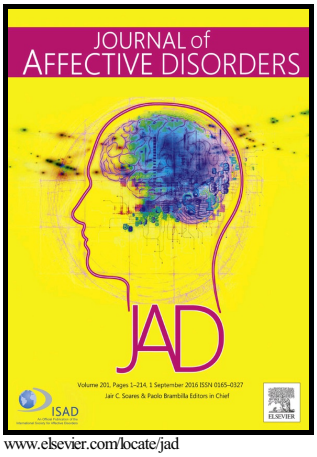

www.elsevier.com/locate/jad

Detecting depression stigma on social media: a linguistic analysis

Ang Li, Dongdong Jiao, Tingshao Zhu

PII: S0165-0327(17)32184-5
DOI: https://doi.org/10.1016/j.jad.2018.02.087
Reference: JAD9633

To appear in: *Journal of Affective Disorders*

Received date: 21 October 2017
Revised date: 4 February 2018
Accepted date: 25 February 2018

Cite this article as: Ang Li, Dongdong Jiao and Tingshao Zhu, Detecting depression stigma on social media: a linguistic analysis, *Journal of Affective Disorders,* https://doi.org/10.1016/j.jad.2018.02.087

Detecting depression stigma on social media: a linguistic analysis

Ang Li [1,2], Dongdong Jiao [3], Tingshao Zhu [2]

[1] Department of Psychology, Beijing Forestry University, Beijing, China

[2] Institute of Psychology, Chinese Academy of Sciences, Beijing, China

[3] National Computer System Engineering Research Institute of China, Beijing, China

Corresponding authors:

(1) Ang Li

Address: No 35 Tsinghua East Rd, Haidian District, Beijing, China (100083); E-mail: angli@bjfu.edu.cn; Tel: +86 13717831110

(2) Tingshao Zhu

Address: No 16 Lincui Rd, Chaoyang District, Beijing, China (100101); E-mail: tszhu@psych.ac.cn; Tel: +86 15010965509

Abstract

Background: Efficient detection of depression stigma in mass media is important for designing effective stigma reduction strategies. Using linguistic analysis methods, this paper aims to build computational models for detecting stigma expressions in Chinese social media posts (Sina Weibo). **Methods:** A total of 15,879 Weibo posts with keywords were collected and analyzed. First, a content analysis was conducted on all 15,879 posts to determine whether each of them reflected depression stigma or not. Second, using four algorithms (Simple Logistic Regression, Multilayer Perceptron

Abbreviations: WUP (Weibo User Pool); API (Application Programming Interface); DSS (Depression Stigma Scale); SCLIWC (Simplified Chinese version of LIWC); GRAE (Gain Ratio Attribute Evaluator); SAE (Significance Attribute Evaluator); CAE (Chi-squared Attribute Evaluator); SLR (Simple Logistic Regression); MLPNN (Multilayer Perceptron Neural Networks); SVM (Support Vector Machine); RF (Random Forest).