

Comparing Emotional Valence Scores of Twitter Posts from Manual Coding and Machine Learning Algorithms to Gain Insights to Refine Interventions for Family Caregivers of Persons with Dementia

Sunmoo YOON^{a,b}, Peter BROADWELL^c, Frederick F SUN^d,
Sun Joo JANG^c and Haeyoung LEE^{c,1}

^a General Medicine, Department of Medicine, Columbia University, USA

^b Data Science Institute, Columbia University, USA

^c Center for Interdisciplinary Digital Research, Stanford University, USA

^d Department of Rehabilitation and Regenerative Medicine, Columbia University, USA

^e Department of Nursing, Chung-Ang University, South Korea

Abstract. We randomly extracted Korean-language Tweets mentioning dementia/Alzheimer's disease (n= 12,413) from November 28 to December 9, 2020. We independently applied three machine learning algorithms (Afinn, Syuzhet, and Bing) using natural language processing (NLP) techniques and qualitative manual scoring to assign emotional valence scores to Tweets. We then compared the means and distributions of the four emotional valence scores. Visual examination of the graphs produced indicated that each method exhibited unique patterns. The aggregated mean emotional valence scores from the NLP methods were mostly neutral, vs. slightly negative for manual coding (Afinn 0.029, 95% CI [-0.019, 0.077]; Syuzhet 0.266, [0.236, 0.295]; Bing -0.271, [-0.289, -0.252]; manual coding -1.601, [-1.632, -1.569]). One-way analysis of variance (ANOVA) showed no statistically significant differences among the four means after normalization. These findings suggest that the application of NLP can be fairly effective in extracting emotional valence scores from Korean-language Twitter content to gain insights regarding family caregiving for a person with dementia.

Keywords. Dementia caregiving, emotional valence, natural language processing

1. Introduction

Dementia is the seventh leading cause of death among Koreans. It is a social norm, strongly rooted in traditional family-centered Korean culture, for family members to provide financial, material, and physical care for a person with dementia [1]. Moreover, the term dementia in the Korean language directly refers to diminished brain function in a demeaning way. Patients with dementia and their families have emotionally suffered

¹ Corresponding Author, Haeyoung Lee, Department of Nursing, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, Republic of Korea 06974; E-mail: id: im0202@cau.ac.kr; Tel.: +82-10-3117-6381.

from the stigmatizing effect of the term. Consequently, psychological distress and burdens have been attributed to family caregivers for a person with dementia among Koreans, regardless of their immigration status.

Emotional valence describes one's affective reaction to an event or a situation's intrinsic positive or badness. Anger and fear are assigned negative valence, while happiness and pleasure map to positive affective valence. Emotional valence can be determined by observing facial expressions, measuring micro-expressions, functional brain imaging, subjective self-reports, and sentiment analysis of one's verbal or written expressions [2]. Although sentiment analysis (Afinn, Syuzhet, Bing) has been commonly applied to conveniently assess sentiment and emotional valence from large corpora of streaming social media data for several years, this method has only recently been widely adopted in health research [2]. Thus there is potential for improving and validating such algorithms in the health context.

The purpose of this study was to compare emotional valence scores of Tweets mentioning dementia/Alzheimer's as determined via machine learning approaches (Afinn, Syuzhet, Bing [2]) to the results of manual coding methods as a foundation for developing future Twitter-based interventions for family caregivers of a person with dementia among Koreans. This study will inform which machine learning approaches (Afinn, Syuzhet, Bing) can be used to measure the effectiveness of such an intervention.

2. Methods

We applied manual scoring and machine learning-based sentiment analysis to calculate emotional valence scores from publicly available Korean Tweets mentioning dementia/Alzheimer's disease, $n=12,413$) posted from November 28 to December 9, 2020 [2,3]. We used the NCapture software for data collection.

First, two independent nursing researchers with dementia caregiving expertise manually determined if the Tweets were relevant to dementia and Alzheimer's disease and manually assigned an emotional valence score to each of the 12,413 Tweets on a scale from -5 [worst] to +5 [best], with 0 being neutral emotional valence.

Second, another bilingual researcher with data science and domain expertise applied natural language processing (NLP) techniques to preprocess the text contents of the Tweets (e.g., stop words removal, translation with the function =*GOOGLETRANSLATE* (*cell*, "*ko*," "*en*")) and machine learning sentiment analysis (Afinn, Syuzhet, Bing) to extract an emotional valence score for each of the Tweets using the R programming language [2]. Next, we compared the aggregated mean emotional valence scores from manual coding to the mean scores produced by the three machine learning sentiment analysis packages listed above, using one-way ANOVA accompanied by visual checking of the four graphs of distributions. The larger study was approved by the Institutional Review Board (IRB). Resources, including analytic R codes and de-identified data, are available on GitHub and OSF.io (<https://osf.io/gruf3>).

3. Results

Among a total of 12,413 Korean-language Tweets mentioning dementia/Alzheimer's disease, approximately one-third of the Tweets ($N=4,364$, 35.16%) were identified as non-relevant to dementia or Alzheimer's disease via manual examination.

Visual examination of the emotional valence distribution graphs revealed that each method produced a unique pattern: 1) Center: While the center of the distribution from manual coding was located around the mid-negative side (-2), the center from the Afinn and Bing algorithms was located at the middle of the emotional score range, indicating neutral valence (0). Similarly, the center of the emotional valence score from the Syuzhet algorithm tended towards a slight negative mean valence, but this was less negative than the mean from manual coding; 2) Spread: A large volume of observations was assigned negative emotional valence status in manual coding, with a smaller volume in the positive direction, but a similar negative-skewed pattern was also found in results from Syuzhet. By contrast, similar volumes of positive and negative valence counts were observed in the outputs from Afinn and Bing; 3) Shape: Manual coding produced a Christmas tree-shaped pattern, with more negative scores enlarging the bottom region, while Afinn and Bing showed symmetrical patterns across the middle (0) point; 4) Unusual features: the results from Bing showed a very large volume of neutral emotional valence (Figure 1).

The aggregated mean emotional scores from the NLP approaches were mostly neutral, while the mean score from manual coding was negative (Afinn 0.029, 95% CI [-0.019, 0.077]; Syuzhet 0.266, [0.236, 0.295]; Bing -0.271, [-0.289, -0.252]; manual coding -1.601, [-1.632, -1.569]). One-way analysis of variance (ANOVA) showed no statistically significant difference among the four means after normalizing the scales of the distributions.

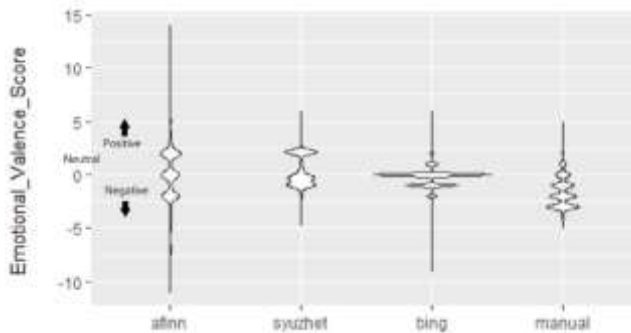


Figure 1. Distribution of emotional valence scores of Korean-language Tweets mentioning dementia/Alzheimer's disease using manual coding and machine learning (Afinn, Syuzhet, Bing)

4. Discussion and Conclusions

This study explored the differences in emotional valence scores assigned to Korean-language Tweets mentioning dementia and Alzheimer's disease using manual versus machine learning approaches.

We found that the mean score from manual coding was negative. Surprisingly we found that the negative use of the term *death* (59 times) was commonly used within the corpus (e.g., a person [with dementia or a family member] should have died, willing to suicide [upon diagnosis]). Considering the social norm and belief of family-centered Korean culture and the stigmatized effect of the term, it seriously concerns the level of burdens and psychological distress among Korean family members of a person with dementia [1].

In terms of our approach, consistent with the findings from other similar studies regarding the novel application of machine learning for sentiment assessment [2,3,4], we found that machine learning extracted emotional valence scores from a large corpus that were generally similar to the results from human coding. In addition, we found that the output from the machine learning algorithms from Afinn and Bing lacked some of the volumes of negative observations extracted from human coding. Consequently, it is reasonable to point out that the quality of the emotional valence scores from the machine learning approaches has room to improve compared to the human coding, considering the following: 1) their lack of medium-scale volumes of negative emotional valence scores (Afinn, Bing) and 2) likely miscategorization of neutral status as positive emotional status (Syuzhet) and negative emotional status as neutral status (Bing) in the context of dementia and Alzheimer's disease.

Nevertheless, machine learning approaches like those provided in the Syuzhet R package are recommended to analyze the sentiment of a fairly large corpus due to its cost-effectiveness for projects with limited resources (e.g., time and budget) [3,4]. It is estimated that a maximum of one week was needed to apply machine learning via Syuzhet to perform an analogous NLP task in a corpus containing 3,798- kilobytes of text (this spanned 462,727 words), while approximately three months were needed to perform the manual emotional valence coding described above (cost estimate: 10 hours * 1 analyst * 90 USD/hour = 900 USD per project for sentiment discovery; 120 hours * 2 coders * 30 USD/hour = 7,200 USD per project for manual content analysis).

In conclusion, the emotional valence scores from applying manual coding vs. automated machine learning approaches from Afinn, Syuzhet and Bing to Korean-language tweets mentioning dementia and Alzheimer's diseases were mostly similar. Nevertheless, the distribution produced via Afinn was slightly skewed away from the negative emotional valences identified via manual coding, while Bing categorized a large volume of negative observations as neutral. These findings suggest that the application of NLP can be fairly effective in extracting emotional valence scores from qualitative data of Korean Twitter to gain insights regarding family caregiving for a person with dementia, and thus can provide a foundation for developing future Twitter-based interventions for Korean dementia caregivers. The accuracy of our findings is limited by using Google Translate.

Acknowledgments

This research was supported by the US federal grant TweetS2 R01AG060929 (PI: Yoon) and Chung-Ang University Research Grants in 2020 (PI: Lee).

References

- [1] Korea Ministry of Health and Welfare. National implementation plans to the responsibility of dementia. Sejong: Korea Ministry of Health and Welfare. 2017.
- [2] Jockers M. Package 'syuzhet'. Available at: <https://cran.r-project.org/web/packages/syuzhet>. 2020.
- [3] Yoon S. What can we learn about mental health needs from Tweets mentioning dementia on World Alzheimer's Day?. *Journal of the APNA*. 2016;22(6):498-503.
- [4] Liu B. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*. 2012 May 22;5(1):1-67.