

# Research on Chinese Short Text Emotional Polarity Recognition Based on Specific Immune Recognition

J. MA & H. QIAO

*School of Management science and Engineering, Shandong Normal University, Jinan, China*

**ABSTRACT:** This paper presents a new method about Chinese short text sentiment polarity analysis. Introducing a mechanism of biological immunology into the traditional Chinese word sentiment analysis. Artificial immune system is designed to simulate the related features of biological immune system. In order to solve the problems of the ambiguity classification of Chinese words and needs too much manual work always in the traditional ways. The principle of the system in biology is antibody specific recognize the antigen and interact with each other. This article put forward a new model——immune specificity sentiment analysis to calculating the Chinese short tests' emotional polarity which learned from the artificial immune model of limited resources. The model not only can be used for one-time cluster learning, also show the strong ability of continuous learning. It is very suitable for natural language emotional polarity classification. By analyzing the experimental data, we get a more accurately emotional tendency analysis in order to achieve more accurate tendency positioning.

**KEYWORD:** Specific immune recognition; Emotional polarity classification; Antigen; Antibody

## 1 INTRODUCTION

The speech tagging is to make a correct tag on the context which was according to the sentence in the artical. It is the foundation of the further processing. The speech tagging is an important part in many applications such as the text classification, the speech synthesis, the corpus processing and so on. The traditional emotional polarity analysis model mainly divided into two: the method based on rules and the method based on statistics. Rule-based labeling sentiment analysis system is strong associated with the language ability of the system designer, and the language ability of the designer directly reflects the rules. However, it is a very difficult also time consuming task that to make a structure of the rules for various of language, and we may face a variety of choices when judge speech of a word according to the rules. The next on is the methods based on statistical like hidden Markov model (HMM) [3]. When the model calculate the results of most likely part-of-speech tagging, that is namely to consider binary or ternary probability parameters of the context. Statistical model has strong robustness, but it also has a lot of deficiencies, such as it needs a lot of training corpus to achieve a high accuracy with an annotation. And

the traditional analysis methods without combining the existing language knowledge.

Immunology is a relatively young discipline and humans study on the natural immune dates back to 300 years ago. In the 17th century, some medical scientists in our country has creatively developed "acne" against smallpox[2]. In 1976, Edward Jenner, a British doctor invent "cowpox" which replaced the smallpox vaccine. From then on, the modern immunology began. In 1974, Jerne Farmer, a Nobel Prize winner in the United States, put forward the immune network theory. They establish an effective computer system which based on immune principle and intelligent system. Varela in 1989 put forward the ideas of the immune network convergence in some way and the immune system produce different antibodies to adapt to the new environment[5]. Then a new research field generated which called the artificial immune system.

The internet, like weibo, twitter, facebook produce lots of information everyday which contains a lot of emotional passage. In this paper, we use finite artificial immune system resources which had been improved a specific immune treatment for goods review essay. Specific immune response refers to the body generate immunoglobulin antibodies and the immune lymphocytes cells after the antigen stimulated by viruses, microorganisms.

When contact with the specific antigen, the body produce antibodies and inspire the specific reaction to protect the body. The ultimate goal of reaction is make the antigen lost the activity. The immune system has strong ability to adaptability, diversity, learning, identify, and memory etc. So we can draw lessons from operation mechanism and the features of the immune system, to design a corresponding strategy to deal with the problems which needs a lot of artificial participation and the text classification elements which are too single.

## 2 ARTIFICIAL IMMUNE SYSTEM

In the 17th century, medical scientist in our country has creatively developed smallpox to prevent people. In 1796, a British doctor Edward Jenner invented vaccinia, which is regarded as the beginning of the modern immunology. After 300 years of development, the immunology has developed from microbiology chapter into an independent discipline.

### 2.1 Antigen and Antibody

Biological immune system is a intelligence system which has many characteristics. Such as learning, identify, and memory. It consists of many immune organs and a variety of lymphoid cells. The lymphocytes mainly include two types of T cells and B cells, which are distributed throughout the body and plays a main role in the immune. Biological immune macro is described in figure 1 below:

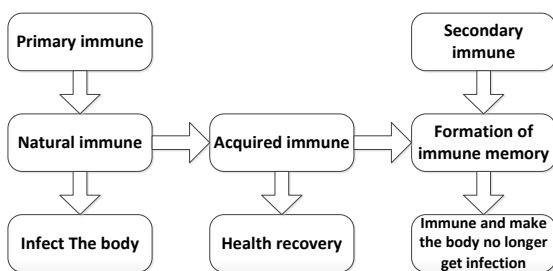


Figure1.The description of biological immune macroscopic

Antigen is point to the things that induce the body's immune system and stimulate the system produces an immune response. The products of the corresponding immune reaction are in vivo or in vitro[1]. In the artificial immune system, antigens represent a pattern which needs to be solved. In this article the antigens is point to the words collected from the internet needs emotional judgment.

Antibodies refers to a protein molecules produced by plasma cells, it also be called immunoglobulin molecules. Antibodies are in combination with external invasion antigens like source of infection or toxic substances[1]. Then on their own or with other elements of the immune system's help, the

antibodies destroy these antigens, eliminate the threat the human body. In the artificial immune system, antibodies on behalf of recognize patterns or the candidate solution of the problem.

In this article, we put the antigens into the system which have been established. There are some preset eligible antibodies exist in the system. The antigens identified by the antibodies to form suitable antibody group. The suitable group deal with the text which needs to be processed.

### 2.2 The process of biological immune

Immune Response refers to the immune effect that immune cells identify antigen and activate, differentiation and proliferation to clear the process of antigens. Primary immune response in the immune system happened in some kind of antigen invasion for the first time, the immune system produce antibodies to infect and help remove antigen out of the body. Immune response process is finished by the antigenic determinant combined with antibody cell. As shown in figure 2:

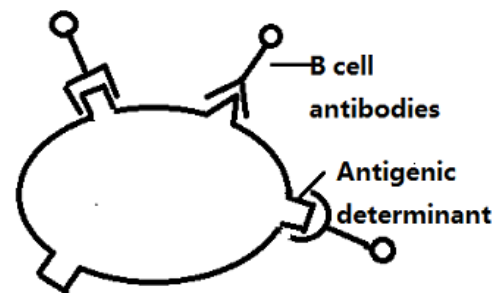


Figure 2. The recognition between antibody and antigen

In the immune response, the immune system encounter with foreign material and remove it in vitro, at the same time retain a certain number of B cells in the immune system as an immune memory cells. This makes the immune system react quickly when they meet foreign material again in body and fight back[7]. This process is the secondary immune response. Secondary response does not need to relearn, and response more quickly.

## 3 A SYSTEM BASED ON THE SPECIFIC IMMUNE RESPONSE STRATEGY

The Visualization Resource Limited Artificial Immune System is put forward by Timmis in 1999, the model not only can be used for one-time cluster learning, at the same time can show the ability of continuous learning[12]. Also it is very suitable for Chinese emotional words polarity classification in natural language processing. Timmis inspired by Perelson's research about identify the Ball, and then put forward the Artificial Recognition Ball (ARB)

concept. Within the space form, a finite number of antibodies can represent an infinite number of antigen and an antibody can identify all the epitope. In the form of space there is a small area called the response area, with said  $V_e$ . Each ARB in the network can represent many of the same B cells, as shown in figure 3:

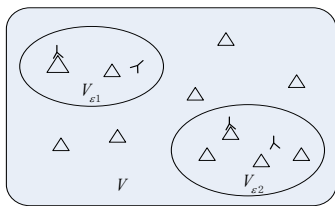


Figure 3. The representation of ARB

RLAIS is composed by a set of ARB and the contacts between them. ARB's competition in RLAIS on behalf of many B cells to produce antibodies to adapt the environment. RLAIS algorithm flow chart 1 as show:

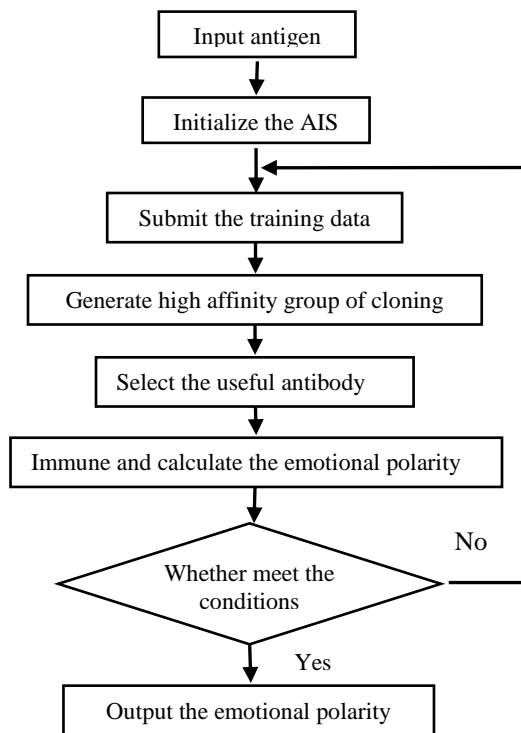


Chart 1. RLAIS algorithm flow chart

Training data continuous submitted to the network showed in the chart 1, an ARB experience stimulate calculation using the formula (1) below[7]:

$$\text{stimulation} = \sum_{x=0}^a (1 - pd) + \sum_{x=0}^a (1 - dis_x) - \sum_{x=0}^a dis_x \quad (1)$$

Among them, the  $a$  is the one have exposed antigen in ARB,  $pd$  is defined as the ARB regularization space and distance between the first  $x$  antigen,  $0 \leq pd \leq 1$ ,  $dis_x$  is a distance started from the B cells to the first  $x$  neighbor.

Constantly presenting training data during the learning cycle in cloning, use the formula (2):

$$e_x = k \cdot SI_x \quad (2)$$

The  $k$  is a constant number be used to limit the cloned cells.  $SI_x$  is an ARB stimulus level.

B cells is a relationship between RLAIS model nodes, encode collection of the antigens. In the condition of network initialization, B cells are often with low inhibition of stimulation level. A variety of antigens are present at the same time, constantly stimulate and screening on B cells, then remove the weakest link. Highly affinity B cells connected between each other into a network, form a group that specificity for the antigen antibody. The system will reached the polarity classification of emotional words through the antibody polarity identification.

#### 4 EXPERIMENT AND RESULT ANALYSIS

In this article, the experiment was conducted to evaluate the performance of RLAIS specific text sentiment analysis. And we try to compare the classifier performance of sentiment analysis with traditional Support Vector Machine (SVM).

This experiment adopts corpus for Esteelauder cosmetics essay on Tmall shopping. We collect 2000 comments text from taobao shopping about the lock water day cream. Converse the texts into a uniform format. Then divide all the corpus into two parts, of which 800 article as the training corpus, the remaining 1200 as test corpus collection T. Emotional polarity clustering is through the analysis of the existing emotional system, and tagging emotional tendencies (positive or negative). The results in 368 are opposite tendency text, 432 are the positive tendency of the text.

Calculated each test sentence by function formula (3) and standard polarity similarity calculation. Choose the results which highest similar to the baseline as the polarity of the sentence [14].

$$\text{Sim}(S_1, S_2) = \alpha * \varphi_1 + \beta * \varphi_2 + \gamma * \varphi_3 \quad (3)$$

$\varphi_1$  considering the word characteristics, embodies the information of surface polarity word;  $\varphi_2$  considering the meaning of characteristics, embodies the polarity of the context information;  $\varphi_3$  considering the syntactic characteristics, embodies the dependent relationship between words, and get more understanding of the sentence. The experiment is combined with the results to adjust the parameters, select  $\alpha=0.4$ ,  $\beta=0.2$ ,  $\gamma=0.1$  as the weights of similarity calculation.

Evaluation index valued by precision rate P(%) and recall rate R(%) and F. To illustrate the effectiveness of the method, this article is compared

with the traditional methods of obtaining statistics of (3) which based on polarity word. The results as shown in table 1:

Table 1. Comparative analysis of experimental data

	P(%)	R(%)	F
SVM	65.7	61.5	63.5
RLAIS	78.3	79.1	77.5

The experimental results show that the precision of this method reached 78.3%, the F value reached 77.5%. The RLAIS's F value significantly increased performance compared with polarity which based on word frequency statistics by nearly 14%.

The experimental data and the analysis shows the method based on word frequency statistics recall rate is low, the reason is that the method calculate a simple superposition of the positive and negative polarity. This kind of emotional polarity simple superposition by the polarity of the word plus or minus stack and led negative sentences tagging error. Such as "The effect of the cosmetic is not bad". The method we proposed not only considers the polarity word itself, but also capture the polarity information according to the syntax analysis from the context and syntax level. Even it can distinguish the word dynamic polarity from a certain extent. So it can improve the classification accuracy.

But some neutral statements are not well identified considering the complexity of the algorithm. Such as the sentence "The price of this cream neither high nor low ". The local of the words with polarity, the polarity of "not high" is positive, and "not low" is negative. When the machine judge the polarity, it thought the emotion of the text is neutral. In Chinese, this kind of neutral language is quite common, and the application is very extensive. Different people have different emotional tendency of neutral language. So the language emotional polarity judgment has great difficulties. Algorithm in this paper default that the sentence will be positive polarity, but in fact the overall polarity to neutral. In addition, the algorithm also did not deal with tone, the classification accuracy needs to improve further.

## 5 SUMMARY

This article put forward a new specific classification strategy of emotions analysis about Chinese short text which based on Immune reaction mechanism. Adopt specific immune response mode to enhance the precision classification of the short text. The algorithm can fast convergence to the global optimal solution. In this process, we try to keep the diversity of population to improve the rate of convergence of the global optimal. The model solved many disadvantages of the traditional Chinese text

classification methods in theoretically. Such as text annotation needs a lot of artificial participation, polarity text categorization is not clear and the error is bigger. Not only improve the efficiency of classification, and puts forward a new train of thought about Chinese essay in classification.

## REFERENCES

- [1] Yan Ren, Zonghai Chen: *Artificial immune system and its application in the field of control* edited by Information and Control, 2003, 32(1): 45-50.
- [2] Jun Wang, Xiyu Liu: *Artificial immune system and its model analysis* edited by Computer Technology and Development, 2006 16(7):105 -107.
- [3] Keming Xie, Gang Xie, Hongbo Guo: *Artificial immune system and its algorithm* edited by Journal of Electronics and Information Technology, 2005, 27(11):1839 -1844.
- [4] Yihong Zhao, Zhengwen Li, Qisi He: *Biological information processing system--genetic algorithm* edited by Journal of Chengdu University of Technology, 2004, 31(5):531 -534.
- [5] Jun Wang: *Based on data mining and application of artificial immune model*, 2004, TP18:10045.
- [6] Weigang Zhang, Jianhao Tan: *Text classification based on artificial immune system in the network research* edited by Science Technology and Engineering, 2006, 6(22):3621 -3623.
- [7] Fahong Chen: *Polarity text classification based on kernel method research* edited by The channel of science, 2011, 8(56):17 -21.
- [8] Jianbo Liu: *Facing the emotional essay this meaning string of discovery and analysis algorithms* edited by Journal of Wut, 2011, 33(5):742 -745.
- [9] Na Fan, Wandong Cai: *Chinese text emotional topic sentence analysis and extraction* edited by Journal of Computer Applications, 2009, 29(4).
- [10] Yuan Ma, Li Fu: *Short text sentiment analysis technology research*, 2011(14).
- [11] Ziqiong Zhang, Qiang Ye, Yijun Li: *The review on the Internet goods comments sentiment analysis research* edited by Journal of management sciences in China, 2010, 13(6): 84 -95.
- [12] Tianfang Yao, Decheng Lou: *Research on semantic orientation analysis for topics in Chinese sentence* edited by Journal of Chinese Information Processing, 2007, 21(5): 73-79.
- [13] Qiang Ye: *Automatically measuring subjectivity of Chinese sentences for sentiment analysis to reviews on the internet* edited by China Journal of Information Systems, 2007.1(1):79-91.
- [14] Peiying Zhang: *Method of Chinese word segmentation based on directed graph* edited by Computer Engineering and Applications, 2009.
- [15] Xiaojuan Zhu, Tefang Chen: *Research about Chinese word segmentation based on statistic* edited by Biosystem, 2007, 3(78): 1671 -1041.
- [16] Yanji Liu, Tili Sun: *Research of ambiguity algorithm for Chinese word segmentation based on dictionary* edited by China Hownet, 2009(5).
- [17] Y. Jiang, F. Shang: *Research on Multibit-Trie Tree IP Classification Algorithm* edited by International Conference on Communications, 2006. 18 (14):121 -130.