VisTA: From Topic Probability Dissimilarity to a Latent Feature Space for Understanding Twitter User Stances

Sikata Sengupta

November 2017

Abstract

Detecting sentiments and stances in a conversation is one of the developing areas of Natural Language Processing (NLP). Traditional supervised methods for stance detection are resource-intensive and often focus on classifying statements in isolation. However, to understand the nature of a discourse, a fine-grained understanding of a participant's stances and preferences is needed. This participant-centric view provides the motivation for seeking both user level data and unsupervised methods to discover latent feature spaces associated with the user. Representing users as topic probability vectors and using a particular information-based distance in the probability space, I explore the landscape of user diversity. In practice, for some query targets, the latent feature space is useful for further classification tasks. More importantly, the geometry reveals that user polarization is not binary, but rather graded, allowing one to detect users who are potentially open to engagement in constructive dialogues.

1 Introduction

Over the past election season, there has been quite a bit of controversy surrounding social media. While some posit that it empowers its users, providing information about current events and society, others criticize it because it creates echo-chambers [25], where one simply reiterates his or her own view. Given the popularity and convenience of social media, it is unlikely that one will forego its use for the sake of preventing bias. Thus, my research strives to create a system for social media that develops spheres for positive and constructive engagement. From this system, users would be able to learn and consider controversial topics with other points of view.

In order to take on such a challenging task, one must first understand the structure of the community, consisting of individual users, each with their own set of topics and opinions. Such an endeavor weaves together two threads of research in Natural Language Processing (NLP), stance detection [33, 28, 20, 34, 9, 29] and topic modeling [6, 13, 14, 11, 8, 3]. Stance detection, complementing sentiment analysis, has been a growing area of investigation during the last decade, because of its many applications, particularly in retrieval and in summarization. The methods of choice in stance detection naturally come from supervised learning techniques. Topic modeling has a longer history, discovering hidden semantic structures in documents. This area mostly draws on unsupervised methods for finding latent features in texts.

Stance detection is the task of deciding whether an author's attitude towards a target topic is 'in favor of', 'against' or 'neither against nor in favor'. This task is distinct from sentiment analysis which surmises the emotional state of the author in the particular text. For example, one could be in favor of 'Feminist Movement' and express the sentiment of being 'happy' or 'angry'. Early work on the subject focused on congressional or online debate [33, 28, 20, 34, 9, 29]. A twitter dataset for stance detection —[17], basis of SemEval 2016 task 6 [18] and subsequent work [19] have prompted efforts to detect stance from short texts. This dataset and my subsequent collection of related data form the basis of my study.

The issues with supervised stance detection, *e.g.* from individual tweets, are as follows:

- A single tweet can provide only so much information. Users do not simply discuss or have opinions on one issue. Rather, they have multiple thoughts and stances on different topics. To fully understand their viewpoints, one must gather more data on their word statistics from multiple tweets.
- It suffers from inflexibility, having to manually decide on the topics and categories

rather than letting the data decide what is important. With supervised methods, one has to categorize text into fixed, pre-determined labels. This method is problematic when dealing with a large number of topics and more nuanced stances that do not fit into discrete classes.

- Accurately labeling stances for large datasets is time consuming and such labeling is not easily available for many applications. When dealing with Twitter users, there are no provided labels–hashtags tend to be misleading and do not fully provide information on the topic and stance of the user. Thus, one has to manually create large labelled datasets of the population of Twitter users who discuss controversial political issues.

Given these issues, there is a serious need to develop unsupervised and semisupervised methods for characterizing users. For the starting point of such an approach, topic modeling is one obvious candidate.

Topic modeling originated from efforts, over many decades, to analyze hidden similarity between documents. Representing documents by vectors with entries as weighted frequencies has already been in use from the works of Gerard Salton and his colleagues [27]. Latent Semantic Analysis (LSA) [6] has its origin in finding a low rank representation of matrices associated with a corpus, with each document being a row vector. This representation, using singular value decomposition, allows for efficient document recovery. A step towards a probabilistic generative model of documents was taken by Hofmann's probabilistic Latent Semantic Analysis (pLSA)[11], which amounts to a nonnegative low rank factorization of the matrix [15]. A full fledged Bayesian model with priors favoring topic sparsity, namely Latent Dirichlet Allocation (LDA), generalizes pLSA and controls overfitting [3]. LDA is a very popular topic modeling tool and is central to my study.

In the age of social media, both sets of methods face new difficulties to overcome. The typical dataset consists of a large number of user-generated texts, each of which is much shorter than typical documents like news reports, journal articles or books that were analyzed by the methods mentioned above [12]. These methods falter badly because words frequencies are very poorly estimated in short texts. Hence, researchers have looked for newer methods for topic discovery for short texts [35, 30]. True to many works on topic modeling, most of these new studies are focused on the quality of topics, their coherence, interpretability and other desirable aspects. Although a few studies on evaluating topic models depart from the norm [4, 1], relatively less effort has been spent on the utility of using topic structure to find useful low dimensional representations of documents or users.

I propose that such a representation provides insights into the structure of the relevant user

community. I hypothesize that this low-dimensional representation could contain relevant information for stance detection, could lead to similar or better performance compared to classifiers based on far more complex text features. I also speculate that a representation based on topic choice similarity between users is useful for understanding the degree of polarization in the community as well as for practical applications like, content recommendation systems.

During this study, at first, I explored baseline supervised methods for stance classification at the single tweet-level, using both a linear support vector machine (LinSVM) (with character ngrams and word ngrams as features) and a Naïve Bayes (NB) model (with a binary unigram feature space). The performance of these classifiers informs us of what can be achieved with painstaking annotation and hand-crafted feature engineering . Then, under the advice of my mentor, I started considering semisupervised and unsupervised approaches based solely on word frequencies. I used k-means clustering on tf-idf representations of the tweets as well as on representations using word-averaged word2vec, GloVe embeddings. I found that k-means cluster identity has some predictive value for the original target topics of the tweet, but has little information on the stance. These experiments led me to conclude that I needed to create a feature space at the user level. I decided to model topics in an users twitter stream, build a feature space from there, and only then, study its appropriateness for stance classification and other tasks.

To create my dataset, I collected a group of users interested in politically controversial topics from [18]. Then I collected tweets corresponding to each individual users, creating pseudo-documents. I applied both Latent Dirichlet Allocation and word clustering by k-means to this dataset in order to see what latent topics were generated. Both methods generate some coherent topics. These topics have overlaps with the original targets topics, but are not identical. I show that, in the case of LDA, there is a natural way to create feature space representing users. Using probabilistic distances and approximately distance preserving embeddings, I explore the landscape of the space of user and show that stance classification is possible in this feature space. I call this method VisTA: **V**isualization by **T**opic **A**llocation. The overall scheme for embedding users in a multidimensional space is shown in figure 1. This representation provides additional information on the community structure, like the graded polarizations of users on controversial issues.

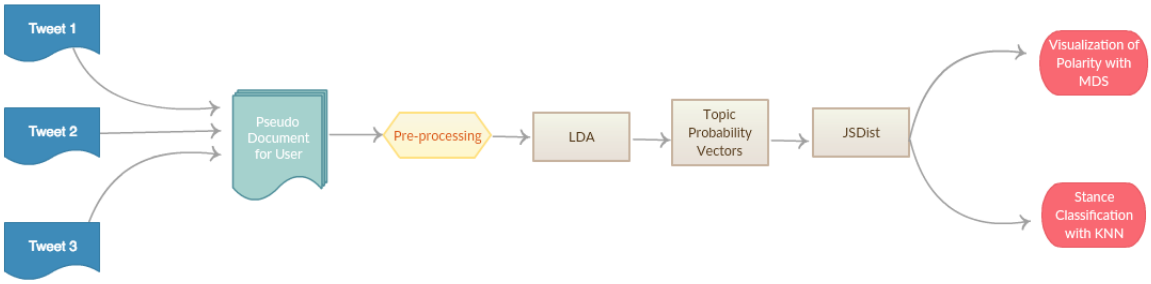

Figure 1: Flow chart for VisTA (Visualization by Topic Allocation): discovering and visualizing user feature space.

2 Methods

Note: All programming was performed in Python, primarily using Numpy, Scipy and Scikit-learn libraries [22]. Figures are plotted with Matplotlib from Jupyter notebooks. The source code is available at GitHub [7].

The project progressed in four phases as described below:

Phase 1: Supervised Approaches to Stance Detection: This part of the project analyzed the SemEval 2016 stance detection dataset [18]. The authors provided annotated results of five query target topics: 'Atheism', 'Climate Change is a Real Concern' ('Climate'), 'Feminist Movement' ('Feminism'), 'Hillary Clinton' ('Clinton') and 'Legalization of Abortion' ('Abortion'). Tweets were labeled 'FAVOR', 'AGAINST' and 'NONE'. Of the 4163 labeled tweets, 2914 was for training and 1249 for test.

My work took off from my mentor's effort to replicate the results of Mohammad *et al.* [19]. Since the code for the published work was not publicly available, he/she used an in-house preprocessor for the tweets and then implemented a linear support vector machine (with a feature space of char-1,2,3 grams and n-1,2,3 grams), as described in Mohammad *et al.* [19]. I adapted the code to Python 3. For a different project for a chrome extension, I modified the code so that the classifier could be saved and be used to score new tweets. I also ran a Bernoulii Naïve Bayes (NB) classifier on the tweets (with NLTK list of English stop words removed) from the SemEval 2016 dataset, based solely on word unigrams. The feature space of the SVM classifier included both character and word n-grams. In contrast, the NB classifier simply included binary variables indicating the presence of absence of a word.

Throughout this work, I will measure the informative capacity of the method in terms of a normalized mutual information [31] between true labels and predicted labels. If T and P are

the true and the predicted labels, respectively, then normalized mutual information is given by

$$\mathrm{NMI}(T, P) = \frac{I(T, P)}{\sqrt{H(T)H(P)}} \tag{1}$$

where $I(T, P)$ is the mutual information and $H(T), H(P)$ are the entropies of the two variables. This framework could be used for supervised methods, reporting correlation between true classes and predicted classes. It could also be used to investigate relationships between ground truth and labels from unsupervised methods, like cluster identities. One thing to note is that if the predicted label has zero entropy, this $\mathrm{NMI}(T, P)$ would come out as NaN. That result is, therefore, an indication of poor performance.

Phase 2: K-Means Clustering for Tf-idf: Following my mentor's suggestion, I used a tf-idf representation [27] of the tweets in the SemEval dataset and clustered these vectors using k-means clustering. I then explored the contingency table of cluster identity and target topic as well the contingency table of cluster identity and stance label for particular target topics. I studied the association between clusters and target topics/stances by χ^2 tests. Additionally, I also computed normalized mutual information as an alternative measure of association. I also computed tweet-averaged word2vec and GloVe embeddings as alternatives to tf-idf and passed the outputs through the clustering and association pipeline.

To display the results of clustering, I use singular value decomposition (SVD). For high-dimensional sparse vectors, like tf-idf, computing principal components becomes problematic, since subtracting the mean makes it into a low information dense vector. In many cases, removing the leading singular vector component (SVD0) has a similar effect as removing the mean. I therefore leave the leading component and use the next three components (SVD1, SVD2, SVD3) to make 3D projections.

Phase 3: Unsupervised Learning and Topic Modeling: To create my dataset, I collected a group of users interested in politically controversial topics from SemEval 2016 dataset [18]. I removed punctuations and other unimportant characters by using RegExp no punc tokenizer from the Natural Language Toolkit (NLTK) [2] and then preprocessed the filtered sentences by a Python 3 program based on the Stanford NLP Twitter PreProcessor originally written in Ruby [21]. The tokens were stemmed by NLTK lemmatizer and corrected by a spell check program.

I organized tweets belonging to the same users into pseudo-documents, generating 705 of them. I then looked for topics in the documents by two different methods. A baseline approach was to study vector space embeddings of all words in this corpus and use clustering

to find topics [1, 30]. The second approach was to use LDA on the pseudo-documents [12, 35].

Using precomputed GloVe vectors for Twitter [24] obtained from the Stanford NLP website [23]. I originally fit a 5-class Gaussian Mixture Model (GMM). I did not train corpus specific word vectors as was done in previous work [30], but unlike these authors, I used full covariance matrices. I also used k-means clustering on the word vectors, following the baseline method of Bhatia *et al.* [1]. I used 100 clusters and 10 cluster solutions.

For LDA, I chose to run a 10-topic model on the 705 pseudo-documents using gensim library's implementation (gensim.models.ldamodel.LdaModel) [26] of the online algorithm from Hoffman *et al.* [10]. I also extracted the estimate of topic probabilities for individual documents/users.

Phase 4: From Topic Probability Distance to User Feature Space: I will denote the the topic probability distribution, $P(\text{topic}|\text{user})$ as a vector p_u, where u is the user. With 10 topics, the vector is 10-dimensional with elements adding up to one. To define a natural distance for probability distribution, there are measures related to information theory, like Kullback-Leibler (KL) divergence [16]

$$\text{div}_{\text{KL}}(p_1||p_2) = \sum_t p_1(t)\log(p_1(t)/p_2(t)). \tag{2}$$

Since KL divergence is not symmetric, it is preferable to use the Jensen-Shannon divergence (JSDiv).

$$\text{div}_{\text{JS}}(p_1, p_2) = \frac{1}{2}\big[\text{div}_{\text{KL}}(p_1||m) + \text{div}_{\text{KL}}(m||p_2)\big] \tag{3}$$

where $m = \frac{p_1+p_2}{2}$, is the average of the two probability distributions. Using JSDiv between probability distributions one can define the Jensen-Shannon distance (JSDist) between two probability distributions p_1, p_2:

$$\text{dist}_{\text{JS}}(p_1, p_2) = \sqrt{2 * \text{div}_{\text{JS}}(p_1, p_2)}. \tag{4}$$

The distance between users u_1, u_2 is given by $\text{dist}_{\text{JS}}(p_{u_1}, p_{u_2})$. I decided to use this distance to describe the geometry of the latent feature space I am looking for to embed the users in.

To see if this distance is useful, I considered the original labeled tweets in the SemEval 2016 data set from the 705 users whose stream provided the pseudo-documents analyzed by LDA. I explore whether it is possible to predict the stances for any of the target topics based on JSDist. I performed a leave-one-out study to check if the stance label can be retrieved based on the labels of the neighbors. A natural choice for distance based classification is

the k-nearest neighbor (KNN) method. I used the 4 nearest neighbors to predict labels. Since, depending on the target, there are very unbalanced class sizes. a method is needed for weighing classes differently [5, 32]. I chose a uniform weight for each class i, proportional to $1/C_i^\alpha$ where $\alpha < 1$ and C_i is the number of training data points from class i are available. I also used multidimensional scaling (MDS) with JSDist for users who have labeled tweets on a particular target topic. The embedding provided by MDS allows me to study how labels relate to the geometry of neighborhoods in the low dimensional space.

To go further, I took 50 labeled tweets, at random, on the target 'Atheism' and labeled them from 1 to 5, depending upon how polarized the opinion was (of course, without looking at the labels or the MDS emebedding). Below are examples of texts corresponding to each manual rating category:

Rating 1: "May my only happiness be to please Thee, O Infinite Goodness. #Catholic #prayerJesus, Brightness of eternal Light, have mercy on us. Pray the Holy Name"

Rating 2: "is the ache <emojis: lips, red heart>2have my eyes closed 1 minute lost in thoughts+prayers Then 2open my eyes+look up And immediately see a shooting star Divine Timing Maybe<emojis: cross, black heart,..>"

Rating 3: "It's wrong to say please love me too cause I know you never do - King GiradoThe only in my life "

Rating 4: "It doesn't matter how you feel, far logical to argue for billions than a few k. Also it's secular and I'm not a socialist."

Rating 5: "#atheist #humanism #atheism@ianthebeard My point is that the Bible is best interpreted as fiction@ianthebeard ... but when you're putting together a fictional universe, the more you strain credulity, the more you lose readersSeason 1 of the Bible jumps the shark with its Jonah-Whale subplot."

I associated the scores +1 with 'FAVOR', 0 with 'NONE' with -1 for 'AGAINST'. I then created a computed average for each point based on the scores of its 4-nearest neighbors, taking the average over the nearest neighbors with the weights used for KNN. I plot the rating against the computed score and study correlation.

3 Results and Discussions

Results from Phase 1: For stance classification performance of the classifiers for different targets see table 1. Not surprisingly, with a more complex feature space, the SVM outperforms the baseline NB classifier. Although the classifiers succeed partially in predicting stances, particular stances on a subset of target topics, such as with the target topic of 'Hillary Clinton', remain difficult to detect for either classifiers perform quite poorly.

Classifier \ Target	Atheism	Climate	Feminism	Clinton	Abortion
Naive Bayes	NaN	0.024	0.048	NaN	0.046
Linear SVC	0.101	0.066	0.069	0.132	0.079

Table 1: Classifier performance for each topic with both Naive Bayes and Linear SVC, based on normalized mutual information. Linear SVC seems to hold more information than Naive Bayes, but more importantly, these numbers provide a point of comparison for my future results. Note that NaN is produced when the method does not produce more than one labels, an indication of poor performance.

Results from Phase 2:

I found that k-means clustering worked somewhat well in determining the topics of the tweet, see figure 2, but not very well on the stance of the tweets. These experiments led me to conclude that I needed to create a feature space on the user level to improve upon topic modeling and stance classification. The tf-idf clusters held statistically significant topics, as measured by the normalized mutual information (see table 2). If I was to compute a χ^2 test, ignoring zeros and low frequency terms, the p-value would be extremely small: 1.6×10^{-251}. Although the clusters hold significant information about the topics, they are not in one to one correspondence with each other.

Cluster \ Target	Atheism	Climate	Feminism	Clinton	Abortion
Cluster 0	282	320	302	255	319
Cluster 1	0	0	0	125	4
Cluster 2	0	3	0	107	70
Cluster 3	6	168	2	7	44
Cluster 4	122	40	12	17	45

Table 2: Contingency table for tf-idf clusters for all tweets and target topics, normalized mutual information=0.19, apparent χ^2 association p-value = 1.6×10^{-251}. Tf-idf clusters hold significant information on the topic of the tweet.

However, clustering tweets on a particular target topic does not reveal much about stances.

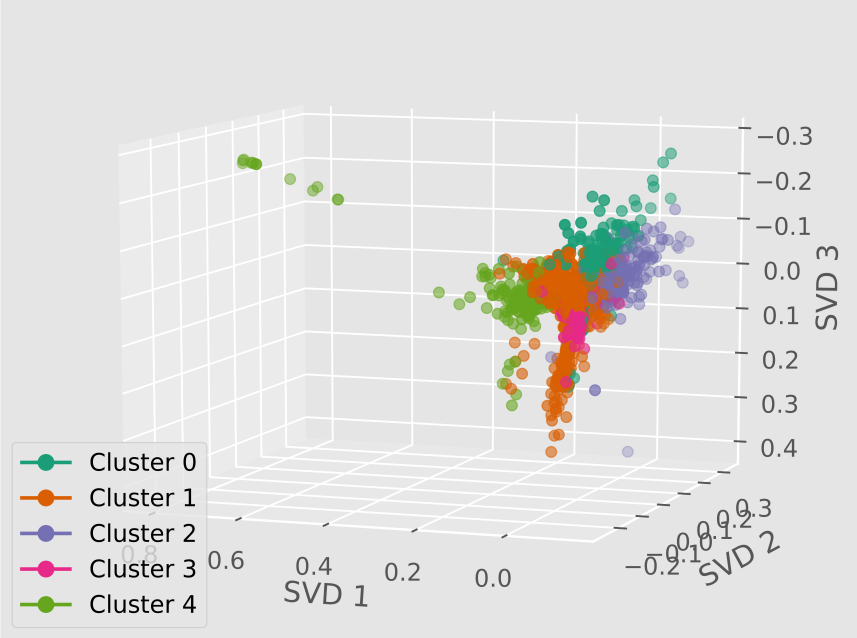

Figure 2: Clustering to find topics of tweets (3D projection via SVD) tf-idf. This figure shows some level of structure for the different clusters, suggesting that tf-idf captures some distinctive features on a tweet's topic.

For example, for the 'Atheism' target, I applied k-means to generate 3 clusters and looked at the contingency table (table 4) between the labels and the cluster identities. Although the χ^2 association p-value is 0.015, normalized mutual information is just 0.0143, indicating very little information about stances being present in the clusters identity. I also tried similar unsupervised techniques on the tweet level with averaged-word2vec and averaged-GloVe vectors, instead of tf-idf embedding and found that they have less information on target (see, *e.g.* contingency table for word2vec results: table 3).

Results from Phase 3: Some topics generated by the 5-class GMM looked plausible while others made very little sense, but trying to assign topics to tweets or to pseudo-documents following [30] produced poor results. It was was dominated by one large cluster, without much coherence. Three to four of the clusters from k-means clustering are easily interpretable and are stable under changes in the number of clusters. An example of a good topic, represented by the most common words looks as follows {god, lord, amp, psalm, good, proverb, one, heaven, love,..}. However assigning topics to users was still a problem. The

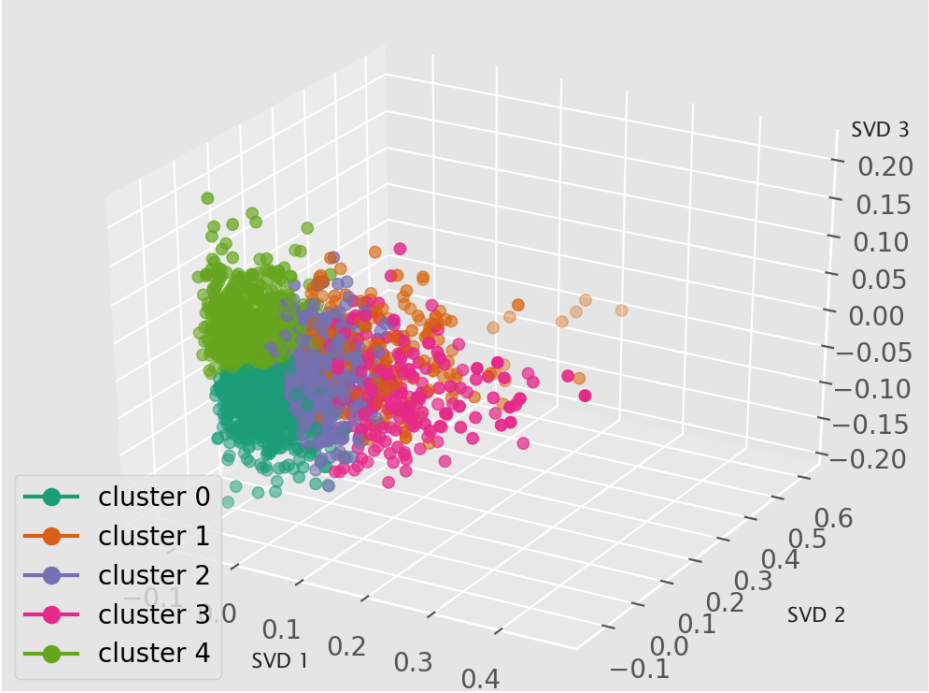

Figure 3: Clustering to find topics of tweets (3D Projection) using tweet-averaged word2vec embedding. Average word2vec vectors do not seem to hold distinctive features for each tweet as the figure does not show much structure in its separation of clusters.

majority of the words were associated with what seems like relatively incoherent topics such as: {co, thank, cloud, n, say, storm, para, sand, indeed, abide, viewed, ...}. Distinguishing users, without hand-picking relevant clusters, seems to be problematic.

LDA, on the other hand produced, many interpretable topic (see figure 4). With a natural estimated topic probability vector for each user, I could go ahead and compute Jensen-Shannon distance JSDist (equation 4) between each pair. This distance forms the basis of the work in Phase 4.

Results from Phase 4: The modified k-nearest neighbors method, using JSDist shows some ability to recover original labels. If I treat the 3×3 confusion matrix as a contingency table (table 5), I get normalized mutual information=.062, χ^2 association p-value=.0027. Of course, The results are much better than trying to predict the stance from tf-idf clusters. KNN performs better even compared to the Bernoulli NB classifier, which is also based on the bag of words model. Only the LinSVM performs somewhat better, most probably because of its richer feature space. Going beyond stance classification performance, I show that the distances describe the geometry of the continuum of stances in the following.

I now show the results of two dimensional and three dimensional embeddings, via MDS, of

Cluster \ Target	Atheism	Climate	Feminism	Clinton	Abortion
Cluster 0	88	101	71	108	133
Cluster 1	42	69	46	69	40
Cluster 2	22	46	46	78	36
Cluster 3	128	160	80	135	100
Cluster 4	130	155	73	121	173

Table 3: Contingency table for averaged-word2vec clusters for all tweets and target topics, normalized mutual information=0.19, apparent χ^2 association p-value = 1.6×10^{-251}. Averaged word2vec holds less information on the topic of the tweet compared to tf-idf.

Cluster \ Stance	AGAINST	NONE	FAVOR
Cluster 0	12	53	37
Cluster 1	28	81	24
Cluster 2	33	103	39

Table 4: Contingency table for k-means of 'Atheism' tweets, normalized mutual information=.0143, χ^2 association p-value=.015. Tf-idf does not hold particularly distinctive features on the stance of the user given the topic of a tweet.

the users with 'Atheism' related tweets. The color is related to the labels, as described in the legends. One can see that users in favor of 'Atheism' form a smaller cluster intermingling with a much wider cluster of opponents.

When I visualize the manual labels in the 3D embedding, the polarization gradation relates to, a direction in this space (see figure 6). I also see a strong correlation between the manual score and the predicted score based on the neighborhood (figure 7). The correlation coefficient is 0.544 with a p-value of = 1.19×10^{-5}. These observations, together, suggest that the geometric representation is of value.

4 Conclusion and Future Work

In this work I show that the topic probability vectors for each user do hold features distinguishing the users. Using an appropriate measure of dissimilarity of these probabilities, one can create a feature space for users to gain insight into community structure. In this space, there is a continuous spectrum of polarization, rather than distinct clusters of stances. On the basis of the observations presented, at least for some target topics, this probability based distance provides an unsupervised approach to a feature space facilitating stance

```
[(0,
  '0.009*"science" + 0.008*"twitter" + 0.007*"ii" + 0.006*"based" + 0.006*"human" + 0.006*"law" + 0.005*"course" + 0.
005*"tedcruz" + 0.005*"photo" + 0.005*"education"'),
 (1,
  '0.011*"woman" + 0.010*"feminist" + 0.009*"tweet" + 0.006*"speech" + 0.006*"pouts" + 0.006*"Boston" + 0.005*"men" +
0.005*"going" + 0.005*"feminism" + 0.005*"protest"'),
 (2,
  '0.041*"co" + 0.023*"via" + 0.006*"Nazi" + 0.006*"video" + 0.005*"profile" + 0.005*"play list" + 0.005*"bickering"
 + 0.004*"Barcelona" + 0.004*"us" + 0.004*"trump"'),
 (3,
  '0.017*"connecting" + 0.017*"co" + 0.009*"la" + 0.005*"_tmf" + 0.005*"gustavorejivik" + 0.005*"citron cockatoo" +
 0.005*"friends\'" + 0.005*"eddarrell" + 0.005*"sapienthetero" + 0.005*"drwaheeduddin"'),
 (4,
  '0.024*"scripture" + 0.008*"mufti" + 0.006*"syromalabar" + 0.006*"0.005 + "⇀t*"digitalkalyan" + 0.005*"indiancathol
ic" + 0.004*"Kemp" + 0.004*"hjfelder" + 0.004*"frannydascani" + 0.004*"montimai"'),
 (5,
  '0.016*"amp" + 0.013*"people" + 0.013*"trump" + 0.010*"like" + 0.007*"know" + 0.007*"would" + 0.007*"realdonaldtrum
p" + 0.007*"right" + 0.007*"white" + 0.007*"need"'),
 (6,
  '0.172*"co" + 0.008*"thanks" + 0.005*"new" + 0.005*"connect" + 0.005*"f" + 0.005*"time" + 0.005*"like" + 0.005*"x"
 + 0.004*"go" + 0.004*"live"'),
 (7,
  '0.008*"co" + 0.008*"you tube" + 0.007*"Canada" + 0.007*"wt" + 0.007*"one" + 0.006*"fucking" + 0.006*"school" + 0.0
05*"oh" + 0.005*"remember" + 0.004*"game"'),
 (8,
  '0.019*"life" + 0.016*"pray" + 0.011*"day" + 0.011*"church" + 0.008*"new" + 0.007*"love" + 0.007*"work" + 0.006*"pr
ofile" + 0.006*"every" + 0.005*"may"'),
 (9,
  '0.039*"Jesus" + 0.033*"god" + 0.017*"bible" + 0.012*"love" + 0.011*"flt" + 0.010*"awesome" + 0.010*"catholic" + 0.
008*"Christianity" + 0.007*"free" + 0.007*"prayer"')]
```

Figure 4: LDA topics are in parentheses, with word probabilities 'multiplying' the corresponding words. Some of these topics are interpretable by humans, including political, religious, and feminist themes.

Predicted \ True	AGAINST	NONE	FAVOR
AGAINST	87	22	6
NONE	6	9	4
FAVOR	13	7	2

Table 5: Contingency table for KNN Predictions for 'Atheism', normalized mutual information=.062, χ^2 association p-value=.0027. The KNN predictions suggest that these probability topic vector representations do hold information on the stance of a user.

classification.

This study compares and contrasts traditional supervised stance detection approaches with unsupervised methods for finding appropriate representations of user level data that is informative of stances. The empirical results are currently limited to users appearing in a particular dataset. To establish the utility of this approach, it needs to be applied to many other datasets. It is possible that the efficacy of the method would be different for different target topics.

Beyond classification, this feature space appears to also provide a graded measure of polarization. The multidimensional embedding has directions other than the polarization axis. The interpretation of these directions remain to be discovered.

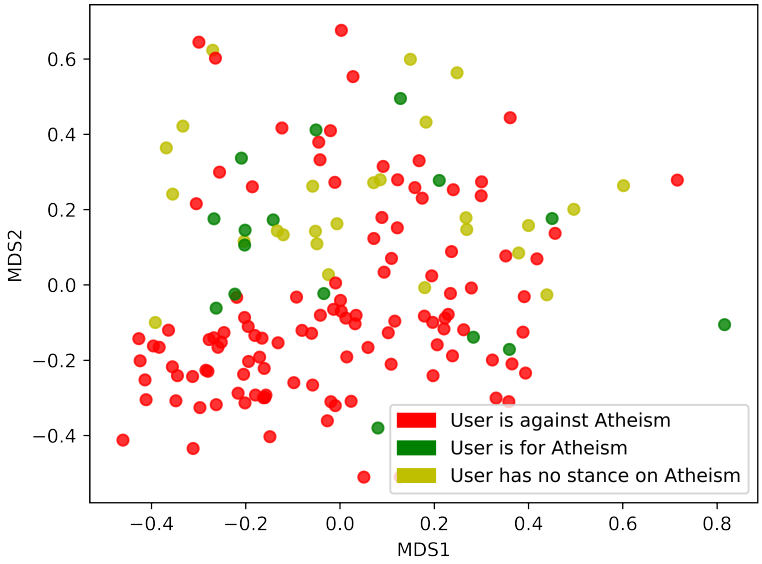

Figure 5: Atheism users, 2D embedding from multidimensional scaling. Here, one can see separation between the red, green, and yellow users, both enabling the visualization of community structure and suggesting that the topic probability vectors hold information on stances.

My current approach centers around users, but ultimately, it should focus on the structure of conversations and the complex network of interactions among users. Various social media platforms provide rich structured data for this task. Designing appropriate feature spaces for representing the stances and flow of influence is an interesting challenge. Lastly, one needs to build recommendation systems based on this geometric representation and get user feedback in order to validate this whole enterprise.

This approach, therefore, brings a new perspective to task of stance detection and topic modeling, furthering progress in Natural Language Processing and in Machine Learning. It has the potential to radically alter how we gain information and interact with each other on social media. Ultimately, changing the collective dynamics of knowledge acquisition could transform the nature of political participation.

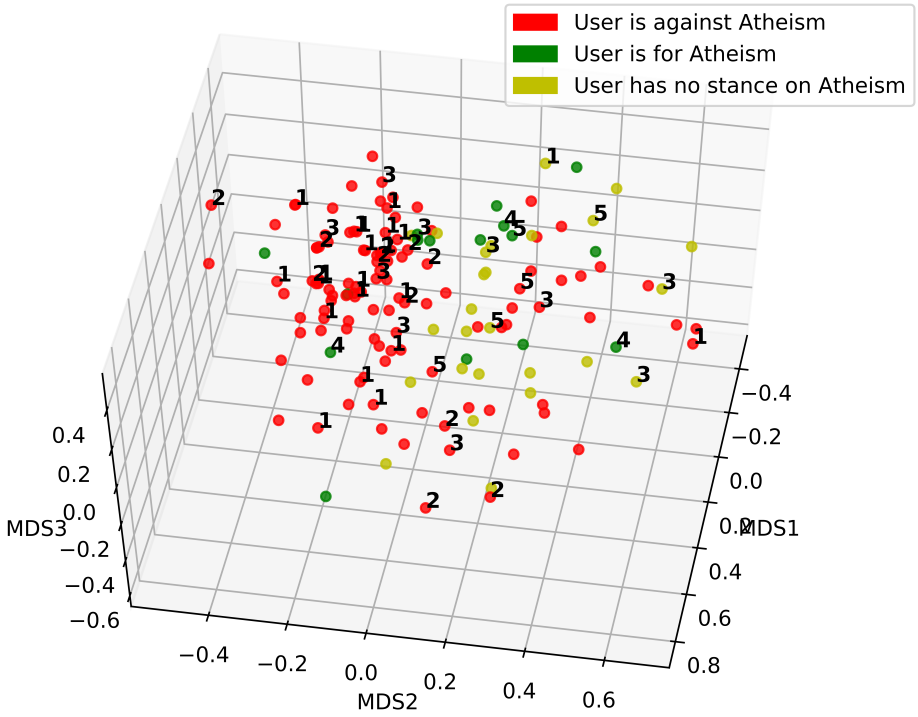

Figure 6: Labeled manual score for data points of users conversing on 'Atheism' plotted on embedding generated by 3D multidimensional scaling. This figure suggests that the topic probability representations of users also holds information on the polarization of the user as there is a 'spatial' structure to the assigned labels within this figure.

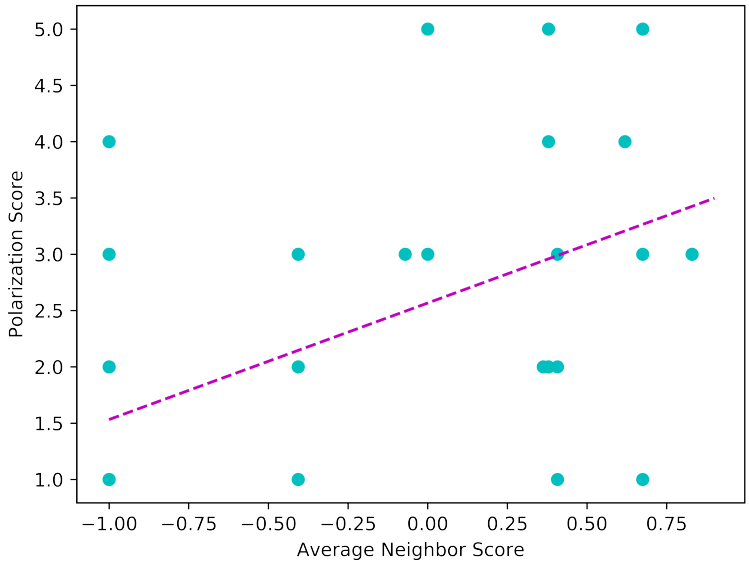

Figure 7: Manual labeled score vs. average neighbor score for 'Atheism' users, correlation coefficient=.544, p-value=1.19×10^{-5}. The average neighbor score holds statistically significant information on the polarity of the user. The dashed line is the linear regression fit.

GitHub login information is provided under the entry for Competition Entrant

References

[1] Shraey Bhatia, Jey Han Lau, and Timothy Baldwin. An automatic approach for document-level topic model evaluation. *arXiv preprint arXiv:1706.05140*, 2017.

[2] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[3] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[4] Jonathan Chang and David M Blei. Relational topic models for document networks. In *International conference on artificial intelligence and statistics*, pages 81–88, 2009.

[5] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[6] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391, 1990.

[7] Competition Entrant. VisTA: Visualization by Topic Allocation, 2017. GitHub repository, Username=Outis2017, Password=Find_VisTA_Code2017, published at https://github.com/Outis2017/VisTA.git.

[8] Thomas L Griffiths and Mark Steyvers. A probabilistic approach to semantic representation. In *Proceedings of the Cognitive Science Society*, volume 24, 2002.

[9] Kazi Saidul Hasan and Vincent Ng. Stance classification of ideological debates: Data, models, features, and constraints. In *IJCNLP*, pages 1348–1356, 2013.

[10] Matthew Hoffman, Francis R Bach, and David M Blei. Online learning for latent dirichlet allocation. In *Advances in neural information processing systems*, pages 856–864, 2010.

[11] Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57. ACM, 1999.

[12] Elıas Jónsson and Jake Stolee. An evaluation of topic modelling techniques for twitter.

2016.

[13] Thomas K Landauer and Susan T Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211, 1997.

[14] Thomas K Landauer, Peter W Foltz, and Darrell Laham. An introduction to latent semantic analysis. *Discourse processes*, 25(2-3):259–284, 1998.

[15] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.

[16] Christopher D Manning, Hinrich Schütze, et al. *Foundations of statistical natural language processing*, volume 999. MIT Press, 1999.

[17] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiao-Dan Zhu, and Colin Cherry. A dataset for detecting stance in tweets. In *LREC*, 2016.

[18] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiao-Dan Zhu, and Colin Cherry. Semeval-2016 task 6: Detecting stance in tweets. In *SemEval@ NAACL-HLT*, pages 31–41, 2016.

[19] Saif M Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3):26, 2017.

[20] Akiko Murakami and Rudy Raymond. Support or oppose?: classifying positions in online debates from reply activities and opinion expressions. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 869–875. Association for Computational Linguistics, 2010.

[21] Romain Paulus and Jeffrey Pennington. Original Ruby code for preprocessing tweets found at http://nlp.stanford.edu/projects/glove/preprocess-twitter.rb.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[23] Jeffrey Pennington, Richard Socher, and Christopher Manning, 2014. Precomputed GloVe vectors could be downloaded from http://nlp.stanford.edu/projects/glove/.

[24] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[25] Walter Quattrociocchi, Antonio Scala, and Cass R Sunstein. Echo chambers on facebook. 2016.

[26] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA.

[27] Gerard Salton, Anita Wong, and Chung-Shu Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, 1975.

[28] Swapna Somasundaran and Janyce Wiebe. Recognizing stances in online debates. 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 1-Volume 1*, pages 226–234. Association for Computational Linguistics, 2009.

[29] Dhanya Sridhar, Lise Getoor, and Marilyn Walker. Collective stance classification of posts in online debate forums. In *Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media*, pages 109–117, 2014.

[30] Vivek Kumar Rangarajan Sridhar. Unsupervised topic modeling for short texts using distributed representations of words. In *VS@ HLT-NAACL*, pages 192–200, 2015.

[31] Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research*, 3(Dec):583–617, 2002.

[32] Songbo Tan. Neighbor-weighted k-nearest neighbor for unbalanced text corpus. *Expert Systems with Applications*, 28(4):667–671, 2005.

[33] Matt Thomas, Bo Pang, and Lillian Lee. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 327–335. Association for Computational Linguistics, 2006.

[34] Marilyn A Walker, Pranav Anand, Robert Abbott, and Ricky Grant. Stance classification using dialogic properties of persuasion. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 592–596. Association for Computational Linguistics, 2012.

[35] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1445–1456. ACM, 2013.