An Improving Deception Detection Method in Computer-Mediated Communication

Hu Zhang
School of Computer & Information Technology, Shanxi University, Taiyuan 030006, China
Email: zhanghu@sxu.edu.cn

Zhuohua Fan
Assets and Laboratory Administration Division, Shanxi University, Taiyuan 030006, China
Email:fzh@sxu.edu.cn

Jiaheng Zheng and Quanming Liu
School of Computer & Information Technology, Shanxi University, Taiyuan 030006, China
Email: {jhzheng, liuqm}@sxu.edu.cn

Abstract—Online deception is disrupting our daily life, organizational process, and even national security. Existing deception detection approaches followed a traditional paradigm by using a set of cues as antecedents, and used a variety of data sets and common classification models to detect deception, which were demonstrated to be an accurate technique, but the previous results also showed the necessity to expand the deception feature set in order to improve the accuracy. In our study, we propose a novel feature selection method of the combination of CHI statistics and hypothesis testing, and achieve the accuracy level of 86% and F-measure of 0.84 by using the novel feature sets and SVM classification models, which exceeds the previous experiment results.

Index Terms—deception detection, Chinese text, feature selection, linguistics cues

I. 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 multimedia 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, the attention paid to deception detection mainly focus on English resource, but the researches of Chinese resource being the second bigger information carrier are still on the initial stage on the whole. So the recent researches mainly include two aspects: the deception theory research and the deception detection experiment research for English resource.

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 R. Daft and R. Lengel (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 perspectives; ②Social presence theory(SPT) [4,6] was developed by J. 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 J.R. Carlson and R.W. Zmud (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 D. Buller and J. Burgoon (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 three studies of J.F. George, K. Marett and P. Tillery (2004), J.F. George and K. Marett (2004), K. Marett and J.F. George (2005) investigated the effect on the occurrence of deception and detection in different communication forms [9-11]. The conclusions are listed as follows: ① Suspicious receivers have a higher accuracy of deception detection, and the more suspicious receivers will accept the less deceptive information; ② People have facticity bias when they detect deception, and think truthful information is more than deceptive information; ③ In CMC, it is easier for deceiver to release more deceptive information.

The other three researches of L. Zhou etc. (2003), L. Zhou and D. Zhang (2004) and J.T. Hancock etc. (2005) emphatically analyzed effective deceptive cues [12-14]. The conclusions are drawn: ① In deceptive communication people use more words, for example, senders use more words in each sentence, and receivers ask more questions. ② Senders use less singular personal pronoun and more third personal pronoun. ③ Senders 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.

Some initial success in deception detection has created a new wave of applying intelligent technologies to support deception detection [15-18]. Those studies treated deception detection as a text classification problem that depends on the extraction of a predefined set of features (also called cues to deception). Such a dependency between deception detection approaches and cues to deception introduces several problems. First, the establishment of cues to deception faces its own challenges posed by the influence of various possible moderating factors on the effectiveness of cues. Second, research on cues to deception detection is still at its infancy. Third, the manual extraction of cues from free text is labor intensive, and it is not yet feasible to extract some cues (for example, plausibility) automatically with modern technologies.

Automated English text-based deception detection has been introduced as an alternative to these other methods [12, 16-18]. This technique has shown accuracy of up to 72% when trained on a very small sample of real-world lies for English. This method has previously relied on the common classification methods of logistic regression, decision trees, discriminant analysis, and multilayer perceptron neural networks (MLP). Studies using these methods have found the MLP to be the most accurate classifier for English text features cues. In addition to the MLP, four additional classification techniques not previously used in this research stream were chosen along with three cue sets in order to identify the model which maximizes the accuracy of classification using these cues.

This study extends prior work on deception detection by analyzing the linguistics cues and investigating three classification models [18]. Firstly, we reconstruct the deceptive and non-deceptive Chinese text corpora, and analyze their characteristics and differences between the deceptive and non-deceptive corpora. Secondly, we propose a novel feature selection method, and respectively use the naïve Bayes classifier, KNN classifier and SVM classifier to conduct deception detection, and then test the method effective.

The rest of the paper is organized as follows. Section 2 describes the proposed experimental model. Section 3 introduces two feature selection methods. Section 4 discusses the experimental results. Finally, the concluding remarks are given in Section 5.

II. DECEPTION DETECTION MODEL

A. Date Set 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. 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.

Now we construct the corpora summing up 3 500 000 Chinese characters by downloading the deceptive and non-deceptive texts from Internet based on the following principles.

① Deception 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;

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

③ The 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;

④ The 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;

⑤ In 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.

B. Model Description

The experimental model includes four main parts: pretreatment, feature selection, model training and deception detection.

Pretreatment includes constructing the deceptive corpora, Chinese word segmentation, etc.

Feature selection intends to select the most effective and representative linguistics cues for deception detection. Model training gets the parameters of the classifier by conducting the experiments on the training corpus.

Deception detection flags the measured samples into deceptive or non-deceptive by using the classifier on the test corpus.

The process is showed in Fig.1:
According to the experimental method that we proposed, the deception detection question can be transformed into a two-classified question, which also is to flag the measured sample into deceptive or non-deceptive. In this paper we separately use the naive Bayes classifier, KNN classifier and SVM classifier to carry on the deception detection.

III. FEATURE SELECTION

Feature selection is a process commonly used in machine learning, wherein subsets of the features available from the data are selected for application of a learning algorithm. The best subset contains the least number of dimensions that most contribute to accuracy; we discard the remaining, unimportant dimensions. This is an important stage of preprocessing and is one way of avoiding the curse of dimensionality. In our study, we propose a novel linguistics feature selection method based on hypothesis testing, and simultaneity use the representative CHI Statistics method to select the most effective text words features for deception detection.

A. Linguistics Cues Extraction Based On Hypothesis Testing

In the prior research of deception detection on English text, Zhou et al (2003) used 27 linguistics cues: word, verb, noun phrase, etc., and the cues were clustered into nine linguistics constructs: quantity, diversity, complexity, specificity, expressivity, informality, affect, uncertainty, and non-immediacy. After analyzing Chinese text linguistics features and the prior linguistics cues we created 17 linguistics cues as dependent variables, and the cues were clustered into seven linguistics constructs: quantity, complexity, non-immediacy, diversity, expressiveness, informality, and certainty.

After analyzing the experiment corpus including 700 deceptive texts and 1 000 non-deceptive texts, we conducted the following statistics with being shown as Table I according to the linguistic cues hypotheses.

<table>
<thead>
<tr>
<th>Linguistic Cues</th>
<th>All texts</th>
<th>non-deceptive texts</th>
<th>deceptive texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total of noun</td>
<td>330112</td>
<td>134198</td>
<td>195914</td>
</tr>
<tr>
<td>Total of verb</td>
<td>214900</td>
<td>85420</td>
<td>129480</td>
</tr>
<tr>
<td>Chinese character words</td>
<td>1697951</td>
<td>675969</td>
<td>1021982</td>
</tr>
<tr>
<td>different words</td>
<td>1083347</td>
<td>422388</td>
<td>660960</td>
</tr>
<tr>
<td>misspelled words</td>
<td>8617</td>
<td>297</td>
<td>569</td>
</tr>
<tr>
<td></td>
<td>71801</td>
<td>26508</td>
<td>45203</td>
</tr>
<tr>
<td></td>
<td>31474</td>
<td>12057</td>
<td>19417</td>
</tr>
<tr>
<td></td>
<td>1140</td>
<td>431</td>
<td>709</td>
</tr>
<tr>
<td></td>
<td>915</td>
<td>295</td>
<td>620</td>
</tr>
<tr>
<td></td>
<td>14492</td>
<td>6338</td>
<td>8154</td>
</tr>
<tr>
<td>First personal pronoun</td>
<td>5797</td>
<td>1466</td>
<td>4331</td>
</tr>
<tr>
<td>Second personal pronoun</td>
<td>841</td>
<td>113</td>
<td>728</td>
</tr>
<tr>
<td>Third personal pronoun</td>
<td>9738</td>
<td>2716</td>
<td>7022</td>
</tr>
</tbody>
</table>

Based on the statistical data in Table I, we can compute each linguistics cues. During the calculation process we need note that: the size of two kinds of corpus is different, and then it is necessary to take normalization according to the bigger one, so the following formula (1) is created to calculate each linguistic cues value:

\[
X_{N_D/D} = \frac{TDT \cdot CNDT}{TNDT \cdot CDT}
\]

Where, the detailed descriptions of the symbols in (1) are showed as follows:

TDT is the total number of characters in deceptive texts;

TNDT is the total number of characters in non-deceptive texts;

CDT is the value of corresponding cues in deceptive texts;

CNDT is the value of corresponding cues in non-deceptive texts.

The ratio value can be computed by the formula (1), which is showed in Table II.
According to the ratio to non-deceptive texts and deceptive texts, we can sort the linguistics cues descending: Name, lexical diversity, Average sentence length, First personal pronoun, and Second Personal Pronoun. Obviously, the gap between the ration value and 1 is bigger, the corresponding cue of the ratio is more important. Therefore, the feature value can be defined as in (2):

\[ FV = |X_{R/F} - 1| \quad (2) \]

Based on the above formula (2) we may get the feature value of the linguistics cues with being shown as Table III:

<table>
<thead>
<tr>
<th>Feature Item</th>
<th>Feature Value</th>
<th>Feature Item</th>
<th>Feature Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPP</td>
<td>0.767</td>
<td>Pausality</td>
<td>0.246</td>
</tr>
<tr>
<td>FPP</td>
<td>0.488</td>
<td>ANCS</td>
<td>0.241</td>
</tr>
<tr>
<td>TPP</td>
<td>0.415</td>
<td>Redundancy</td>
<td>0.219</td>
</tr>
<tr>
<td>Name</td>
<td>0.381</td>
<td>Sentence</td>
<td>0.099</td>
</tr>
<tr>
<td>LD</td>
<td>0.349</td>
<td>Word</td>
<td>0.034</td>
</tr>
<tr>
<td>TR</td>
<td>0.318</td>
<td>Emotiveness</td>
<td>0.006</td>
</tr>
<tr>
<td>ASL</td>
<td>0.266</td>
<td>Verb</td>
<td>0.003</td>
</tr>
<tr>
<td>Certainty</td>
<td>0.264</td>
<td>Character</td>
<td>0.000</td>
</tr>
<tr>
<td>AWL</td>
<td>0.249</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### B. CHI Statistics

CHI method measures the relativity between word and documents category \( c \). Having supposed there is \( \chi^2 \) distribution with a first-order freedom degree between \( t \) and \( c \), the \( \chi^2 \) statistics of the word represents the word contribution to the category. The \( \chi^2 \) 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. Therefore, we use the CHI statistics to carry on the text feature selection, and the total number of the candidate features is 100 855 in this study. After computing the CHI value of every feature word respectively, we list some feature words and their corresponding feature values according to the feature value order in Table IV, among them the feature items are described in the Chinese Pinyin corresponding to the Chinese character.

Based on the three researches of L. Zhou and D. Zhang (2004), J.T. Hancock etc. (2005) and L. 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.

### IV. EXPERIMENTAL RESULTS AND ANALYSIS

In our experiment we downloaded the most corpus confirmed to be deceptive from Internet, which includes 1493 texts. Simultaneously, we also collected the non-deceptive corpus related to deceptive texts content, which includes 2191 texts. Obviously the corpora scale summing up 3 500 000 Chinese characters is not very satisfying, but in fact we have been trying our best to collect the corpus confirmed to be deceptive as much as possible. In this study, we randomly select 700 deceptive texts and 1000 non-deceptive texts as the training corpora, and 793 deceptive texts and 1191 non-deceptive texts as the test corpora, and respectively use two types feature sets, including CHI statistics and the combination of CHI statistics and Hypothesis Testing, to conduct deception detection experiment.

### A. Evaluation Measures

For each test sample the classifier have two kinds of possibility decision results as shown Table V:

<table>
<thead>
<tr>
<th>Sample Set</th>
<th>Detected non-deceptive texts</th>
<th>Detected deceptive texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-deceptive</td>
<td>TP</td>
<td>FN</td>
</tr>
<tr>
<td>Deceptive</td>
<td>FP</td>
<td>TN</td>
</tr>
</tbody>
</table>

Where, the detailed descriptions of the symbols in Table V are showed as follows:

- TP is the number of the non-deceptive samples that are detected for the non-deceptive texts;
- FP is the number of the deceptive samples that are detected for the non-deceptive texts;
- FN is the number of the non-deceptive samples that are detected for the deceptive texts;
- TN is the number of the deceptive samples that are detected for the deceptive texts.

Therefore, the recall, precision and F-measure for deceptive texts can be computed by the following formulas (3), (4) and (5) with the numbers:

\[ \text{Recall} = \frac{TN}{(TN + FP)} \quad (3) \]
\[ \text{Precision} = \frac{TN}{(TN + FN)} \quad (4) \]
\[ \text{F-measure} = \sqrt{\text{Recall} \cdot \text{Precision}} \quad (5) \]

### B. SVM Experimental Results Based on the CHI Statistics Feature Sets

In this deception detection experiment, we choose the words whose feature value is bigger than 3 as the feature items [18]. 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 open test results are shown in Table VI.

### Table VI. SVM Experimental Results

<table>
<thead>
<tr>
<th>Kernel Function</th>
<th>Parameter Open Test</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>c</td>
<td>g</td>
<td>r</td>
<td>P</td>
</tr>
<tr>
<td>polynomial</td>
<td>1</td>
<td>200</td>
<td>100</td>
<td>62.5%</td>
</tr>
<tr>
<td>RBF</td>
<td>185</td>
<td>1</td>
<td>0</td>
<td>61.3%</td>
</tr>
<tr>
<td>sigmoid</td>
<td>400</td>
<td>2800</td>
<td>0</td>
<td>79.6%</td>
</tr>
</tbody>
</table>

The experimental results show the precision and recall of the deception detection algorithm by using sigmoid kernel are the highest, reaching 79.6% and 75% in the open test respectively.

### C. Experimental Results Based on the Novel Feature Sets

#### (1) Experimental Results

In this experiment we select the words whose CHI statistics feature value is bigger than 3 and the linguistics cues whose feature value is bigger than 0.3 as the feature items, and separately use Bayes classifier, KNN classifier and SVM classifier to conduct deception detection. According to the experimental results the numbers of the different types of samples are counted in Table VII.

### Table VII. Experimental Results

<table>
<thead>
<tr>
<th>Classification Model</th>
<th>Open Test</th>
<th>Closed Test</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes</td>
<td>881</td>
<td>119 27</td>
<td>673</td>
<td>TP</td>
<td>906</td>
<td>285</td>
<td>159</td>
<td>634</td>
<td></td>
</tr>
<tr>
<td>KNN</td>
<td>909</td>
<td>91 35 665</td>
<td>949</td>
<td>TP</td>
<td>242</td>
<td>197</td>
<td>596</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>974</td>
<td>36 29 671</td>
<td>1082</td>
<td>TP</td>
<td>109</td>
<td>141</td>
<td>652</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using the statistical data and the formulas (3), (4) and (5), we compute the precision (P), recall (R) and F-measure (F) for deceptive texts, which are showed in Table VIII.

### Table VIII. Experimental Results for Deceptive texts

<table>
<thead>
<tr>
<th>Classification Model</th>
<th>Open Test</th>
<th>Closed Test</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes</td>
<td>85.0%</td>
<td>96.1%</td>
<td>0.90</td>
<td>TP</td>
<td>69.0%</td>
<td>79.9%</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KNN</td>
<td>88.0%</td>
<td>95.0%</td>
<td>0.91</td>
<td>TP</td>
<td>71.3%</td>
<td>75.2%</td>
<td>0.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td>94.9%</td>
<td>95.9%</td>
<td>0.95</td>
<td>TP</td>
<td>85.7%</td>
<td>82.2%</td>
<td>0.84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### (2) Experimental Results Comparison

Furthermore, we choose the words whose CHI statistics feature value is bigger than 3 and the linguistics cues whose feature value is bigger than 0.2 as the feature items, and then conduct the open test using SVM classifier. Fig.2 separately shows the recall, precision and F-measure values, which obtained by using SVM classifier with three kinds of feature sets on our 793 deceptive texts and 1 191 non-deceptive texts.

![Figure 2. Three Feature Sets Experimental Comparison.](image)

**Figure 2.** Three Feature Sets Experimental Comparison. Obviously, compared to three feature sets experimental results, the precision and recall of deception detection changes significantly, with the precision from 0.79 to 0.85, and the recall from 0.75 to 0.84. Simultaneity, the result shows the precision is the highest as using the second feature set. In the experiments the other important problem found is that the proportion and scale of training corpus and test corpus have an effect on the experimental results, so in order to test the effect we choose a different proportion of training corpus and test corpus for experiments, including 2 to 8, 4 to 6, 5 to 5, 6 to 4 and 8 to 2 of the proportion of training corpus and test corpus. To five types corpus we all conduct the homologous experiments by using SVM classifier and the same feature set including the words whose CHI statistics feature value is bigger than 3 and the linguistics cues whose feature value is bigger than 0.3. Fig.3 shows the trends of the recall and precision of the experiments using five types corpus.

![Figure 3. The Trends of the Recall and Precision of five types corpus](image)

**Figure 3.** The Trends of the Recall and Precision of five types corpus

Obviously, the trends of the recall and the precision change stably as a whole, but the trends change significantly with the proportion of 2 to 8 and 8 to 2. In a word, based on the former experiment our method using the linguistics cues features acquires a satisfying result compared to the former method based on the only text features.

### V. CONCLUSIONS

The experimental results indicate that the proposed feature set is feasible and effective, so the method in this paper can improve significantly the accuracy of deception detection for Chinese text. In order to acquire the better experimental results we need conduct the following researches in the future:

⑴In this paper we use texts features and linguistics cues features to conduct the deception detection. In the
following researches we will use the other linguistics clues including the context, the relationships between words in the text and so on, expecting to obtain the better experimental result.

(2) The experimental results show the recall and the precision change significantly with the experimental corpus proportion of 2 to 8 and 8 to 2. Thereinto one important reason is that the used classifiers do better in balanced corpus than in non-balanced corpus, and another reason is that there is the small probability event in statistical natural language processing. Therefore, it will be an important research content to solve the problems in the following work.

(3) The scale of the corpus has an important effect on the experimental results. At present it is very difficult to construct the corpora, so the scale of the used corpora summing up 3 500 000 Chinese characters is a little small. In the future we will try our best to enlarge the scale of the corpora by making full use of more measures of constructing the corpora, and conduct the deception detection experiments by them.

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers who made constructive comments. This work was supported in part by NSFC (No.60775041, 61005053, 61100138).

REFERENCES


Hu Zhang, born in 1979, earned a B.S. and M.S. degree in school of computer & information technology from Shanxi University, in 2002 and 2005 respectively. After graduate school, he joined Shanxi University in 2005. He is a lecturer at the Shanxi University currently, and is pursuing his Ph.D. degree in Shanxi University. His current research interests include natural language processing, information security.

Dr. Zhang is a member of China Computer Association.

Zhuohua Fan was born in Yuanping, Shanxi Province, China on 1980. She received his B.S. degree, M.S. degree in school of computer & information technology from Shanxi University, in 2003 and 2007 respectively. She has been working at Shanxi University, Taiyuan, China since 2007. Her current research interests include natural language processing and artificial intelligence.

Jiaheng Zheng, born in 1948, is a professor at the Shanxi University currently. Her current research interests include natural language processing, artificial intelligence.

Quanming Liu, born in 1973, is an associate professor at the Shanxi University currently. His current research interests include information security, artificial intelligence.