Study Correlation Between Media Focus and COVID Spread Using NLP with Spatial Data Approach

ZHUOCHENG SHANG, CS 297 (Fall 2020), University of California Riverside, USA

The widespread COVID-19 is not only a health problem but also a political argument. Especially while approaching the president election, media outlets on the right and left high lighting COVID-19 on politics from totally different angles. However, media on social network would have certain influence on people's attitude to COVID-19. This paper purpose to analyze correlation between media reporting and subscriber's or commenter's reaction while dealing COVID-19 outbreak between September and November over social media, such as Twitter. This research will export spatial location and attitude from news media which reporting COVID-19 or related news. Further, the narrative from the media, comments and other tweets from followers will be analyzed using sentiment analysis in order to summary general attitudes and sharpness. The research is aiming to to visualize correlation between news context, partisan difference, community's reaction, and COVID-19 spread tendency for each specific county.

1 PROBLEM

The global COVID-19 outbreak has introduced new argument on partisan debate. This tendency is also shown in news media clearly, media would report the pandemic from a more political angle before the president election. For example, one side may assume the virus is already less harmful, encourage economy reopen or focus on political conspiracy. While the other side may still focus on the COVID cases growth and health issues coverage. In consequence, the media subscribers will response to the COVID in different ways through this social processes and networks.[5]. To improve the understanding of the correlation between news media's narrative, media commenter's/follower's behavior and COVID-19 spread, this research will purpose three general approaches and focus on Fox News narrative and related TV-show scripts:

- **NLP on spatial data with machine learning** - This project will attempt to learn a model on one set of gazetteer data and apply SVM or CNN methods to generalize the model for exporting correct spatial location from Twitter content.
- **Sentiment analysis over text context** - The model will be trained on sentiment dataset and will attempt to use it to discriminate between positive and negative attitude towards COVID based on news media and Twitter context.
- **Analyze correlation** - This article will study the correlation between media narrative, follower's behavior and COVID spread tendency over county level geo-location.

Input:
All data shall focus between September and November, inclusively.

- Fox News media reports, Fox News talk show narrative and scripts;
- commenter's/follower's social media content (e.g.Twitter)
- COVID cases, weather data

Desired Output:

- Fox News attitude and narrative navigation toward COVID.
- Commenter's attitude and social distancing behavior toward COVID.
- Commenter's Geo-location extracted from Geotagged Twitter content.
- Compare attitude transferring for both media and people before and after the First Debate.

Author's address: Zhuocheng Shang, zshan011@ucr.edu, CS 297 (Fall 2020), University of California Riverside, 900 University Ave, Riverside, California, USA, 92507.

- Predict COVID spread at county-level based on both media focal point and people's responding behavior.
- Study correlation between media narrative navigation and people's responding health activity.

Challenges:

- Extract accurate county level geo-location from free text document.(E.g Twitter followers location)
- Merge COVID cases data and media narrative tendencies with efficient spatial join method.
- Correctly export attitude and sentiment information towards virus for both news media and individual social media context.

2 PREVIOUS AND RELATED WORK

Applying NLP over spatial data is a hot topic aiming at extracting accurate geographic location, mainly based on social media context. Although social media such as Twitter would contain Geo-tags, there has two limitations. Geo-tag is privacy context for most Twitter; further, these tags may differ from real locations where Twitter posted[7]. Hence, it is necessary for applying NLP to discover precise spatial location from pure text data.

How to stream Geo location from free text document is the first attempt. Time stamps and location labels could be extracted through Krovetz stemmer[4]. Typically, NLP is approached with SVM or Neural networks techniques. For SVM, several paper start with training gazetteer data,feature selection for model[2][9], then it used SVM to classify location information from text article[10]. Neural networks have been commonly applied to embed NLP and spatial data. For example, combine von Mises-Fisher distribution and UnicodeCNN[1] to export GPS locations from Twitter context. Another paper developed Character-level Convolutional Networks[11] in order to realize text classification.Another way for generating geotagged Twitter was applying tree based structure[3] with Jaccard coefficient.

Other than generating one specific location foe each Twitter ID, several papers focused on neighborhood information as well. For dealing with spatial join, an article applied R-tree structure to store spatial index which efficiently processing spatial searches[5].

For sentiment analysis, paper produced by Michael J. introduced that the accuracy of such analysis could be improved by applying sentiment classification with Alchemy API[8], because this method focus more on entity rather than overall score. SVM models was also developed for text mining, it is suitable for analysing real-time events. For an example, SVM could be designed to determine the topic of a Tweet, if the content is related to Earthquake or not.[6]

However, most paper paid attention to individual social media context rather than famous news media such as Fox or CNN news. Under COVID-19 pandemic situation, news media attitude navigation would make influence on how people react on pandemic from different area. This project will also research how news media from different location describe differently, and how will their context influence COVID spreading tendencies and people's health behavior.

3 DATA AND TOOLS USED

3.1 Datasets

This research project plan to use the news media context, Twitter, COVID-19 spread data.
Following are potential dataset could be applied to this project.

(1) POIs [1] data for world location.
(2) Tree-base location hierarchy Yago[2].

[1]https://www.factual.com
[2]https://yago-knowledge.org

(3) Obtained text data (follower's tweets) through Twitter Streaming API.
(4) Episode transcripts, LexisNexis (e.g. Fox News script data)[3]
(5) GDELT Television Compare API, news face interface, which provides statistics on relative frequency for user-provided search queries.

3.2 Tools

This research plans to use the Python, Twitter API, scikit-learn frameworks and tensor-flow in our project. scikit-learn and tensor-flow will be primarily used for implementations works, while Twitter API used for gathering data.

4 CONTRIBUTIONS

Planned contributions to this body of knowledge are as follows:

(1) Obtain media narrative (Fox News), script and their commenter's/follower's COVID-19 related tweets between September and November.
(2) Obtain county level spatial information from Geo-tagged micro-blog document.
(3) Apply SVM models over text documents in order to study media attitude and people's reaction towards COVID during September to November.
(4) Merge media narrative, commenter's/follower's tweets and COVID cases based on spatial index, especially with R-tree data structure.
(5) Visualize correlation between media focus tendencies and pandemic outbreak situation.

REFERENCES

[1] Mike Izbicki, Vagelis Papalexakis, and Vassilis Tsotras. 2019. Geolocating Tweets in Any Language at Any Location. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (Beijing, China) *(CIKM '19)*. Association for Computing Machinery, New York, NY, USA, 89–98. https://doi.org/10.1145/3357384.3357926

[2] C. Lee, H. Yang, and S. Wang. 2009. A Location Based Text Mining Approach for Geospatial Data Mining. In *2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC)*. 1172–1175. https://doi.org/10.1109/ICICIC.2009.23

[3] G. Li, J. Hu, J. Feng, and K. Tan. 2014. Effective location identification from microblogs. In *2014 IEEE 30th International Conference on Data Engineering*. 880–891. https://doi.org/10.1109/ICDE.2014.6816708

[4] Qiaozhu Mei, Chao Liu, Hang Su, and ChengXiang Zhai. 2006. A Probabilistic Approach to Spatiotemporal Theme Pattern Mining on Weblogs. In *Proceedings of the 15th International Conference on World Wide Web* (Edinburgh, Scotland) *(WWW '06)*. Association for Computing Machinery, New York, NY, USA, 533–542. https://doi.org/10.1145/1135777.1135857

[5] Hsien-Wen Meng Dapeng Li Ken R. Smith James A. VanDerslice Ming Wen Feifei Li Quynh C. Nguyen, Suraj Kath. 2016. ALeveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. Elsevier.

[6] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, 851–860. https://doi.org/10.1145/1772690.1772777

[7] Axel Schulz, Aristotelis Hadjakos, H. Paulheim, Johannes Nachtwey, and M. Mühlhäuser. 2013. A Multi-Indicator Approach for Geolocalization of Tweets. In *ICWSM*.

[8] Michael J. Widener and WenWen Li. 2014. Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Applied Geography* 54 (Oct. 2014), 189–197. https://doi.org/10.1016/j.apgeog.2014.07.017

[9] J. Witmer and J. Kalita. 2009. Extracting Geospatial Entities from Wikipedia. In *2009 IEEE International Conference on Semantic Computing*. 450–457. https://doi.org/10.1109/ICSC.2009.62

[10] Wei Zhang and J. Gelernter. 2014. Geocoding location expressions in Twitter messages: A preference learning method. *J. Spatial Inf. Sci.* 9 (2014), 37–70.

[11] Xiang Zhang, Junbo Zhao, and Yann LeCun. 2016. Character-level Convolutional Networks for Text Classification. arXiv:1509.01626 [cs.LG]

[3] https://www.lexisnexis.com/en-us/gateway.page