Twitter and political culture: Short text embeddings as a window into political fragmentation

Amar Budhiraja
Amar.Budhiraja@microsoft.com
Microsoft Research, India

Joyojeet Pal
jopal@microsoft.com
Microsoft Research, India

ABSTRACT

Mapping polarization and relationships in political discourse on social media is challenging since politicians' positions and relationships can be hard to pin down. In this paper, we attempt to use politicians' tweets as a metric of their affinities using representation learning, by modifying the Word2Vec method such that politicians are directly encoded into a Euclidean space. Our analysis of Indian politicians shows that the relatively populous, linguistically more homogeneous northern states are cohesively clustered based on their party affiliations, whereas southern states cluster based on geography. We propose that computational methods can be useful in examining the tensions of regionalist tendencies against dominant national political narratives.

CCS CONCEPTS

• **Information systems** → *World Wide Web*; **Social networks**; • **Applied computing** → *Sociology*.

KEYWORDS

Political Social Networks; Representation Learning

ACM Reference Format:
Amar Budhiraja and Joyojeet Pal. 2020. Twitter and political culture: Short text embeddings as a window into political fragmentation. In *ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '20), June 15–17, 2020, , Ecuador*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3378393.3402276

1 INTRODUCTION AND HYPOTHESIS

Representation learning has been employed in Natural Language Processing (NLP) in tasks like sentiment analysis [2] and sentence completion [5]. The notion of representation learning in NLP helps compute what words are likely to appear near each other due to contextual co-occurrence. Thus, 'cat' and 'dog' are *closer*, more likely to co-occur than 'dog' and 'spa'. Representation learning has been used in computational social sciences to study phenomena such as and social network analysis [7].

In this paper, we explore representation learning as a means to understand political discourse. Political polarization and populist nationalist politics often go hand in hand, and there are fears such politics are growing worldwide [1, 9]. In parts of the Global South, there have been fears that polarization and nationalistic tendencies may have wide-ranging impacts distracting away attention from social and economic priorities by undermining institutions and harming the plurality of societies [3]. A feature of such politics is that populist movements tend to try and centralize the discourse in a uniform national narrative, to the detriment of marginal politics of smaller linguistic or regional units within a nation state [10]. Systematically studying polarization in groups presents significant challenges and need for human annotation [8].

To explore the use of computational methods in identifying political communication, we attempt to understand centralizing tendencies in the political discourse. We select India, a pluralistic society, where much recent press has focused on the increasing dominance of a populist government with an emphasis on a strong national narrative over federalism. We study party activity on Twitter across states to understand if the centralization is reflected in the political discourse.

2 METHODOLOGY

We took a list of approximate 6 million most recent tweets from an annotated list of 13,111 Indian politicians from over 25 parties from an archive tracking Indian elections. Our goal was to study how similar or dissimilar politicians sound based on party or state. We represent each politician as a dense vector based on the content in the tweets that they have written. This enables distance computation between any politicians and the distance is indicative of how similar their tweets are. We then visualize them by parties and states.

Such representations can give insight about the overlaps in the issues and language used by politicians. These learnt representation for Indian Politicians are used to validate the following hypothesis *-do politicians specific state sound more like other politicians from their own states, or do they sound like their parties at the national level when they tweet.*. We use this as a proxy to understand whether politicians are "statists" in the federal sense - ie concerned primarily with their local issues, or does a "national" political ideology has a more important role to play on self representation on social media.

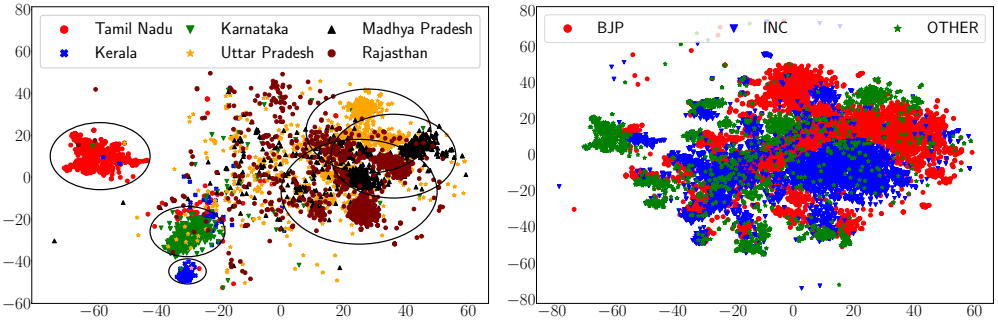
Using the learning setting of Word2Vec [6], we define the a model to learn the embeddings of politicians. In Word2Vec, the goal of the model (in Skipgram) is - given a word, predict the context around the word. We modify this learning setting and define it as given a politician (his/her Twitter handle) and a tweet written by him/her, predict random words from the tweets with the politician id as the input. At the end of training, similar to Word2Vec, we get dense vector representations for the input politicians.

3 ANALYSIS AND RESULTS

To validate our hypothesis, we projected the learnt politician representations into 2-D space for visualization using t-SNE [4]. Figure 1 (a) shows the 2-D projections for politicians from all parties for six states (3 Northern States: Rajasthan (RJ), Madhya Pradesh (MP)

COMPASS '20, June 15–17, 2020, , Ecuador

ACM ISBN 978-1-4503-7129-2/20/06.
https://doi.org/10.1145/3378393.3402276



(a) (b)

Figure 1: Politicians colored according to their states in (a) and parties in (b).

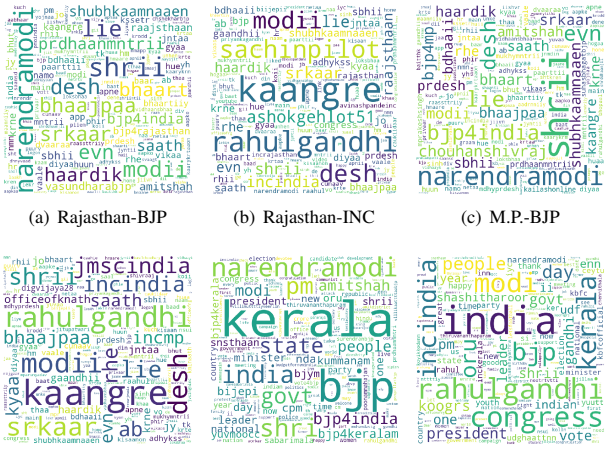
State-Party	State-Party	Corr.
KL BJP	MP BJP	0.7684
KL BJP	RJ BJP	0.7958
MP BJP	RJ BJP	0.9694
KL BJP	KL INC	0.9675
KL INC	MP INC	0.7396
KL INC	RJ INC	0.828
MP INC	RJ INC	0.9485
MP INC	MP BJP	0.9096
RJ INC	RJ BJP	0.9161

Table 1: Correlation of common words

and Uttar Pradesh (UP); 3 Southern States: Kerala (KL), Tamil Nadu (TN) and Karnataka (KT)). In Figure 1 (a), it can be observed that politicians from Southern Indian States are cohesively clustered based on their states whereas Politicians from the Northern States do not have comparable state based clustering; and are rather spread out and have overlapping structure over each other. We only show 6 states in this visualization but our data shows consistent observations of close amalgamation in all northern states, whereas all give southern states are more cohesively associated geographically. The non-cohesive states (meaning they sound like each other) are those where the dominant nationalist party, the BJP is in clear majority. In Figure 1(b), we plotted the politicians of the same six states and it can be seen that the politicians in the northern states are well segregated based on their party affiliations: showing that their authorship on Twitter is more influenced by their party than their home state. This indicates that "statism" is relatively stronger in Southern, non-BJP states.

To further understand this, we plotted word clouds of tweets from the politicians in KL, MP and RJ for INC and BJP. Figure 2 shows these word clouds. It can be seen that BJP-MP and BJP-RJ word clouds look similar to each other; and so do INC-MP and INC-RJ. KL-BJP word cloud does not look like BJP-MP, but rather similar to INC-KL, the word cloud of its opposition party. Essentially while the party handles of northern states have affinity, in the southern state of Kerala, which is traditionally a non-BJP state, the party sounds more like its rival than its northern siblings.

The learnt politician embeddings and the word clouds indicate of Statism in the South and Federalism in the North. To validate our findings, we compute Pearson's correlation coefficient of the frequencies of the common words for the 3 states discussed above for both INC and BJP. Higher correlation indicates usage of the same words in the same amount and, since we have only considered common words between any two State-Party combination, it is language agnostic. Table 1 shows a subset of these correlations. We can see that in MP and RJ, the words used by politicians in tweets sounds alike, whereas in KL, the parties sound like each other across party lines. In this paper, we have shown that while AI for social good is typically seen primarily from an instrumental perspective of deploying initiatives, sophisticated ML techniques can be usefully deployed to better understand sociological phenomena that are important to our collective political and social future.



(a) Rajasthan-BJP (b) Rajasthan-INC (c) M.P.-BJP

(d) M.P.-INC (e) Kerala-BJP (f) Kerala-INC

Figure 2: Word Clouds for politicians of respective states: BJP MP and BJP RJ word clouds look similar to each other; and so do INC MP and INC RJ. KL BJP word cloud does not look like MP BJP but is rather similar to KL INC; so is the Kerala INC word cloud.

REFERENCES

[1] Timothy J Conlan and Paul L Posner. 2016. American federalism in an era of partisan polarization: The intergovernmental paradox of Obama's "New Nationalism". *Publius: The Journal of Federalism* (2016).

[2] Maria Giatsoglou, Manolis G Vozalis, Konstantinos Diamantaras, Athena Vakali, George Sarigiannidis, and Konstantinos Ch Chatzisavvas. 2017. Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications* (2017).

[3] Kanishka Jayasuriya and Kevin Hewison. 2004. The antipolitics of good governance: from global social policy to a global populism? *Critical Asian Studies* (2004).

[4] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *JMLR* (2008).

[5] Oren Melamud, Jacob Goldberger, and Ido Dagan. 2016. context2vec: Learning generic context embedding with bidirectional lstm. In *CoNLL*.

[6] Tomas Mikolov et al. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[7] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *KDD*. ACM.

[8] Markus Prior. 2013. Media and political polarization. *Annual Review of Political Science* (2013).

[9] Anton Shekhovtsov and Andreas Umland. 2014. The Maidan and Beyond: Ukraine's Radical Right. *Journal of Democracy* (2014).

[10] Mark R Thompson. 2016. The moral economy of electoralism and the rise of populism in the Philippines and Thailand. *Journal of Developing Societies* (2016).