



HAL
open science

Depression Tendency Recognition Model Based on College Student's Microblog Text

Jie Qiu, Junbo Gao

► **To cite this version:**

Jie Qiu, Junbo Gao. Depression Tendency Recognition Model Based on College Student's Microblog Text. 2nd International Conference on Intelligence Science (ICIS), Oct 2017, Shanghai, China. pp.351-359, 10.1007/978-3-319-68121-4_38 . hal-01820904

HAL Id: hal-01820904

<https://inria.hal.science/hal-01820904v1>

Submitted on 22 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Depression Tendency Recognition Model Based on College Student's microblog Text

Qiu Jie¹, GAO Junbo²

¹Qiu Jie, master graduate. College of Information Engineering, Shanghai Maritime University Shanghai 201306, China

²Gao Junbo (1972 -), male, associate professor, doctor, CCF senior member, main research direction: computing intelligence, data mining.
qiujie1125@163.com

Abstract: In order to solve the issue of identifying depression tendency hidden in microblog text, a depression emotional inclination recognition model based on emotional decay factor is proposed. Make the self-rating depression scale, collect students' microblog text and ask the psychology specialist to annotate the microblog artificially. Construct the depressive emotion dictionary, and then build a depression emotion classifier based on support vector machine. Considering the continuity of depression mood swing, the mathematical model of emotional decay factor is constructed to realize the continuity of discrete emotional state. The experimental results show that the model can effectively identify the depression of user for a period of time, the recognition accuracy rate is 83.82%.

Key words: Depression tendency recognition; emotional decay factor; support vector machine(SVM); depression emotional dictionary; sina micro-blog

1 Introduction

Web2.0 era, micro-blog style format because the free, easy to use and a large number of users to publish their own thoughts and feel. Many researchers use microblogging text for emotional analysis research ^[1], but the research is mainly for a specific thing ^[2], such as film critics, product reviews, etc., and for the text researched on depression emotion a bit less. How to timely and effectively identify the user through the microblogging tendencies, to prevent the long-term spread of depression has become an urgent problem that to be solved.

In the world, depression is one of the most common mental that illness people are faced with ^[3]. With the development of social networking platform such as microblogging, many researchers use the characteristics of user network to determine their psychological depression ^[4]; literature ^[5] based on the frequency of maternal postpartum social networks, to analyze language style and establish a statistical model of maternal depression. Wang ^[6] and so on. The depressive patient is regarded as a node, and a graph network is constructed as the center. According to the attributes of the adjacent nodes in the network and the weight of the connection, the model is given to calculate the depression. ^[7] The method of using brain-imaging study starting at the

resting of the changes of brain function in patients with depression. ^[8] In terms of time users send microblogging, number of fans and a number of concerns such as depression and to analyze the situation of the user.

Existing studies have focused on the microblogging network and external factors to analyze user behavior characteristics of depressive tendencies situation, ignoring the role of micro-blog text Tech in terms of expression. In this paper, we will study the tendency of depression in the students' microblogging text, which integrate the emotional decay factor and time factor into the change of emotion, and describe the fluctuation of the depressive mood of the students more effectively and effectively. Hospital staff to identify patients with depression and provide aid.

2 Related work

2.1 Depression

Depression is a common psychological disease and the cause is very complex, which the researchers on the pathogenesis of depression put forward many theoretical hypotheses ^[9]. Psychological and medical researchers also proposed a variety of Depression Diagnostic Scale, which provides an important experimental basis for the relevant practice. Zung ^[10] proposed depression self-rating scale with a high degree of operability and adaptability; many medical institutions also use this scale to measure the degree of depression in patients. It divides depression into four categories according to the score, [20, 41] indicates normal, [42, 49] indicates mild depression, [50, 57] indicates moderate depression, [58, 80] indicates severe depression. This article produces an online depression self-rating scale (<https://sojump.com/jq/9743549.aspx>) to obtain a score for each student's depression.

3 Depression Emotional Tendency Recognition Model

In this paper, the depression and emotional model is shown in Figure1.

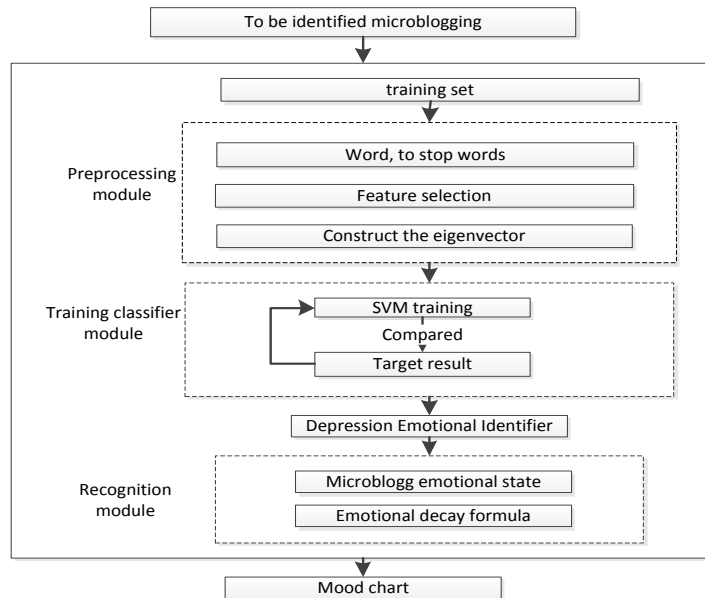


Figure 1 Depression Emotional Tendency Identification Model

The model uses the support vector machine algorithm to construct the depressed emotion classifier, and uses the pretreatment training sample training classifier to obtain the reliable depression emotion recognizer. The identification module that the decay factor into emotions and time factor or the discrete continuous depressive emotional state. The emotional chart obtained by Chart efficient user who determines the depression for a period of time.

3.1 Preprocessing module

Before the microblogging text is used to identify the depression tendency, the acquired raw data must be converted into a data form that the computer can understand.

In order to solve the high dimensionality of feature space in text categorization, this paper uses CHI for feature selection^[11], and then calculates the weight using the classical TF-IDF method^[12] in the field of text processing. In order to solve the sparseness problem of microblogging feature, we use (L T: W) to represent each microblogging, where L is the label of each microblogging, T is the characteristic item, and W is the weight of the characteristic item. Such as: "I really love you, close your eyes, that I can forget, but shed tears, or did not deceive themselves," get five feature words uttered that word feature selection, are "cheat", "Love ", " Tears ", " close your eyes ", " shed ". Therefore, this sentence can be expressed as "1.0 28: 0.4528 39: 0.2295 49: 0.3215 862: 0.5811 1832: 0.54878", where 1.0 is the label, 28 is the characteristic word "cheat" index number and 0.4528 is the weight of the characteristic word.

3.2 The Construction of Emotion Classifier

Based on SVM, the depressive emotion classifier is constructed. Firstly, the nonlinear mapping function is constructed to convert from the original microblogging data x to the high dimensional space H . Then, the linear classifier is used to classify the microblogging emotion in the high dimensional space H . The Gaussian kernel function is used to linearize the nonlinear problem, and the Gaussian kernel function is shown in equation (1).

$$K(x \cdot x_i) = \exp \frac{-\|x-x_i\|^2}{2\delta^2} \quad (1)$$

In the data samples may be time-linear high-dimensional space, SVM obtains the optimal hyperplane by calculating the minimization. At the same time, taking into account the number of data points may deviate from the hyperplane, it is slack variables introduced in the existing conditions based on the constraints, constraints as shown in equation (2).

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad i = 1, 2, \dots, n \quad (2)$$

At this point if you take any value, then any of the super-plane are in line with the requirements. Therefore, it is necessary to add an item after the original minimization function, so that all and the smallest, the new objective function as formula (3).

$$\min \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \right\} \quad 0 \leq \alpha_i \leq C \quad (3)$$

Where C is a parameter that controls the weight between the two above-mentioned objective function. Further, the constraint is added to the objective function, Lagrange function is configured, the solution was expressed constraints corresponding Lagrange multiplier. Corresponding classification function as shown in equation (4).

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x \cdot x_i) + b \quad (4)$$

Where χ is the text of the microblog to be classified, χ_i is the support vector for the text, and y_i is the corresponding classification of χ_i . When $f(x) > 0$, there χ has a tendency to depression; $f(x) \leq 0$ is normal.

3.3 Mathematical model of depression and emotional decay

3.3.1 Basic assumptions

Depression emotion classifier can identify whether each of these two micro-blog text depressive emotional state, and cannot effectively portray the evolution of these two emotional states. As a thoughtful individual, the person's depression is constantly changing over time, depression decay before giving formula, we first proposed the following basic assumptions.

1. The uncertainty of a person is not a depressed patient; this depression only assumes that individuals within any a period of time tend to be fluctuating mood.
2. An initial state is assumed that any normal individuals are emotional state, represented by state c . In this experiment, the state is the result of emotion classifier, there are two states, where $c = 1$ for the depression tendency state and $c = 0$ for the

normal emotional state. The state of a microblogging can only explain the individual's emotional state, and cannot describe the degree of individual state, the corresponding state value.

3. The state definition of a value greater than 0 indicates predisposition to depression, state 0 indicates a value less than pleasant state, emotional state value equal to 0 indicates an equilibrium state. Moreover, the larger the state, the higher status individuals tend to some degree bias. The happy state is a subset of the normal state.

3.3.2 Depression Emotional Decay Formula

The structural decay formula is (5).

$$f(t) = f(t-1) + (-1)^n e^{-\lambda t} \quad (5)$$

Where $f(t)$ represents time, the micro-blog corresponding depressive emotional state value; $f(t-1)$ is the depressive emotion state value corresponding to the microblogging text of the last time. λ is the emotional decay factor, which indicates the decay speed of the emotion, $\lambda = 0.5$, that is to say that the depression emotion is in accordance with the half-life law; the value of n is related to the state of microblogging at two adjacent time points, and according to the above basic assumption 3, the value of n is as follows (6).

$$n = \begin{cases} 2 & \forall c \in \{0,1\} \rightarrow c = 1 \\ 1 & \forall c \in \{0,1\} \rightarrow c = 0 \end{cases} \quad (6)$$

In view of the limited number of students' microblogging text, the time of microblogging has no regularity. In this paper, the time t which in the above-mentioned emotional decay mathematical model is defined as the time interval of two adjacent microblogs, then the range of t is: $T = 0, 1, 2, 3 \dots, n$, and the initial state of any individual $f(t=0) = 0$.

Basic assumptions above 2 (3) specifically explained in emotion decay formula. When two or more consecutive 0 state occurs, when the state of the next moment $c = 1$, then t_i . On the contrary, when the two consecutive or the above state one, if the next time t of the state $c = 0$, then the value of t is not set, but then the next time in turn increase. In the above two kinds of state alternation process, the value of $f(t-1)$ remains unchanged, still the state value of the last moment.

To quantitatively characterize the condition of the individual depression tendency over time, individuals taking herein mean value of the emotional state of a period of time for the metric, as shown in equation (7). Where $t = i$ represents the depressive state value of the i -th microblog from the i -th microblogging, and $f(t = i)$ represents the mean value of depression from the i -th microblog to the n -th microblog.

$$Avg = \frac{\sum_{t=i}^n f(t=i)}{n-i+1} \quad (7)$$

Above mathematical model of emotional decay, this paper presents a mathematical model that identifies whether the individual is depressed over a period, as in equation (8).

$$E(Avg) = \begin{cases} normal, & Avg \in [-1.6, 0.2] \\ depression, & Avg \in [0.2, 2] \end{cases} \quad (8)$$

Equation (8) shows that the average state of individual Yu state in the $[-1.6, 0.2]$ interval, the emotional state is normal; depressive state of the average in the $[0.2, 2]$ interval, there is a tendency to depression.

3.3.3 Case Analysis

Twitter test set is selected from the nickname "Fuji child package" student embodiment, with the emotional classifier on its pre-processed microblogging text classification, in view of the limited space, only select the first five microblogging content and classifier output. The emotional state presented as shown in Table 1 below. The state sequence of the output is $\{1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0\}$. According to Hypothesis 3, the state value $c = 0$ at time $t = 0$, that is. After the initial state, the $f(t = 0) = 0$ sequence is $\{0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0\}$, according to the emotional depression Decay the mathematical model that get the emotional state values ,as shown in Table 2 below. According to the different emotional status values at all times, you can draw depressive mood charts.

For this case, from $t = 0$ to $t = 15$, calculate the mean value of depression $Avg = \sum_{t=0}^{15} f(t) / 16 = 0.7631$. According to the formula (8), it can be judged that the student is in a depressed state during that time.

label	content
1	Did not expect stomach pain into the hospital
0	Meet the Tianzi Square
1	To TM computer, brush more than 200 sheets of A4 and close to the collapse of the end of the period.
0	Headphones placed in the draft of Stefanie and always can cause a lot of resonance and back
0	How the end of the holiday began a little homesick

Table 1 microblogging content and emotional classifier output status
Table 2 emotional state value corresponding to the emotional state sequence

$f(t = 0) = 0$	$f(t = 8) = f(t = 7) + (-1)^1 e^{-0.5 \times 3} = 0.9957$
$f(t = 1) = f(t = 0) + (-1)^2 e^{-0.5} = 0.6065$	$f(t = 9) = f(t = 8) + (-1)^1 e^{-0.5 \times 4} = 0.8604$
$f(t = 2) = f(t = 1) + (-1)^2 e^{-0.5 \times 2} = 0.2387$	$f(t = 10) = f(t = 9) + (-1)^1 e^{-0.5 \times 5} = 0.7783$
$f(t = 3) = f(t = 2) + (-1)^2 e^{-0.5 \times 3} = 0.4618$	$f(t = 11) = f(t = 10) + (-1)^2 e^{-0.5} = 1.3848$
$f(t = 4) = f(t = 3) + (-1)^1 e^{-0.5 \times 4} = 0.3264$	$f(t = 12) = f(t = 11) + (-1)^1 e^{-0.5 \times 2} = 1.0169$
$f(t = 5) = f(t = 4) + (-1)^1 e^{-0.5 \times 5} = 0.2444$	$f(t = 13) = f(t = 12) + (-1)^1 e^{-0.5 \times 3} = 0.7938$
$f(t = 6) = f(t = 5) + (-1)^2 e^{-0.5} = 0.8509$	$f(t = 14) = f(t = 13) + (-1)^2 e^{-0.5} = 1.4003$
$f(t = 7) = f(t = 6) + (-1)^2 e^{-0.5 \times 2} = 1.2188$	$f(t = 15) = f(t = 14) + (-1)^1 e^{-0.5 \times 2} = 1.0324$

4 Experimental results and analysis

4.1 Data acquisition and annotation

From December 2016 to March 2017, data collected from 271 students' microblogging, microblogging that get 7321 text. From the scale score results, 80 people tend to depression, normal 191 people. Ratio of 3: 1, were randomly selected from two types of training with examples and test students, the number of training set and test concentration shown in Table 3 below. After microblog corpus manually labeled, received a total of depression microblogging 1512 and 3786 normal microblogging. There is 1154 test set microblogging. This experiment is carried out in win10 32-bit system, using python language.

Table 3 training set and test set number and microblogging number

	Training	Test
Depression	60	20
normal	143	48
total	5298	1154

4.2 Result analysis

Experiment 1 test concentration Twitter 1154, using the depression and emotional depression without using dictionaries and dictionary emotion experiment, the results shown in Table 4 below.

Table 4 uses the experimental accuracy of depression and emotional dictionary

TRANING SAMPLE	NO ADOPTED DICTIONARY	ADOPTED DICTIONARY
2000	70.62%	80.58%
3000	73.83%	85.68%
4000	75.85%	87.36%
5298	78.45%	90.04%

As can be seen from Table 4, the artificial depiction of emotional depression that a single microblogging recognition accuracy significantly improved.

Experiment 2 uses the Depression Emotional Classifier to classify the microblogging text of each student in the test set, and obtain the intelligence status and classification accuracy of each microblogging emotion of 68 test cases. Some of the data are shown in Table 5 below.

Table 5 different students' single microblogging classification accuracy

	Microblog name	Number of microblog	total	Ac
Normal	盛夏 SQH	15	150	0.96
	谭华要做江南小清新	9	224	0.9821
	小疯子白日梦想家	9	159	0.9811
	夜夜夜先森	7	54	0.9629
	真菌菌凹	20	85	0.8118
Depression	我是一把宝剑	3	27	0.9630
	胖胖是笑声	26	103	0.7865
	Arioster	10	58	0.8793
	图拉拉树袋熊	11	42	0.9048

Using the mathematical model of depressive emotion decay to draw out the emotional trend graphs of 68 test cases, limited to the limited length of the articles, the trend graphs of some test cases were randomly selected from two types of test sets, as shown in Fig. 2 and Fig. 3.

As can be seen from Fig. 2 and Fig. 3, a normal emotional state of the individual curves are generally fluctuations in the χ axis, the majority of the individuals are in a state of pleasure. On the contrary, the emotional curves of individuals with depression tend to be above the χ axis; two types of individuals over time curve mood swings are likely to exceed the critical emotional point $f(t) = 0$, which is also a composite of the real situation of human emotional changes.

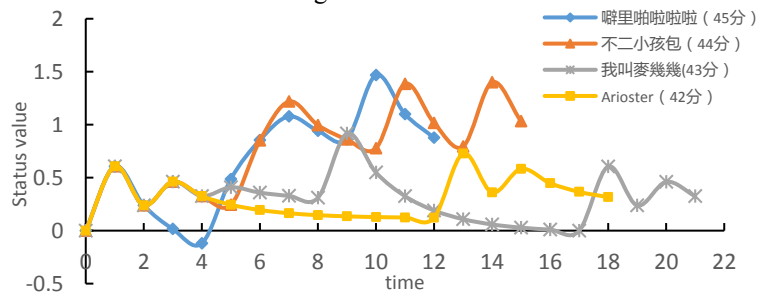


Figure 2 shows the emotional trend of students with depression

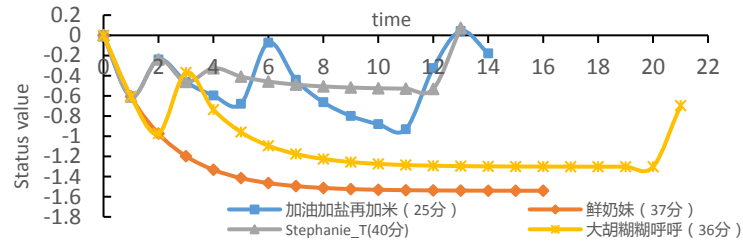


Figure 3 Normal student's emotional chart

Experiment 3 randomly selected two periods for each test case in 68 test cases, with 15 microblogs as a period. Among them, from 20 depressed tendencies of the students which in the depression trend chart selected 40 time periods, from 48 emotional normal students of which the depression trend chart selected 96 time periods, respectively, using the formula (7) (8) to determine. Results are shown in table 6. Recognition accuracy than the normal emotions of depressive tendencies rate is low, probably due to the students in the normal state that issued microblog contains words related to depression.

Table 6 Test results for the test set

	Total	Correctly identify	Accuracy
Depression tendency	40	34	85.00%
Normal mood	96	80	83.75%
total	136	114	83.82%

5 Conclusion

Based on the microblogging text, this paper constructs a model of depression and emotion tendency recognition. The experiment shows that the model can identify the tendency of the individual in a period of time. This paper only considers the interaction of depression at adjacent time points, but does not consider the broader time interval factor. At the same time, the value of emotional decay factor is related to various factors such as environment and personality. It is only assumed that the emotional decay meets the half-life. Regularly, follow-up will be the above aspects of research, with a view to a better experimental result.

References

- [1] Liu Nan, He Yanjie. Emotional analysis for microblogging essay [D]. Wuhan: Wuhan University, 2013: 24-91.
- [2] Liu Li, Wang Yongheng, Wei Hang. Fine-grained Emotion Analysis for Product Reviews [J]. Journal of Computer Applications, 2015, 35 (12): 3481-3486.
- [3] World Health Organization, <http://www.who.int/en/>.
- [4] Youn S J, Trinh N H, Shyu I, et al. Using Online Social Media, Facebook, in Screening for Major Depressive Disorder among College Students [J]. International journal of clinical and health psychology, 2013.13(1):74-80.
- [5] De Choudhury M, Counts S, Horvitz E. Predicting Postpartum Changes in Emotion and Behavior via Social Media[C]//Proceedings of the 2013 ACM annual conference on Human factors in computing systems. ACM, 2013:3267-3276.
- [6] Wang Xinyu, Zhang Chunhong, Ji yang et al. A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network, in Proc. PAKDD-2013 Workshop on Data Analysis for Targeted Healthcare, 2013, in press.
- [7] Qiu Tian-shuang, Dai Rui-jiao, Liu Ya-jie. Study on the Relationship between Depression and Functional Brain Region Based on Restricted FMRI Low Frequency Amplitude [J]. Chinese Journal of Data Acquisition and Processing, 2015, 30 (5): 940-947.
- [8] Li Pengyu, Yu Guang. Microblogging social network in the student depression identification method [D]. Harbin: Harbin Institute of Technology, 2014: 17-60.
- [9] Huang Shijing, Chen Yuxia, Zhang Ying. Magnolia treatment of depression and antidepressant mechanism [J]. World Journal of Integrated Traditional and Western Medicine, 2015 (07).
- [10] Zung W W K, Richards C B, Short M J. Self-rating Depression Scale in an Outpatient Clinic: Further Validation of the SDS [J]. Archives of general psychiatry, 1965, 13(6): 508.
- [11] Aizawa A. An Information-theoretic Perspective of tf-idf Measures [J]. Information Processing & Management, 2003, 39(1): 45-65.
- [12] Vapnik V. The Nature of Statistics Learning Theory[M]. New York, USA: Springer Verlag, 1995.