Analyzing and Visualizing Location Based Social Media Data

by **Sagnik Mukherjee**

This work aims to add to the corpus of analyzing tweets related to the EU migration crisis while using novel techniques to capture opinions and events. Tweets in six major European languages were studied and the results indicate that these techniques can be used to overcome language barriers for event detection while also safeguarding user privacy. However further work is required to refine the methodology and make the results more conclusive then they are now.

Using HyperLogLog

HyperLogLog (HLL) is an algorithm used to cryptographically hash social media data for analysis while safeguarding user privacy [1]. Using the privacy aware data structure developed from HLL, both quantitative and qualitative analysis can be performed. However, using this data format does have certain drawbacks.

Privacy safeguarding and low error rates for quantitave estimation are achieved at the cost of losing context of the data. Be it the temporal information or groupings of hashtags for certain posts, the processing into the privacy aware format leaves the privacy aware data quite lacking when it comes to qualitative analysis. Of course, the union and intersection set operations to allow for qualitative analysis on data and could be used to find patterns of hashtags/emoji co-occurences among other things. But, since the data itself is so divorced from the original raw twitter data while being converted to the privacy aware format, these analysis techniques are limited in their capacity to provide meaningful information. This should be the goal for further development.

Dashboard

The visualizations used throughout this work has been dependent on the facet of the data to be visualized [2]. These visualizations were derived mainly from the raw dataset.

The individual visualizations for the facets paint a partial picture of the entire dataset. When these visualizations across the facets are combined, a better holistic picture of the dataset appears. This is why making a dashboard was necessary.

Elements of narrative visualization were incorporated into the dashboard to foster a sense of involvement into the dataset. User choice was integrated to break down the complex data into understandable parts and also to explore the dataset to their wishes.

However, their remained areas of improvements for the dashboard. The effectiveness of using both elements of narrative visualization and a traditional dashboard could not be tested. There were also parts of the dashboard which required more clarity for the user.

Figure 2. Spatial extent of the tweets

Hashtags and the Crisis

Using hashtags and their typiclaity [3], a number of interesting observations pertaining to the reactions of people on twitter can be obtained. First, the typicality of specific hashtags during the time period of the dataset helps to identify the events in the world to which the users have reacted.

For example, the hashtag muslimban had been used to react to Donald Trump's signing of the executive order barring Muslims from entering the United States. Once the event is identified, using co-occuring hashtags of the principal hashtag an opinion analysis can be conducted.

This is especially helpful considering the multi-lingual nature of the dataset. A case where this was used, was for the hashtag refugeeswelcome. The co-occuring hashtags pointed to the support extented to the plight of refugees coming into Europe.

These cases still need to be further researched. Due to time constraints, only the German and English tweets were throughly investigated. Therefore conclusive remarks about the Crisis can't be made.

Conclusion

The EU Migration Crisis is a complex and still ongoing pan European issue which is reflected in the complexity of the twitter data. The methodology of this work could potentially be used for many other cases but it needs to be further developed to make conclusive remarks.
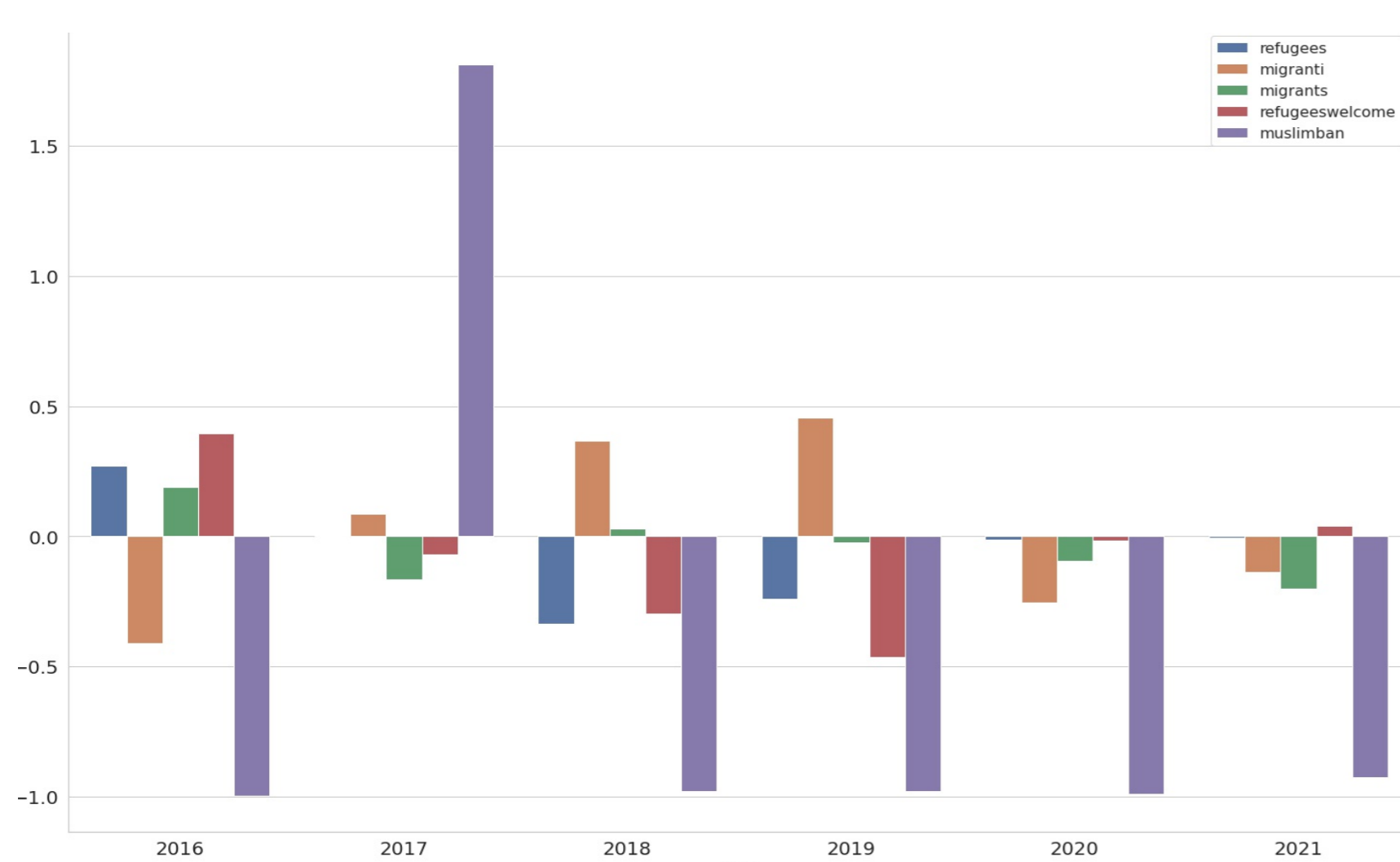

Figure 1. Typicalities of the six most popular hashtags with respect to the time period of the dataset

References

[1] Dunkel, A., Löchner, M., & Burghardt, D. (2020). Privacy-aware visualization of volunteered geographic information (VGI) to analyze spatial activity: A benchmark implementation. ISPRS International Journal of Geo-Information, 9(10)

[2] Dunkel, A., Andrienko, G., Andrienko, N., Burghardt, D., Hauthal, E., & Purves, R. (2019). A conceptual framework for studying collective reactions to events in location-based social media. International Journal of Geographical Information Science, 33(4), 780-804.

[3] Hauthal, E., Dunkel, A., & Burghardt, D. (2021). Emojis as Contextual Indicants in Location-Based Social Media Posts. ISPRS International Journal of Geo-Information, 10(6), 407.