Article

Social Science Computer Review
2022, Vol. 40(3) 598–617

DOI: 10.1177/0894439320914505
journals.sagepub.com/home/ssc
SAGE

Mapping Social Distress: A Computational Approach to Spatiotemporal Distribution of Anxiety

Yong Suk Choi[1], Hansung Kim[1], and Dongyoung Sohn[1]

Abstract

Anxiety is a pervasive emotional state that tends to arise in situations involving uncertainty due partly to social and contextual issues including competition, economic disparity, and social insecurity. Thus, distribution of aggregate emotions, such as in anxiety, may reveal an important picture of otherwise invisible social processes in which individuals interact with local and global opportunities, constraints, and potential threats. The aim of this study is to present a computational approach to the dynamic distribution of anxiety extracted from natural language expressions of users of Twitter, a popular global social media platform. We develop an unsupervised machine learning procedure based on a naive Bayes model to classify contents of anxiety, estimate the degree of anxiety, and construct a geographic map of spatiotemporal distribution of anxiety. To validate our mapping results, a multilevel statistical analysis was performed to examine how anxiety distribution is correlated with other district-level sociodemographic statistics such as rates of birth and early divorce. Implications for further research and extension are discussed.

Keywords

anxiety, spatiotemporal distribution, machine learning, social media, computational social sciences

Anxiety is an emotional state characterized by feelings of unease, tension, worry, and nervousness (Beuke et al., 2003). Unlike fear, which describes the feeling toward known, realistic threats, anxiety refers to an unpleasant psychological response to vague, mostly unlikely, events or situations. In general, anxiety is a normal mental reaction to stress resulting from an uncertain environment and future, which, arguably, can be useful for survival. Nevertheless, psychologists have often treated anxiety as an abnormal mental condition requiring therapeutic intervention (Rachman, 2009). Implicit in such psychiatric approaches is an assumption that anxiety is a personal emotional problem. This assumption obscures the underlying sociocultural roots of anxiety (Mirowsky & Ross, 1986). In fact, anxiety is such a pervasive feeling that everyone experiences it to some extent, and it

[1] Hanyang University, Seoul, Korea

Corresponding Author:
Dongyoung Sohn, Hanyang University, 222, Wangsimni-ro, Seongdong-gu, Seoul 04763, Korea.
Email: dysohn@hanyang.ac.kr

reportedly affects the choices and behaviors of everyday life. Individuals with higher anxiety levels are more likely to lack coping skills (Raffety et al., 1997), to avoid social interactions and communication (Turner, 1988), to be less generous (Rodebaugh et al., 2016; Rodebaugh et al., 2013), and to make selfish decisions in cooperative settings (Wu et al., 2013). Furthermore, it has been extensively documented that death anxiety triggered by thoughts of one's own mortality leads to in-group favoritism and derogation of out-group members (Burke et al., 2010; See & Petty, 2006).

A recent survey conducted by the American Psychiatric Association shows a society-wide increase in anxiety, with over 40% of respondents in the United States reporting feelings of higher anxiety in comparison to the previous year. This rise has been attributed to widespread usage of social media, exacerbating an endless flow of unpleasant news and gossip, causing people to feel more distressed and insecure (Willams, 2017). The spike was more salient among women, racial minorities, and those with a lower socioeconomic status, suggesting that anxiety is not limited to personal matters but is deeply associated with the circumstantial aspects of one's life. Furthermore, increased anxiety and social distress can have a significant impact on many aspects of life insofar as anxiety gives rise to various social problems such as over-competition for education and jobs, collapse of trust, polarization of public opinion, and increases in violence and suicide rates, among many other examples. These problems may compound over time through path-dependent social processes, possibly leading to disintegration of communities (DiMaggio & Garip, 2012; Page, 2015; Scherer, 2005). Alleviating such potential problems through systematic interventions requires monitoring of distributions and changing trends of social distress across various sociodemographic segments and geographic regions over time, which has been attempted mainly through self-reported measures and interviews.

Despite their numerous advantages, conventional methods, including self-reported measures in surveys, are limited in capturing the sentiments of individuals because they measure post hoc cognitive appraisals rather than actual emotional states at any given moment (Dasborough et al., 2008; Scherer, 2005). In addition, due to reasons such as privacy protection and social acceptability, people are often reluctant to disclose their true emotions and feelings. Furthermore, emotions change over time, often in a dramatic fashion. Even with multiphase surveys, it has proven to be extremely cumbersome to track all the subtle dynamics of emotional trends. Therefore, to overcome these limitations, a method must be developed to observe the emotional expressions of individuals in a less obtrusive manner than direct questioning. As an alternative, we present a computational approach to examining the dynamic distribution of anxiety, extracted from natural language expressions of users of the popular social media platform Twitter. We chose Twitter mainly because users have options to disclose spatiotemporal information in their posted messages, known as *tweets*, more than on other popular platforms such as Facebook. We developed a classifier based on a naive Bayes model to estimate the degree of anxiety implied in user messages.

In the field of affective computing and sentiment analysis of natural language expressions, for the sake of performance improvement, there have been recent attempts to apply advanced deep learning models such as word embedding (Tang et al., 2015), convolutional neural networks (dos Santos & Gatti, 2014; Kalchbrenner et al., 2014), recurrent neural networks (P. Chen et al., 2017; Zhao et al., 2017), memory networks (Hazarika et al., 2018), and attention mechanisms (Majumder et al., 2019). While deep learning models usually require much more training data than traditional machine learning algorithms, agglutinative languages (including Korean, the target language of our study) tend to contain huge numbers of words and word combinations, making it considerably difficult to collect sufficient amounts of data to train deep learning models. Considering the difficulties inherent to the Korean language and the relatively small size of our training data set (i.e., <1 million), instead of a deep learning model, we adopted a simple but relatively robust machine learning model, a naive Bayes model, as a baseline with which future applications of complex deep learning models can be compared.

Combined with type of spatiotemporal information revealed in discrete tweets of users, the number of tweets classifying a user as anxious was computed and converted into an index to indicate the degree of anxiety for a geographic region or district of Korea. Subsequently, all the anxiety indices were visually cast on a national map using color gradients. To validate our mapping results, we utilized multilevel statistical analysis to examine the extent to which the distributions of anxiety indices correlate with existing district-level sociodemographic data. This visual mapping of anxiety or social distress enables intuitive understanding of the current distribution of emotional states for small or large regions and the ability to track spatiotemporal changes and dynamics of anxiety over time, which has not been possible with conventional data collection methods. We believe that the distribution of anxiety quantified and visualized as such stands to complement conventional survey data for various social scientific inquiries and the development of relevant social policies. In the following sections, the study background, proposed procedures, and preliminary results are described in detail.

Anxiety as a Social Indicator

Human societies can be characterized based on temporal patterns or regularities in complex interactions among numerous actors, which in turn shape and constrain their behaviors (Page, 2015). This *emergent* nature of human societies highlights the notion that the feelings and experiences of individuals are deeply intertwined with those of social neighbors and are integral to the larger distribution of psychological experiences (Barbalet, 1998). Living in an area with higher crime rates, for example, may be similar to life in a hostile neighborhood full of distrust for and conflict with neighbors and strangers. These feelings of insecurity may, in turn, force community members to act in a selfish and overly defensive manner. The story may not end there—ample empirical evidence shows that emotions, whether positive or negative, are highly contagious (Bond et al., 2012; Coviello et al., 2014; Kramer et al., 2014). Thus, escalation of personal circumstances has the potential to result in macro-social consequences such as disintegration of community.

Despite having crucial, inseparable associations with macro-social outcomes, human emotions have seldom been at the center of social scientific inquiry (Barbalet, 1998; Turner & Stets, 2006). The main reasons for this neglect are 2-fold. First, emotions, including anxiety, are typically regarded as deeply personal experiences subject mainly to psychological inquiries and interventions. Second, in lacking proper methods or instruments for capturing the transient nature of feelings, it has been viewed as far-fetched to attribute certain social phenomena to underlying emotions or sentiments (Turner & Stets, 2006). Consequently, instead of treating emotions explicitly, the long-standing approach of social scientists has been to translate emotions into calculable outcomes such as utility, uncertainty, and risk (Schoemaker, 1982). As a result, human behaviors have largely been portrayed as rational, cognition-driven processes, devoid of irrational or emotional components (Barbalet, 1998; Slovic, 1995).

This has been changing as a massive amount of trace data, including textual online content, has become available due to the global proliferation of social media. As of 2018, for example, there were an estimated 330 million Twitter users worldwide, producing 500 million tweets containing various emotion-laden words per day. This abundance of textual data opens a new venue for studying the roles of emotions in social processes (Cambria, 2016). Various attempts have been made to apply sentiment analysis to social media content to explain and predict social consequences including stock market changes (Bollen et al., 2011; Nguyen et al., 2015), drug reactions (Korkontzelos et al., 2016), brand management (Mostafa, 2013), and many other examples. Most prior attempts, however, have been confined to identifying the implicit tone and valence of individual texts, whether positive or negative, and to linking personal posts to social issues of interest.

Relatively little attention has been paid to identifying particular types of emotions in social media, including anxiety.

Defined as an emotional response to unspecified threats, anxiety tends to arise in uncertain circumstances due to social and contextual issues such as competition, economic disparity, and social insecurity. Prior studies have found that increased anxiety might cause individuals to express anger in response to stressful social exchanges (McClure et al., 2007), to become more selfish (Rodebaugh et al., 2016; Rodebaugh et al., 2013), and to be less generous (Wu et al., 2013) when cooperation is needed. Furthermore, an overall increase in anxiety might correlate to a society-wide increase in depression (Horwitz, 2010), possibly worsening other serious social problems including violence, suicide rates, substance abuse, and obsessive-compulsive behaviors. Preventive interventions for these types of social problems require close monitoring of societal distributions and changes in anxiety and other relevant emotions. In the context of conventional observation methods such as surveys, a major obstacle to effective analysis is that anxiety is a latent emotion that is difficult to detect with direct questioning. Anxious feelings often manifest as other emotional expressions including heightened sensitivity to surroundings, nervousness, anger, panic, and/or frustration.

This suggests that we need, as an alternative, some indirect, unobtrusive approaches to detecting and extracting the signs or symptoms of anxiety from natural language expressions and behaviors of individuals. Social media platforms offer an excellent environment containing natural behavioral traces from which meaningful patterns can be identified through the application of machine learning procedures. This does not imply that conventional, interview-based data collection methods are inappropriate or obsolete. Instead, the main premise of this study is that such an indirect, unobtrusive approach to detecting emotions stands to reveal an important view of otherwise invisible social processes in which individuals interact with local and global instances of opportunities, constraints, and threats. That is, the spatiotemporal distribution of aggregate emotions, such as anxiety, if properly captured and quantified, may serve as a valuable source of information by complementing existing social statistics and opinion polls. Furthermore, visual mapping of the spatial distribution of anxiety over time will allow researchers to examine the longitudinal dynamics and trends of aggregate emotions, which can be applied to the diagnosis of regional problems and issues as well as the study of relevant social phenomena and policy development.

Extracting Anxiety

Parameter Estimation and Morphological Analysis

In statistical terms, parameters refer to certain numerical values that represent population characteristics or a data generating process to be estimated from sample data assuming a random variable with a family of probability distributions. From among many methods of parameter estimation (J. V. Beck & Arnold, 1977), we employ the two well-known methods of maximum likelihood (ML) estimation and maximum a posteriori (MAP) estimation, calculated as follows.

$$\begin{aligned} L(\theta;\, x_1, x_2,\ \ldots,\, x_n) &= L(\theta; X), \\ &= f(x_1, x_2,\ \ldots,\, x_n \mid \theta) = f(X \mid \theta), \end{aligned} \tag{1}$$

$$\arg\max_{\theta} L(\theta; X) = \arg\max_{\theta} f(X \mid \theta)\,, \tag{2}$$

$$\arg\max_{\theta} f(\theta \mid X) = \arg\max_{\theta} \frac{f(X \mid \theta) f(\theta)}{f(X)}\ . \tag{3}$$

ML estimation is a method of estimating the parameters of a statistical model by finding the parameter values that maximize the likelihood of the observations (Guindon & Gascuel, 2003).

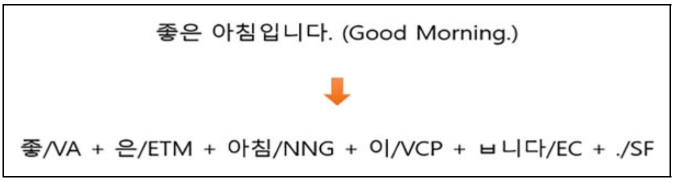

Figure 1. Example of a morphological analysis using Korean Morphological Analyzer.

Suppose that there is a sample $X = (x_1, x_2, \ldots, x_n)$ of n observations, from a distribution with an unknown probability density function $f_0(\cdot)$. The likelihood function can be defined as in Equation 1, and then the ML estimation finds the parameters that maximize the likelihood function as in Equation 2. While ML estimation depends heavily on observations (Sowell, 1992), MAP estimation finds an estimate of an unknown quantity that is equal to the mode of the posterior distribution based on Bayes's theorem. If a priori information is available, then MAP estimation might yield better results than ML. In the MAP estimation shown in Equation 3, $P(\theta)$ is an a priori probability, and $P(\theta\,|\,x)$ is an a posteriori probability.

The naive Bayes classifier, one of the most popular probabilistic classifiers, is based on Bayes's theorem with naive independence assumptions among features (Bayes, 1763; McCallum & Nigam, 1998; Vapnik, 1998). The naive Bayes classifier can be trained very efficiently in supervised learning. In many practical applications, parameter estimation for naive Bayes models uses the method of ML with no assumption of Bayesian probability or methods.

$$\begin{aligned} \hat{c}_{ML} &= \underset{c \in C}{\arg\max}\, P(x_1, x_2, \ldots, x_n \,|\, c)\,, \\ &= \underset{c \in C}{\arg\max} \prod_{i=1}^{n} P(x_i \,|\, c)\,, \end{aligned} \tag{4}$$

$$\begin{aligned} \hat{c}_{MAP} &= \underset{c \in C}{\arg\max}\, P(c \,|\, x_1, x_2, \ldots, x_n), \\ &= \underset{c \in C}{\arg\max} P(c)\, P(x_1, x_2, \ldots, x_n \,|\, c), \\ &= \underset{c \in C}{\arg\max} P(c) \prod_{i=1}^{n} P(x_i \,|\, c). \end{aligned} \tag{5}$$

In text classification, the term θ (in Equations 1–3) is classified by ML estimation in Equation 4 and MAP estimation in Equation 5. The value of C denotes a set of classes, and denotes words composing a sentence. The class that maximizes the equation is. The naive Bayes classifier assumes that all attributes of the examples are independent of each other given the context of the class. Equation 4 and Equation 5 are derived based on this assumption. Although this assumption is unrealistic in many real-world tasks, the naive Bayes model is very effective in performing text classification (Lewis, 1998; Rennie et al., 2003).

Morphological analysis for understanding the linguistic characteristics and structure of words is fundamental to natural language processing (Kao & Poteet, 2007; Kim et al., 1994; Manning et al., 2014). In the context of this study, each tweet is broken down into several *morphemes*, which are the smallest meaningful units of a language with a part of speech (POS) tag such as a noun, verb, or adjective. Because the tweets for the current study were collected in Korea, we employed the well-known Korean Morphological Analyzer (KOMORAN) to analyze tweets.

Table 1. Frequency Dictionary of Nonanxious and Anxious Tweets.

Word/Part of Speech	Nonanxious	Anxious
word_A/NNG	200	30
word_B/VV	100	10
word_C/VA	200	0
word_D/NNG	100	20
word_E/VV	0	30
word_F/MAG	400	10
Total	1,000	100

Note. NNG = noun; VV = verb; VA = adjective; MAG = adverb.

Figure 1 exemplifies how a sentence in Korean is composed of multiple morphemes. On the left of the slash (/), "좋," "은," "아침," "이," "ㅂ니다," and "." are morphemes, while "VA," "ETM," "NNG," "VCP," "EC," and "SF" are POS tags. Irrelevant parts of speech for sentimental analysis were filtered out, and only NNG (noun), VV (verb), VA (adjective), MM (determiner), and MAG (adverb) were used in morphological analysis.

Naive Bayes Text Classification

After the sets of tweets were morphologically analyzed, a naive Bayes classifier (A. Y. Ng & Jordan, 2002) was constructed using a frequency dictionary of word morphemes as obtained from the collection of tweets. Table 1 shows a template of the word morphemes/POS tags together with their frequencies from sample sets of nonanxious and anxious tweets. Unlabeled tweets can be further classified based on the frequency dictionary.

As an illustrative example, suppose a frequency dictionary is constructed from sample tweets initially tagged as anxious (anxiety) or not (nonanxiety), as in Table 1. With ML estimation, the probability that a new sentence consisting of word_A, word_B, and word_D is classified as anxiety is $P(\text{word}_A, \text{word}_B, \text{word}_D \mid \text{anxiety}) = \frac{30}{100} \times \frac{10}{100} \times \frac{20}{100} = .006$, while the probability of being classified as nonanxiety is $P(\text{word}_A, \text{word}_B, \text{word}_D \mid \text{nonanxiety}) = \frac{200}{1{,}000} \times \frac{100}{1{,}000} \times \frac{100}{1{,}000} = .002$. In this way, the sentence is classified as anxiety because the probability of such is greater than that of being classified into the nonanxiety condition.

If a sentence consists of word_B, word_D, and word_F, then the probability of being classified as anxiety is $P(\text{word}_B, \text{word}_D, \text{word}_F \mid \text{anxiety}) = \frac{10}{100} \times \frac{20}{100} \times \frac{10}{100} = .002$, while the probability of being assigned to the nonanxiety condition is $P(\text{word}_B, \text{word}_D, \text{word}_F \mid \text{nonanxiety}) = \frac{100}{1{,}000} \times \frac{100}{1{,}000} \times \frac{400}{1{,}000} = .004$. Accordingly, the sentence is classified as nonanxiety.

With MAP estimation, on the other hand, classification results might differ slightly. For the same sentence consisting of word_A, word_B, and word_D, the probability of being classified as anxiety is $P(\text{anxiety}) \cdot P(\text{word}_A, \text{word}_B, \text{word}_D \mid \text{anxiety}) = \frac{100}{1{,}000+100} \times \frac{30}{100} \times \frac{10}{100} \times \frac{20}{100} \approx .00055$, while the probability of being classified as nonanxiety is $P(\text{nonanxiety}) \cdot P(\text{word}_A, \text{word}_B, \text{word}_D \mid \text{nonanxiety}) = \frac{1000}{1{,}000+100} \times \frac{200}{1{,}000} \times \frac{100}{1000} \times \frac{100}{1{,}000} \approx .00182$. Thus, the sentence is classified as nonanxiety because of the higher probability of nonanxiety than of anxiety.

If there is a sentence consisting of word_A, word_C, and word_E, then the probabilities of both classes are zero because $P(\text{word}_C \mid \text{anxiety}) = P(\text{word}_E \mid \text{nonanxiety}) = 0$. To solve this "zero frequency" problem, we used add-one smoothing (also known as Laplace smoothing), which assigns nonzero probabilities to words that do not occur in a collection. For each class, every word is given

Table 2. Performance Comparison of ML and MAP Estimations.

Method	Recall of *Anxiety*	Recall of *Nonanxiety*	Accuracy
ML	.8550	.8292	.8317
MAP	.5069	.9641	.9371

Note. ML = maximum likelihood; MAP = maximum a posteriori.

an extra one frequency, and the total frequency is added to the size of the word vocabularies (S. F. Chen & Goodman, 1999; Pang et al., 2002; Peng & Schuurmans, 2003; Yuan et al., 2012). Consequently, our example sentence has the following probabilities, which result in classification into the anxiety condition: $P(\text{word}_A, \text{word}_C, \text{word}_E \mid \text{anxiety}) = \frac{30+1}{100+6} \times \frac{0+1}{100+6} \times \frac{30+1}{100+6} \approx .000806$ and $P(\text{word}_A, \text{word}_C, \text{word}_E \mid \text{nonanxiety}) = \frac{200+1}{1{,}000+6} \times \frac{200+1}{1{,}000+6} \times \frac{0+1}{1{,}000+6} \approx .000039$.

Experimenting With Twitter Data

Twitter Data Collection

Using a tweet crawler (Twitter4j) distributed publicly as an open application programming interface,[1] we collected over 2.7 million (2,711,807) tweets containing both spatial and temporal information (i.e., tweets specifying the time and user location when posted) from February 2016 through November 2017 in South Korea. All data were collected and used according to Twitter's terms of service and privacy conditions. To prepare the initial data for training the anxiety classifier, 91,269 tweets (3.37%) from the total collected were randomly sampled and tagged by human coders either as anxiety related or not anxiety related. Four graduate students in human communications at a large metropolitan university in Korea were recruited to operate as the coders.

Anxiety might manifest in the guise of other relevant emotions such as "nervousness," "perplexity," "worry," "frustration," and/or "discomfort." Thus, the coders were instructed to consider expressions of not only explicit anxiety but also other relevant emotions insofar as they could be symptomatic of the state of "being anxious,"[2] particularly when the emotions appeared in association with things that are difficult to control, such as future events, environmental uncertainty, or the presence/behaviors of other people (i.e., social anxiety). Among the tweets tagged by the human coders, only 81,873 (89.7%) showed consistent tagging results among all four coders (i.e., all four coders assigned the tweets the same tags). Only these tweets were incorporated in the next step and were segmented into sets of 75,051 training and 6,822 testing tweets. In the training set, the number of anxious tweets was 7,366. Thus, $P(\text{anxiety}) = \frac{7366}{75051} \approx .098$ and $P(\text{nonanxiety}) = \frac{67685}{75051} \approx .902$.

From the tagged training set, we constructed our frequency dictionary (enlisting word morpheme frequencies in anxious tweets and nonanxious tweets) using KOMORAN. A naive Bayes classifier was built using the frequency dictionary to estimate whether newer tweets were classified as anxious.

Comparative Analysis of ML and MAP Estimations

Using the testing set of 6,822 tagged tweets, we first evaluated the performance of our classifier by applying ML and MAP estimations.

As shown in Table 2, the MAP estimation performed better than the ML estimation in terms of accuracy. In terms of recall of anxiety, however, MAP estimation was worse. Because the training set is greatly biased toward nonanxiety data (i.e., $P(\text{nonanxiety}) \gg P(\text{anxiety})$), the results of classification applying MAP estimation are mostly Nonanxiety, and MAP leads to poorer performance in the recall of anxiety but better performance in accuracy. In contrast, ML estimation shows reasonable performance overall because it is not influenced by any prior distribution of $P(\text{anxiety})$ and $P(\text{nonanxiety})$. Because our purpose is to identify anxious tweets, our emphasis is on recall rather than accuracy. Thus, ML estimation appears to be more appropriate than MAP estimation for our purpose. Hence, we applied ML estimation to a naive Bayes classifier for further analyses (Chawla, 2005; Frank & Bouckaert, 2006; Mladenic & Grobelnik, 1999).

To improve the performance of the naive Bayes classifier applying ML estimation, we formulated a criterion that, if Equation 6 is true, then a sentence, as tweeted, is classified into the anxiety condition.

$$\frac{P(\text{Tweet}|\text{anxiety})}{P(\text{Tweet}|\text{noanxiety})} > \text{Threshold}. \tag{6}$$

A basic naive Bayes classifier is supposed to classify tweets with threshold = 1, which is not optimal for our application. Instead, we performed another experiment after changing the value of the threshold (V. Ng & Cardie, 2002; Schneider, 2005). Figure 2 shows the results of this experiment. Here, we used products of “recall of anxiety” and “accuracy” to reflect both measures at the same time in terms of geometric mean. The product of the two measures had a maximum value of 0.7294 at threshold = 2.5. Consequently, we adopted a threshold of 2.5 for Equation 6.

Spatiotemporal Mapping of Anxiety

The degree of anxiety estimated and normalized was then mapped with gradient colors onto all the districts in Korea as shown in Figures 3 and 4.[3] The regions in darker red colors had greater proportions of anxious tweets among all the tweets therein, while the opposite is true for the regions in darker blue colors (with diagonal patterns overlaid in the grayscale images). Figure 3 depicts a spike of anxiety in the county of *Seongju* in the larger province of *Gyeongsangbuk-do* during July 2016 and August 2016. This was when military forces from the United States and Korea announced deployment of the Terminal High Altitude Area Defense antiballistic missile system in the region. Aggressive protests in the region occurred as the residents strongly opposed deployment of the missile system in their region, which put them at risk of becoming an important military target for potential enemies. Note that the recorded level of anxiety was heightened only in the county, even though news about the protests was broadcast nationwide. This result partly indicates that the computational mapping presented here effectively captures and depicts dynamic emotional changes occurring in geographic regions in the context of ongoing important social events.

Figure 4 shows another example—distribution of anxiety for the month of March 2017 during which the then president of Korea, Park Geun-hye, was impeached. The provinces in dark red colors are known political hotbeds in Korea—the province on the right-hand side, *Gyeongsangbuk-do*, is arguably the most politically conservative region and is regarded as the political hometown for the impeached former president and her party. In contrast, the region on the left-hand side, *Jeollabuk-do*, is among the most politically liberal/progressive provinces in Korea. The figure visually captures times during which, regardless of ideological stance, levels of anxiety were notably higher in the two provinces due to political upheaval in Korea. Taken together, the figures suggest that spatiotemporal mapping of anxiety shows results with at least some surface validity.

Figure 5 maps the counties or districts with the highest (i.e., red) and lowest (i.e., blue) levels of anxiety. At first glance, these regions might not seem qualitatively different in that most of them are small suburban or rural areas. The most anxious regions, however, were those with a greater proportion of elderly people (i.e., mean = 30.85%) in comparison to the least anxious regions (i.e., mean = 23.05%). Furthermore, the mean birth rate for the most anxious regions was higher

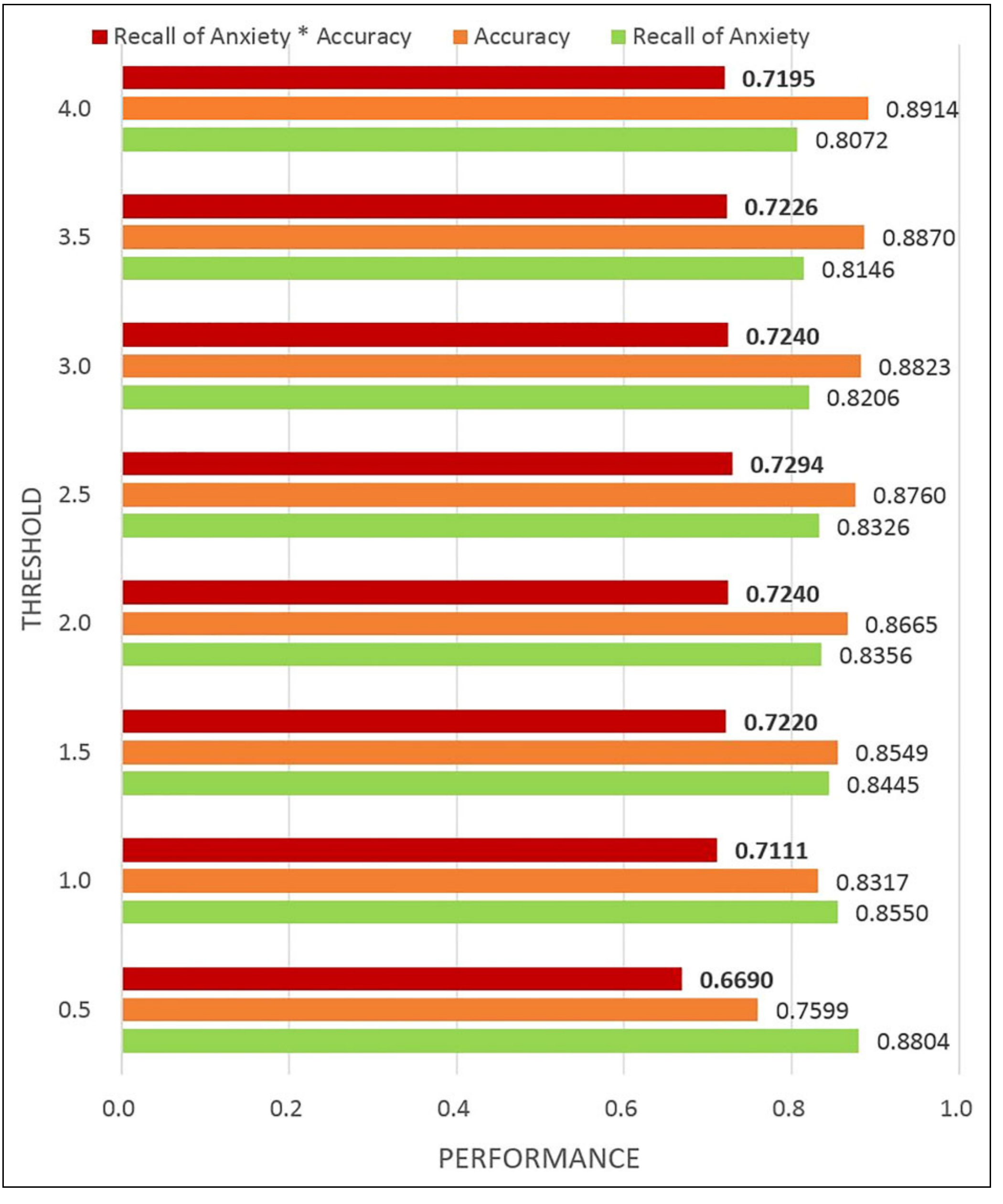

Figure 2. Performance of the classifier over threshold values.

(i.e., mean = 1.48) than that for the least anxious regions (i.e., mean = 1.29). Also of interest is that the most anxious regions were concentrated in southwestern provinces, namely *Jeolla-do*, while the least anxious regions were spread nationwide. Finding explanations for these patterns is beyond the scope of this study. It is certain, however, that there are more factors involved in these correlations than our first observations revealed, which merits further systematic attention.

The mapping of anxiety can also be conducted across time, allowing examination of temporal shifts in the distribution of anxiety. As an example, Figures 6 and 7 exemplify temporal fluctuations

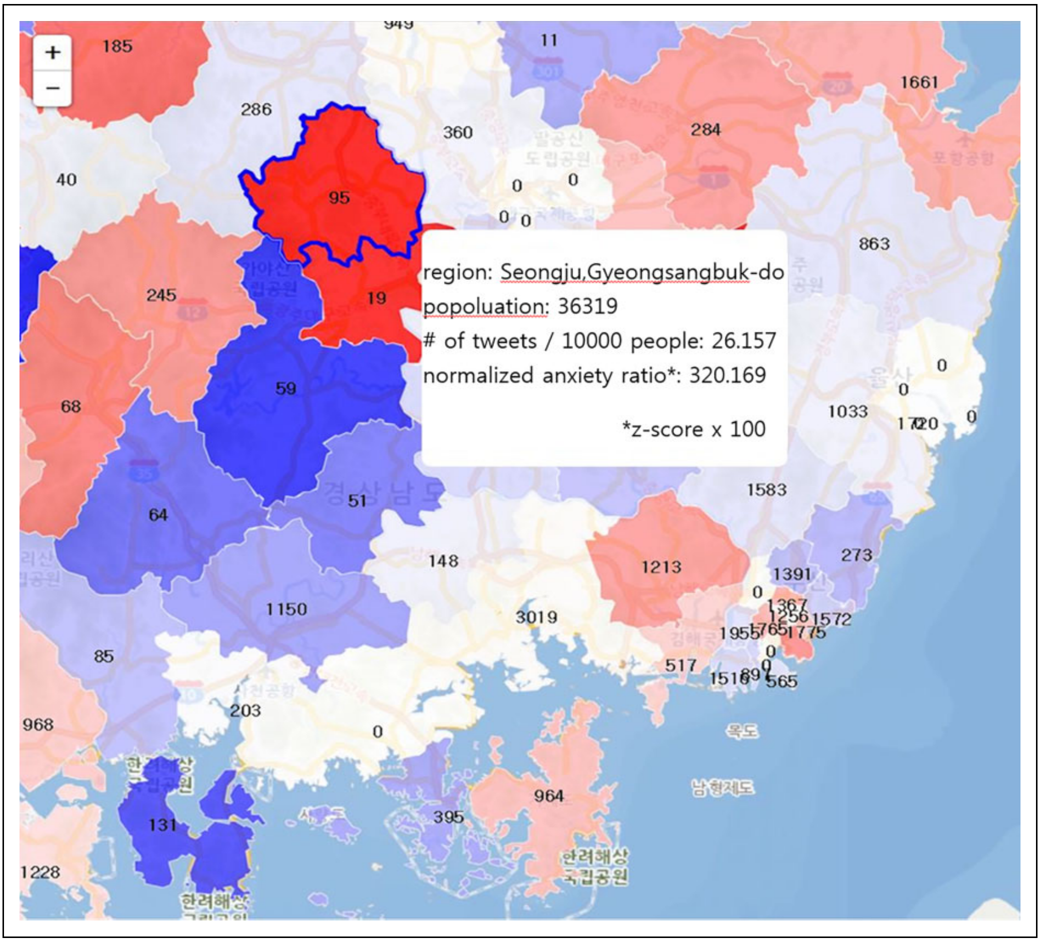

Figure 3. Anxiety mapping result example (July 2016–August 2016).

in total and anxious tweet counts in a selected region as well as in the anxious tweet counts between regions of interest and the entire nation, respectively. The figures show how both total number of tweets and degree of anxiety significantly changed in the province of *Jeollanam-do* and nationwide around the periods of two large political events occurring in Korea—namely massive civil demonstrations for the impeachment of a president (August 2016–January 2017) and a presidential election (May 2017).

Validation

An important emerging question is whether the anxiety indices extracted from tweets, as shown above, can be considered valid indicators of true levels of anxiety across geographic regions. In an effort to answer this question, we attempted to test the criterion validity. One approach for criterion validation is the determination of the extent to which a construct is related to other constructs in a predictable manner (O'Leary-Kelly & Vokurka, 1998). In advanced industrial societies, anxiety has built up around various issues such as fear of environmental degradation, incurable disease, and economic crisis (Hollway & Jefferson, 1997; Smith, 1988; Ungar, 2001). According to U. Beck (1992), modern society is a "catastrophic society" where risks are largely unpredictable,

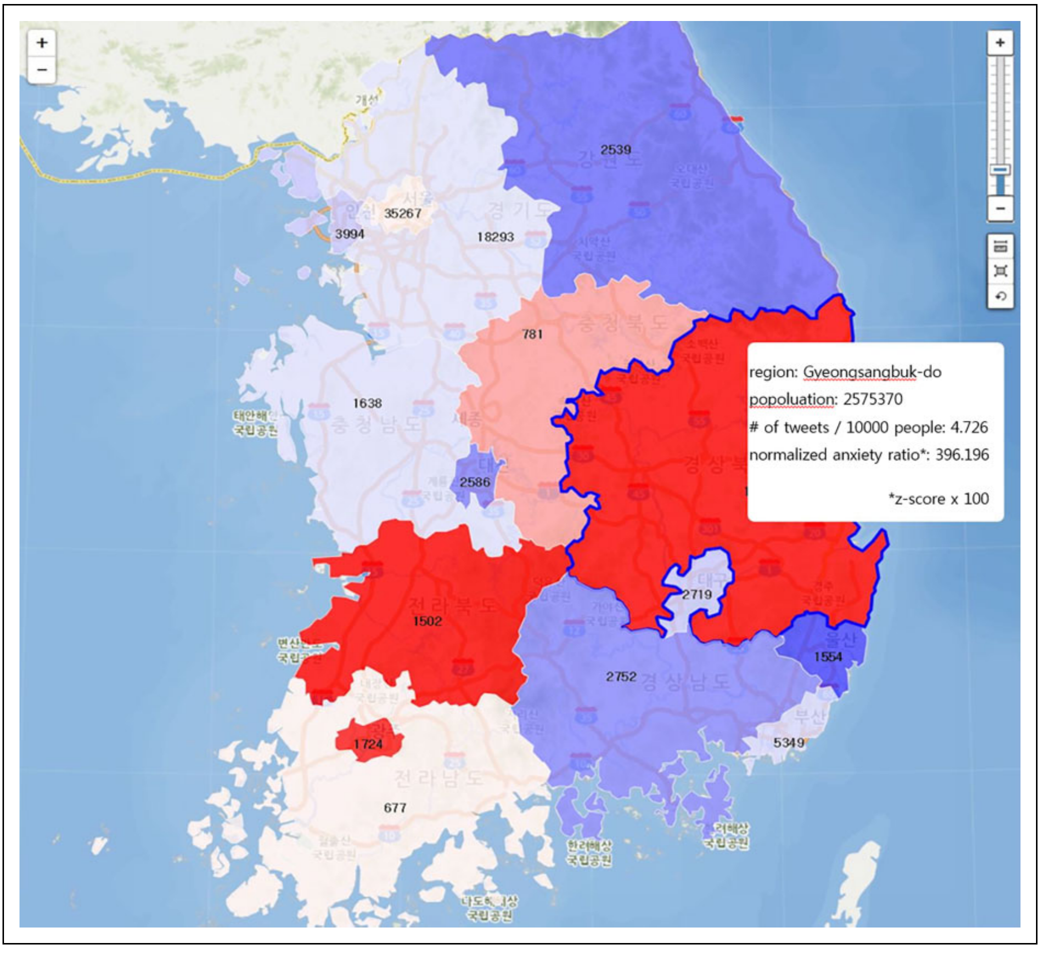

Figure 4. Anxiety mapping example (March 2017).

incalculable, and complex in cause (U. Beck, 1992; Ungar, 2001). In addition, anxiety is an occasional experience that is common to all and one that should be understood as a product of social processes (Wilkinson, 2001).

From this perspective, anxiety is related to position or location within the structure of society together with levels of commitment toward predominant cultural values (Wilkinson, 2001). As the ecological approach suggests, observable variations in human attitudes and behaviors across areas can be explained by distinctions of compositional and contextual characteristics (Macintyre & Ellaway, 2000). Therefore, aggregate anxiety at the community level is expected to be associated with various other community characteristics. Prior studies using traditional survey data in Korea have shown that young women, people living in poverty (Macintyre & Ellaway, 2000), and elderly people (Hong et al., 2006) tend to have relatively higher levels of anxiety. Along similar lines, an analysis of Google search queries has found that searches for anxiety and anxiety-relevant topics tend to be more frequent in places with a lower median income and lower education levels and in rural areas in comparison to urban metropolitan cities (Lee, 2003). Thus, to be valid indicators, our anxiety indices ought to be correlated with these indicators to some extent.

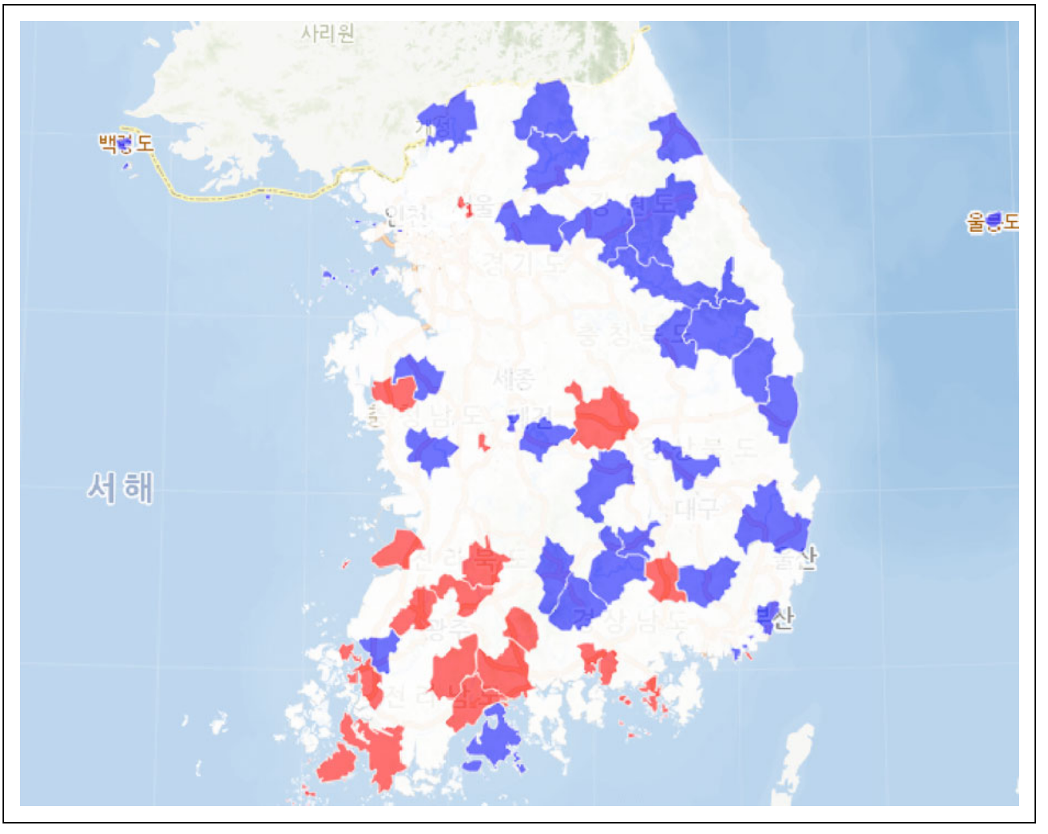

Figure 5. Regions with highest and lowest anxiety levels (February 2016–November 2017).

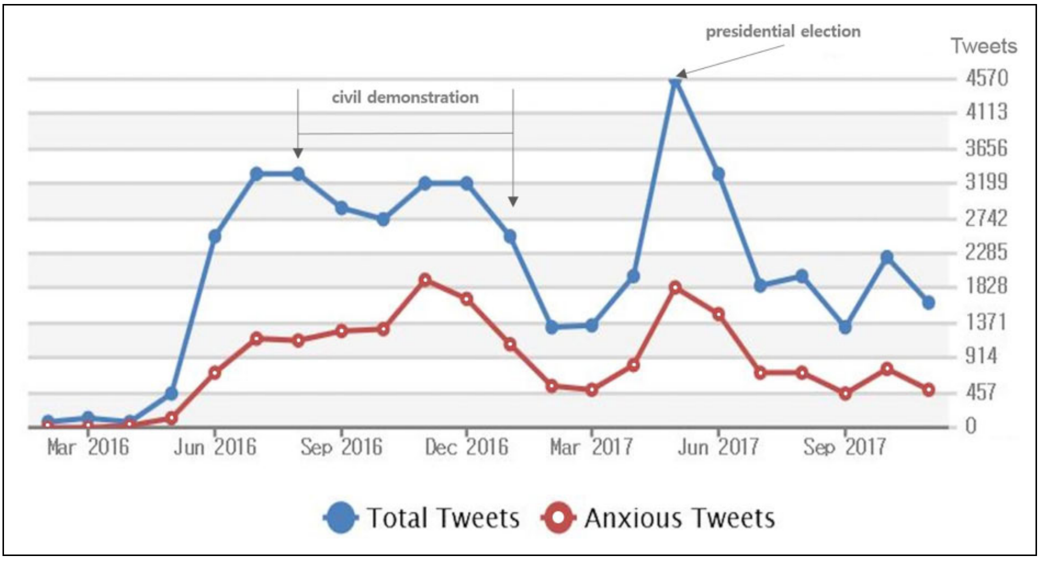

Figure 6. Comparison of total tweets and anxious tweets in *Jeollanam-do.*

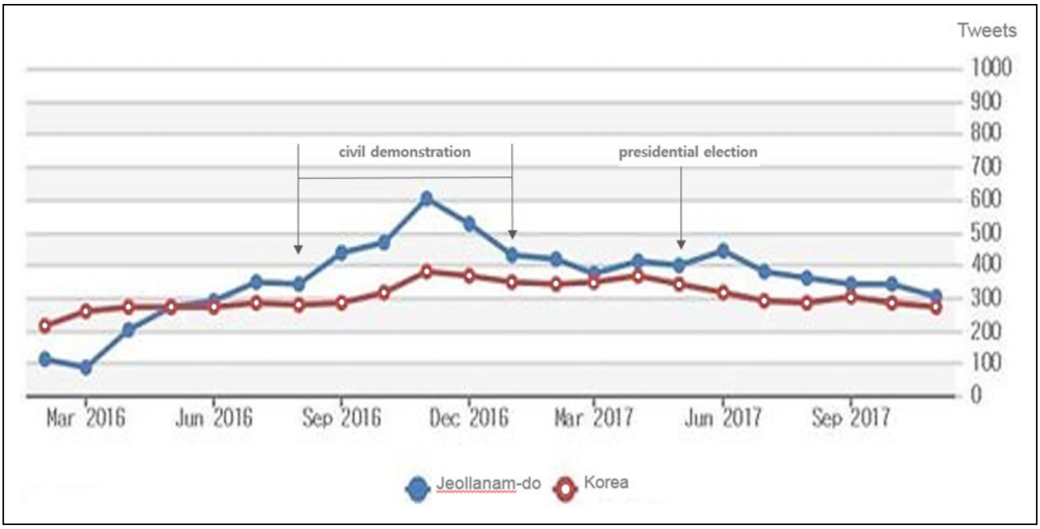

Figure 7. Comparison of regional and nationwide anxiety levels (anxious tweets per 1,000 tweets).

Table 3. Generalized Multilevel Modeling Results.

	Model I	Model II	Model III
Intercept	–1.14 (.04)***	–1.13 (.04)***	–1.11 (.04)***
Birth rate		0.05 (.02)*	0.05 (.02)*
Divorce rate		–0.00 (.02)	–0.02 (.02)
Elderly population ratio		0.04 (.02)	0.07 (.02)**
Employment rate		–0.03 (.02)	–0.03 (.02)
Birth trend [a]			–0.03 (.02)
Divorce trend [a]			–0.04 (.02)*
Migration trend [a]			0.03 (.02)
Social capital deficiency [b]			–0.04 (.03)
Akaike information criteria	–449.16	–434.73	–403.08
Log likelihood	227.58	224.36	212.54
Variance (intercept)	.01	.01	.01
Variance (residual)	.01	.01	.01

Note. Model I is a baseline model with a random intercept; standard errors are in parentheses.
[a]The trend variable shows a difference between the first and last years in the period, indicating an increasing or decreasing trend across years. [b]The social capital deficiency variable indicates the proportion of survey respondents in a district who reported having insufficient social capital to mobilize when in need.
*p .05. **p .01. ***p .001.

Thus, we employed district-level statistical data for 3 years (i.e., 2015–2017), as published on the Korean Statistical Information Service website, to examine relationships in the data with our anxiety indices spanning the years of 2016 and 2017. Secondary data contained various sociodemographic characteristics for 230 districts nested within 16 metropolitan cities or provinces, including rates of fertility, marriage, divorce, migration, and employment, among many other dimensions—the data were inherently hierarchical. In addition, the distribution of anxiety, which was to be used as a dependent variable, was skewed to the right, failing to meet the normality assumption (i.e., Shapiro–Wilk $W = .95$, $p < .001$). Accordingly, we carried out generalized multilevel modeling of the Gaussian distribution with a log-link to account for city-/province-level variances in data. All the predictor variables included in the models were normalized due to wide heterogeneity between their measurement scales. Table 3 summarizes the multilevel modeling results.

To test whether there was a nonnegligible variance between the metropolitan cities and provinces, we first carried out a likelihood ratio test for two models with a fixed or random intercept and found a statistically significant difference, $\chi^2(1) = 27.39$, $p < .001$. The value of the intraclass correlation, showing the proportion of between-city/between-province variance to the total variance, was 0.4726. This means that the between-state/city variance accounted for 47.26% of the total variance present in the data. Model I is a baseline model with a random intercept across cities or provinces. Based upon the initial correlation analyses of the variables in the secondary data, several seemingly relevant variables were selectively included in Models II and III as predictors.

In Model II, the only statistically significant predictor of anxiety level was birth rate. As shown therein, the level of anxiety increases by 5.1% as the district's birth rate increases by one standard deviation from the national average ($e^{.05} = 1.051$). Subsequently, more variables were added as in Model III. Therein, however, both the proportion of elderly people and divorce rate were statistically significant, whereas the effect of birth rate remained unchanged. That is, in Model III, a district's anxiety level increased by 7.3% per one standard deviation increase in elderly population ratio ($e^{.07} = 1.073$) but decreased by 3.9% as early divorce rates showed an upward trend ($e^{-.04} = .9607$). Interestingly, any present divorce rate had little effect on anxiety, while the early divorce trend—whether increasing or decreasing—did significantly differentiate anxiety. Deviating from our expectations, employment rate and social capital deficiency were not statistically significant in terms of levels of anxiety. In summary, anxiety indices based on Twitter were greater in regions with higher birth rates, with larger elderly populations, and with no increasing trend in early divorce rate.

These results appear to be largely in line with previous findings wherein elderly populations tend to have higher levels of anxiety because of economic hardship following retirement and their own failing health. Our findings are also consistent with previous findings that rural regions with larger proportions of elderly people tend to have higher anxiety levels than metropolitan areas (Stephens-Davidowitz, 2016). Furthermore, it is also conceivable that younger parents raising children are likely to be exposed to more stressful circumstances due to uncertainty regarding the future and safety of their children (Lee, 2003). Perhaps the increasing burden on younger married couples partly explains why there is a negative relationship between early divorce trend and anxiety. An early divorce might be viewed as a (desperate) move to avoid or reduce increasing levels of pressure or distress. With more individuals in a region acting this way, the resulting anxiety levels therein would subsequently decrease, which might explain why anxiety is associated with the trend rather than with the present rate of divorce. That is, an upward trend in early divorce may be a symptom of how individuals cope with distress or anxiety in their life circumstances.

While not significant here, the ups and downs of employment rates and social capital might be correlated with anxiety trends. Indeed, there may be many other psychological, social, and cultural factors that influence the extent to which people experience social distress. Although the results examined here are far from complete in revealing the true underlying causes of the distribution of anxiety, our findings are at least consistent with some major findings of prior studies. This suggests that the digital trace-based anxiety index developed here is a valid indicator that can be used to reveal some important, otherwise unobservable aspects of collective emotional processes. From data obtained in this way, we may begin probing a wide variety of intriguing social problems and phenomena.

Discussion

Anxiety is an essential part of our lives in terms of its importance for survival in increasingly complex social environments. In psychology, it has long been recognized that a person's mental state and behavior is a function of his or her psychological field, often referred to as *life space*, a mixture of personal and environmental factors or forces (Lewin, 1939). What a person does is largely shaped by how she or he subjectively makes sense of the surrounding environment. In an environment full of uncertainty, for example, one's life space or psychological field may be disorderly, producing tension or anxiety. In an effort to maintain an appropriate level of anxiety, we attempt to control important decisions including whether to get a new job, to pursue more education or higher income, to get married or divorced, to have arguments with others on contentious issues, or to migrate. Thus, in some meaningful sense, the series of choices or behaviors across our life course may be viewed as arising from an *anxiety coping mechanism*. This suggests that anxiety is a critical link between individual psychology and environmental dynamics (Hirsh et al., 2012). Population statistics for various social issues, including public opinion, employment, migration, suicide, and marriage, can thus be viewed (in a sense) as reflectors of the collective emotional processes in which numerous individuals try to cope with uncertainty in their own life spaces.

Prior studies in anxiety have mostly treated the condition as an atypical mental disorder requiring therapeutic interventions, while devoting relatively less attention to the universal nature and sociocultural roots of anxiety (Mirowsky & Ross, 1986; Rachman, 2009). For this reason, instruments for measuring anxiety, broadly referred to as standard patient-reported outcome measures (PROMs), have been developed primarily for therapeutic contexts and are limited in application to more general populations and circumstances. While structured metrics such as PROMs warrant efficient collection of data, they are not flexible enough to capture subtleties in the way individuals express their emotions in response to ever-changing environments. As an alternative or supplementary approach, a hybrid of computational sentiment analysis together with PROMs, also known as *sentic PROMs*, has been suggested (Cambria et al., 2012). Sentic PROMs allow not only the measurement of responses to structured questions but also the analysis of semantic relationships in unstructured expressions. Such hybrid approaches are, however, primarily utilized for symptom diagnosis in therapeutic contexts. Thus, the approaches remain obtrusive in nature insofar as they are based on patient responses to given questions.

This study showcases an attempt to capture anxiety in a more general, system-level context through an unobtrusive, observational approach to emotion-laden natural language expression. Drawing on the entropy model of uncertainty, in which an individual is portrayed as a self-organizing microsystem trying to maintain its own integrity in a continual dialogue with the surrounding environment (Hirsh et al., 2012), our goal is to lay a groundwork for placing anxiety—conventionally treated as a purely personal experience—as a key link between person and social–environmental dynamics. Our main premise is that, to better understand the causes and effects of anxiety, it is necessary to look at the big picture, the dynamic processes in which individuals reveal their emotions in an effort to cope with their local environments. As Schelling (1978), the Nobel laureate in economics, aptly remarked, "people are responding to an environment that consists of other people responding to their environment, which consists of people responding to an environment of people's responses" (p. 14).

This notion highlights the crucial interdependence between human behaviors and collective social dynamics. The reason why we experience anxiety may be in close conjunction with what others have chosen to do or would do in response to their own life circumstances. Accordingly, personal feelings and emotions may not be personal at all—rather, they are the consequences of a bigger social process emerging from numerous chosen behaviors. Meaningful associations between anxiety distributions and other regional statistics found in the current study provide evidence of how spatiotemporal mapping of anxiety distribution can provide an important lens for examining the hidden social processes of collective anxiety coping—that is, the macro-social processes through which individuals across geographic regions cope with various degrees of anxiety in differing life circumstances. Although still preliminary, we believe that continuous efforts to establish a big picture of collective emotions through computational data collection and analysis will allow establishment of a better micro–macro link, essential for grasping the intricate relationships between human emotions and evermore complex societies.

Conclusion

Explosive growth in digital trace data has enabled social scientists to inquire into an unprecedented range of social, economic, political, and cultural issues. This study is an attempt to extend the methodology to human emotions, particularly to feelings of anxiety or social distress, which has crucial implications for various social inquiries as well as policy development. While some prior research based on sentiment analysis has been conducted, to the best of our knowledge, this is the first attempt in mapping anxiety distributions extracted from social media data. In our estimation, capability in tracking the spatiotemporal distribution of anxiety may allow social scientists across disciplinary borders to better grasp links among human emotions, behaviors, and emergent macro-level social problems. Furthermore, it may also be possible to examine the interactions and communication processes between major news media and the public—for example, the relationships between anxiety-triggering news and temporal fluctuations of anxiety distributions—toward an end of improving people's social–emotional wellness. With proper empirical validation, the procedures presented in the current study can be extended to classifying and visualizing other types of emotions, such as happiness, anger, or fear, which will help us better elucidate the complex relationships between human emotions and social phenomena in a more systematic fashion.

Data Availability

All collected Twitter data are downloadable at "http://166.104.143.105:61000/main"

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government Ministry of Science and ICT (No. 2018R1A5A7059549).

Software Information

The developed software prototype is publicly available as a web service at http://166.104.143.105:62000/main

Notes

1. The java library (twitter4j) for the Twitter application programming interface can be found at http://twitter4j.org/archive/twitter4j-4.0.7.zip
2. As a reference, the coders were provided a list of 18 anxiety-relevant emotions, derived from the State-Trait Anxiety Inventory (Spielberger, 1983). The list comprises "nervousness," "perplexity," "worry," "excitement," "restlessness," "frustration," "apprehension," "discomfort," "fear," "turmoil," "yearning," "depression," "gloom," "hostility," "desperation," "dismay," "petulance," and "malaise."
3. This visualization tool can be accessed at http://166.104.143.105:62000/main

References

Barbalet, J. (1998). *Emotion, social theory, and social structure: A macrosociological approach*. Cambridge University Press.

Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical transactions of the Royal Society of London*, *53*, 370–418.

Beck, J. V., & Arnold, K. J. (1977). *Parameter estimation in engineering and science*: John Wiley & Sons.

Beck, U. (1992). *Risk society: Towards a new modernity* (Vol. *17*). SAGE.

Beuke, C. J., Fischer, R., & McDowall, J. (2003). Anxiety and depression: Why and how to measure their separate effects. *Clinical Psychology Review*, *23*(6), 831–848. https://doi.org/10.1016/S0272-7358(03)00074-6

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295–298.

Burke, B. L., Martens, A., & Faucher, E. H. (2010). Two decades of terror management theory: A meta-analysis of mortality salience research. *Personality and Social Psychology Review*, *14*(2), 155–195.

Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, *31*(2), 102–107.

Cambria, E., Benson, T., Eckl, C., & Hussain, A. (2012). Sentic PROMs: Application of sentic computing to the development of a novel unified framework for measuring health-care quality. *Expert Systems with Applications*, *39*(12), 10533–10543.

Chawla, N. V. (2005). Data mining for imbalanced datasets: An overview. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 853–867). Springer.

Chen, P., Sun, Z., Bing, L., & Yang, W. (2017). *Recurrent attention network on memory for aspect sentiment analysis* [Paper presentation]. Conference on Empirical Methods on Natural Language Processing (EMNLP-2017), Copenhagen, Denmark.

Chen, S. F., & Goodman, J. (1999). An empirical study of smoothing techniques for language modeling. *Computer Speech & Language*, *13*(4), 359–394. https://doi.org/10.1006/csla.1999.0128

Coviello, L., Sohn, Y., Kramer, A. D., Marlow, C., Franceschetti, M., Christakis, N. A., & Fowler, J. H. (2014). Detecting emotional contagion in massive social networks. *PLoS One*, *9*(3), e90315.

Dasborough, M., Sinclair, M., Russell-Bennett, R., & Tombs, A. (2008). Measuring emotion: Methodological issues and alternatives. In N. M. Ashkanasy & C. L. Cooper (Eds.), *Research companion to emotion in organizations* (pp. 197–208). Edward Elgar Publishing.

DiMaggio, P., & Garip, F. (2012). Network effects and social inequality. *Annual Review of Sociology*, *38*, 93–127.

dos Santos, C. N., & Gatti, M. (2014). *Deep convolutional neural networks for sentiment analysis for short texts* [Paper presentation]. International Conference on Computational Linguistics (COLING-2014), Dublin, Ireland.

Frank, E., & Bouckaert, R. R. (2006). Naive Bayes for text classification with unbalanced classes. In J. Fürnkranz, T. Scheffer, & M. Spiliopoulou (Eds.), *Knowledge discovery in databases: PKDD 2006* (Vol. *4213*, pp. 503–510): Springer.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, *52*(5), 696–704. https://doi.org/10.1080/10635150390235520

Hazarika, D., Poria, S., Mihalcea, R., Cambria, E., Cambria, E., & Zimmermann, R. (2018). *ICON: Interactive conversational memory network for multimodal emotion detection* [Paper presentation]. Empirical Methods in Natural Language Processing (EMNLP-2018), Brussels, Belgium.

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychological Review*, *119*(2), 304–320.

Hollway, W., & Jefferson, T. (1997). The risk society in an age of anxiety: Situating fear of crime. *The British Journal of Sociology*, *48*(2), 255–266. https://doi.org/10.2307/591751

Hong, Y. O., Song, K. J., Park, S. A., Lee, H. J., & Lee, J. C. (2006). Social anxiety in Korean society. *Korean Journal of Psychological and Social Issues*, *12*(1), 129–160.

Horwitz, A. V. (2010). How an age of anxiety became an age of depression. *The Milbank Quarterly*, *88*(1), 112–138.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). *A convolutional neural network for modelling sentences* [Paper presentation]. Annual Meeting of the Association for Computational Linguistics (ACL-2014), Baltimore, Maryland, USA.

Kao, A., & Poteet, S. R. (Eds.). (2007). *Natural language processing and text mining*. Springer Science & Business Media.

Kim, D. B., Lee, S.-J., Choi, K. S., & Kim, G. C. (1994). *A two-level morphological analysis of Korean*. Proceedings of the 15th Conference on Computational Linguistics, Kyoto, Japan.

Korkontzelos, I., Nikfarjam, A., Shardlow, M., Sarker, A., Ananiadou, S., & Gonzalez, G. H. (2016). Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. *Journal of Biomedical Informatics*, *62*, 148–158. https://doi.org/10.1016/j.jbi.2016.06.007

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences of the United States of America, *111*(29), 8788–8790, 10779–10779.

Lee, C. W. (2003). The life instability of the elderly in Korea. *Korean Studies*, *18*, 37–67.

Lewin, K. (1939). Field theory and experiment in social psychology: Concepts and methods. *American Journal of Sociology*, *44*(6), 868–896.

Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In C. Nédellec & C. Rouveirol (Eds.), *Machine learning: ECML-98* (Vol. *1398*, pp. 4–15): Springer Berlin.

Macintyre, S., & Ellaway, A. (2000). Ecological approaches: Rediscovering the role of the physical and social environment. In L. F. Berkman & I. Kawachi (Eds.), *Social epidemiology* (pp. 332–348). Oxford University Press.

Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A., & Cambria, E. (2019). *DialogueRNN: An attentive RNN for emotion detection in conversations* [Paper presentation]. Association for the Advancement of Artificial Intelligence (AAAI-2019), Honolulu, Hawaii, USA.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). *The Stanford Corenlp natural language processing toolkit* [Paper presentation]. Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, MD, USA.

McCallum, A., & Nigam, K. (1998). *A comparison of event models for naive Bayes text classification* [Paper presentation]. AAAI-98 Workshop on Learning for Text Categorization, Budapest, Hungary.

McClure, E. B., Parrish, J. M., Nelson, E. E., Easter, J., Thorne, J. F., Rilling, J. K., Ernst, M., & Pine, D. S. (2007). Responses to conflict and cooperation in adolescents with anxiety and mood disorders. *Journal of Abnormal Child Psychology*, *35*(4), 567–577.

Mirowsky, J., & Ross, C. E. (1986). Social patterns of distress. *Annual Review of Sociology*, *12*(1), 23–45. https://doi.org/10.1146/annurev.so.12.080186.000323

Mladenic, D., & Grobelnik, M. (1999). *Feature selection for unbalanced class distribution and naive Bayes* [Paper presentation]. International Conference on Machine Learning, Bled, Slovenia.

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, *40*(10), 4241–4251. https://doi.org/10.1016/j.eswa.2013.01.019

Ng, A. Y., & Jordan, M. I. (2002). *On discriminative vs. generative classifiers: A Comparison of logistic regression and naive Bayes* [Paper presentation]. Advances in Neural Information Processing Systems, Vancouver, British Columbia, Canada.

Ng, V., & Cardie, C. (2002). *Improving machine learning approaches to coreference resolution* [Paper presentation]. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, Pennsylvania, USA.

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, *42*(24), 9603–9611.

O'Leary-Kelly, S. W., & Vokurka, R. J. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, *16*(4), 387–405. https://doi.org/10.1016/s0272-6963(98)00020-5

Page, S. E. (2015). What sociologists should know about complexity. *Annual Review of Sociology*, *41*(1), 21–41. https://doi.org/10.1146/annurev-soc-073014-112230

Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up? Sentiment classification using machine learning techniques* [Paper presentation]. Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, USA.

Peng, F., & Schuurmans, D. (2003). Combining naive Bayes and n-gram language models for text classification. In L. Azzopardi, B. Stein, N. Fuhr, P. Mayr, C. Hauff, & D. Hiemstra (Eds.), *Advances in information retrieval* (Vol. *2633*, pp. 335–350): Springer.

Rachman, S. (2009). Psychological treatment of anxiety: The evolution of behavior therapy and cognitive behavior therapy. *Annual Review of Clinical Psychology*, *5*(1), 97–119. https://doi.org/10.1146/annurev.clinpsy.032408.153635

Raffety, B. D., Smith, R. E., & Ptacek, J. T. (1997). Facilitating and debilitating trait anxiety, situational anxiety, and coping with an anticipated stressor: A process analysis. *Journal of Personality and Social Psychology*, *72*(4), 892–906. https://doi.org/10.1037/0022-3514.72.4.892

Rennie, J. D. M., Shih, L., Teevan, J., & Karger, D. R. (2003). *Tackling the poor assumptions of naive Bayes text classifiers* [Paper presentation]. Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington DC, USA.

Rodebaugh, T. L., Heimberg, R. G., Taylor, K. P., & Lenze, E. J. (2016). Clarifying the behavioral economics of social anxiety disorder: Effects of interpersonal problems and symptom severity on generosity. *Clinical Psychological Science*, *4*(1), 107–121.

Rodebaugh, T. L., Shumaker, E. A., Levinson, C. A., Fernandez, K. C., Langer, J. K., Lim, M. H., & Yarkoni, T. (2013). Interpersonal constraint conferred by generalized social anxiety disorder is evident on a behavioral economics task. *Journal of Abnormal Psychology*, *122*(1), 39–44. https://doi.org/10.1037/a0030975

Schelling, T. C. (1978). *Micromotives and macrobehavior*. W. W. Norton & Company.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729.

Schneider, K.-M. (2005). Techniques for improving the performance of naive Bayes for text classification. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing* (Vol. *3406*, pp. 682–693). Springer.

Schoemaker, P. J. (1982). The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature*, *20*(2), 529–563.

See, Y. H. M., & Petty, R. E. (2006). Effects of mortality salience on evaluation of ingroup and outgroup sources: The impact of pro-versus counterattitudinal positions. *Personality and Social Psychology Bulletin*, *32*(3), 405–416.

Slovic, P. (1995). The construction of preference. *American Psychologist*, *50*(5), 364–371. https://doi.org/10.1037/0003-066X.50.5.364

Smith, T. W. (1988). A report: Nuclear anxiety. *The Public Opinion Quarterly*, *52*(4), 557–575.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, *53*(1–3), 165–188. https://doi.org/10.1016/0304-4076(92)90084-5

Spielberger, C. D. (1983). *State-Trait Anxiety Inventory*. Consulting Psychologists Press.

Stephens-Davidowitz, S. (2016). Fifty states of anxiety. *New York Times*. https://www.nytimes.com/2016/08/07/opinion/sunday/fifty-states-of-anxiety.html

Tang, D., Qin, B., & Liu, T. (2015). *Learning semantic representations of users and products for document level sentiment classification* [Paper presentation]. Annual Meeting of the Association for Computational Linguistics (ACL-2015), Beijing, China.

Turner, J. H. (1988). *A theory of social interaction*. Stanford University Press.

Turner, J. H., & Stets, J. E. (2006). Sociological theories of human emotions. *Annual Review of Sociology*, *32* (1), 25–52. https://doi.org/10.1146/annurev.soc.32.061604.123130

Ungar, S. (2001). Moral panic versus the risk society: The implications of the changing sites of social anxiety. *The British Journal of Sociology*, *52*(2), 271–291. https://doi.org/10.1080/00071310120044980

Vapnik, V. N. (1998). *Statistical learning theory*. John Wiley & Sons.

Wilkinson, I. (2001). *Anxiety in a risk society*. Routledge.

Willams, A. (2017). How anxiety became society's prevailing condition. *The Independent*. https://www.independent.co.uk/news/long_reads/anxiety-prozac-nation-depression-mental-health-disorder-america-panic-usa-memoirs-self-help-book-a7785351.html

Wu, T., Luo, Y., Broster, L. S., Gu, R., & Luo, Y.-j. (2013). The impact of anxiety on social decision-making: Behavioral and electrodermal findings. *Social Neuroscience*, *8*(1), 11–21. https://doi.org/10.1080/17470919.2012.694372

Yuan, Q., Cong, G., & Thalmann, N. M. (2012). Enhancing naive Bayes with various smoothing methods for short text classification. *Proceedings of the 21st International Conference on World Wide Web* (pp. 645–646), ACM, New York, NY, USA.

Zhao, Z., Lu, H., Cai, D., He, X., & Zhuang, Y. (2017). *Microblog sentiment classification via recurrent random walk network learning* [Paper presentation]. Internal Joint Conference on Artificial Intelligence (IJCAI-2017), Melbourne, Australia.

Author Biographies

Yong Suk Choi is a professor in the Department of Computer Science and Engineering, Hanyang University, Seoul, Korea. His research areas include artificial intelligence and computational natural language processing (email: cys@hanyang.ac.kr).

Hansung Kim is an associate professor in the Department of Sociology, Hanyang University, Seoul, Korea. His research areas include social work and disability studies (email: hsk@hanyang.ac.kr).

Dongyoung Sohn is a professor in the Department of Media and Communication, Hanyang University, Seoul, Korea. His research areas include media psychology and computational modeling of social dynamics (email: dysohn@hanyang.ac.kr).