

TECVis: A Visual Analytics Tool to Compare People's Emotion Feelings



Ilya Nemtsov, MST Jasmine Jahan, Chuting Yan, and Shah Rukh Humayoun

Department of Computer Science
San Francisco State University, USA

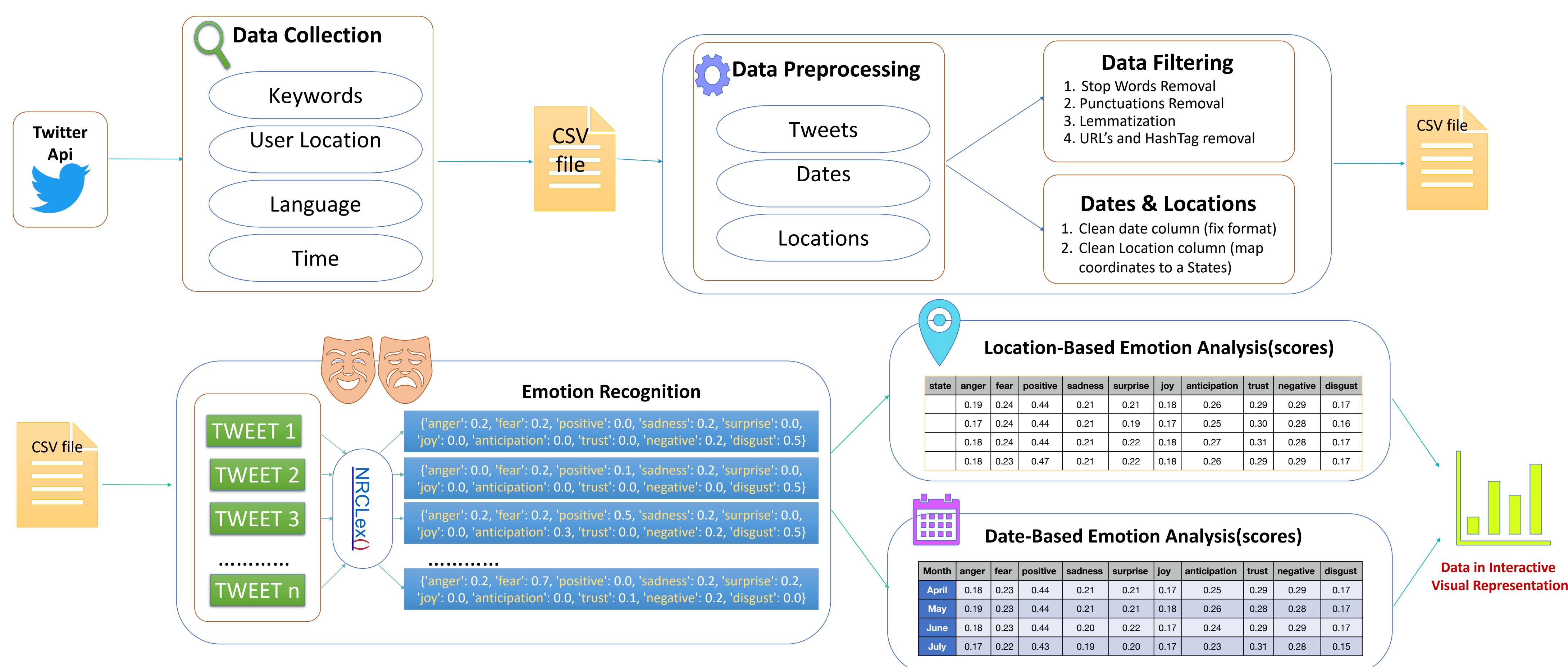
nemtsovilya@gmail.com, jammy.sust@gmail.com, christinayan1019@gmail.com, humayoun@sfsu.edu



Introduction

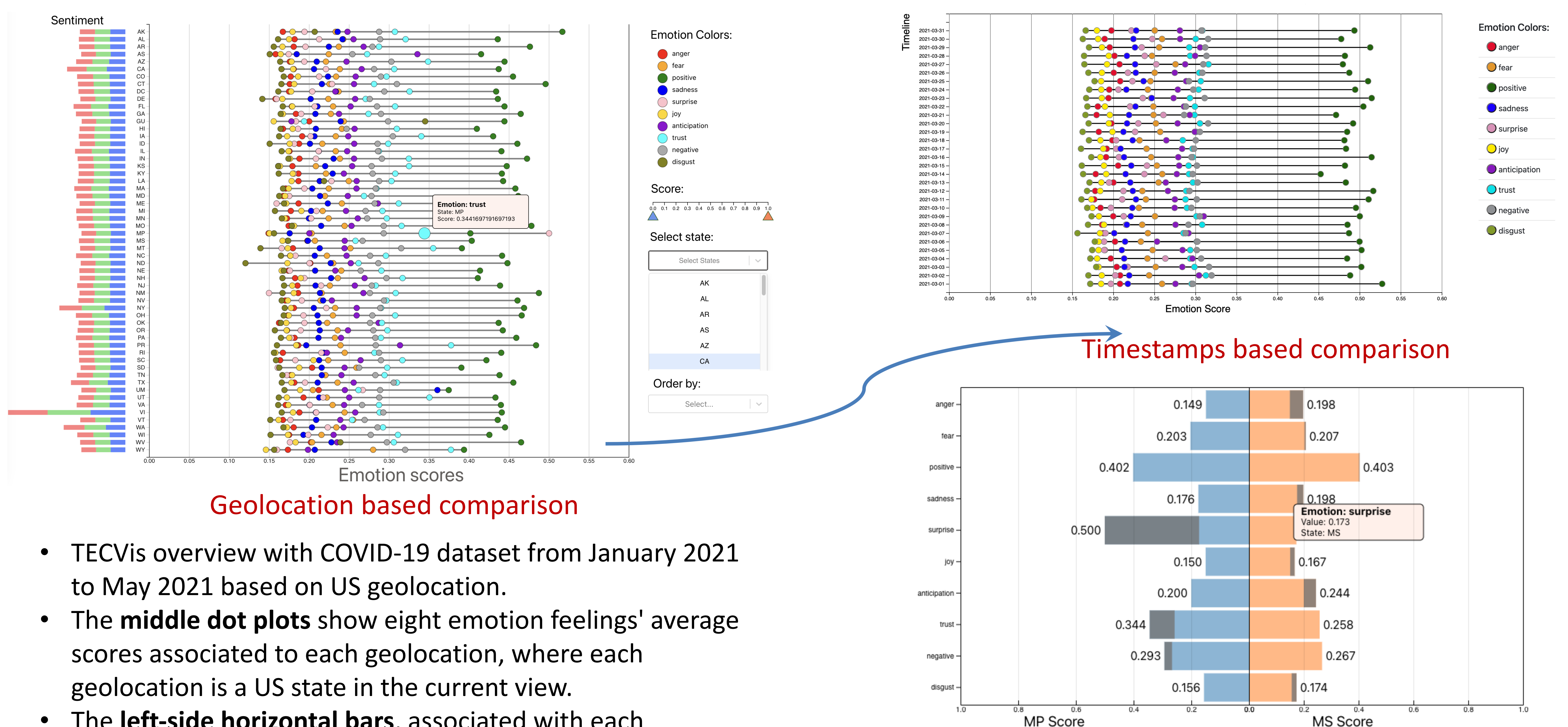
- Motivation:** Twitter (renamed as **X**) is one of the popular social media platforms where people share news or reactions towards an event or topic using short text messages called "tweets". Emotion analysis in these tweets can play a vital role in understanding peoples' feelings towards the underlying event or topic.
- Our Solution:** We present our visual analytics tool, called **TECVis** (Tweets' Emotions Comparison Visualizer), 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.
- The Dataset:** As a proof of concept, we targeted the COVID-19 topic and 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 final dataset consisted of 574,806 tweets (only English tweets and US geolocations) from an initial 952,278 collected tweets.

Data Collection and Pre-processing



- NLTK** (Natural Language Toolkit) library is used to tokenize the keywords.
- VADER** (Valence Aware Dictionary and sEntiment Reasoner) library is used to get sentiment polarities of tweets and their confidence scores.
- NRC Emotion Lexicon** library is used to get emotions for each tweet.
- Eight basic emotions are categorized into *positive emotion feelings* (i.e., **trust, anticipation, surprise, and joy**) and *negative emotion feelings* (i.e., **anger, fear, sadness, joy, and disgust**) as mentioned earlier, to the given text with association scores ranging from 0 to 1 and two sentiments (i.e., negative and positive).

TECVis: Tweets' Emotion Comparison Visualizer



- TECVis overview with COVID-19 dataset from January 2021 to May 2021 based on US geolocation.
- The **middle dot plots** show eight emotion feelings' average scores associated to each geolocation, where each geolocation is a US state in the current view.
- The **left-side horizontal bars**, associated with each geolocation, show the tweets count and their sentiment polarities.
- The **right-side panel** provides filtering options.

- Each side in this Tornado chart shows emotion feeling scores of one selected geolocation or timestamp.
- TECVis highlights the difference value using a darker color in high score side.