HHS Public Access

Author manuscript

Stud Health Technol Inform. Author manuscript; available in PMC 2018 March 01.

Published in final edited form as:

Stud Health Technol Inform. 2017 ; 245: 1292.

Comparison of Different Algorithms for Sentiment Analysis: Psychological Stress Notes

Sunmoo Yoon[a], **Faith Parsons**[b], **Kevin Sundquist**[b], **Jacob Julian**[b], **Joseph E. Schwartz**[b,c], **Matthew M. Burg**[d], **Karina W. Davidson**[b], and **Keith M Diaz**[b]

[a]School of Nursing, Columbia University Medical Center, New York, NY, USA

[b]Center for Behavioral Cardiovascular Health, Columbia University Medical Center, New York, NY, USA

[c]Department of Psychiatry and Behavioral Science, Stony Brook University, Stony Brook, NY, USA

[d]Departments of Internal Medicine and Anesthesiology, Yale University School of Medicine, New Haven, CT, USA

Abstract

To visualize and compare three text analysis algorithms of sentiment (AFINN, Bing, Syuzhet), applied to 1549 ecologically assessed self-report stress notes obtained by smartphone, in order to gain insights about stress measurement and management.

Keywords

natural language processing

Introduction

Psychological stress is linked to all six of the most common causes of death in the U.S. In psychology, content analysis methods derived from paper-and-pencil surveys have been applied to patient records to improve mental health outcomes. With the advance of technology, there is an increasing volume of patient generated free-text data reporting mental health symptoms and context. As a result, natural language processing-computer lingustics has been successfully applied to patient-generated free-text to gain insights from symptom and emotion management. A sentiment analysis package, 'Syuzhet', for processing free-text data has recently become publicly available. However, few studies have applied this package to free-text stress notes or diaries extracted from smartphone-based ecological momentary assessments [1].

Address for correspondence: 630W 168 street, mail code 6, New York, NY, 10032 USA, sy2102@columbia.edu.

This study aims to visualize and compare three algorithms for sentiment analysis (Syuzhet, AFINN, Bing) applied to 1549 ecologically assessed self-report stress notes using smartphones to gain insights into how the analysis of large volumes of stress diaries might inform emotion management.

Methods

We extracted 1549 free-text notes describing self-reported momentary stressful occurrences, which were collected daily from Jan 2014 to April 2015 from sixty participants. Natural language processing was applied using three sentiment analysis algorithms (Syuzhet, AFINN, Bing) [1]. Pearson correlations were calculated between each algorithm and the participant's concurrently self-reported stress rating (0–10 scale).

Results

Figure 1 displays the pooled emotion scores from 1549 stress notes, each applying a different sentiment analysis. Pearson correlation coefficients among the three algorithms and self-rated stress scores are shown in Table 1. The correlations among the three algorithms are moderately high, but the correlations of algorithm scores with self-ratings are low. Positive emotion (lack of negative feeling) was deteced from half of the corpora of stress notes. (e.g., "Excitement!" Syuzhet emotion score +1, Self-report stress score −4).

Conclusion

Application of sentiment analysis natural language processing and visualization techiques provide insights for research teams regarding large volumes of daily self-report stress notes. The positive emotion scores detected by sentiment analysis algorithms from qualitative data (free text) provide quantified descriptive contextual information on low level self-rated stress scores.

Acknowledgments

National Institute of Health grant # R01 HL115941; NSF Institute for Pure and Applied Mathematics

References

1. Jockers M. Package Syuzhet. 2016 V 1.0.

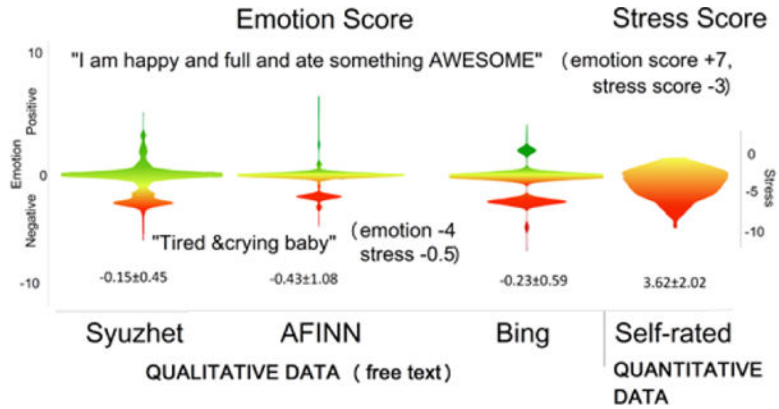

Figure 1.

Visualization of Distribution of Emotion Scores of Daily Stress Notes applying Different Algorithms

Table 1

Correlations among Three Sentiment Algorithms

Algorithms	Syuzhet	AFINN	Bing
Syuzhet	1		
AFINN	0.73**	1	
Bing	0.83**	0.67**	1
Self-Report Score	0.04	0.03	0.03

** p< 0.01, N=1549 notes