

Affective Computing Model for the Set Pair Users on Twitter

Chunying Zhang¹, Jing Wang²

¹ College of Science, Hebei United University
Hebei, Tangshan, China

² College of Science, Hebei United University
Hebei, Tangshan, China

Abstract

Affective computing is the calculation about sentiment, sentiment generated and the aspects of affecting the sentiment. However, the different factors often cause the uncertainty of sentiment expression of the users. Today twitter as the information media of real-time and timely has become better sentiment expression vector for users themselves. Therefore, in allusion to the diversity of sentiment form of twitter information to express sentiment, this paper constructs affective computing model, starting from the differences of the constituted form of Twitter based on set pair theory to make analysis and calculation for user sentiment, from the text, emoticon, picture information and other multi-angle to analyze the positive, negative and uncertain emotion of the users for the signal twitter, consolidating the weight of various parts in emotional information, building hierarchical set pair affective computing model for twitter users, to offer more favorable data support for the relevant departments and businesses.

Keywords: *Affective computing, hierarchical model, set pair analysis, the constituted form of Twitter, Set Pair Users.*

1. Introduction

With the rapid development of the Web2.0, the role of users themselves in the network have undergone a rapid change; initially, from the information recipient only browse the page content to the information producer, publisher and communicator of publishing their own thoughts, viewpoints. The appearance of twitter truly marks the personal Internet age has arrived, by virtue of the autonomous access to information, quickness, breadth and content of dapper has won the favor of people. People on twitter who through their favorite form of text to express their own sentiment, and to share and exchange with the social friends who have the common interests, or one social hot topic of real-time for discussing, which are accustomed to express their own sentiment orientation by

means of the social network. In addition, with respect to the relevant departments, they can obtain the "voices" of the people through the sentiment orientation that the user expressed on twitter event. Therefore, they can give back and solve the problem in the shortest time. The affective computing commits to analyze the text to mine the user's sentiment orientation. "Affective computing" [1, 2] was initially proposed in 1995 by the R.Picard professor of MIT Laboratory. He published the monograph - Affective computing in 1997, in the book, his gave the definition that "the affective computing is associated with the sentiment, from sentiment or the calculation that can exert influence on the sentiment". Affective Computing [3-5] is the basis for text orientation analysis. In essence, it is the sentiment analysis and mining for the text, its main purpose is to give the computer ability of understanding and cognition to distinguish the human sentiment and tonality to express sentiment. Currently, in the existing sentiment analysis [6-7] research, most of them are based on the sentiment analysis study of the ordinary text, work [8] divided sentiment into positive sentiment, neutral sentiment, negative sentiment based on the rule method of sentiment dictionary, only by the comparison of the number of sentiment vocabulary of ordinary text to get the user's sentiment orientation, and also mentioned emoticon rule method in the text, the difference of the number of emoticon of different sentiment polarity to get sentiment orientation, one work mentions the latest manifestation on twitter are some pictures and links, but do not start the detailed analysis; The work [9] mentioned a form of expression of the new network word related to the twitter content, but did not expand affective computing analysis according to their characteristics. Now, the sentiment analysis on twitter is only from one particular form of expression to give the sentiment orientation of the users, or just referred to a part, not give the summarized analysis. In fact, with the constantly updated and the development of the network, the picture, new word of network,

emoticon, links, etc of representing sentiment is emerging. Therefore, bring the text that be able to express sentiment information integrated analysis is very important, for different events users held views is different, and the level of opinion is not the same, there is uncertainty. Set pair analysis is one method to resolve the uncertainty, the significance of the contact number is that link the number and scope of this number, contacting one specific number and the uncertainty with certainty within its scope to expand trend analysis. This paper integrates the constituted form of twitter that the user can express their own sentiment. Based on set pair analysis theory to construct the model set pair affective computing, so that we can via calculating to get the sentiment orientation for users.

2. The analysis on the constituted form of twitter.

Twitter as a platform which information sharing, dissemination and access to is based on user relationship, through WEB, WAP, and a variety of clients, users can set up personal communities, then update information about 140 words, and greatly promote the dissemination and sharing of the information, its great commercial value began to show and highlight the commercial advantages in the areas of crisis public relations, public opinion speculation and web promotion. With the popularity of the Internet, The constituted form of twitter are constantly update, before there is no word, phrase, sentence as a network of new words appear in network communication, with a circulation speed and fresh are welcomed by the people.

Well, first of all, we give one twitter intercepted on twitter client, the twitter forms of expression including the ordinary text, the new network word, emoticon, and pictures information mentioned in this article. Figure 1 is intercepted from the QQ twitter.



Fig. 1 An entry QQ twitter.

Therefore, this paper bring the diversification of the constituted form of twitter information into consideration,

set the text messages, emoticon, pictures into together, and design to achieve better computational analysis. We will give a brief description for each component as follows.

2.1 The text information analysis

Now the sentiment analysis method of text mainly has two classification model, on one hand, using the method that combine the sentiment dictionary with rules, the number of positive sentiment words and negative sentiment words contained in the text to make sentiment classification; On other hand, those who use machine learning methods to select some of the characteristics of the text to train and test the set, the main method are classifiers that are Naïve Bayes^[10], maximum Entropy, support vector machine. The work [8] contraposes the sentiment analysis of Chinese, based on the sentiment dictionary rules to determine the sentiment polarity of the sentiment word ,it mentioned select the word number of positive affect and negative affect to get sentiment polarity ,the sentiment polarity of the words in certain under any circumstances characterization, but some part of speech is uncertain, the polarity of the sentiment expressed in different contexts is different, so this part uncertain vocabulary polarity analysis is particularly important; The analysis above rarely involved in description and classification for the vocabulary polarity. This paper divides into three categories: positive sentiment vocabulary, negative sentiment vocabulary, uncertainty sentiment vocabulary. As in figure 1, the twitter contains two kinds of text: ordinary text and new network words, the text contain the new network words: dark reddish purple, porridge, drops, and me. The dark reddish purple means like this, the porridge means like. In the calculation of this text, we will convert the new network word into ordinary word to analyze calculation.

2.2 The emoticon analysis

In the paper “Using the to make sentiment classification can be reduce the dependence of the machine learning techniques.”, the foreign scholar Jonathon Read elaborate that one can overcome geographical, subject and time can be as annotations of sentiment, And to achieve the visualization of sentiment states, and this has led to the birth of the emoticon. In the paper, the author used lots of sentiment in Twitter API, thus point out that “smile” usually as a positive text, and “frown” usually show that negative text. And finally using the ways of sentiment to do the sentiment classification, and then reach the accurate rate of 70%. The domestic scholars Xie LiXing has done more hierarchical structure strategy method for the twitter sentiment analysis and applies the emoticon rule method, in view of the expression symbols that sina twitter provides to make classification of the positive and

negative emoticon, and extracted the expression symbols of positive and negative in the text with analysis, the difference size of positive and negative expressions symbol as the classification of sentiment polarity. The analysis method of corpus automatic tagging in work, it uses two kinds of sentiment knowledge (emoticon and sentiment words) for large-scale not label samples to conduct automatic tagging and to get the training samples. For the classification of the emoticon also use the way of work [11]: positive sample and negative sample.

For the analysis of a series of emoticons, they all have been given a clear polarity, in fact, the representation of the emoticons are constantly updated, for example, on behalf of positive sentiment : 🥰 (nod) 😄 (excited) 😊 (happy), though the affective polarity all represent the positive, the three sentiment belonging to the degree of positive expression is inconsistent, Therefore, in the rest of this analysis are given different weights of the polarity degree of emoticons to reflect such sentiment differences. The cartoon faces 🤖 in the crawl will display the text meaning "rotation"; the sentiment orientation is different when such category emoticons appear in different contexts. We consider them as uncertain sentiment emoticon to analyze. Based on the above analysis of emoticons, we will give three categories of sentiment symbol, positive emoticon, uncertain emoticon and negative emoticon. In figure 1 included in four emoticon, wherein each emoticon corresponds its homologous text, for example, symbol 🤔 on behalf of the doubt is an uncertainty; emoticon 😓 from the appearance point of view is a sweating state, combined with the text meaning represents a negative sentiment emoticon; growl 😊 is pleased state belong to positive emoticon. In the following article, through its text explanation to analyze the generic of the emoticon.

2.3 The picture information analysis

For the picture information from two aspects to consider, on the one hand, the picture is already on the network, users directly upload to enrich the content of their own twitter, a class of pictures accompanied by text attachment, the meaning of the words express the sentiment polarity; Another type is only picture without any text message, abstracting the meaning of the expression from picture content for an sentiment polarity to analyze. The other hand, analyzing the pictures that taken by the user on the ground through the communication tools to express their seen and heard, and this kind of pictures on twitter from users themselves microscopic point. The pictures show a true and more complete world. In discusses the application of twitter on media, the spread of twitter and newspaper illustrates the broad scope of twitter, showing a new trend

in the new media environment, and also mentioned the application of expressing their sentiment [12].

For the picture information processing, this paper classify positive picture, uncertain picture, negative picture. For the each picture classification, we adopt the manner consistent with emoticon, which we can also consider the generic weight value of belonging to certain types of sentiment pictures. At the end of figure 1 gives one picture, which can be seen from the analysis of the picture contents, the expression of sentiment orientation in this picture is the positive sentiment pictures.

In view of the above explanation of three parts, the following we utilize the set pair theory to expand affective computing and trend analysis for the form content of twitter.

3. The set pair affective computing and trend analysis on twitter

3.1 The set pair analysis method

Set pair analysis^[13-14] was first proposed by Professor Zhao Keqin in 1989 in Baotou that convening the National System Science and Regional Planning Symposium, And this relationship of certainty and uncertainty related influence and mutual restraint, even under certain conditions are met can be mutually converted^[15], so the contact can be expressed as:

$$U = A + Bi + Cj \quad (1)$$

Thereinto A, B, C are non-negative real number, $j = -1, i \in [-1, 1]$ depending on the situation to have value in the range, commanding $N = A + B + C$, N is the contact norm, let N divided the two sides of formula (1), then taking

$$u = \frac{U}{N}, a = \frac{A}{N}, b = \frac{B}{N}, c = \frac{C}{N}$$

So

$$u = a + bi + cj \quad (2)$$

Therefore, the formula (2) is called the contact number expression.

Thereinto $a, b, c \in [0, 1], a + b + c = 1, j = -1, i \in [-1, 1]$,

a is the same degree, b is the difference degree, c is the confrontation degree, i is the difference coefficient, j is the confrontation coefficient. Thereinto, the trend is the concept that reflecting the size relations order of the same degree (a), the difference degree (b), the confrontation degree (c), and can be divided into the same potential, the balance potential and the anti-potential. In addition, each potential can also be subdivided by the degree of size.

3.2 The set pair affective computing on Twitter

Based on the above analysis on the constituted form of twitter, combined with the characteristics of contents form classification, this paper adopts set pair analysis theory method to calculate from the three aspects: the text, emoticon, pictures information, and the calculation process of each part as follows.

3.2.1 The set pair affective computing on Twitter

A message that a user issues on twitter is generally constituted by multiples of vocabularies, including general vocabulary and sentiment vocabulary. These vocabularies can be divided into positive sentiment words, negative sentiment words and uncertain sentiment words, according to their parts of speech, then adopt the set pair analysis method to conduct affective computing for the single twitter.

Supposing one twitter TW_k has x_1 positive sentiment vocabularies, x_2 negative sentiment vocabularies, x_3 uncertain sentiment vocabularies, recording as $x_1 + x_2 + x_3 = N_1$, so, the basic model of text set pair affective computing for this twitter, as follows:

$$A_1(TW_k) = \frac{x_1}{N_1} + \frac{x_2}{N_1}i + \frac{x_3}{N_1}j \quad (3)$$

Wherein $j = -1$, $i \in [-1,1]$ depending on different circumstance to get value, in the formula (3), when $i = -1$, uncertain sentiment vocabulary is converted to the negative sentiment vocabulary, that is, at this time, the number of negative sentiment vocabulary is $x_2 + x_3$, the number of positive vocabulary x_1 ; when $i = 0$ the twitter does not exist uncertain sentiment vocabulary, the number of positive and negative sentiment vocabulary have no change; when $i = 1$, the uncertainty of the sentiment vocabulary convert into positive sentiment vocabulary, at this time, the number of positive sentiment vocabulary is $x_1 + x_3$, the number of negative sentiment vocabulary is x_2 .

In the actual sentiment analysis, the extent of the sentiment orientation of each vocabulary is not identical. We via the weight to improve the above set pair affective computing model.

Supposing one twitter TW_k has x_1 positive sentiment vocabularies, and the weight values is f_1 , the Product of vocabulary and its corresponding weights is expressed by c_1 in the contact number; Supposing one twitter TW_k has x_2 negative sentiment vocabularies, and the weight values is f_2 , the Product of vocabulary and its corresponding

weights is expressed by b_1 in the contact number; Supposing one twitter TW_k has x_3 uncertain sentiment vocabularies, and the weight values is f_3 , the product of vocabulary and its corresponding weights is expressed by c_1 in the contact number; Therefore, updating the text set pair affective computing basic model of twitter as follows:

$$\begin{aligned} A_1(TW_k) &= f_1 \frac{x_1}{N_1} + f_2 \frac{x_2}{N_1}i + f_3 \frac{x_3}{N_1}j \\ &= a_1 + b_1i + c_1j \end{aligned} \quad (4)$$

Through the sum of the products of the expressions of each word and the corresponding weight to be better reflect the degree of sentiment differences among the different vocabulary, making the analysis results more accurate.

We expand calculation in figure 1 in the text, using equation (4) model to conduct the affective computing:

$$\begin{aligned} A_1(TW_k) &= (\frac{3}{12} \times 0.2 + \frac{1}{12} \times 0.4 + \frac{1}{12} \times 0.6) + (\frac{1}{12} \times 0.2 + \frac{3}{12} \times 0.4)i + (\frac{3}{12} \times 0.4)j \\ &= \frac{2}{15} + \frac{7}{60}i + \frac{1}{10}j \end{aligned}$$

The data obtained from the analysis of the text, we can get $a_1 > b_1 > c_1$, thus the sentiment orientation of the text analysis on twitter is positive.

3.2.2 The affective computing of the emoticon

Supposing one twitter TW_k has y_1 positive sentiment emoticon, y_2 negative sentiment emoticon, y_3 uncertain sentiment emoticon, recording as $y_1 + y_2 + y_3 = N_2$, so the basic model of emoticon set pair affective computing for this twitter, as follows:

$$A_2(TW_k) = \frac{y_1}{N_2} + \frac{y_2}{N_2}i + \frac{y_3}{N_2}j \quad (5)$$

Wherein $j = -1$, $i \in [-1,1]$ depending on different circumstance to get value, in the formula (5), when $i = -1$, uncertain sentiment emoticon is converted to the negative sentiment emoticon, that is, at this time, the number of negative sentiment emoticon is $y_2 + y_3$, the number of positive emoticon y_1 ; when $i = 0$ the twitter does not exist uncertain sentiment emoticon, the number of positive and negative sentiment emoticon have no change; when $i = 1$, the uncertainty of the sentiment emoticon convert into positive sentiment emoticon, at this time, the number of positive sentiment emoticon is $y_1 + y_3$, the number of negative sentiment emoticon is y_2 .

For emoticon of strong expressive, and each emoticon belong to each category is different. Therefore, in view of this situation, we have different emoticon

calculated to give each a weight value. Giving the updated emoticon connection degree expression is as follows:

$$A_2(TW_k) = w_1 \frac{y_1}{N_2} + w_2 \frac{y_2}{N_2} i + w_3 \frac{y_3}{N_2} j$$

$$= a_2 + b_2 i + c_2 j \quad (6)$$

In the formula (6), w_1, w_2, w_3 is the corresponding weight value for the each category sentiment emoticon.

3.23 The affective computing of the picture information

This paper adopts the research process that similar with text, emoticon for picture information to conduct affective computing.

Supposing one twitter TW_k has z_1 positive sentiment picture information, z_2 negative sentiment z_3 , uncertain sentiment z_3 , recording as $z_1 + z_2 + z_3 = N_3$, so the basic model of picture information set pair affective computing for this twitter, as follows:

$$A_3(TW_k) = \frac{z_1}{N_3} + \frac{z_2}{N_3} i + \frac{z_3}{N_3} j \quad (7)$$

Wherein $j = -1, i \in [-1, 1]$ depending on different circumstance to get value, in the formula (7), when $i = -1$, uncertain sentiment picture information is converted to the negative sentiment picture information, that is, at this time, the number of negative sentiment picture information is $z_2 + z_3$, the number of positive picture information z_1 ; when $i = 0$ the twitter does not exist uncertain sentiment picture information, the number of positive and negative sentiment picture information have no change; when $i = 1$, the uncertainty of the sentiment picture information convert into positive sentiment picture information, at this time, the number of positive sentiment picture information is $z_1 + z_3$, the number of negative sentiment picture information is z_2 .

Therefore, we give a weight value for the picture information calculating for different pictures. The expression of picture information connection degree is as follows:

$$A_3(TW_k) = g_1 \frac{z_1}{N_3} + g_2 \frac{z_2}{N_3} i + g_3 \frac{z_3}{N_3} j$$

$$= a_3 + b_3 i + c_3 j \quad (8)$$

In the formula (8), g_1, g_2, g_3 is the corresponding weight value for the each category sentiment picture information.

3.24 The comprehensive affective computing on twitter

Thus we can conclude that the constituted form of twitter mainly including the form of the above three aspects, the proportion of the effect of sentiment expression for each form is not the same, therefore need to give different weight value for each type of the constituted form,

respectively is α, β, γ , and the size of three weight value is also depending on the specific circumstances. The specific expression is as follows:

$$A = \alpha A_1 + \beta A_2 + \gamma A_3$$

$$= \alpha(a_1 + b_1 i + c_1 j) + \beta(a_2 + b_2 i + c_2 j) + \lambda(a_3 + b_3 i + c_3 j)$$

$$= (\alpha a_1 + \beta a_2 + \lambda a_3) + (\alpha b_1 + \beta b_2 + \lambda b_3) i + (\alpha c_1 + \beta c_2 + \lambda c_3) j$$

$$= a + b i + c j \quad (9)$$

The above formula (9) is the analysis calculated combination of the three type constituted form, a is behalf of the positive sentiment part, it includes vocabulary, emoticon, pictures information; b is behalf of the uncertain sentiment part; c is behalf of the negative sentiment part, thereby obtaining comprehensive twitter sentiment formula.

4. Application Example

The twitter of figure 1 contains the text, emoticon, and pictures information as the example to verify the validity of the method, the vocabulary divided into the general vocabulary and sentiment vocabulary, according to the number and weight to summarize the sentiment vocabulary; for emoticon and picture information, under the number and weight to summarize, that are shown in the following Table 1:

Consolidating the data in Table I and the weight value of the vocabulary, emoticon, picture information of this twitter to calculate, giving the size is $\beta = 0.4 > \gamma = 0.3 > \alpha = 0.2$ from the expressed result, calculating the size of a, b, c according to the formula (9).

Then get the relation is: $a = \frac{68}{300}, b = \frac{19}{300}, c = \frac{24}{300}$, that is $a > c > b$.

Comprehensive analysis by the method of this article, this user expressed sentiment orientation for positive sentiment orientation. We also obviously see that we selected this entry twitter held attitude support by the data that given on table one. Corresponding to the twitter users to express the sentiment state is pleasure, Secondly, we also need to be taken into account for the components of polar uncertain, to some extent these factors are also occupy a certain share ratio on the results of the problem. Through I in the interval of $[-1, 1]$ to get the different value to get the different types of sentiment expressions.

Table 1: Summarization of Vocabulary, Emoticon, Picture Information

Polarity Weight Range	Positive			Uncertain			Negative		
	v	e	p	v	e	p	v	e	p
(0-0.2]	3	1		1					
(0.2-0.4]	1		1	3	1		3		
(0.4-0.6]	1	1						1	

(In the above table v= vocabulary, e= emoticon, p= picture information)

5. Conclusion and next step

In this paper, to fully consider the constituted form of twitter, which to be updated with the network development. From the angel of sentiment orientation to divide the constituted form of Twitter into text, emoticon, and picture to analyze. And assign weight to solve the magnitude of belonging to a certain type of sentiment polarity; at last, giving the affective computing model of the three constituted form of integration by set pair analysis method to obtain the good results. The next work: (1) According to the above analysis and conduct affective computing trend analysis for the users. (2) Considering the relationship between twitter first posts its thread, affective computing and trend analysis of single twitter extended to multiple twitter.

Acknowledgments

This work is supported by Hebei Province Natural Science Fund Project (No.F2012209019).

References

- [1] Pizard R W. Perceptual Computing Technical Report. Affective Computing[R]. Cambridge: MIT Media Lab, 1995, pp. 57-63.
- [2] Li Jiayuan. Affective Computing Research Cognitive Dilemma. Journal of Dialectics of Nature, Vol.2, 2010, pp.23-28.
- [3] Zhang Chenggong, Liu Peiyu, Zhu Zhenfang, Fang Ming. The sentiment analysis methods based on the polarity dictionary. Shandong University (Natural Science). Vol. 47, No.3, 2010, pp.47-50.
- [4] Aparna Trivedi, Apurva Srivastava, Ingita Singh, Karishma Singh and Suneet Kumar Gupta. Literature Survey on Design and Implementation of Processing Model for Polarity Identification on Textual Data of English Language. IJCSI International Journal of Computer Science Issues, Vol. 8, No. 3, 2011, pp.309-312.
- [5] Nevin VUNKA JUNGUM, Éric LAURENT. Emotions in Pervasive Computing Environments. IJCSI International Journal of Computer Science Issues, Vol.6, No.1, 2009, pp.8-22.
- [6] Xie Lixing, Zhou Ming, Sun Maosong. Based multi-strategy hierarchy of Chinese microblog sentiment analysis and features extraction. The Chinese Information. Vol.26, No.1, 2012, pp. 456-459.
- [7] Pang B, Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. Scott D, ed. Proc. of the ACL 2004. Morristown: ACL, 2004, pp.271-278.
- [8] Du Weifu. Sentiment dictionary build research of text tendentious analysis technology. Harbin Institute of Technology. 2010.
- [9] Akshi Kumar, Teeja Mary Sebastian. Sentiment Analysis on Twitter. IJCSI International Journal of Computer Science Issues. Vol.9, No.3, 2012, pp. 372-378.
- [10] Domingos, P, & Pazzani, M. J. On the optimality of the simple Bayesian classifier under zero-one loss [J].Machine Learning.1997, pp.103-130.
- [11]Jonathon Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification[C]//Proceedings of the ACL Student Research Workshop. Association for Computational Linguistics, Morristown, NJ, 2005, pp.89-92.
- [12]He Xi. Microblog and newspaper picture spread analysis. Business culture (the second half). Vol.5, 2012, pp.236-240.
- [13]Zhao Keqin. Set Pair Analysis and its Preliminary Application [M].Science and technology press.Vol.13, No.47, 1994, pp. 67-72.
- [14]Zhao Keqin. Uncertainty of description and treatment on set pair. Information and control. Vol.24, No.3, 1995, pp.162-166.
- [15]Zhang Chunying, Liang Ruitao, Liu Lu, Wang Jing, 2011,Set pair community mining and situation analysis based on web social network, 2011 International Conference on Advanced in Control Engineering and Information Science, CEIS2011,2011/8/18-2011/8/19,pp.3456-3460,Dali,Yunnan, China(EI).

Chunying Zhang is a PhD in computer application technology of Yanshan University, Master and Bachelor of computer science and technology. She is currently working as a university professor in department of computer science at Hebei United University, and also as the master tutor. She is the members of the experts working group of specialty instruction guidance Committee Subcommittee of the ministry education college and computer science

and technology, a member of the steering committee of computer teaching in Hebei Province; an executive director of the Machine Learning institute in Hebei Province, a director of computer educational research association in Hebei Province, an member of China Computer Federation and other society duties. Her major research directions are: social network analysis, intelligent information processing, and Web mining and other fields. In recent years, she has published more than 50 papers on important domestic, international journals and international conferences, and more than 30 articles are retrieved by the three retrieves.

Jing Wang is doing Master of Science in Applied mathematics from Hebei United University, she has done her bachelor in applied mathematics, and she is currently working in the field of social network analysis.