

Visual Exploration of Emotion Feelings Comparison in Tweet Data

Ilya Nemtsov, MST Jasmine Jahan, Chuting Yan, Shah Rukh Humayoun[†]

Department of Computer Science, San Francisco State University, USA

Abstract

Twitter (renamed as X) is one of the popular social media platforms where people share news or reactions towards an event or topic using short text messages called "tweets". Emotion analysis in these tweets can play a vital role in understanding peoples' feelings towards the underlying event or topic. In this work, we present our visual analytics tool, called TECVis, that focuses on providing comparison views of peoples' emotion feelings in tweets towards an event or topic. The comparison is done based on geolocations or timestamps. TECVis provides several interaction and filtering options for navigation and better exploration of underlying tweet data for emotion feelings comparison.

CCS Concepts

• **Human-centered computing** → **Visual analytics**; • **Computing methodologies** → **Information extraction**;

1. Introduction

The rise of social media platforms (e.g., Facebook, Twitter (X), Instagram, etc.) has transformed the way we access and disseminate information. Among these platforms, Twitter (renamed as X), a microblogging service, has become one of the most influential platforms. The platform enables users to share their feelings and news through short text messages called "tweets". These tweets often come with additional metadata (e.g., user profile, geolocation, timestamp, etc.) creating a valuable information resource for various analytical purposes.

This extensive pool of data at Twitter provides unique opportunities for researchers to uncover patterns and insights that can then be applied to diverse fields such as marketing, political science, public health, etc. Targeting this, different visualization tools have been developed to explore tweet data, e.g., showing the sentimental analysis of tweets (e.g., [HHM*16] [HAAE17] [LHW*15]), analyzing and exploring tweet data based on the geo-spatial information (e.g., [GWS17] [MJR*11] [KWD*13]), or exploring topics and analyzing their evolution over time (e.g., [DGWC10] [HMA21] [ZDWX12]), detecting unexpected events [KRHW12], building impression of a user [KWS14], stance exploration in social media [RSK*17], exploring relationships between the frequent keywords [HAAE17], classifying the sentiment of tweets [dHH14], etc. Researchers have also targeted providing visual analytics support to explore dynamics of conversation over time to analyze speaker interactions and behavior patterns [EAGA*16].

One way to better explore tweet data is about comparison of peoples' reaction towards the same event from different geolocations,

e.g., their views to COVID-19 pandemic. Targeting this, we developed a visual analytics tool, called TECVis (Tweets' Emotions Comparison Visualizer), that focuses on comparison of peoples' feelings towards an event or topic using the emotions in tweets (i.e.: *anger, fear, anticipation, trust, surprise, sadness, joy, and disgust*) and sentiment analysis (i.e.: *negative, neutral, and positive*). The comparison is done based on geolocations or timestamps, where users would be able to explore peoples' feelings towards the same event or topic from different geolocations or from one timestamp to another timestamp. Such comparison visual exploration of tweet data based on geolocations or timestamps, which is missing in above-mentioned previous work, would help us analyzing better public views towards an event or topic.

2. The Dataset

As a proof of concept, we collected the dataset through the Twitter API[†] using the COVID-19 and related keywords, from the time period January 2021 to May 2021 based on US geolocations. The Twitter api access was granted in the academic research mode, approved by Twitter, to collect and use the data for academic and research purposes. The final dataset consisted of 574,806 tweets (only English tweets) from an initial 952,278 collected tweets. We assigned random ID numbers to make users' IDs anonymous. We used the Natural Language Toolkit (NLTK[‡]) to tokenize the keywords. We used Valence Aware Dictionary and sEntiment Reasoner (VADER[§]) library to get sentiment polarities of tweets (i.e., *nega-*

[†] <https://developer.twitter.com/en/docs/twitter-api>

[‡] <https://www.nltk.org/>

[§] <https://github.com/cjhutto/vaderSentiment>

[†] humayoun@sfsu.edu

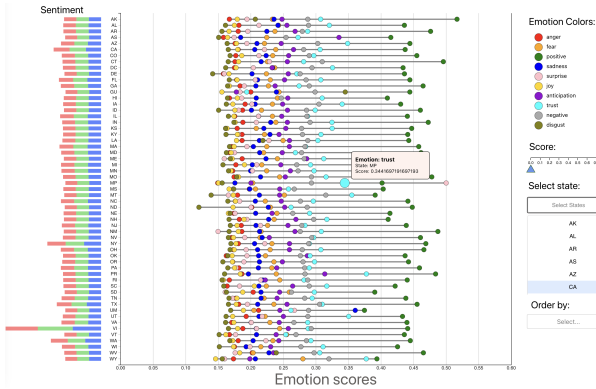


Figure 1: TECVis overview with COVID-19 dataset. The middle dot plots show eight emotion feelings' average scores associated to each geolocation, a US state in this view. The left-side horizontal bars, associated to geolocation, show tweets' count and their sentiment polarities. The right-side panel provides filtering options.

tive, neutral, and positive) and their confidence scores. We used NRC Emotion Lexicon library⁴ to get emotions for each tweet. The library provides eight basic emotions, as mentioned earlier, to the given text with association scores ranging from 0 to 1.

3. TECVis: Tweets' Emotions Comparison Visualizer

TECVis is a web-based tool where the client-side was developed using HTML, CSS, and JavaScript, D3.js library, while the server side was developed using the Node.js. TECVis provides the comparison of peoples' feelings towards an event or topic using the emotions and sentiment polarities in tweets based on different geolocations and timestamps over a period of time.

Figure 1 shows the overview of the TECVis tool based on geolocations comparison. For showing the eight feelings in emotion comparison, we use a dot plot where the x-axis is used to display the average score ranging from 0 to 1 for each feeling. On the y-axis, all the geolocations (i.e., US states in the current dataset) are listed. On the dot-plot, each dot represents the average score of each emotion feeling of all the associated tweets for a particular geolocation. We use different colors for each emotion feeling. We categorize emotion feelings into **positive-feeling** category (i.e.: *anticipation, trust, surprise, and joy*) and **negative-feeling** category (i.e.: *anger, fear, sadness, and disgust*). On the left-side of each geolocation, TECVis uses a horizontal bar to show the associated tweets' count to this geolocation with associated sentiment polarity distribution (i.e., red for negative, blue for neutral, and green for positive).

For a side-by-side comparison of two geolocations or timestamps (users can click to select these on the main view), TECVis uses a Tornado chart (Fig. 2.a) and a Radar chart (Fig. 2.b). In the Tornado chart, each side shows emotion feeling scores of one selected geolocation or timestamp, where the difference between the

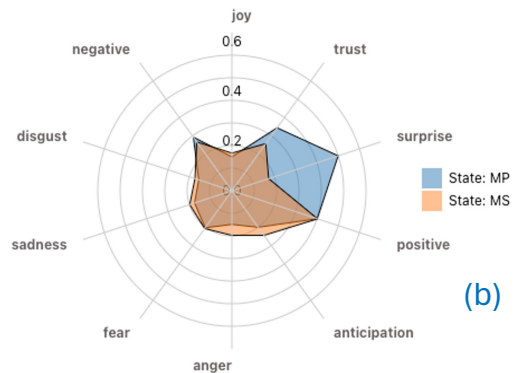
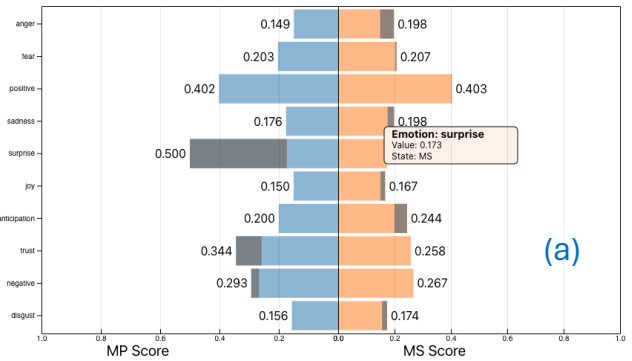


Figure 2: TECVis uses a Tornado chart (a) for the side-by-side comparison of two geolocations or timestamps and a Radar chart (b) for comparison using overlapping polygon on a circular grid.

two sides is highlighted using a darker color in the higher score side. In the Radar chart, each side (geolocation or timestamp) is represented by a polygon on the circular grid. In this case, the common values between both sides are highlighted in a darker color, while the differences are shown in their own colors (see Fig. 2.b).

TECVis provides several interaction, filtering, and navigation options: Users can navigate from geolocation comparison to timestamp comparison and vice versa, e.g., when users select a particular geolocation then TECVis updates the current view with showing the geolocation tweet data with comparison perspective of timestamps, where users can see the comparison by days, weeks, or months. TECVis also provides the facility to filter the data based on selected geolocations (see right-side panel in Fig. 1), timestamps, a particular emotion feeling, or emotion feeling score (see right-side panel in Fig. 1).

In the future, we plan to use latest large language model (LLM) libraries for better analysis of tweet data, as recent advancements in LLM show better sentiment analysis performance compared to classical approaches [ZDL*23]. Then we would conduct user studies to evaluate the tool, both from explorative perspective and from common usability aspects. We also intend to open source the tool so researchers and analysts can use their datasets for exploring different events or topics.

⁴ <https://pypi.org/project/NRCLex/>

References

- [DGWC10] DÖRK M., GRUEN D., WILLIAMSON C., CARPENDALE S.: A visual backchannel for large-scale events. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1129–1138. 1
- [dHH14] DA SILVA N. F., HRUSCHKA E. R., HRUSCHKA E. R.: Tweet sentiment analysis with classifier ensembles. *Decision Support Systems* 66 (2014), 170–179. URL: <https://www.sciencedirect.com/science/article/pii/S0167923614001997>, doi:<https://doi.org/10.1016/j.dss.2014.07.003>. 1
- [EAGA*16] EL-ASSADY M., GOLD V., ACEVEDO C., COLLINS C., KEIM D.: ConToVi: Multi-Party Conversation Exploration using Topic-Space Views. *Computer Graphics Forum* (2016). doi:[10.1111/cgf.12919](https://doi.org/10.1111/cgf.12919). 1
- [GWS17] GODWIN A., WANG Y., STASKO J. T.: Typotweet maps: Characterizing urban areas through typographic social media visualization. *EuroVis '17*, Eurographics Association, p. 25–29. 1
- [HAAE17] HUMAYOUN S. R., ARDALAN S., ALTARAWNEH R., EBERT A.: TExVis: An Interactive Visual Tool to Explore Twitter Data. In *EuroVis 2017 - Short Papers* (2017), The Eurographics Association. 1
- [HHM*16] HOEBER O., HOEBER L., MESEERY M. E., ODOH K., GOPI R.: Visual twitter analytics (vista): Temporally changing sentiment and the discovery of emergent themes within sport event tweets. *Online Inf. Rev.* 40 (2016), 25–41. 1
- [HMA21] HUMAYOUN S. R., MANSOUR I., ALTARAWNEH R.: Tevise: An interactive visual analytics tool to explore evolution of keywords' relations in tweet data. In *INTERACT 2021* (2021), pp. 579–599. 1
- [KRHW12] KRSTAJIĆ M., ROHRDANTZ C., HUND M., WEILER A.: Getting there first: Real-time detection of real-world incidents on twitter. 1
- [KWD*13] KRAFT T., WANG D. X., DELAWDER J., DOU W., YU L., RIBARSKY W.: Less after-the-fact: Investigative visual analysis of events from streaming twitter. In *IEEE LDAV 2013* (2013), pp. 95–103. 1
- [KWS14] KUMAMOTO T., WADA H., SUZUKI T.: Visualizing temporal changes in impressions from tweets. *iiWAS '14*, p. 116–125. 1
- [LHW*15] LU Y., HU X., WANG F., KUMAR S., LIU H., MACIEJEWSKI R.: Visualizing social media sentiment in disaster scenarios. In *WWW '15 Companion* (New York, NY, USA, 2015), ACM, p. 1211–1215. 1
- [MJR*11] MACÉACHREN A. M., JAISWAL A., ROBINSON A. C., PEZANOWSKI S., SAVELYEV A., MITRA P., ZHANG X., BLANFORD J.: Senseplace2: Geotwitter analytics support for situational awareness. In *VAST 2011* (2011), pp. 181–190. 1
- [RSK*17] RAFAEL M. M., SIMAKI V., KUCHER K., PARADIS C., KERREN A.: StanceXplorer: Visualization for the interactive exploration of stance in social media. In *2nd Workshop on Vis. for the Digital Humanities* (2017). 1
- [ZDL*23] ZHANG W., DENG Y., LIU B., PAN S. J., BING L.: Sentiment analysis in the era of large language models: A reality check, 2023. [arXiv:2305.15005](https://arxiv.org/abs/2305.15005). 2
- [ZDWX12] ZHAO J., DONG L., WU J., XU K.: Moodlens: an emoticon-based sentiment analysis system for chinese tweets. In *KDD* (2012). 1