On Estimating Depressive Tendencies of Twitter Users Utilizing their Tweet Data

Sho Tsugawa*, Yukiko Mogi†, Yusuke Kikuchi‡, Fumio Kishino†, Kazuyuki Fujita§, Yuichi Itoh§, and Hiroyuki Ohsaki§
*Graduate School of Economics, Osaka University, Japan
† School of Science and Technology, Kwansei Gakuin University, Japan
‡ Graduate School of Science and Technology, Kwansei Gakuin University, Japan
§Graduate School of Information Science and Technology, Osaka University, Japan
Email: s-tugawa@econ.osaka-u.ac.jp

***Abstract*—In this paper, we investigate the effectiveness of the records of user's activities in Twitter, which is a popular microblogging site, for estimating his/her depressive tendency. We construct multiple regression model to estimate user's depressive tendency from the frequencies of words used by the user. We perform experiments to estimate participants' depressive tendencies using the constructed regression model. Our experimental results show that there exists medium positive correlation (correlation coefficient $r \simeq 0.45$) between the Zung's Self-rating Depression Scale, which is a popular measure for estimating depressive tendency, and estimated score obtained from the regression model.**

***Keywords*-Depression; Zung's Self-rating Depression Scale; Twitter; Multiple Regression Analysis**

I. Introduction

In recent years, depression is one of significant problems in the world [1]. In the survey conducted by World Health Organization (WHO) in 2012, depression is estimated to affect 350 million people in the world [1]. Since depression can lead to suicide, the increase of depressive people is considered to be a serious problem [1].

For the effective treatment of depression, it is important to recognize depressive tendencies of individuals [2, 3]. WHO reported that the majority of people who need the treatment for depression do not receive it [1]. Among the physicians in the primary care, it has been a problem that depression is not always recognized, correctly diagnosed, and often patients with depression do not receive adequate treatment [3]. Hence, for the effective treatment of depression, it is important to recognize depressive tendencies of patients by themselves or the people around them.

In the literature, several measures for estimating depressive tendencies of individuals from the results of questionnaires have been proposed [4, 5]. Zung's Self-rating Depression Scale (SDS) [4] and Hamilton's Depression Rating Scale (HDRS) [5] are popular measures for estimating depressive tendency. Zung's SDS is obtained from the results of 20 questions answered by the patient [4]. In contrast, HDRS is obtained from the results of 17 questions about a patient answered not by the patient, but by other person [5].

Such measures obtained from subjective information (i.e., results of the questionnaires) are widely used, and expected to achieve high estimation accuracy. On the contrary, it is also important to develop a technique to estimate depressive tendencies of individuals from objective information rather than subjective information. Questionnaire needs a certain amount of costs. In contrast, if depressive tendencies of individuals can be estimated from objective information such as records of their daily activities, the estimation can be performed frequently with very low cost. Such technique to estimate depressive tendencies of individuals form objective information should be effective for the recognition of depression in a patient, and effective treatment for him/her.

As the objective information for estimating depressive tendencies of individuals, we focus on large-scale records of users' activities in social media. In recent years, social media has been getting increased attentions, and therefore, large-scale and fine-grained records of users' activities are available [6]. Many researchers are interested in such large-scale data, and such data are utilized for estimating stock market indicators and inferring results of the elections [7, 8]. Since social media is often used for expressing emotions of users, the records of users' activities should be effective for estimating depressive tendencies of the users.

In this paper, we therefore investigate the effectiveness of the records of user's activities in Twitter, which is a popular microblogging site, for estimating his/her depressive tendency. We conduct a survey to a total of 50 Japanese men and women in their 20s using Zung's SDS, and investigate the relation between depressive tendencies of participants and the frequencies of words used in their tweet messages. We perform multiple regression analysis to estimate Zung's SDS scores of participants from frequencies of words used in their tweet messages. Consequently, we examine the effectiveness of frequencies of words used in the tweet messages for estimating depressive tendencies of Twitter users

IEEE Virtual Reality 2013
16 - 20 March, Orlando, FL, USA
978-1-4673-4796-9/13/$31.00 ©2013 IEEE

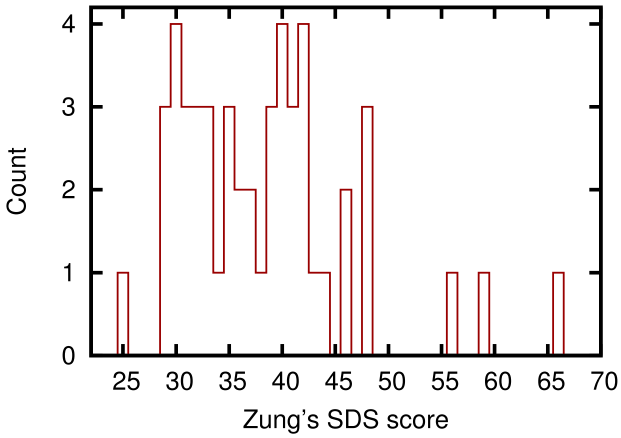

Figure 1: Histogram of Zung's SDS scores of 50 participants

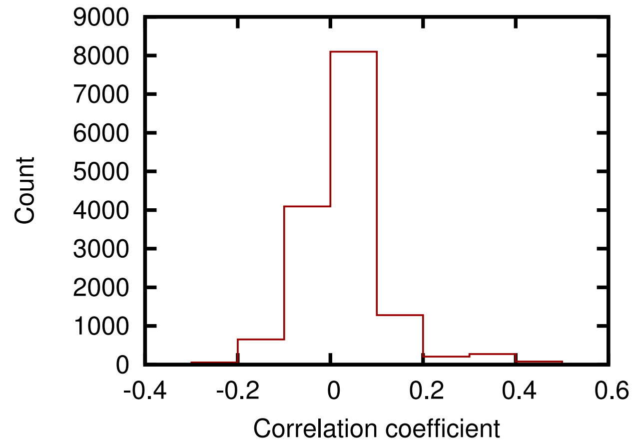

Figure 2: Histogram of correlation coefficients between the Zung's SDS scores of participants and the frequencies of words used in their tweet messages

II. Experiment

As a feature obtained from the records of Twitter users' activities, we focus on the frequencies of words used in their tweet messages, and investigate the effectiveness of the frequencies of words used in the tweet messages for estimating depressive tendencies of Twitter users.

A. Data Acquisition and Morphological Analysis

First, we conducted a survey to 50 Japanese (25 males and 25 females) using Zung's SDS, and obtained Zung's SDS scores of 50 participants. The ages of the participants range from 21 to 27. The participants are recruited from students in our laboratories and our acquaintances who often use Twitter. Zung's SDS score is obtained from the results of 20 questions answered by the participant. Each question is scored on a scale of 1 through 4. The sum of the scores of 20 questions are the Zung's SDS score. Figure 1 shows the histogram of Zung's SDS scores of 50 participants.

We next obtained tweet messages of each participant by using the Twitter API. We obtained the tweet messages within one week after we conducted the survey to obtain Zung's SDS scores. Because of the limitation of the Twitter API, we obtained at most 3,200 tweet messages of each participant.

Finally, we obtained the frequencies of words used in the tweet messages of each participant with Japanese language morphological analysis. We used a popular Japanese morphological analyzer called MeCab [9], and excluded words that are particles, auxiliary verbs, adnominal adjectives, and symbols. Moreover, we excluded words that are used only by one participant among 50 participants. We obtained 14,757 words by the above procedures. Since the number of tweet messages are different among 50 participants, we obtained frequencies of words used by a participant by normalizing the number of occurrences of each word by the number of occurrences of all the words in his/her tweet messages.

B. Correlation Analysis

We first investigate the relation between the Zung's SDS score of each participant and the frequency of each word used in his/her tweet messages. Figure 2 shows the histogram of correlation coefficients between the Zung's SDS scores of participants and the frequencies of words used in their tweet messages.

From Fig. 2, we can find that while frequencies of most words have little correlation with Zung's SDS scores, there exist a few words that have medium correlation with Zung's SDS scores. For instance, there exist 99 words whose absolute values of correlation coefficients with Zung's SDS scores are larger than 0.4. Frequencies of such words are expected to be effective for estimating depressive tendencies of Twitter users.

For analyzing relations among frequencies of words used in tweet messages, we next obtained correlation matrix of frequencies of words used in the tweet messages. Due to space limitation, we only show a correlation matrix of top 10 words whose correlation coefficients with Zung's SDS scores are high (Tab. I). Table I also contains correlation coefficient between frequency of each word and Zung's SDS score.

Table I shows that there exist strong correlations among frequencies of words used in participants' tweet messages. Hence, when we construct a regression model to estimate Zung's SDS score, simply using frequencies of all words

Table I
CORRELATION MATRIX OF ZUNG'S SDS SCORE AND TOP 10 WORDS WHOSE CORRELATION COEFFICIENTS WITH ZUNG'S SDS SCORE ARE HIGH

	Zung's SDS	*even if*	*low fever*	*very*	*workplace*	*hopeless*	*disappear*	*too much*	*sickness*	*bad*	*hospital*
Zung's SDS	1.00	0.59	0.59	0.59	0.57	0.57	0.57	0.55	0.55	0.54	0.53
even if	0.59	1.00	0.91	0.83	0.90	0.55	0.81	0.85	0.87	0.67	0.83
low fever	0.59	0.91	1.00	0.87	0.98	0.50	0.94	0.86	0.98	0.72	0.89
very	0.59	0.83	0.87	1.00	0.87	0.51	0.78	0.78	0.84	0.65	0.75
workplace	0.57	0.90	0.98	0.87	1.00	0.47	0.94	0.88	0.97	0.72	0.87
hopeless	0.57	0.55	0.50	0.51	0.47	1.00	0.36	0.60	0.42	0.51	0.56
disappear	0.57	0.81	0.94	0.78	0.94	0.36	1.00	0.80	0.97	0.71	0.82
too much	0.55	0.85	0.86	0.78	0.88	0.60	0.80	1.00	0.84	0.66	0.76
sickness	0.55	0.87	0.98	0.84	0.97	0.42	0.97	0.84	1.00	0.73	0.88
bad	0.54	0.67	0.72	0.65	0.72	0.51	0.71	0.66	0.73	1.00	0.75
hospital	0.53	0.83	0.89	0.75	0.87	0.56	0.82	0.76	0.88	0.75	1.00

as the independent variables is not adequate. Moreover, we can find from Tab. I that among the words that have high correlations with Zung's SDS scores, there exist negative words such as *hopeless*, *disappear*, *sickness*, and *bad*.

C. Multiple Regression Analysis

We next construct multiple regression model to estimate user's depressive tendency from the frequencies of words used by the user, and evaluate the effectiveness of the constructed model for estimating user's depressive tendency.

We divided the participants into two groups (i.e., training group and testing group), constructed a regression model by using data of the training group, and estimated Zung's SDS scores of participants in the testing group by using the constructed regression model. Each group consisted of randomly selected 25 persons. Candidate words for the independent variables in the regression model are chosen from the top 1 word to the top 40 words whose correlation coefficients with Zung's SDS score are high. We changed the number of candidate words, and constructed regression models. As we have discussed in the previous section, since there exist strong correlations among independent variables, it is not adequate to use all the words as independent variables. We therefore constructed regression models with stepwise selection. We generated 1,000 pairs of the training and the testing groups. In each pair, we constructed a regression model from the data of the training group, and calculated correlation coefficient between actual Zung's SDS scores of participants in the testing group and estimated scores obtained from the regression model.

By applying a regression model constructed from the training group to the testing group, we evaluate the effectiveness of the regression model for estimating depressive tendencies of Twitter users whose Zung's SDS scores are unknown. Note that we obtained candidate words for independent variables from all data (i.e., data from the training and the testing groups). Since the words that have high correlations with Zung's SDS scores are expected to be not so different among different samples, we select candidate words from 50-participants data for the simplicity of the experiments.

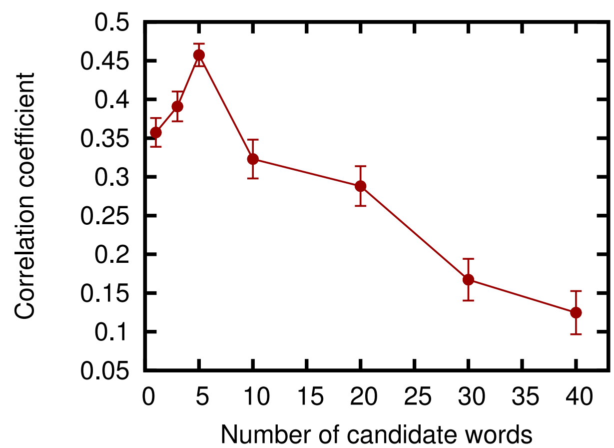

Figure 3: Average correlation coefficients between actual Zung's SDS scores and estimated scores obtained from the regression models when changing the number of candidate words for the regression models

Figure 3 shows average correlation coefficients between actual Zung's SDS scores and estimated scores obtained from the regression models when changing the number of candidate words for the regression models. 95% confidence intervals are also shown in the figure.

Figure 3 shows that when the number of candidate words is 20 or less than 20, there exist medium positive correlation between actual Zung's SDS scores of participants in the testing group and estimated scores obtained from the regression models, which are constructed from the training group.

The medium positive correlation between estimated and actual Zung's SDS scores suggests the effectiveness of the regression models. Correlation coefficient between estimated and actual Zung's SDS scores is approximately 0.45 when

Table II
REGRESSION COEFFICIENTS AND STANDARDIZED REGRESSION COEFFICIENTS IN THE MOST ACCURATE REGRESSION MODEL IN THE EXPERIMENTS WHEN THE NUMBER OF CANDIDATE WORDS IS FIVE (ADJUSTED COEFFICIENT OF DETERMINATION $R^2 = 0.48$, $**$:$p < 0.01$)

	regression coefficient	standardized regression coefficient
low fever	196.01**	0.43**
hopeless	68.09	0.36
(intercept)	34.61**	

the number of candidate words is five. This correlation is not so high. However, this result is considered to be a positive result since the data available for constructing regression models are very small scale. We expect that more accurate regression models can be constructed if more data are available.

We can also find from Fig. 3 that correlation coefficient between actual and estimated Zung's SDS scores takes low values as the number of candidate words is increased more than five. The cause of this is that the stepwise selection does not produce good regression models when too many candidate words are used. The constructed regression model is too optimized for the data of the training group.

We finally investigate the independent variables and their regression coefficients in the constructed regression models in order to identify words that contribute to the estimation of depressive tendencies of Twitter users. Due to space limitation, we only show a regression model when the number of candidate words is five. Table II shows regression coefficients in the most accurate regression model among the constructed regression models. The correlation coefficient between actual Zung's SDS scores of participants in the testing group and scores estimated from the regression model shown in Tab. II is the highest among constructed regression models in our experiments when the number of candidate words is five.

From Tab. II, we can find that frequencies of *low fever* and *hopeless*, which have negative mood, positively affect Zung's SDS score in the regression model. From the standardized regression coefficients, we can find that the frequencies of *low fever* and *hopeless* are almost equally affect the Zung's SDS score.

From above results, it is suggested that frequencies of words used in tweet messages are effective for estimating depressive tendencies of Twitter users. Our experiments are not so large scale. However, we obtained positive results that there exist medium correlation between estimated and actual Zung's SDS scores. We are planning to conduct more large-scale experiments for more detailed evaluations.

III. CONCLUSION AND FUTURE WORKS

In this paper, we have investigated the effectiveness of the records of user's activities in Twitter for estimating his/her depressive tendency. We have constructed multiple regression model to estimate user's depressive tendency from the frequencies of words used by the user. We have performed experiments to estimate participants' depressive tendencies using the constructed regression model. Our experimental results have shown that there exists medium positive correlation (correlation coefficient $r \simeq 0.45$) between the Zung's Self-rating Depression Scale, which is a popular measure for estimating depressive tendency, and estimated score obtained from the regression model.

We are planning to conduct more large-scale experiments for the detailed evaluations. We are also planning to investigate the effectiveness of other features for estimating depressive tendencies of Twitter users.

REFERENCES

[1] M. Marcus, M. T. Yasamy, M. van Ommeren, D. Chisholm, and S. Saxena, "Depression: A global public health concern," tech. rep., WHO Department of Mental Health and Substance Abuse, 2012. Also available as http://www.who.int/mental_health/management/depression/who_paper_depres%sion_wfmh_2012.pdf.

[2] G. Simon, S. Goldberg, B. Tiemens, and T. Ustun, "Outcomes of recognized and unrecognized depression in an international primary care study," *General Hospital Psychiatry*, vol. 21, no. 2, pp. 97–105, 1999.

[3] B. Tiemens, J. Ormel, J. Jenner, K. Van der Meer, T. Van Os, R. Van Den Brink, A. Smit, and W. Van den Brink, "Training primary-care physicians to recognize, diagnose and manage depression: does it improve patient outcomes?," *Psychological Medicine*, vol. 29, pp. 833–845, July 1999.

[4] W. W. K. Zhung, "A self-rating depression scale," *Archives of General Psychiatry*, vol. 12, no. 1, pp. 63–70, 1965.

[5] M. Hamilton, "Development of a rating scale for primary depressive illness," *British Journal of Social & Clinical Psychology*, vol. 6, no. 4, pp. 278–296, 1967.

[6] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, pp. 1320–1326, May 2010.

[7] X. Zhang, H. Fuehres, and P. Gloor, "Predicting stock market indicators through twitter "I hope it is not as bad as I fear"," *Procedia-Social and Behavioral Sciences*, vol. 26, pp. 55–62, 2011.

[8] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM'10)*, pp. 178–185, May 2010.

[9] T. Kudo, "Mecab: Yet another part-of-speech and morphological analyzer." http://mecab.sourceforge.net/.