Using GIS to harness the power of Twitter for resilience analysis of Alberta's rural highways

Sina Abolhoseini †
Phani Kumar Patnala †
Babak Mehran †
Jonathan D. Regehr *,†
† University of Manitoba

Abstract

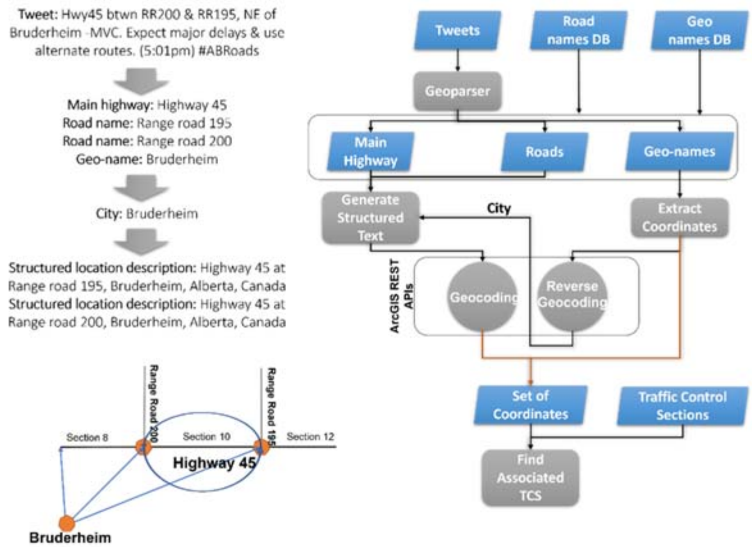
Social networks, such as Twitter, are becoming valuable sources of historical and real-time traffic data [1]. For example, in 2022, approximately 410 million tweets were posted by 290 million users each day [2]. Many agencies operate Twitter accounts dedicated to reporting traffic-related incidents. One such account, the Alberta 511 account, has been tweeting since 2016. A shortcoming with this data is that only 1-2% of all tweets on Twitter have geographical coordinates [3]. Instead, traffic-related tweets describe locations of incidents briefly in a semi-structured text format. Moreover, the Alberta 511 account only reports confirmed incidents and updates the event status when the incidents are cleared. Therefore, this account has potential for providing data for resilience analysis of the Alberta highway network. However, a methodology is needed to extract geographical information from its tweets. In this research, GIS techniques are used to develop a methodology for extracting information regarding motor vehicle collisions (MVCs) from the Alberta 511 account to be used for resilience analysis.

The figure below presents the methodology as a flowchart. The proposed methodology requires four resources: the Tweets archive was fetched from Twitter API [4]; geo-names were obtained from Canada Open Data portal [5]; road names were extracted from the Shapefile layer of National Road Network from Alberta Open Data portal [6]; and the Traffic Control Sections Shapefile layer were created from Alberta Traffic Data Mapping website [7]. After obtaining the required resources, tweets were parsed into road-names and geo-names. The first road-name mentioned in each tweet was assumed to be the highway that the collision happened on and was called the main highway. This resulted from investigating the structure of 511 tweets. After that, the coordinates of geo-names were entered into ArcGIS reverse geocoding API [8] to get the city of the geo-names. Then, structured location descriptions were generated based on the main highway, mentioned road names, and geo-named cities. City names were added to deal with uncertainties. This structure is compatible with the ArcGIS geocoding API [9]. Coordinates were extracted for each description by entering them into the ArcGIS geocoding API. At this stage, for each tweet we had a set of points. After converting points into a Shapefile layer, GIS was used to find the main highway's TCS that had the minimum cumulative distance to each set of points. This TCS was considered as the location of a tweet.

To investigate the performance of the methodology, it was deployed on Alberta 511 tweets in September 2022. Results were compared with manual geo-tagging. From the overall 112 tweets, we were able to manually geo-tag 106 tweets with certainty. From the 106 MVC tweets, 71% (75 tweets) were geocoded correctly, 17% (18 tweets) were geocoded on an adjacent TCS, and the rest of the tweets were geocoded incorrectly.

This methodology is neither limited to MVCs nor to the Alberta 511 Twitter account. Different provinces use a similar structure to post tweets across Canada, and we can use this methodology to get nationwide data coverage for resilience analysis.

* jonathan.regehr@umanitoba.ca



Flowchart of the proposed methodology for extracting geographical information from contextual content of the Alberta 511 Twitter account (with an example on the left side)

Keywords: social media, motor vehicle collision, text mining, natural language processing, geographical information system, geocoding

References

[1] Y. Gu, Z. S. Qian and F. Chen, "From Twitter to detector: Real-time traffic incident detection using social media data," Transportation research part C: emerging technologies, vol. 67, pp. 321-342, 2016.

[2] "Visualizing A Decade Of Twitter's Evolution: Jan 2012 – Nov 2022 & Musk's Purchase," 17 11 2022. [Online]. Available: https://blog.gdeltproject.org/visualizing-a-decade-of-twitters-evolution-jan-2012-nov-2022-musks-purchase/.

[3] "Advanced filtering with geo data," Twitter, [Online]. Available: https://developer.twitter.com/en/docs/tutorials/advanced-filtering-for-geo-data. [Accessed 27 03 2023].

[4] "Download Geographical Names Data," Government of Canada, [Online]. Available: https://natural-resources.canada.ca/earth-sciences/geography/download-geographical-names-data/9245. [Accessed 27 3 2023].

[5] "National Road Network (NRN) - AB, Alberta," Government of Alberta, 6 6 2016. [Online]. Available: https://open.alberta.ca/opendata/gda-6aa3584c-e1c1-4d9e-879c-082686587608. [Accessed 27 3 2023].

[6] "Traffic Data Mapping," Government of Alberta, [Online]. Available: https://www.transportation.alberta.ca/mapping/. [Accessed 27 3 2023].

[7] "Traffic Data Mapping," Government of Alberta, [Online]. Available: https://www.transportation.alberta.ca/mapping/. [Accessed 27 3 2023].

[8] "Reverse Geocode," Esri, [Online]. Available: https://developers.arcgis.com/rest/geocode/api-reference/geocoding-reverse-geocode.htm. [Accessed 27 3 2023].

[9] "Geocode Addresses," Esri, [Online]. Available: https://developers.arcgis.com/rest/geocode/api-reference/geocoding-geocode-addresses.htm. [Accessed 27 3 2023].