American Politicians Diverge Systematically, Indian Politicians do so Chaotically: Text Embeddings as a Window into Party Polarization

Amar Budhiraja[*†1], **Ankur Sharma**[†2], **Rahul Agrawal**[3], **Monojit Choudhury**[4], **Joyojeet Pal**[4]
[1]Myntra Designs Pvt. Ltd., [2]Microsoft, USA, [3]Bing Ads, Microsoft, [4]Microsoft Research, India.

Abstract

Conversations on polarization are increasingly central to discussions of politics and society, but the schisms between parties and states can be hard to identify systematically in what politicians say. In this paper, we demonstrate the use of representation learning as a window into political dialogue on social media through the tweets authored by politicians on Twitter. Using a short-text based embedding technique, we visualize statements by politicians in a space such that their output embedding vectors represent the content similarity between the two politicians based on their tweets. The learnt embeddings for politicians of India and the United States show two trends. In the US case, we find a clear distinction between Democrats and Republicans, which is also reflected in the coalescing of the states that lean towards each party placing likewise in a graphical space. However, in the Indian case, the federal structure, multiparty system, and linguistic differences manifest in the coalescing political discourse in the largely monolingual north and the scattered regional states. Our work shows ways in which machine learning methods can offer a window into thinking about how polarized party discourses manifest in what politicians say online.

Introduction

Making sense of large-scale social media discourse is an active area of research (Nikfarjam et al. 2015; Shen and Kuo 2015; Tang and Liu 2010) and several attempts have been made to extract important information from ongoing social media posts such as medical abuse (Sarker et al. 2016), complaints of products (Jin and Lu 2020; Kursar and Gopinath 2013), and trend estimation (Mall et al. 2019; Park, Ciampaglia, and Ferrara 2016). In this paper, we map Twitter communication by party politicians. This involves charting how party, state, and individual choice may be factors in impacting what a politician says. The challenge with this problem is estimating the interplay between these factors, and how, in turn, these can be reflected in a graphical space to provide an understanding of politicians' individual and collective leanings.

There are challenges in mapping a politician's positions using short text social media formats, given that their positions need to be seen as a reflection of the entirety of their communication. But the same short text formats also allow a window into a cogent form of messaging in which a politician chooses their words to project a position. To this end, we aim here to visualize and understand party affiliation as a driving factor of what politicians write on Twitter. We propose to use unsupervised representation learning to answer this question by creating embeddings of politicians based on the tweets written by them and analyzing these embeddings to understand whether the topical communities are driven clustered by political affiliation.

Text-based embedding learning has been previously employed to understand phenomenon on social media (Alam, Joty, and Imran 2018; Chen, McKeever, and Delany 2019; Yan et al. 2020; Bahgat, Wilson, and Magdy 2020), including politics on social media (Oliveira et al. 2018; Hemphill and Schöpke-Gonzalez 2020). Some of the existing methods in the literature, which aim at embedding social media users based on the written content, concatenate all the posts of a user as a single document and then train a document level embedding model(Ding, Bickel, and Pan 2017, 2018; Benton, Arora, and Dredze 2016)[1]. Such approaches could lead to the introduction of noise in the training process since different posts may talk about different topics, and their concatenation would disturb the word distributions. To address this concern, we propose an embedding method which learns the embedding of a user by considering one tweet at a time, whereby each tweet contributes to the training, and there is no concatenation noise. We observe that the proposed method performs better than the concatenation approaches, showing this method was able to extract more relevant signal directly from the data.

To understand if these learnt embeddings, which do not use any party information during the training, are correlated with the party affiliation of the politician, we train a logistic regression model on the learnt embeddings with labels of only a handful of politicians and test it on the remaining set to see if this content representation of a politician can predict the affiliation of a politician. Our results demonstrate that using a simple Word2Vec model can achieve as high as 91% accuracy in party prediction for the USA, indi-

*Corresponding Author: amar.budhiraja1@gmail.com
†Work done while at Microsoft Research, India.

[1]We refer the reader to the survey paper by Pan *et. al.* (Pan and Ding 2019) for a detailed literature review

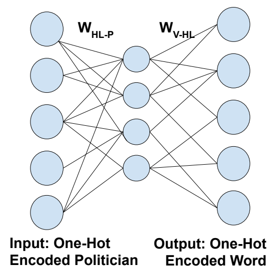

Figure 1: Model Architecture to learn politician embeddings

cating that there is indeed a very strong party influence on what politicians write on Twitter in the USA. However, the party prediction task did not lead to the same high accuracy for India's case ($\approx 85\%$ for the best approach), indicating that party may not be the most affluent factor contributing to what politicians write on Twitter. Upon further analysis, we observed that in the case of India, geography (the state to which the politician belongs) along with party affiliation also plays a very significant role in what they write on social media. Through this paper, we want to demonstrate that the machine learning methods like representation learning can help us in understanding complex social phenomena.

The rest of the paper is structured as follows. We present the proposed method of embedding politicians and party prediction experiment in Section 2. We discuss the observations around political affiliation and topical communities in Section 3, followed by a discussion in Section 4. We conclude and give directions to future work in Section 5.

Proposed Method and Party Prediction Task

Proposed Method

The learning setting of the proposed model is such that given a politician and a tweet they have written, predict 'K' sequence words from the tweet with the politician ID as input. Intuitively, the model tries to learn the consistency of the words written by the politician in a tweet, based on their own past tweets – ergo, *does this sound like they wrote it*. This has obvious challenges since we don't know if politicians write their own tweets, or if those who write their tweets are necessarily consistent. That said, because these are public pages, and often the primary means of outreach for politicians, they have an important representational value, which makes the experiment valuable in studying their positions.

The assumption we go with is that politicians who talk about similar topics should have similar embeddings. This is generally true since similar word distributions are used for similar topics, and if multiple politicians are talking about the same topics, the model will output words from the same distribution. Thus, the function that the model tries to approximate is $\Phi(P_i, w_j) = \{0, 1\}$ where P_i is the unique ID of the politician, w_j is a word and the label $\{0, 1\}$ is decided whether that word w_j occurred in the sampled 'K' words of the tweets that the politician has authored.

To approximate Φ from the given data, we propose to use a shallow neural network. Figure 1 shows the outline of the neural network architecture. The input layer corresponds to the politicians and is of the size of the number of politicians.

Count/Country	USA	India
Politician Handles	4422	13111
Tweets	2767344	5637474
State Annotations	1500	13111
Party Annotations	598	13111

Table 1: Dataset Statistics

The hidden layer is the size of the embedding dimension that is desired. The output layer is for the output words and is of the size of the vocabulary to enable prediction of words for the input politician. The model is trained to maximize the likelihood of prediction of the 'K' words sampled from the tweet. This is stated as follows,

$$\mathcal{L}(\theta) = \frac{1}{M}\frac{1}{K}\sum_{j=1}^{M}\sum_{i=1}^{K}\log(Pr(w_i, P_{t_j}) \mid \theta) \qquad (1)$$

where P_{t_j} is the politician who is the author of t_j^{th} tweet, *Pr* is the probability of predicting the word w_i out of the K words sampled from a tweet written by P_{t_j}, M is the total number of tweets in the dataset, and θ denotes the parameters of the model.

Party Prediction Task

To understand whether what a politician writes is influenced by their political party, we conduct an empirical experiment to predict the party of a politician using the learnt embeddings as features in a logistic regression model. High accuracy on this task indicates a high degree of correlation (and vice versa) between the content posted by politicians and their respective parties.

We create four types of feature vectors per politician using the tweets for the logistic regression classifier - 1) using TF-IDF (Salton and Buckley 1988) by constructing TF-IDF weights considering each politician as a document of concatenated tweets written by the respective politician; 2) using Word2Vec (Benton, Arora, and Dredze 2016); 3) using User-DBOW model (Ding, Bickel, and Pan 2017, 2018); and 4) using the proposed method. The experiment using TF-IDF is important to understand if the raw word distributions are sufficient enough to predict the parties of politicians, indicating no need for complex embedding methods. We observe that in the case of the USA, TF-IDF based performs very close to the embedding methods. We also conduct a hyper-parameter sweep over embedding size (50, 100, 200, 300 and 400) and window size (3,5,8). For evaluation purpose, we measure standard classification metrics - Accuracy, and F1-score (macro) for US dataset since it is a two-class classification problem and Accuracy, Precision, Recall and F1-score (macro) for India dataset since it is a multi-class classification problem.

For our experiments, we use the publicly available Nivaduck dataset (Panda et al. 2020) which contains state and party annotations for India and only state tags for the United States of America. Table 1 states the description of the dataset used. We use all accounts for training the embeddings and creating the TF-IDF vectors.

Method	Accuracy	F1-Score
Proposed Method	**0.9795 (0.0066)**	**0.9793 (0.0066)**
User-DBOW (Ding, Bickel, and Pan 2017)	0.9616 (0.0081)	0.9613 (0.0082)
Word2Vec Baseline (Benton, Arora, and Dredze 2016)	0.9131 (0.0169)	0.9126 (0.0170)
TF-IDF Baseline	0.9036 (0.0278)	0.9026 (0.0286)

Table 2: Results of Party Prediction Task for USA Dataset. Standard deviation for each metric is shown in brackets.

USA Dataset: For the USA dataset, we use the tagged dataset of Governors, Senators, and Congress from the Civil Service USA data[2] as our labelled instances. We further label some more political accounts to go beyond congressional handles, and our total dataset for this experiment is 598 politicians with 319 democrat and 279 republican accounts prior to the 2019 presidential election. We train a logistic regression on the embeddings with labels as the party and do 15-fold testing of the data. For training each fold, we randomly sample 100 politicians from each party, and remaining politicians act as the test users.

Table 2 shows the average values of metrics across different folds. It can be seen that TF-IDF gives a 90% accuracy, which is improved slightly by the method proposed in (Benton, Arora, and Dredze 2016). The User-DBOW approach proposed in (Ding, Bickel, and Pan 2017) improves the results significantly by approximately 5% followed by the proposed method, which can predict most accurately and has a gain of more than 1.5% over User-DBOW approach. This shows that the correlation between the word distribution and the party of a politician is significant enough that a TF-IDF feature vector with linear classifier can identify it 90% of the time. Interestingly, our model also predicted *Janet Mills* (Twitter handle: JanetMillsforME) as a democrat, whereas she was wrongly tagged as a republican in the Civil Services USA dataset.

India Dataset: There are multiple parties in India and, hence, for the Indian dataset, this is a multi-class problem. We consider politicians belonging to the most frequent top-9 parties in the dataset and government handles (which are expected to be neutral). The top-9 parties that we considered are: AAP, AIADMK, BJD, BJP, CPIM, DMK, INC, NCP, and SP. The final filtered dataset contained 11976 handles out of the initially available 13111 handles. Similar to the USA dataset, a logistic regression model is trained with input as the embedding of the politician and output as the party. We do a 15-fold experiment such that for each party at each fold, we select 50 politicians from each party and train the model. Thus, the training dataset is 500 politicians at each fold, and test dataset size is 11476. We select a fixed number of politicians for training from each party to avoid the problem of class imbalance.

Table 3 shows the results for the India dataset where each metric is averaged across the 15-folds. In the case of India, it can be seen that TF-IDF is not performing as high as the USA, but it is still able to predict the right party 65% of the time. However, Word2Vec based model (Benton, Arora, and Dredze 2016) significantly improves upon the TF-IDF baseline, and the trend continues with User-DBOW and the proposed method where the proposed method performs better 3.5% and 7.5% compared to User-DBOW (Ding, Bickel, and Pan 2017) and Benton *et. al.* (Benton, Arora, and Dredze 2016) approaches respectively. This shows that in India, while the embeddings which are learnt via the content of the tweets of politicians can predict the parties, the results are not as accurate as of the USA dataset. For example, the F1-Score in Table 3 is 0.699 compared to 0.97 in the case of the USA. This indicates that it is possible that in India, there may be factors beyond the party of a politician which influences their writing on social media, which we discuss in the next section.

[2]https://github.com/CivilServiceUSA/

Political Affiliation and Topical Communities

From our analysis in the previous section, we can see that the content of the tweets can effectively predict party of the politicians to a great extent, but in the case of India, we still see a large error of approximately 15% even in the most robust embedding training method. To further understand the topical communities amongst politicians, we employ tSNE (Maaten and Hinton 2008) to reconstruct the higher dimension neighbourhood in two dimensions of the embeddings learnt from the proposed method. Through the 2D plots of these embeddings, we visualize these clusters as a means of helping generate hypotheses on what these clusters indicate.

In the case of India, we observe a distinct pattern of bifurcation between the Northern and the Southern states. In India, the northern states are generally BJP-majority states whereas southern India tends to be dominated by regional and ethnolinguistic parties. Figure 2 (a) shows the scatter plot of the 2D representations of the learnt embeddings of six Indian states - Tamil Nadu (TN), Karnataka (KA), Madhya Pradesh (MP), Kerala (KL), Uttar Pradesh (UP) and Rajasthan (RJ). It can be seen that MP, UP, and RJ, all dominated by the BJP in the national parliament have significantly overlapping clusters - meaning the three states "sound-alike". In contrast, in the southern states of KL, TN, and KA, the political clusters are more cohesive. In a nutshell, a politician from UP, MP, or RJ could sound like one from another one of those states, irrespective of which party they belong to. However, in the southern states, even the BJP politicians sound more like other politicians from their state than their party members from neighbouring states. This is an interesting insight into the politics of distinction, and it shows that geography (state) of the politician is also a crucial factor in determining what they write on social media.

For the USA dataset, we use 1500 politicians according to their states and as well as if they are Congressional, Senatorial, or State legislators' handles. We then plot Congressional handles according to their political affiliation and other political handles (such as Governors) based on their states. Figure 2 (b) shows the congressional handles (Democrats in blue circles and Republicans in red crosses) and some political Twitter handles from six states - three generally blue states (MA, NY, and WA), and three generally

Method	Accuracy	Precision	Recall	F1-Score
Proposed Method	**0.8518** (0.0100)	**0.9206** (0.0061)	**0.6062** (0.0160)	**0.6990** (0.0146)
User-DBOW (Ding, Bickel, and Pan 2017)	0.8151 (0.0156)	0.9053 (0.0083)	0.5712 (0.0168)	0.6651 (0.0163)
Word2Vec Baseline (Benton, Arora, and Dredze 2016)	0.7677 (0.0143)	0.8533 (0.0067)	0.5004 (0.0159)	0.5756 (0.0160)
TF-IDF Baseline	0.6527 (0.0238)	0.8110 (0.0127)	0.4745 (0.0274)	0.5315 (0.0233)

Table 3: Results of Politician Polarization for Indian Dataset. Standard deviation for each metric is shown in round brackets.

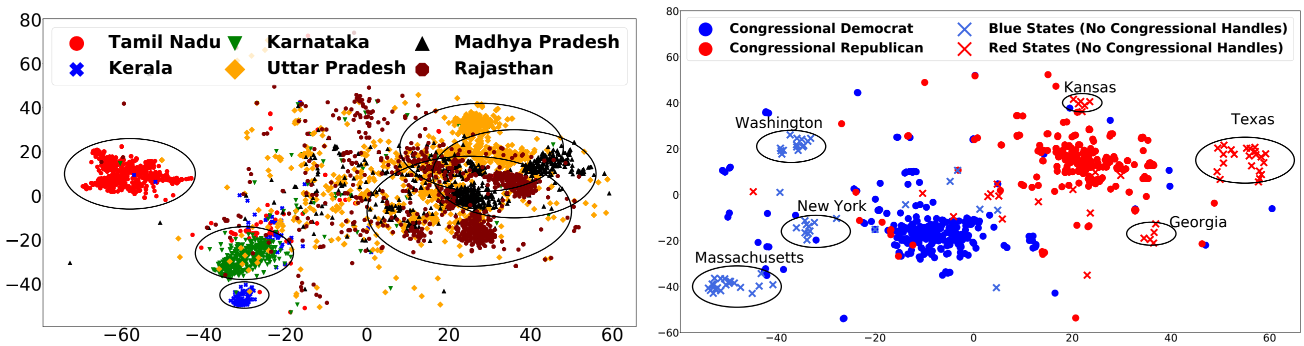

Figure 2: (a) Scatter plot of embeddings for six states in India - Each state is colored by a different color. (b) Scatter plot of the embeddings for USA - Blue circles represent Twitter handles of Congressional Democrats, Red Circles represent Twitter handles Congressional Republicans. Blue Crosses represent blue states, and Red Crosses represent red states.

red states (GA, KS, and TX). A non-Congressional politician handle from these six states is shown as a blue or red ‘x’. It can be seen in the figure that there are two big clusters - red and blue pertaining to Republican Congressional handles and Democratic Congressional handles respectively. Blue states (MA, NY, and WA) are closer to the Blue cluster, and Red states (GA, KS, and TX) are closer to the Red cluster. This shows a clear party-based affinity from state politicians as well as congressional handles. Interestingly, we also observe that there are clear democratic and republican clusters within each state, which explains why the party prediction task had very high accuracy. The high classification metrics from the party prediction in the previous section and the corresponding 2D tSNE representations show evidence that party affiliations could be the driving force of their tweets for USA politicians.

Discussion

Our methodology gives us a way to think about polarization in political systems. As we see in the graphical representations, party polarization in the US is very stark. Republicans and Democrats sit virtually on two separate planes, and the states that veer towards one or another, closely mirror that separation in politicians’ online discourse.

The Indian multi-party case shows us the complexity of cleanly visualized distinctions when mapping the discourse of politicians. We see that the Hindi-speaking northern states are highly clustered and overlapping, representing the hegemonic control of the ruling Bharatiya Janata Party in those states, while the non-Hindi speaking states with their regional languages as well as their separate political parties are strikingly separated into their own clusters.

Essentially, while Republicans and Democrats are consistent in their output irrespective of the state in the United States, the Twitter feed of politicians in India is truer to the state of origin than it is to the party, with the exception of the one integrated cluster of North Indias states. Here, our method gives us a unique ability to show how the discourse of this dominant region is coalescing into a large and relatively uniform graphical space, while the peripheral states indeed appear like a periphery – in compact, separate clusters apart from the center. These visualizations reflect the nuances of polarizing politics in the two countries.

Conclusion and Future Work

Recent events including the polarization among politicians and the general public over various protests movements in India and USA have laid bare the importance of understanding the contours of division in our political discourse. We demonstrate that representation learning gives us a window into how polarized our politics are, and think about the risks that brings. Yet the same methods can can also be used to understand what the boundary objects and issues are, where in turn there may be scope for conversation.

As a part of future work, we plan to further improve the proposed model by considering sequence and attention model instead of the bag of words approach. We also plan on investigating how to include temporal signals to understand if there is topical migration across states and parties.

References

Alam, F.; Joty, S.; and Imran, M. 2018. Graph based semi-supervised learning with convolution neural networks to classify crisis related tweets. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Bahgat, M.; Wilson, S.; and Magdy, W. 2020. Towards Using Word Embedding Vector Space for Better Cohort Analysis. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Benton, A.; Arora, R.; and Dredze, M. 2016. Learning multiview embeddings of twitter users. In *Association for Computational Linguistics*.

Chen, H.; McKeever, S.; and Delany, S. J. 2019. The use of deep learning distributed representations in the identification of abusive text. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Ding, T.; Bickel, W. K.; and Pan, S. 2017. Multi-view unsupervised user feature embedding for social media-based substance use prediction. In *Empirical Methods in Natural Language Processing*.

Ding, T.; Bickel, W. K.; and Pan, S. 2018. Predicting Delay Discounting from Social Media Likes with Unsupervised Feature Learning. In *ASONAM International Conference on Advances in Social Networks Analysis and Mining*. IEEE.

Hemphill, L.; and Schöpke-Gonzalez, A. M. 2020. Two Computational Models for Analyzing Political Attention in Social Media. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Jin, J.; and Lu, L. 2020. Service complaint identification in hotel social media: A two-step classification approach. *The International Journal of Electrical Engineering & Education* 0020720920928467.

Kursar, B.; and Gopinath, J. 2013. Validating customer complaints based on social media postings. US Patent App. 13/646,548.

Maaten, L. v. d.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* .

Mall, U.; Matzen, K.; Hariharan, B.; Snavely, N.; and Bala, K. 2019. Geostyle: Discovering fashion trends and events. In *International Conference on Computer Vision*.

Nikfarjam, A.; Sarker, A.; O'connor, K.; Ginn, R.; and Gonzalez, G. 2015. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association* .

Oliveira, L. S.; de Melo, P. O. V.; Amaral, M. S.; and Pinho, J. A. G. 2018. When Politicians Talk About Politics: Identifying Political Tweets of Brazilian Congressmen. In *AAAI Conference on Web and Social Media*.

Pan, S.; and Ding, T. 2019. Social media-based user embedding: a literature review. *arXiv preprint arXiv:1907.00725* .

Panda, A.; Gonawela, A.; Mishra, D.; Mohapatra, Mugdha Chandrasekaran, R.; and Pal, J. 2020. PoliTwictionary with NivaDuck - A scalable database of political Twitter handles for India and the United States. In *Conference on Social Media & Society*.

Park, J.; Ciampaglia, G. L.; and Ferrara, E. 2016. Style in the age of instagram: Predicting success within the fashion industry using social media. In *Conference on computer-supported cooperative work & social computing*.

Salton, G.; and Buckley, C. 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management* .

Sarker, A.; O'Connor, K.; Ginn, R.; Scotch, M.; Smith, K.; Malone, D.; and Gonzalez, G. 2016. Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter. *Drug safety* .

Shen, C.-w.; and Kuo, C.-J. 2015. Learning in massive open online courses: Evidence from social media mining. *Computers in Human Behavior* .

Tang, L.; and Liu, H. 2010. Community detection and mining in social media. *Synthesis lectures on data mining and knowledge discovery* .

Yan, M.; Lin, Y.-R.; Hwa, R.; Ertugrul, A. M.; Guo, M.; and Chung, W.-T. 2020. MimicProp: Learning to Incorporate Lexicon Knowledge into Distributed Word Representation for Social Media Analysis. In *Proceedings of the International AAAI Conference on Web and Social Media*.