AMCIS 2022 Proceedings

SIG ED - IS in Education, IS Curriculum, Education and Teaching Cases

Aug 10th, 12:00 AM

Opportunities and Challenges of Remote Learning in the COVID Era – A Study Based on Sentiment Analysis of Geotagged Tweets

Asish Satpathy
Arizona State University, asish.satpathy@asu.edu

Riya Jignesh Brahmbhatt
Arizona State University (SSO), rbrahmb1@asu.edu

Srijan Satpathy
University of California, Los Angeles, satpathy1923@g.ucla.edu

Opportunities and Challenges of Remote Learning in the COVID Era – A Study Based on Sentiment Analysis of Geotagged Tweets

Emergent Research Forum (ERF)

Srijan Satpathy
University of California, Los Angeles
satpathy1923@g.ucla.edu

Riya Jignesh Brahmbhatt
Arizona State University
rbrahmb1@asu.edu

Asish Satpathy
Arizona State University
asish.satpathy@asu.edu

Abstract

With the onset of COVID-19, remote learning has become a new standard and is slowly being embraced by students and teachers. This research article deduces positive and negative sentiment towards remote learning during the COVID-19 era by conducting sentiment analysis on Tweets geolocated in New York City (NYC), NY, USA. Furthermore, we explore potential associations between local geospatial information (such as demographic and psychographic data) and Twitter sentiment. In our preliminary analysis, we articulate a model that describes how these sentiments are correlated with individuals' demographic information, socioeconomic status, and psychographic traits. Finally, based on our findings, we plan to explore the opportunities and challenges of remote learning in this community.

Keywords

Sentiment Analysis, Geospatial Analysis, Remote Learning, Geotagged Tweets, COVID-19 Pandemic

Introduction

The actual societal, economic, and medical ramifications of the COVID-19 pandemic will not be fully understood for many months. However, the pandemic has taken a significant toll on our education system, with near-total closures of schools and universities worldwide (Srivastava, 2020). The year 2020 marked a complete operational shift as classrooms were hastily vacated and schools were abandoned in favor of remote learning; teaching America's youth suddenly became a graver challenge. Due to 94% of urban residents and 51.6% of rural residents in the United States having access to high-speed internet as of 2018 (Lai, 2020), remote learning seemed like the natural segue. Zoom became a prominent medium of classroom learning, and online education boomed in an environment demanding its successful implementation. However, as with any monumental shift, the concept of fully implemented digital learning was met with praise and criticism alike, as some described being satisfied with school districts' support mechanisms while others mentioned struggles to maintain motivation in this new learning environment (Garbe, 2020). Thus, it is paramount to understand the intricacies of public sentiment towards remote learning in a time of increased digital reliance in the educational sector. While public sentiment can be studied in many ways, Twitter presents personal and rich client-side sentiment data. Using hashtags and location parameters can yield a plethora of relevant Tweets that can be broken down further to determine sentiment (Sarlan, 2014). Data collected during the pandemic is quite rich, as students were all engaged in remote learning and shared their opinions on Twitter. For this study, we focus on a geographic area that was hit hardest in the U.S. by the initial wave of COVID-19, with the epicenter of NYC seeing a sharp rise in supremely ill and ventilated patients (Stylianos, 2020). As a result, the sentiments captured via Twitter are likely to be accurate and raw reflections of remote learning perceptions at the time. Since this perception can depend on context & needs (Nallaperuma, 2021), we believe a more profound understanding can be gained by linking geotagged Tweets with objective location characteristics such as demographic information and socioeconomic status.

We use a supervised learning mechanism to uncover the location variables influencing one's positive or negative perceptions of remote learning. We hope to utilize the results to help bridge the inequality in education. Our research may also help policymakers identify geographic areas of concern and provide potential solutions for remote learning pitfalls. Future research may demonstrate the efficacy of any such analysis we present.

Related Work

Twitter sentiment analysis has been used in many prior settings, so we have much to look to when conducting our research. For example, Twitter data and associated multimedia have been used to understand public opinion on natural disasters, leading to exciting discoveries about the inaccuracies of some text to sentiment classification methods, which may play a role in our study (Alfarrarjeh, 2017). Additionally, a recent paper explores local users' sentiments extracted from geotagged Tweet data with a spatial and temporal perspective (Hu, 2019). Another study shows that geotagged Twitter data can be used to understand how demographic and socioeconomic factors relate to the number of Twitter users at a county level (Y. Jiang, 2019). As can be seen, sentiment can be linked to the location of Tweets as an additional dimension of our overall understanding of remote learning perception. More specifically, one study aimed to determine American sentiment towards reopening normal activities and restarting the economy during the COVID-19 pandemic. The authors cross-referenced pertinent socioeconomic data across the 50 states of the USA with sentiment levels derived from Twitter data and presented relevant conclusions, thus demonstrating a successful marriage of sentiment and location (Rahman, 2021). A recent study also sheds light on how Twitter data can be quite valuable in studying public sentiments on education during the COVID-19 pandemic (Cheeti, 2021). Inspired by prior works, we aim to answer the following research question: are there associations between remote learning sentiment (derived from Twitter during the COVID-19 pandemic) and user demography / other location-based attributes?

Data Collection and Methodology

The primary data sources for our study are Twitter and the Environmental Study and Research Institute (Esri). We restrict our study area to New York City only. To measure people's sentiment towards remote learning, we use the Twitter search Application Programming Interface (API)'s full-archive endpoint to extract sentiment data (geo-Tweets) for the NYC location. We apply filters on the search script to get Tweets for the desired time frame, location, and hashtag (Sloan, 2015). The extracted data has the following metadata: *id*, *text*, *hashtags*, *created_at*, *geo*, *like_count*, *quote_count*, *reply_count*, and *retweet_count*. We pull Tweets from two time periods: (1) COVID era (March 2020-December 2021) and (2) pre-COVID era (May 2018 to February 2020). Only about 1% of the Tweets are geotagged (with latitude and longitude information). To restrict our sample to NYC only, we apply advanced filters such that the geo-Tweets are located within 25 miles of the centroid location at -74.0060 latitude and 40.7128 longitudes. The data is filtered on the hashtags *#RemoteLearning*, *#OnlineLearning*, *#OnlineSchool*, *#OnlineClass*, and *#ZoomUniversity* (along with a pure character string search for the words used in the hashtags). The final data sample has 5391 Tweets from the COVID era and 235 Tweets from the pre-COVID era.

Sentiment Analysis

We used the sentiment lexicon with information about which words and phrases are positive and which are negative (Bonta, 2019). We conducted this sentiment analysis using TextBlob, a Python text processing library that enables the calculation of a piece of text's polarity, which is a measure of positive, neutral, or negative sentiment (Loria, TextBlob: Simplified Text Processing, 2020). To clean and pre-process the Tweets, the text was first made fully lower case (to prevent potential proper noun bias, which was later determined to be a non-factor in sentiment calculation). Mentions (@), hashtags, retweets (RT), and hyperlinks were stripped from the scraped text. Finally, phrases such as "remote learning," "zoom," and "online school" were removed to emphasize the polarity of the other phrases behind those concepts in the TextBlob calculation. The sentiment function of TextBlob outputs a pair of values, one being subjectivity (which does not concern our work) and the other being polarity. Polarity ranges from -1 (very negative) to 1 (very positive), with a score of 0 being set as neutral (or words not being found in the library's training set). Additionally, the TextBlob sentiment derivation implements the Natural Language Toolkit (NLTK) in Python to return an aggregate polarity score for all words in the text based on a Naïve Bayes trained lexicon set (Loria, 2020). For COVID era Tweets, our model depicted 51.15% as positive and 28.98% as negative (the rest were neutral), while pre-COVID era Tweets were 57.26% positive and 16.67% negative.

Neighborhood's Demographic, Socioeconomic, & Lifestyle Factors

Esri gathers its geographical, demographic, and lifestyle data from third-party resources, such as the U.S. Census reports and American Community Surveys, which can be found on its cloud-based platforms like ArcGIS Online (Esri, 2022). We created a one-mile radius buffer around each Tweet location and geo-enriched each buffer ring with 30 different demographic, behavioral, and socioeconomic data attributes. These aggregated data attributes are chosen based on prior studies (Chew, 2020) and our additional hypotheses. With each Tweet now linked to its sentiment polarity, and 30 other location-based data attributes, the further analysis assumes independence within the enriched data attributes.

Data Visualization

When the locations of the geotagged Tweets are mapped throughout NYC, the sentiment towards remote education seems to have some spatial patterns that we would like to investigate further. This also motivates us to explore potential correlations between geographically linked attributes (geo-enriched in the one-mile radius buffer from the Tweet location) and public sentiments towards remote learning. Finally, we hope to shed light on potential opportunities and challenges for remote education in NYC and its surrounding areas.

Model Building

We use a binary logistic regression model to identify factors influencing positive or negative sentiments towards remote learning (Rahman, 2021). With our assumptions that some of the demographic and behavioral features might contribute to sentiment, we test our model with all 30 variables to learn what factors significantly contribute to the positive sentiments towards remote learning.

Results from our rather preliminary analysis are summarized in Table 1. However, we understand it is too early to draw concrete conclusions from these results.

Preliminary Results:

Our preliminary results of the logistic model on COVID era data indicated the influence of independent variables on the positive sentiment of remote learning, as shown in Table 1. Out of 30 independent variables, only 11 variables remained significant with p-value less than 0.05, shown in rank order – most important to least important in representing positive sentiment. The log-likelihood method is used to calculate the parameter estimates (coefficients of the model). A negative parameter estimate indicates a reverse association with positive sentiment. The obtained model accuracy is 63.77%. We understand that association does not entail causation.

We did a similar analysis with pre-COVID era Tweets to check for associations between sentiment and other independent variables. Unfortunately, the sample size of the pre-COVID era dataset was far too small for us to develop any stable prediction model. Therefore, we will continue to explore more options to compare our results with data in the pre-COVID era.

Based on our analysis, the population with higher spending on online toys, games, digital book readers, and musical instruments / accessories exhibit negative sentiments towards remote learning. Acquaintance with online technology does not guarantee positive sentiment towards remote learning.

Similarly, individuals spending more money towards college tuition and different kinds of finance charges are less likely to exhibit positive sentiments towards remote learning. The general perceptions about the cost to value ratio of remote education play an essential role here.

Neighborhoods with a larger population of African American adults are likely to exhibit negative sentiments towards remote learning. This shows a clear gap in public policy that we believe should be rectified if we are serious about remote learning as a complementary tool in our education system.

On the other hand, higher average household income, more significant number of households with internet access, higher number of available educational services, and spending more money on education are all factors likely to exhibit positive sentiments towards remote learning. It is also interesting that communities with higher student loans are more likely to show positive sentiments towards remote learning. This concludes that affluent communities with internet access are an easy target to promote remote education. Therefore, public policy towards remote learning could be prioritized to reduce gaps in public sentiment.

Table 1: Results of binary logistic regression model on COVID era data.

Parameters	**Estimate**	χ^2	**P >** χ^2
1. Bought any online toys or games in the last 12 months	-0.00039	14.11	0.0002
2. Average spending on musical instruments/accessories	-0.29569	12.62	0.0004
3. Average Spending on digital book readers	-0.25104	12.23	0.0005
4. Average Spending on College Tuition	-0.02559	11.23	0.0008
5. Average Spending on Education in general	0.01967	10.39	0.0013
6. Black population who are 18+ years old	-0.00001	9.72	0.0018
7. Available Educational Services (business)	0.00398	7.44	0.0060
8. Average Household Income	0.00014	7.55	0.0064
9. Average Spending on finance/late/interest charge on student loan	-0.01526	6.70	0.0097
10. Number of Households with internet	5.4×10^{-5}	6.47	0.0110
11. Have Education personal loan (Student Loan)	0.000284	3.85	0.0497

Conclusion and Future Work

Despite some known limitations on choice of test method and data sources, our analysis looks promising, and we remain committed to developing a thorough study on the topic to make a recommendation for any public policy changes. For example, the Esri data chosen for our study has data aggregation lag of about 6 months, so not quite in line with the timeframe of this study, although the impact from this limitation could be ignored because the demographic data are not expected to change that rapidly at any given location. Furthermore, the Python sentiment derived using the TextBlob library does not account for emoticons/emojis in tweets, which might have broader implications in determining an individual's sentiment via a Tweet.

The preliminary research demonstrates that demographic and behavioral factors are essential variables to consider when proposing ideas and policies to address issues with remote learning. A complete understanding of the short-term implications of COVID-19 may not be known soon; however, a long-term implication can be assessed by wholly understanding remote learning sentiment in our education system.

We saw a significant increase in traffic on Twitter during the COVID era with the public sharing thoughts on remote learning. Based on our preliminary analysis, we draw the following conclusions:

1. Availability of and acquaintance with technology is necessary but not sufficient to draw positive sentiments on remote learning.
2. People invested in education, those with access to educational services, and the general affluent community tend to have positive sentiments towards remote learning and could be an easy target for promoting pertinent public policy.

3. Remote learning could be a challenge in neighborhoods with a larger population of adult African Americans. This could be an interesting target for assessing the effectiveness of changed public policy to foster higher education.

Currently, we are pursuing several directions to improve our analyses. As such, the enriched data attributes at different locations in our study area could depend on one another. For example, when comparing attribute values to other locations, nearby places may be more similar in value than those in farther locations (Sui, 2008). To ensure independence within the enriched data attributes, we intend to compute the spatial autocorrelation of each variable with itself using Global Moran's I (Goodchild, 1986). To ensure that our geo-enriched buffer rings do not overlap, we plan to calculate the distance between each Tweet location and only select geo-Tweets at least three miles apart from one another. This process is likely to compromise data samples to work with and remains to be tested.

REFERENCES

Alfarrarjeh, A., Agrawal, S., Kim, S. H., & Shahabi, C. (2017, October), Geospatial multimedia sentiment analysis in disasters. In *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 193-202). IEEE.

Bonta, V., & Janardhan, N. K. N. (2019). A comprehensive study on lexicon-based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, *8*(S2), 1-6.

Cheeti, S. S. (2021). Twitter based Sentiment Analysis of Impact of COVID-19 on Education Globally. *International Journal of Artificial Intelligence and Applications (IJAIA)*, *12*(3).

Chew, B., Satpathy, A., & Wong, E. (2020). Geospatial analyses to determine academic success factors in California's K-12 education. *Annals of GIS*, *26*(2), 81-100.

Esri (2022) ArcGIS Online: https://www.esri.com/en-us/arcgis/products/arcgis-online/overview/.

Garbe, A., Ogurlu, U., Logan, N., & Cook, P. (2020). COVID-19 and remote learning: Experiences of parents with children during the pandemic, American *Journal of Qualitative Research*, *4*(3), 45-65.

Goodchild, M. F. 1986. *Spatial Autocorrelation*. Norwich: Geo Books.

Hu, T., She, B., Duan, L., Yue, H., & Clunis, J. (2019). A systematic spatial and temporal sentiment analysis on geo-Tweets, IEEE *Access*, *8*, 8658-8667.

Lai, J., & Widmar, N. O. (2020). Revisiting the Digital Divide in the COVID-19 Era, Applied *economic perspectives and policy*, 10.1002/aepp.13104. Advance online publication.

Loria, Steven. (2020) TextBlob: Simplified Text Processing. https://textblob.readthedocs.io/en/dev/.

Nallaperuma, D., Nawaratne, R., Bandaragoda, T., Adikari, A., Nguyen, S., Kempitiya, T., ... & Pothuhera, D. (2019). Online incremental machine learning platform for big data-driven smart traffic management. *IEEE Transactions on Intelligent Transportation Systems*, *20*(12), 4679-4690.

Rahman, M. M., Ali, G. M. N., Li, X. J., Samuel, J., Paul, K. C., Chong, P. H., & Yakubov, M. (2021). Socioeconomic factors analysis for COVID-19 US reopening sentiment with Twitter and census data. *Heliyon*, *7*(2), e06200.

Sarlan, A., Nadam, C., & Basri, S. (2014, November). Twitter sentiment analysis. In *Proceedings of the 6th International Conference on Information Technology and Multimedia* (pp. 212-216). IEEE.

Sui, D. 2008. "First Law of Geography." In Encyclopedia of Geographic Information Science, edited by K. K. Kemp, 147. Thousand Oaks, CA: SAGE Publications, Inc.

Sloan, L., & Morgan, J. (2015). Who Tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. *PloS one*, *10*(11), e0142209.

Srivastava, S., & Agarwal, N. (2020). Psychological & social effects of pandemic Covid-19 on the education system, business growth, economic crisis & health issues globally. *An International Journal of Management & IT A Refereed Research Journal*, *11*, 40-45.

Stylianos, S., Mesa-Jonassen, A. E., Albanese, C. T., Bacha, E. A., Stark, N., Guida, S. J., ... & Sun, L. S. (2020). The perioperative services response at a major children's Hospital during the peak of the COVID-19 pandemic in New York City. *Annals of Surgery*, *272*(3), e199.

Y. Jiang, Z. Li, and X. Ye, "Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level," Cartography Geographic Inf. Sci., vol. 46, no. 3, pp. 228–242, 2019.