I Stand With You: Detecting and Characterizing Expressions of Solidarity in Social Media

4^{th} International Conference on Computational Social Science IC^2S^2
July 12-15, 2018
Northwestern University's Kellogg School of Management
Evanston, IL. USA

Anonymous

February 06, 2018

keywords: NLP, human behavior modeling, solidarity, social media, deep learning

1 Introduction

On November 13th 2015, attackers carried out suicide bombings and mass shootings at two separate locations in Paris, France. Over a hundred people were killed and scores more were injured in these incidents, now known as the November 2015 Paris attacks [1]. People all over the world took to social media to express their reactions and convey their thoughts about this event. Expressions of solidarity formed a large proportion of these expressions made on social media. Solidarity is a sociological concept and a prosocial behavior, characterized by Bayertz [1999] in relation to "complementary terms such as community spirit or mutual attachment, social cooperation or charity". The collective enactment of online behaviors, including prosocial behaviors such as solidarity, has been known to directly affect political mobilization and social movements [Tufekci (2014); Fenton (2008)]. There is thus a pressing need to understand what drives and, more importantly, what characterizes the convergence of a global public in online social networks, especially in the immediate aftermath of crisis events. We report on the first study to (a) create a corpus of social media posts annotated with respect to expressions of solidarity; and (b) develop and test computational models that automatically detect expressions of solidarity. We present initial analysis with respect to the content of solidarity expressions. Our approach and findings will help advance research in the dynamics of online mobilization.

2 Data Collection and Method

To ensure that our corpus is appropriate for the analysis we are interested in, we collected 2MM tweets posted in the 24 hour period immediately following the November terrorist attacks in Paris (containing any mention of the word "paris") using the Twitter GNIP service. Social media, due to its increasingly pervasive nature, permits a sense of immediacy [Giddens (2013)]

[1]https://tinyurl.com/pb2bohv

- a notion that Fenton [2008] argues would produce high degree of identification among politicized citizens of the web, especially in response to crisis events. From the corpus 2MM tweets collected, we identified 286384 unique English tweets (no retweets). Next, we performed distance labeling [Mintz et al. (2009)] by having two trained annotators assign the most frequent hashtags in our corpus with one of three labels ("Solidarity" (e.g. *#solidaritywithparis*), "Neutral" (e.g. *#breakingnews*) and "Unrelated or Cannot Determine" (e.g. *#rebootliberty*). Using the hashtags that both annotators agreed upon (after achieving a agreement level of $\kappa > 0.65$), we extracted the tweets for the hashtags labeled as expressing solidarity (20465 tweets) and a randomly selected set of equal number tweets which had hashtags annotated as neutral.

We then used Recurrent Neural Networks (RNN) with Long Short Term Memory(LSTM) to distinguish the tweets which contained expressions of solidarity from those that did not. The RNN model incorporates pre-trained GloVe embeddings [Pennington et al. (2014)] and is trained for 20 epochs using Adam optimizer. We also compare the performance of the RNN model with baselines obtained from a Support Vector Machine (SVM) approach using the following different features: (a) Word Bigrams, (b) TFIDF, (c) TFIDF + Bigrams. We found that the deep learning model (RNN+LSTM) outperforms the SVM baselines (88.70% classification accuracy vs. 75.97% average accuracy for the three SVM baselines).

3 Discussion and Future Work

In Figure 1, we show the terms associated with each set of tweets (those expressing solidarity vs. neutral tweets), using the L2-penalized logistic regression coefficients for each term. We find that that the terms most associated with solidarity are other hashtags (*#prayforparis*), while the neutral terms include country names (*france*, *beirut* and *art*). A detailed analysis of the terms is part of future publication.

In future work, we also aim to analyze data with respect to location information to test the hypothesis that solidarity is expressed by individuals *not directly affected by the event*. In addition, we aim to replicate our method in the context of another crisis event (social media postings related to Hurricane Irma), where we expect the expressions of solidarity to be more diffuse over time.

References

Bayertz, K. (1999). Four uses of "solidarity". In *Solidarity*, pages 3–28. Springer.

Fenton, N. (2008). Mediating solidarity. *Global Media and Communication*, 4(1):37–57.

Giddens, A. (2013). *The consequences of modernity*. John Wiley & Sons.

Kessler, J. S. (2017). Scattertext: a browser-based tool for visualizing how corpora differ. *arXiv preprint arXiv:1703.00565*.

Mintz, M., Bills, S., Snow, R., and Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 1003–1011. ACL.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Tufekci, Z. (2014). Social movements and governments in the digital age: Evaluating a complex landscape. *Journal of International Affairs*, pages 1–18.

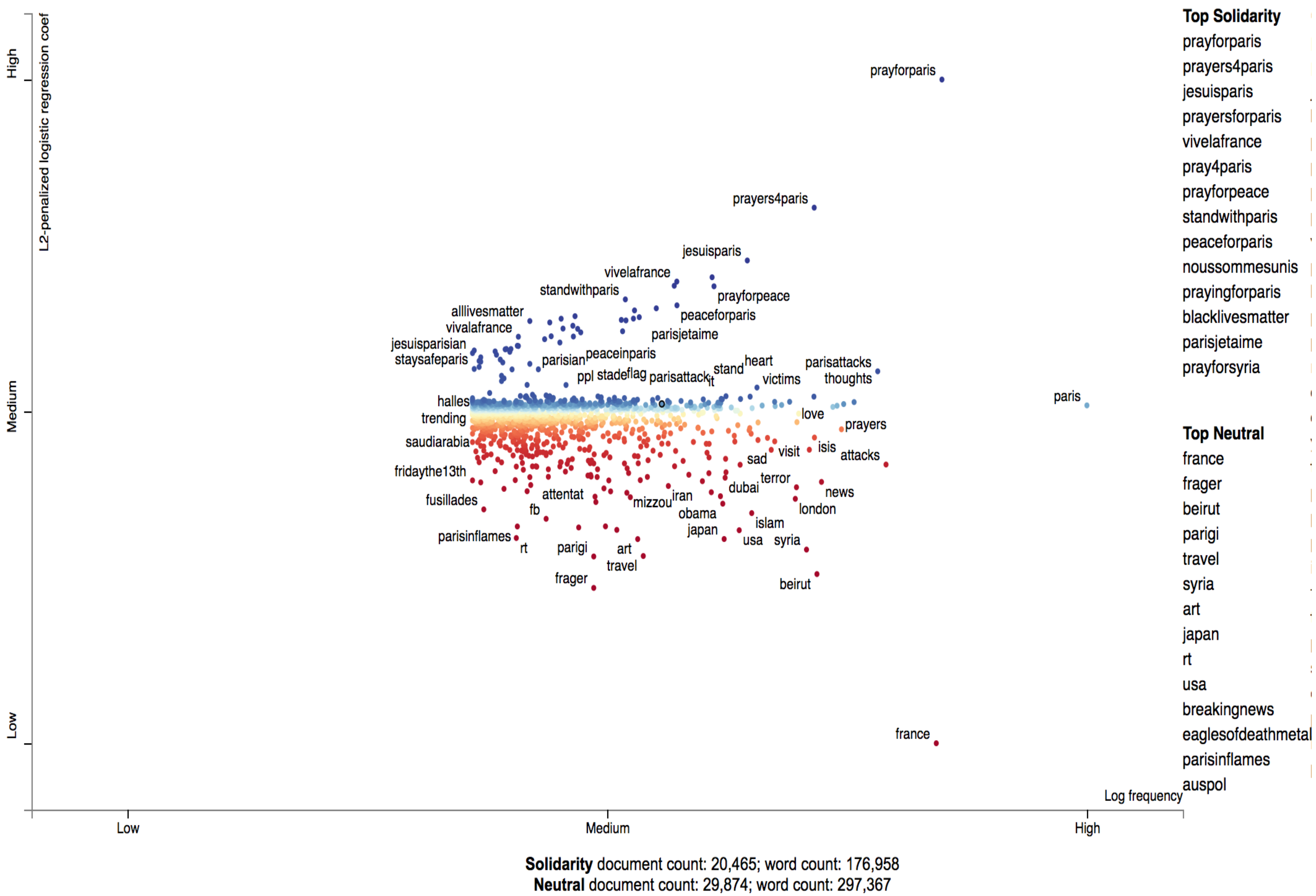

Figure 1: L2-penalized logistic regression coefficients displaying the top solidarity and neutral words generated with the help of the tool developed by Kessler (2017)