for deceiver to release more deceptive information.

The other three researches of Zhou L. and Zhang D. (2004), Hancock etc. (2005) and Zhou etc. (2003) emphatically analyzed effective deceptive cues [12-14]. The conclusions are drawn: ①In deceptive communication people use more words, for example, senders use more words in each sentence, and receivers ask more questions. ②Senders use less singular personal pronoun and more third personal pronoun. ③Senders use more perceptive words and negative words, such as “see”, “listen” and so on. ④ Deceivers use more verbs, modifiers and noun phrases to increase the complexity of the information.

At present the research on deception detection has not begun yet in China, so in our research we firstly construct the deceptive and non-deceptive Chinese text corpora, and analyze their characteristics and differences between the deceptive and non-deceptive corpora, and then propose a novel classification-based deception detection method of Chinese text based on the analysis results. Our method, along with feature selection and parameter adjustment, respectively uses the Bayes classifier and SVM classifier to conduct deception detection. The experimental results indicate the precision rate, the recall rate and the F-value may achieve 79.6%, 75% and 0.77 separately by using SVM classifier in open test.

The rest of the paper is organized as follows. Section 2 introduces the used corpora including corpus scale, construction principle and so on. Section 3 describes the proposed deception detection model. Section 4 introduces two classification models. Section 5 discusses the experimental results. Finally, the concluding remarks are given in Section 6.

2. Corpus Construction

The construction of high-quality and large-scale corpora has always been a fundamental research area in the field of Chinese natural language processing. The corpora have provided the rich language phenomenon for the linguistics researches, and also have provided the full and accurate linguistics information data for computational linguistics scholar who can acquire linguistics knowledge, construct the linguistics model and research NLP technology by them. At present the deceptive and non-deceptive Chinese text corpora have not been reported yet, so in the experiment initial period we constructed the deceptive and non-deceptive corpora.

2.1. Corpus Construction Principle

Presently we construct the corpora by downloading the deceptive and non-deceptive texts from Internet based on the following principles.

①Deception is defined as the active transmission of messages and information to create a false conclusion, so we grasp the “active” and “false conclusion” to collect the corpus;

②The corpora involves the sport, the entertainment, social life and so on;

③The partial corpus come from the downloaded news. We distinguish the deceptive news from the non-deceptive news by the appearance backgrounds and the investigation results of the news;

④The partial corpus come from the BBS. According to the following commentary or some official news we distinguish the deceptive topics from the non-deceptive topics to some nondescript topics;

⑤In order to assure the relativity between the deceptive corpora and the non-deceptive corpora, we collect the content correlation deceptive corpus and non-deceptive corpus.

In the future research we will design the deceptive data artificially. According to some related literatures [12-14], we may use the existing resources to construct some typical deceptive topics data set in our group or our grade as far as possible.

2.2. Corpus Scale

According to the above principle we downloaded the most corpus confirmed to be deceptive from Internet, which is from 2001 to 2008 and includes 153 texts. Simultaneously, we also collected the non-deceptive corpus related to deceptive texts content, which includes 229 texts. Obviously the corpora scale summing up 350 000 Chinese characters is not very satisfying, but in fact we have collected the most corpus confirmed to be deceptive from 2001 to 2008, so the corpus construction, especially artificial construction will play a important role in the future research. In this deception detection experiment, we randomly select 75 deceptive texts and 106 non-deceptive texts as the training corpora, and 78 deceptive texts and 123 non-deceptive texts as the test corpora.

3. Deception Detection Model

3.1. Model Expression

The deception detection model includes three main parts: Feature selection, Model training and Deception detection. The process is showed in the following Fig.1:

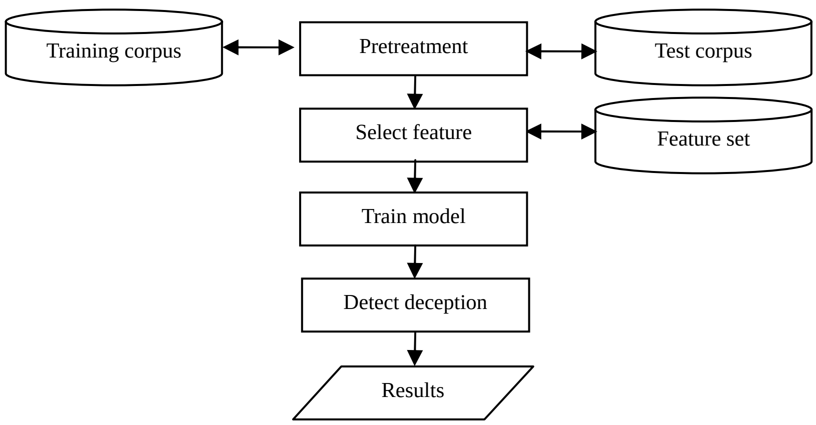

Fig.1 Deception Detection Flowchart

3.2. Problem Transforming

According to the deception detection method that we proposed, the deception detection question can be transformed into a two-classified question with being shown in Fig.2, which also is to flag the measured sample into deceptive or non- deceptive.

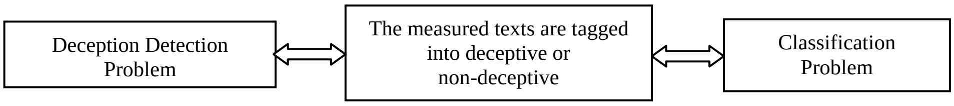

Fig.2 Problem Transforming

In this paper we respectively choose the Bayes classifier and SVM classifier to carry on the deception detection.

4. Classification Model

4.1. Support Vectors Machine

SVM classification essentially is to seek a most superior classification hyperplane, it not only may correctly divide the assigned input samples into two classes, but also make the classes interval be big as far as possible. Suppose we have the training samples $(x_1, y_1), \cdots, (x_l, y_l)$, $x \in R^n, y \in \{+1, -1\}$, where l is the number of the training samples and n is the number of the input dimensions, there must be a hyperplane that make two classes samples to separate completely in the linear separable situation. This hyperplane may be described as follows:

$$(\omega \bullet x) + b = 0$$

There are some useful Kernel Functions of SVM:

①Polynomial Kernel Function: $K(\boldsymbol{x}, \boldsymbol{x}_i) = [(\boldsymbol{x} \cdot \boldsymbol{x}_i) + 1]^d$, where *d* is natural number;

②RBF Kernel Function: $K(\boldsymbol{x}, \boldsymbol{x}_i) = \exp\left\{ -\frac{|\boldsymbol{x} - \boldsymbol{x}_i|^2}{\sigma^2} \right\}$, σ >0;

③Sigmoid Kernel Function: $\tanh(a(\boldsymbol{x} \cdot \boldsymbol{x}_i) + t)$, *a* and *t* are constant, *tanh* is sigmoid function.

Based on Y.Yang et al. (1999) [15] experimental results, the SVM classification performance is better than NNet, Rocchio and LLSF classifier and so on. Therefore, it is feasible theoretically that we choose the SVM method to carry on the deception detection.

4.2. Bayes Classifier

Bayes classifiers are a class of simple probabilistic algorithms which apply Bayes' theorem in order to learn the underlying probability distribution of the data (with a few simplifying assumptions, hence the 'Naive'). In our case, each sample is taken to be a unique variable in the model, and the goal is to find the probability of the sample, and consequently the quote itself, belonging to a certain class: deceptive vs. non-deceptive.

4.3. Feature Selection

The number of text words is as high as thousands or even ten thousands after conducting pretreatment to the training texts, so we have to select more effective features to reduce dimensions. There are two main aims for that:

①Improving the efficiency and the speed of the procedures;

②Improving the accuracy of detection deception.

Different words have the different contributions to deception detection. Some universal words have smaller contribution, but some other words, which are more general in some specific categories and are more infrequent in other categories, have more contributions. Therefore, each text can be expressed by the selected feature items that have more contributions to the deception detection by feature selection. In this paper we use the CHI statistics to carry on the feature selection.

CHI method measures the relativity between word *t* and documents category *c*. Having supposed there is χ^2 distribution with a first-order freedom degree between *t* and *c*, the χ^2 statistics of the word represents the word contribution to the category. The χ^2 statistics value is higher, the independence is smaller and the relativity is stronger between the word and the category, i.e. the word contribution to this category is bigger. The CHI value of word *t* to documents category *c* is computed according to the following formula.

$$\chi^2(t,c) = \frac{N(AD-CB)^2}{(A+C)(B+D)(A+B)(C+D)}$$

Where *t* represents one word and *c* represents one document category; *N* is the total number of the texts in training corpora; *A* is the number of the texts that belong to category *c* and contain word *t*; *B* is the number of the texts containing word *t* but not belonging to category *c*; *C* is the number of the texts belonging to category *c* but not containing word *t*; *D* is the number of the texts that don't belong to category *c* and don't contain word *t*.

After computing the CHI values of all words we remove some words with the smaller CHI value according to the threshold. At the same time the text representation is also the foundation of deception detection, we use the vector space model (VSM) to represent the texts in this paper.

5. Experimental Results and Discussions

5.1. Evaluation Measures

In this paper we use the precision rate(P), the recall rate(R) and the F- value(F) to evaluate the deception detection results, they are defined as follows:

$$P = \frac{\text{Number of the detected correct deceptive texts}}{\text{Number of the detected deceptive texts}}$$

$$R = \frac{\text{Number of the detected correct deceptive texts}}{\text{Number of the deceptive texts}}$$

$$F = \frac{2\times P\times R}{P+R}$$

5.2. Feature Selection

In this paper we use the CHI statistics to carry on the feature selection, the total number of the candidate features is 11 516. After computing the CHI value of every feature word respectively we count the total number of the feature words in the different feature value range. The total number of the feature words whose feature value is grater than 10 is 7, and the more results are showed in the following Table 1.

Table.1 Feature Value Range and the Corresponding Total

Feature value range (Feature value>=)	10	9	8	5	4	3	2	1
Total	7	13	22	174	206	693	923	5561

In the following Table 2 we list some feature words and their corresponding feature values according to the feature value order.

Table.2 Feature Items and Feature Value

ID	Feature Item	Feature Value	ID	Feature Item	Feature Value
1	因为	19.060226440429688	11	专家	9.2769670486450195
2	告诉	12.652788162231445	12	很多	9.2769670486450195
3	所以	10.974848747253418	13	十分	9.1203346252441406
4	东西	10.494000434875488	14	由于	8.8274164199829102
5	打开	10.494000434875488	15	就	8.7441644668579102

6	下来	10.494000434875488	…	…	…
7	了解	10.060426712036133	36	!	6.90209484100342
8	才	9.4416007995605469	…	…	…
9	就是	9.3290081024169922	70	他们	6.3820481300354
10	左右	9.3005027770996094	…	…	…

Based on the three researches of Zhou L. and Zhang D. (2004), Hancock etc. (2005) and Zhou etc. (2003) and the analysis of the prior deceptive corpora we found that the modifier, personal pronoun and punctuations are all the important features that can differentiate the deceptive texts and non-deceptive texts. Hence, we don't remove the stop-words, and the selected feature words include the adverb, personal pronoun, punctuations and so on.

5.3. Bayes Experimental Results

We conducted deception detection experiments by using the different numbers of the features, the results are showed in Table 3.

Table.3 Bayes Experimental Results.

Number of the features	Closed Test			Open Test		
	P	R	F	P	R	F
174	52%	97.5%	0.67826	50%	100%	0.66667
206	57.353%	97.5%	0.72222	50%	100%	0.66667
693	80%	100%	0.88889	50%	100%	0.66667
923	66.667%	100%	0.80000	50%	100%	0.66667
1047	58.824%	100%	0.74074	50%	100%	0.66667

The experimental results show the method acquire better results when the number of the features is 693.

5.4. SVM Experimental Results

In this deception detection process, we choose the words whose feature value is bigger than 3 as the feature items according to the Bayes experimental results; The LIBSVM tools used are simple, wieldy, fast and effective software package that is developed by Lin etc. in Taiwan University. In our experiment we have separately attempted 3 different kernel functions, including polynomial kernel, RBF kernel function and sigmoid kernel function, to conduct the closed test and the open test by adjusting the parameters c and g for every kernel function. The results are showed in Table 4.

Table.4 SVM Experimental Results.

Kernel Function	Parameter			Closed Test			Open Test		
	c	g	r	P	R	F	P	R	F
polynomial	1	200	100	89%	95%	0.92	62.5%	63%	0.63
RBF	185	1	0	87%	96%	0.91	61.3%	66%	0.64
sigmoid	400	2800	0	88%	95%	0.91	79.6%	75%	0.77

The experimental results show two interesting trends. First, the precision and recall of our deception detection algorithm by using polynomial kernel and RBF kernel are promising in the closed test, but they are 61.3% to 62.5% and 63% to 66% in the open test respectively. Obviously, the precision of deception detection is not very promising in open test. Second, the precision and recall of our deception detection algorithm by using sigmoid kernel are 88% and 95% in the closed test, respectively, and 79.6% and 75% in the open test respectively. Obviously, compared to Zhou L. et al. (2003), the precision of deception detection is promising, but the recall has a little insufficiency. In a word, our method acquired a satisfying result compared to Zhou L. et al. (2003) experimental result of deception detection in English text whose precision is 58%-80%.

6. Conclusion and Future Research

The experimental results indicate that the proposed method is feasible and effective. In order to acquire the better experimental results we need conduct the following researches in the future:

①The scale of the corpus has an important effect on the experimental results. At present it is very difficult to construct the corpora, so we use the corpora whose scale is a little small. In the future we will try our best to enlarge the scale of the corpora by making use of more measures of constructing the corpora, and conduct the deception detection tests by them.

②In this paper we conduct the deception detection according to texts feature items in the corpora. In the following researches we will use the other linguistics clues including the POS, the number of the sentences and so on, expecting to obtain the better experimental result.

Acknowledgement

This work is supported by Natural Science Foundation of China (No.60775041, 60473139). The authors are grateful for the anonymous reviewers who made constructive comments.

References

[1] Buller, D., Burgoon, J. Deception: Strategic and Nonstrategic Communication. in J.A. Daly, and J.M. Wiemann (eds.). Strategic Interpersonal Communication. Hillsdale, NJ: Erlbaum, 1994, pp.191-223.

[2] Hancock, J.T., Thom-Santelli, J., Ritchie, T. Deception and design: The impact of communication technologies on lying behavior. Proceedings of Conference on Computer Human Interaction. New York, 2004, pp.130-136.

[3] Carlson, J.R., George, J.F., Burgoon, J.K., Adkins, M., White, C. Deception in Computer-Mediated Communication. Group Decision and Negotiation. 2004, 13 (1): 5-28.

[4] Maksimova Y. Deception and its Detection in Computer-mediated Communication, technical reports in The Human Computer Interaction Program, Iowa State University, 2005

[5] Daft, R., Lengel, R. Organizational information requirements, media richness, and structural design. Management Science. 1986, 32(5): 554-570.

[6] Short, J., Williams, E., Christie, B. The Social Psychology of Telecommunications. New York, NY: John Wiley, 1976.

[7] Carlson, J.R., Zmud, R.W. Channel expansion theory and the experiential nature of media richness persceptions. Academy of Management Journal. 1999, 42 (2): 153-170.

[8] Buller, D., Burgoon, J. Interpersonal Deception Theory. Communication Theory 6. 1996, pp.203-242.

[9] George, J. F., Marett, K., Tilley, P. Deception Detection under Varying Electronic Media and Warning Conditions. In Proceedings of the Proceedings of the 37th Annual Hawaii international Conference on System Sciences (Hicss'04) - Track 1 - Volume 1. HICSS. IEEE Computer Society, Washington, DC, 10022.2, 2004.

[10] George, J.F., Marett, K. Inhibiting detection and its detection. In Proceedings of the Proceedings of the 37th Annual Hawaii international Conference on System Sciences (HICSS'04) , 2004.

[11] Marrett, K. and George, J.F. Group deception in computer-supported environments. Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) , 2005.

[12] Zhou, L., Twitchell, D.P., Qin, T., Burgoon, J.K., Nunamaker J.F. An Exploratory Study into Deception Detection in Text-Based Computer-Mediated Communication. Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03), 2003.

[13] Zhou, L., Zhang, D. Can online behavior unveil deceivers? Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04), 2004.

[14] Hancock, J.T., Curry, L., Goorha, S., Woodworth, M.T. Automated Linguistic Analysis of Deceptive and Truthful Synchronous Computer-Mediated Communication. Proceedings of the 38th Hawaii International Conference on System Sciences(HICSS'05), 2005.

[15] Yang Y M, Liu X. A Re-Examination of Text Categorization Methods. Proceeding of SIGIR-99,22nd ACM International Conference on Research and Development in Information Retrieval, 1999.