Topic Classification of Controversial Subjects across Political Parties' Subreddits

Cristian Mitroi, Aske Holgaard Bentzen, Janus Larsen

May 23, 2019

Abstract

The current work examines the distribution of controversial topics across the two main parties in U.S. politics. We scrape the respective communities on Reddit and then employ topic modelling, word embeddings, and literature research to establish a list of prevalent topics. We then scrape the respective communities dedicated to these topics, and build an LSTM classifier on this dataset. We then apply this classifier on the submissions in the two political parties' subreddits in order to obtain a distribution of topics. We perform a subjective evaluation of the sentences and how they are labeled, based on the confidence scores of the network. We observe some differences in the distribution of share of topics between the two parties that are in line with the literature.

Contents

1 Introduction

As the availability of text data has been increasing given the expansion of social media websites, more research has been looking into what kind of patterns emerge on these sites. Some of this research is based on topic modeling methods, which finds way to distinguish words and sentences in a matter that matches human interpretation. In this paper we look through some of these methods to investigate how political affiliation might explain a variation between topics of discussion.

1.1 Problem Statement

We seek to answer the following questions:

Firstly, what are the differences in terms of topics distributions across political parties on this site? And secondly, how well can a model learn to distinguish among topics when it is trained on data from the social media site Reddit?

We propose to build a classifier system that can learn from data using a sequential approach. We build a machine learning model that can automatically detect topics within political discourse and thus streamline the process of identifying relevant topics for journalists and politicians.

The first part of the project is choosing a list of topics. We use topic modelling and determine controversial political topics based on former literature. We look for similar patterns in the Republican and Democratic communities on Reddit. This is the first dataset built on Reddit. We employ different topic modelling algorithms (LDA, HDP, LSI) to extract salient topics. We also visualize the topics on pre-trained word embedding models, in order to ascertain how close the topics are to each other, and whether there are any relevant clusters. This will give us a visualization of the methods that we use so we are able to determine whether the words match the topics of interest. We also use former literature to set expectations on our findings in the distribution of topics. Out of these analyses we will select a list of n topics, on which we will build a second dataset to train a n-class classifier.

The second part of the project involves building the classifier. We will use an RNN (LSTM) model trained on the submissions in the topics' respective subreddits. This is the second dataset built by scraping Reddit. This provides us a dataset of sentences labeled with the topics they belong to. We test the performance of the model on a held-out validation set.

Finally, we use the classifier model on the first Reddit data set in order to determine which topics are more controversial in each political community. We hypothesize that this approach can reveal or confirm existing research in the field, as outlined in our literature review. The advantage of our approach is that it can be automatized and further improved by extending the training sets to more topics. In this section we also examine the performance of the model in both a quantitative and qualitative assessment, using plots, tables, and examining different examples ourselves.

1.2 Structure

The paper is split into multiple parts. First, in section 2, we discuss the relevant research on the topic and the methods we use. We cover topics related to Natural Language Processing (NLP), Topic Modelling, and Deep Learning (Neural Networks). The third section will include our arguments for why and how exactly we are using the given methods. In section 4, we describe how we create the text database based on Reddit posts. In section 5 we describe the process of handling the data with the given methods. The results will be presented and discussed in section 6.

2 Background & State of the Art

2.1 Earlier Work

Research on text data from blog fora and social media websites is popular mostly due to the accessibility of the data. We will present some of the former work, which sparks the motivation for our analysis. This includes work from sources such as Twitter, Reddit and blogs.

2.1.1 Political Labeling

In Demszky et al. (2019), Twitter posts regarding mass shootings have been investigated. Posts have been identified based on whether they comment on a mass shooting within 2 weeks after they have taken place. Then the author of each Twitter post has been identified as either Republican or Democrat leaning based on the ratio of candidates they follow on Twitter from the respective parties. The amount of posts is then compared based on affiliation in order to determine, which events gain most attention in each community. They conclude that their measure of party affiliation serves as explanation of the way that people frame the given events in their posts (Demszky et al., 2019).

Without considering the author, another approach investigates partiality based on the message text on Twitter. The partiality depends on the context the message is presented in and does not take the truth value into account. They highlight that messages with high sentiment do not always imply high partiality even though these might be related (Zafar, Gummadi, & Danescu-Niculescu-Mizil, 2016). The measure of partiality is validated by using human validation consisting of a group of 10 people.

In contrast to the article by Demszky et al. (2019), we do not consider each author's political affiliation. Instead we investigate affiliation based on which subreddit the post has been written in. Just as affiliation is not necessarily based on the ratio of number of political candidates a profile follows, the posts within either the Democratic or Republican subreddit cannot be considered strictly leaning to one side. We presume however most of the posts are written with the goal of reaching individuals with similar interests.

2.1.2 Controversiality

Similar work is done by Balasubramanyan, Cohen, Pierce, and Redlawsk (2012), where they investigate the level of polarization within blog posts. Polarization is here determined by sentiment which is computed by an externally defined lexicon. They then evaluate this level of polarization by using human labelling to verify the sentiment.

Our analysis is not based on text sentiment directly. However, by utilizing Reddit's rating system, we get something that may work as a proxy for attitude towards the topic. This is only used for selecting posts of interest. We use the 'controversiality' of a submission. An item is controversial when it has both a high amount of votes and the number of up-votes and down-votes are as close as possible. This measure of controversiality might be both affected by the affiliation of the author as well as the sentiment and context of the message.

We are interested in controversiality as we see this project to be a tool. We imagine it could be employed by journalists, in order to know which difficult questions to ask at a political press conference, or by a politician, in order to know which topics are delicate and require handling with care.

2.2 Framing

In an article from Tsur, Calacci, and Lazer (2015) they refer to a phenomenon called 'framing', which is defined by presenting a specific topic in the light of a specific context in order to induce a cognitive bias (Tsur et al., 2015). This is also known as changing people's attitude towards a certain subject by relating it to something known to have a positive or negative connotation in the mind of the receiver (Chong & Druckman, 2007).

According to Tsur et al. (2015) individuals or political parties can possess 'ownership' of certain topics. This means that the handling of the topic at hand is strongly associated with the party. For instance environmental issues is mostly associated with the Democratic party (Dunlap, Xiao, & McCright, 2001).

They introduce a third phenomenon called 'agenda setting'. As well as covering framing it also refers to the level of attention a topic gets. This means that agenda setting can be used from a communicator to shift attention from a sensitive issue to another issue that would gain more support from the listeners (Tsur et al., 2015). Distinguishing 'agenda setting' from

’ownership’ is not relevant to our paper, as we can only look at the joint effect.

In Lin, Xing, and Hauptman (2008), they make a visualization colouring words used in online debates based on which political affiliation they have. Here they find that some words are almost only used by one of the two political sides in the debate. Their results support the hypothesis that the choice of words to some degree reflect the affiliation of the author.

As our selected topics are trained on subcommunities (subreddits) and then used to determine topics within the Democratic and Republican subreddits, we need to keep in mind that the authors within these two political fora might frame the topics differently. This can create uncertainty of the predicted topic as the phrasing can vary across the political spectrum.

Considering the possibility of party-determined ownership of political topics we also expect the focus of topics to vary between the two political parties. As authors on the fora possibly have an agenda behind their post, they may choose to highlight certain aspects of the topic while neglecting others. This could happen whether it is intentional or not. This might result in a focus on different topics between the two political fora. Despite the definition of ’framing’ by Tsur et al. (2015), writing a post on Reddit does not necessarily imply the will to induce a cognitive bias. Our methods do not explain the intention behind the given post.

2.3 Political Topics

When looking through the literature that attempts to model political topics, we find that some choose to focus on very specific topics like the Iraq War (Yano, Cohen, & Smith, 2009) and the Sandy Hook school shooting (Zafar et al., 2016), whereas others find more general topics like the environment and the economy (Balasubramanyan et al., 2012; Tsur et al., 2015; Field et al., 2018). As we are interested getting as much data as possible within each category, we choose to focus on broad topics. Based on the former research as well as a list of political topics provided by the Reddit community (Reddit, 2019a) we construct a list (Appendix E) containing the topics we expect to find with various topic modeling approaches.

2.4 Topic Modelling

In this section we we look at topic modelling and the different approaches that we can use in order to identify the topics within the Reddit posts. All the methods are based on unsupervised machine learning, to locate clusters within the collection of text.

2.4.1 Latent Semantic Indexing (LSI)

Latent semantic indexing (LSI) which analyzes via Singular-Value Decomposition (SVD) uses large amounts of text data to construct a "semantic" space wherein terms and documents that are closely associated, are placed close by. Via SVD we can identify highly associated patterns and remove less important influences from the data set. Through this process, we are able to develop and extend the semantic index by looking at associated patterns. The retrieval process can then decompose and reduce texts by looking at documents in close proximity and return only key topics from this (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990).

2.4.2 Latent Dirichlet Allocation (LDA)

LDA is based on the assumption that to compute a document of words it would need to consist of a certain amount of topics. When this prior distribution is determined for each document, one of these topics is drawn. Since every topic consists of a distribution of words, a word can be drawn based on the given topic. The topic and the word drawn can then be used to infer the conditional distribution (posterior). This results in a distribution of topics for each document based on every word in the document (Blei, Ng, & Jordan, 2003).

2.4.3 Hierarchical Dirichlet Process (HDP)

A Dirichlet process (DP) is defined to be the distribution of a random probability such that, for any finite measurable partition, the random vector is distributed as a finite-dimensional Dirichlet distribution (Teh, Jordan, Beal, & Blei, 2005).

The Hierarchical Dirichlet process is a DP, which refers to the non-parametric Bayesian prior, that can be used in the grouped mixture model setting and sorting the clusters hierarchically.

Both LDA and HDP use unsupervised learning, based on a Dirichlet prior to capture the latent structure in the data (i.e. topics), but unlike LDA, HDP does not need a prefixed number of topics due to its hierarchical structure of clusters (Teh et al., 2005).

2.5 Word Embedding

Word embedding is a method in NLP for generating a vector representation for words. This has been proven useful in a variety of tasks, like question answering (Mohtarami et al., 2018) and sentiment analysis (Dos Santos & Gatti, 2014).

One method for producing these is to use a neural network, where the prediction is the word we want to encode, and the input is a sequence of w words around the word (called the window size). This approach is called the Continuous Bag of Words approach (CBOW). We can do the opposite as well, predicting the window of words around a given input word. This approach is called Skip-Gram (Bojanowski, Grave, Joulin, & Mikolov, 2016). By training this neural network on a corpus we can then extract the representations of the words from the hidden layers of the network.

These dense representations are more efficient than the sparse counter approach [1]. They require fewer weights to be trained in the model that will employ them. They can also capture synonyms. For example, 'car' and 'automobile' will have a small Euclidean distance and be represented closely in the vector space. However, in a sparse representation they would each simply be a one-hot vector with a 1 in different positions. Nothing about those positions could inform us about their relation (Dan Jurafsky, 2018).

2.6 Deep Learning

Neural networks are a development of the basic Perceptron algorithm (Rosenblatt, 1957), with several of these linear learners stacked in order to approximate complex, non-linear functions. Neural networks are highly versatile, having been employed with great success across a vast number of fields: image classification, text generation, self-driving cars etc. The versatility is due to its flexibility. The user can easily stack more layers, and add more cells to a layer, in order to achieve more complex architectures that fit their needs. The user also has a choice of different other techniques to employ during learning: non-linear activation functions, dropout, different learning optimizers (Adam, RMSProp) etc. It is beyond the scope of this project to cover all these.

Learning is achieved by using the gradient descent method on the loss function of the network. The loss function is defined as the error rate of the network. This function is chosen depending on the nature of the task. This is usually either a regression or a classification task. In the former, the user will usually employ Root Mean Squared Error (RMSE). For classification the user will usually employ categorical cross entropy.

The learning is then achieved by propagating the derivative of the loss function across the weights of the network. This is referred to as 'backpropagation'. This can be controlled by a learning factor, which either takes larger steps down the derivative's slope with the risk of over-stepping the optimum, or smaller steps, but with a larger training time required to reach the optimum.

2.6.1 Sequential Deep Learning (RNNs, LSTMs)

Neural networks are limited to one instance of an object at a time, meaning they are oblivious to the context of the element in the sequence of elements. Thus, neural nets are not able to learn sequential or temporal patterns, like financial forecasting, symbolic music generation, or process sentences.

In this case we require the use of Recurrent Neural Networks (Rumelhart, Hinton, Williams, et al., 1988). These work by preserving and passing an internal state to the next time step.

[1] In this case a sentence would be represented by a one-hot vector with 1 in the index of the words that are present in that sentence

These feedback loops hold information about the past steps processed by the network. This in turn can affect how the network processes the given time step.

RNNs are flexible learners, being employed in different tasks:

- one-to-many: the input is a single entity, but the output is a sequence. One example could be image captioning, where the input is an image and the output is a sentence;
- many-to-one: the input is a sequence, output is a label. One example of this is topic classification (in which the current work fits).
- many-to-many: the input is a sequence, the output is also a sequence. An example would be machine translation.

Backpropagation in the case of RNNs is called backpropagation through time. It works similarly to backpropagation, with the added dimension of first unrolling the model and thus obtaining multiple hidden layers. The main problem with RNNs is obvious when the gradient of the backpropagation either vanishes (becomes almost 0) or explodes (becomes very large) as it passes through previous time steps. This can happen if the weights shrink or grow exponentially.

To limit this behavior, we can use Long Short-Term Memory Networks (LSTMs). The LSTM can control the flow of the gradient by using three "gates": the input and forget gates control the input and state of the internal state, while the output gate controls the output of the cell. This model has been employed with great success in numerous sequential models: sentiment analysis (Wang, Huang, Zhao, et al., 2016), music generation (Choi, Fazekas, & Sandler, 2016) etc.

3 Methods & Architecture

In this section we discuss the different algorithms, mechanisms, and models that we use in our work. We outline our reasons for their match to our task. We first discuss each of the techniques. Then we note the methodology for the analysis of the results

3.1 Natural Language Processing (NLP)

Since our data source is posts from human commentators on Reddit, we need a process to convert natural written text into a data set that the computer program can use for analysis. Thus our works also falls within the topic of Natural Language Processing (NLP). Specifically, we use the *Spacy*[2] library for tokenization and lemmatization.

[2] https://spacy.io/

Tokenization refers to the process of splitting the data set into tokens, splitting the string data into individual tokens. This means dividing the strings of text from the collection of documents into individual words. This is useful for identifying stop words, digits, and punctuation. This allows the system to focus exclusively on the meaningful words.

Lemmatization is to group together and remove the inflected forms of words for analysis (Collins, 2019). We thus obtain the basic forms of the words. E.g. 'plays', 'playing', 'played' all become 'play'. We thus restrict our vocabulary in order to ease the learning process of the model.

3.2 Topic Modeling

We employ topic modelling in order to extract relevant words surrounding the topics. We use three different algorithms, as provided by the *gensim* library. These are all methods for clustering words together that belong to a certain field.

3.3 Word Embedding

We use word embedding in order to capture the contextual information of each word, with relation to the corpus. In our case, we want to capture the words related to one of the n topics. Word embeddings, as discussed, are trained on the contextual information of the word, within a window in the sentence or corpus. Thus, we obtain a latent vector of length d representing each word (in the case of FastText, the pre-trained model that we are using (Joulin et al., 2016), it is 300). This vector is not human understandable, but it does carry semantic meaning.

We use word embeddings both in the first phase of the project, where we visualize a selection of words associated with topics from the literature. We employ the tSNE dimensionality reduction technique and plot the words on a 2d scatter plot. This allows us to observe how words are closely related and how they might form a topic.

We then also use word embeddings in the second phase of the project, in training the classifier. In this step we represent each of the words in our corpus with the vector provided by the pre-trained FastText model. This allows the model to consider the semantic context of the words when learning to associate them to one of the topics.

3.4 tSNE

We employ tSNE (t-distributed stochastic neighbour embedding) in order to perform a dimensionality reduction of the word embeddings (Maaten & Hinton, 2008). This allows us to plot the words in a 2d visualization, for easy evaluation of the semantic similarities of words.

3.5 Supervised Learning - Topic Classification

As discussed, topic classification is the task of classifying the entities in a dataset into n separate classes. This is a supervised, classification task, as opposed to non-supervised and regression tasks. We use this technique because of the nature of the dataset. In this case we are referring to the second dataset, the one built by scraping the different subreddits corresponding to each of the topics. We want to obtain a model that can process the input and catalogue it as one of the n topics.

3.6 Sequentiality and LSTMs

We use a sequential Recurrent Neural Network (RNN) model, with LSTM cells. The model matches the structure of written language. In our case we are dealing with sequences that sentences made up of words. This means that the model itself needs to be sequential. As discussed in the background section, LSTMs (the more advanced form of RNNs) allow for processing inputs made up of time steps, with added advantage of being able to maintain an internal state through the use of gates.
Sequentiality is essential to modeling the meaning of sentences, as relations to a specific topic are developed across words. LSTM models are very common in NLP tasks, as outlined in the respective background section. In a similar manner, we hope to capture the semantics of a sentence with sequential LSTM models.

3.7 Deep Learning Techniques

In this section we cover some of the techniques we have used to optimize our neural classifier.

Dropout is a method for improving the performance of a deep neural network by randomly disabling cells during training. This prevents the network from relying too heavily on a certain path inside itself, and evens out the contribution of the cells. It helps prevent overfitting to the training data (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014).

We utilize the Adam gradient descent method (Kingma & Ba, 2014) (Ruder, 2016). It considers the gradients at previous time steps for each of the weights. This greatly improves the learning because it maintains an internal state of the steepness of each of the gradients. It is one of the most popular learners, as highlighted by Karpathy (Karpathy, 2017).

We also employ a method for reducing the learning rate during the training process itself, when the model reaches a plateau. This is achieved using the Keras library (Chollet et al., 2015). If a model stagnates it means that it is slowly reaching a minimum but, because of a large step size, it is overshooting it. We thus reduce the step size if we do not notice an increased performance metric for a given number of epochs.

3.8 Methodology for Analysis

In the first phase, we analyze the results from the topic modeling while considering the topics outlined by the literature in the field. We realize this is a subjective endeavour. We then visualize the Word Embeddings representations of a selection of words that we deem related to the topics from the literature. We analyze this visualization, and, together with the list of topics from the literature, decide on a final list of 10 topics for our classifier.

In the second phase, we analyze the results provided by the LSTM topic classifier model. We provide a quantitative evaluation and a subjective evaluation. We first provide a histogram plot of the distribution of topics per each of the political parties subreddits. We then present a table of the model's confidence scores for predicting each of the topics, per political party. In the subjective evaluation we discuss several examples of submissions from the political parties, with relation to how the model has labeled them and its confidence scores.

4 Dataset

4.1 Using Reddit Data

Reddit has gained popularity in the field of research. In psychology subreddits regarding anxiety disorders were examined in order to study the prevalence of anxiety depending on the narratives in the online posts (Shen & Rudzicz, 2017). This was inspired by the use of Twitter data to determine if the author suffered from anxiety (Pedersen, 2015).

Reddit defines itself as a community consisting of thousands of smaller communities (Reddit, 2019c). In their code of conduct, they encourage behavior that is related to that of blog sites (Reddit, 2019b), hence, we deem Reddit posts comparable to blog posts. Though compared to one-author blog posts, multi-author blogs such as Reddit contain much more diversity in language and framing (Yano et al., 2009).

In an article by Yano et al. (2009), blog posts are described as possessing a "spontaneous, reactive, and informal nature" while also being "rich in argumentative, topical and temporal structure" (Yano et al., 2009, 477). Another benefit of using blog posts is that if the original post contains language far too technical for a human or a machine to understand, the comment section can contain information that may work as clarification and explanation (Yano et al., 2009). This can be helpful for unsupervised topic modelling.

As the choice of words can vary between authors and communities some prefer to separate models based on the language style and format (Balasubramanyan et al., 2012). This is not directly related to our methods, but is important to keep in mind when interpreting the results of the models we use.

In the work of Chow, Reddit comments threads are measured for topical divergence (Chow & Hong, 2016). They also employ a LSTM model to classify submissions into topics.

4.2 Scraping Reddit Communities

4.2.1 Dataset 1

We scrape the Reddit communities for Democrats and Republicans [3], using the Python Reddit API [4]. We take the top submissions sorted by controversiality. We obtain approximately 1000 submissions for each subreddit. The Reddit API returns a maximum of 1000 submissions when asked for top posts in a subreddit. In this dataset we do not extract text from image submissions, as we subjectively observe that the submissions provide sufficient text in general.

4.2.2 Dataset 2

We scrape the respective subreddits associated with the list of topics. We merge them into our dataset for our classifier. When scraping the submissions we only consider the text in the title. We also download the images and attempt to extract the text from the images, if any. For OCR we use the Tesseract library [5]. This is required because in some cases the text in the title is too vague without the context of the image. In the case of the *firearms* subreddit we notice a lot of image macros[6].

We observe from Appendix **A** that the war, religion, and immigration subreddits have the fewest users. The largest subreddits are the politics (by a large margin), economics, and environment. It is important to note that despite the difference in numbers our LSTM classification model achieves above 80% accuracy on most of the classes.

4.3 Preprocessing

We preprocess the text data from the communities using the *Spacy* Natural Language Processing library for Python. We remove stop words, punctuation, numbers, and pronouns. We reduce the words to their lemma form.

[3] `https://www.reddit.com/r/democrats` and `https://www.reddit.com/r/republican`

[4] `https://praw.readthedocs.io/en/latest/`

[5] `https://github.com/tesseract-ocr/tesseract`

[6] e.g. `https://www.reddit.com/r/Firearms/comments/aisp0l/lol/`,
`https://www.reddit.com/r/Firearms/comments/65f7gv/yup_sounds_about_right/`,
`https://www.reddit.com/r/Firearms/comments/5mknhz/fair_point/`

5 Experiments

In this section we provide a detailed overview of the specific steps of our project. We discuss the process we have taken to test our problem statement.

5.1 Topic Modelling

We employ several methods to obtain a list of topics for the submissions in the Reddit communities. We experience that the model focuses heavily on the names of politicians, due to their high density throughout the corpus. To alleviate this problem, we attempt to filter these names, along with other words that do not carry any meaningful results.

We use the implementations provided by the *gensim* package (Řehůřek & Sojka, 2010).

5.1.1 LSI

We use Latent Semantic Indexing on the corpus and obtain the list of topics, trimmed to the top 15. This list is found in Appendix **B**. Here we can observe topics related to 'gay', 'gay rights', and 'marriage' in topic 1. We also observe topics related to 'unemployment' (topic three and four), 'taxation' (topic four), and 'war' (topic eight and nine). We notice that the model does not seem to cluster words around a certain topic, with most of these being a combination of words on different subjects.

5.1.2 HDP

We also employ the Hierarchical Dirichlet Process. We obtain a list of top 10 topics provided in Appendix **C**. We observe that this approach also does not seem to yield coherent topics. We notice the presence of 'gay_right' again, as in the LSI results. Topic that relate to the list observed in the literature seem to be: 'sexism', 'tolerant' (if we think of racism), 'islamophobe' and 'pro_life'.

5.1.3 LDA

Finally, we employ Latent Dirichlet Allocation and obtain a list of 10 topics provided in Appendix **D**. We observe the following words that seem to be related to the topics identified in the literature: 'foreign' (if we relate it to immigration), 'wall_street', 'trade', 'gun_control', 'corrupt' and 'war'.

Overall we see that the topic modelling methods employed do not produce consistent and coherent topics. Thus, we need to employ further methods to obtain a final list of topics for our classifier.

5.2 Word Embeddings of Topics

We extract a list of words that could be considered topics (or relevant to a given topic) and perform a 2d t-SNE dimensionality reduction of the words embeddings. We then visualize these, along with a subset of the corpus. We use the Facebook FastText model pre-trained on Wikipedia for our word embeddings (Joulin et al., 2016). We believe this approach can highlight the connections and overlap between topics. See Figure 1 for an overview of some of the relevant areas of this visualization. We notice that it does reveal an overlap between some of the topics.

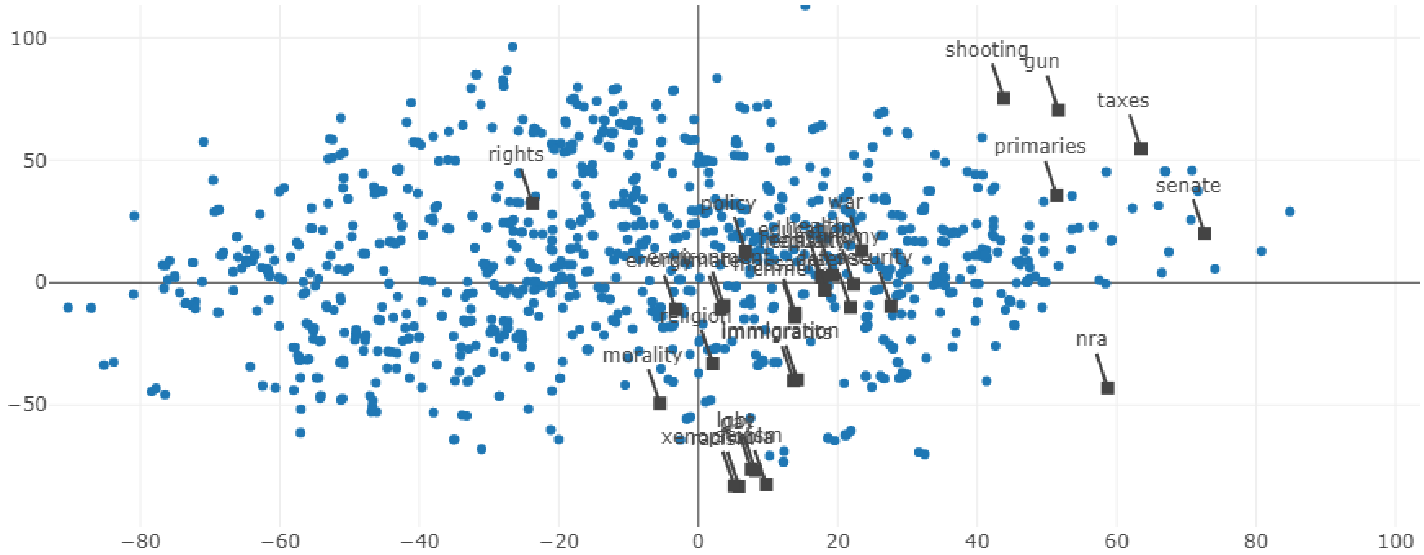

Figure 1: tSNE plot of topics along with corpus, based on FastText

In the area in Figure 2 we can see the connection between 'LGBT', 'sexism', 'racism' and 'xenophobia'. These could all be catalogued as different forms of 'discrimination', 'civil rights' and 'hate speech'.

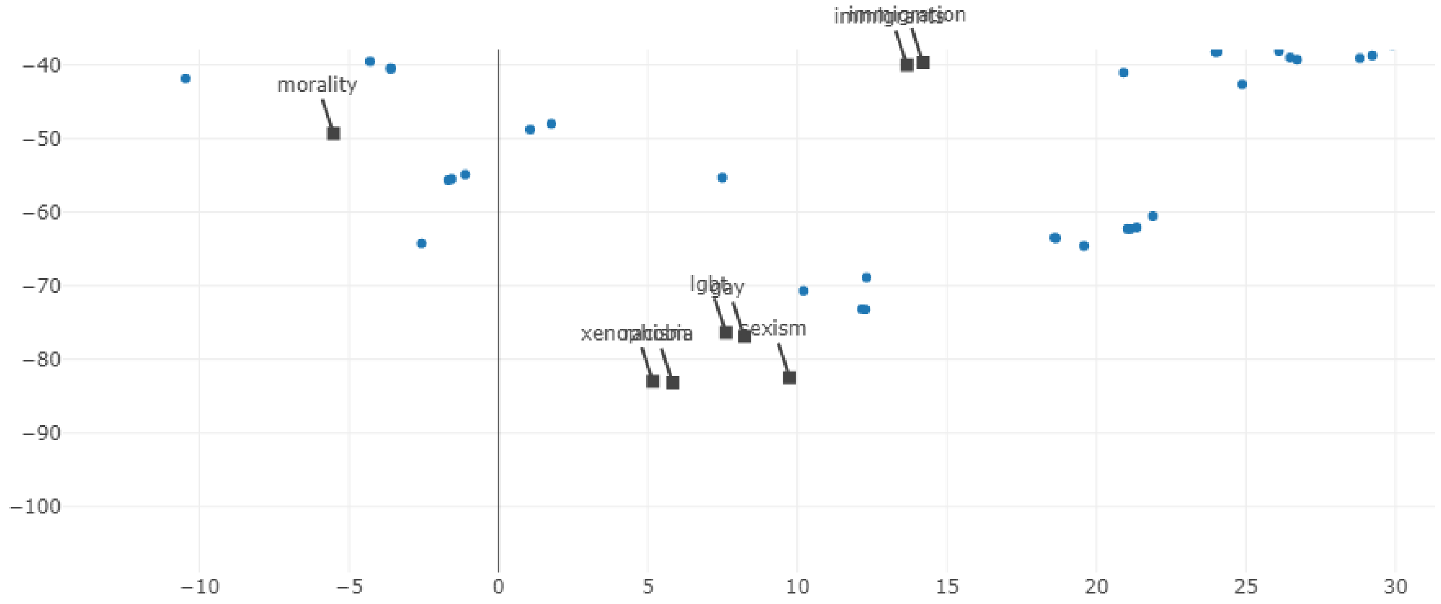

Figure 2: ’Southern’ area in tSNE plot

In Figure 3 we can see a connection between ’economy’, ’inequality’, ’health’ and ’education’. There are also more obvious connections highlighted here, like the one between ’environment’ and ’climate’.

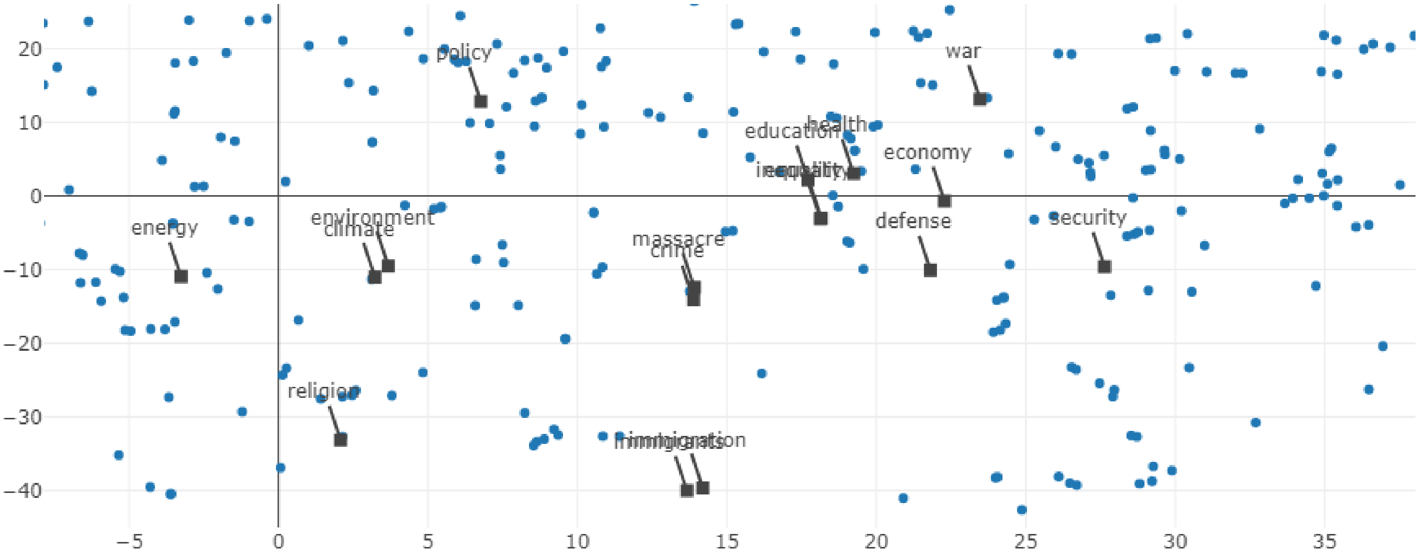

Figure 3: 'South-eastern' area in tSNE plot

5.3 Final List of Topics and Dataset

Based on these discoveries and the list previously mentioned (see Appendix E) we settle for the following list of topics. Another factor that we considered here was the size of the respective subreddits, in terms of users. They should be large enough to contain meaningful submissions.

In order to increase the consistency of each of the topics we also choose to merge different related subreddits into one.

- environment - *https://www.reddit.com/r/environment*
- economy - We decided to merge economics *https://www.reddit.com/r/economics* and business *https://www.reddit.com/r/business*
- religion - *https://www.reddit.com/r/religion*
- education - *https://www.reddit.com/r/education*
- immigration - *https://www.reddit.com/r/immigration*

- lgbt - merging the lgbt subreddit *https://www.reddit.com/r/lgbt* with the gay subreddit *https://www.reddit.com/r/gay*
- war - *https://www.reddit.com/r/war*
- feminism - *https://www.reddit.com/r/feminism*
- guns - *https://www.reddit.com/r/firearms*
- politics - *https://www.reddit.com/r/politics*. This can be considered as a *neutral* choice. We choose this because we observe that most submissions on the two subreddits for the parties seem to be related strictly to politics, e.g. elections, debates around candidates etc.

We scrape these and form the second dataset. For all the subreddits we cannot say for certain if all the text is related to its community name. Even democratic-leaning people can join a discussion on the Republican forum and vice versa. However, as the community names guide people towards the topics they want to read and write about, the categories should work well enough for labeling the political posts.

5.4 Classifier

Finally, we build the LSTM topic classification model and train it on the previously mentioned Reddit dataset. We preprocess the dataset in a similar way as to when we employed topic modelling, but without manually removing any words, apart from stop words, punctuation, pronouns, and digits. We limit the sentences to a maximum of 40 words.

The architecture of the model itself is as follows:

- input layer of 40 words, each word represented by a 300-length vector;
- two LSTM layers, with 32 cells and 40% dropout rate each
- dense layer, with 100 cells, 40% dropout rate, and relu activation
- output layer, with 10 cells, and softmax activation

We employ the Adam gradient descent optimizer and choose an initial learning rate of 0.005. We also employ a method for reducing learning rate when the accuracy stagnates. We use a batch size of 32 and train for 40 epochs. The train/validation ratio is 0.8/0.2. We do not extract a separate test set, as that would decrease the number of training samples, and thus we would have a weaker model overall.

The model obtains an f1 score of 0.89 on the validation data.

Below in figure 4 we observe the training history of the model. We can see that the model reaches a plateau almost immediately, at about 5-10 epochs. Since the training accuracy keeps increasing and the validation accuracy does not decrease, we choose to let the model train for the remaining epochs, until 40, anyway.

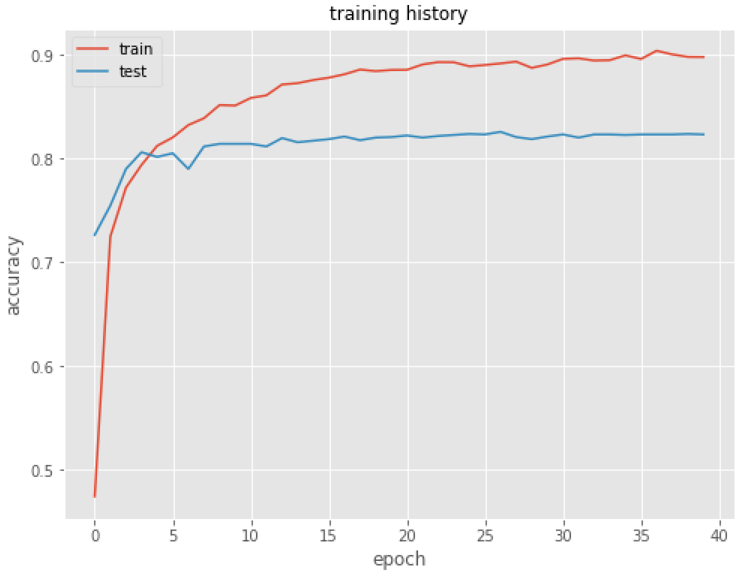

Figure 4: Training history of classification model

In figure 5 we observe the confusion matrix on the validation set:

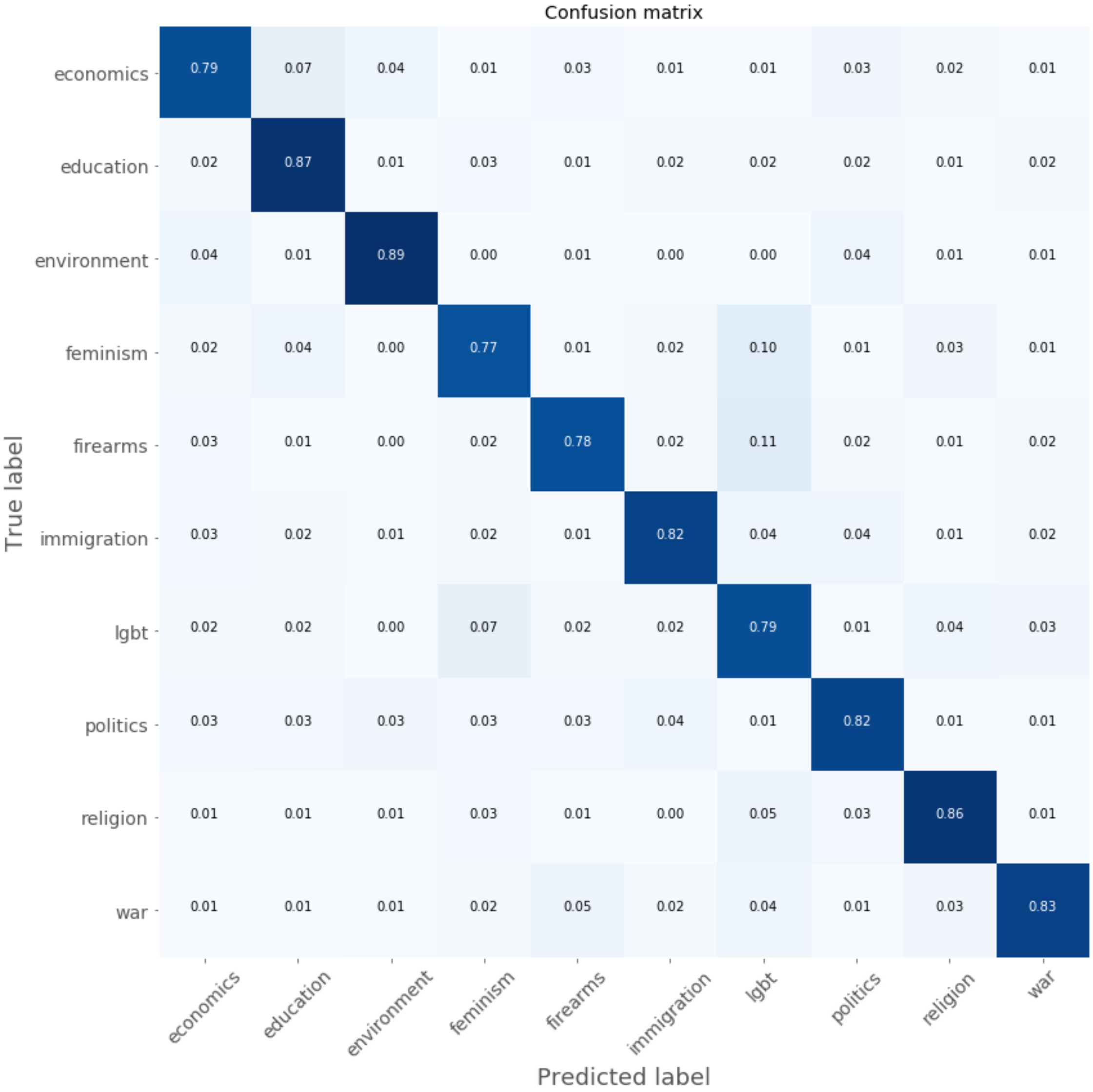

Figure 5: Confusion matrix of classification model

The confusion matrix gives us insight on what topics tend to be confused with each other. We see that for the topic 'politics' it has among the lowest confidence in predicting this topic, despite the prevalence of it in the histograms above. However the confidence is somewhat evenly distributed among the other topics. The topic 'feminism' has the lowest confidence of all the topics, along with 'firearms'. This is primarily driven by a risk of classifying this text as 'lgbt' related. This makes sense, as both are related to issues based on gender and equality. However, 'firearms' also seems to be confused with 'lgbt'. This does not have a clear explanation as the relation between 'lgbt' and 'feminism'. We will investigate this in the following section, when we apply the model to the two party-affiliated subreddits. We do this in order to evaluate whether the misunderstandings are due to the sentences being ambiguous or due to the quality of the model.

6 Results & Discussion

In this section we analyze the results from our model. We provide a subjective discussion of the results and examples, and also include figures, plots, and different relevant statistics.

6.1 Subjective evaluation

Below we see histograms for each topic for the Democratic and Republican subreddits respectively:

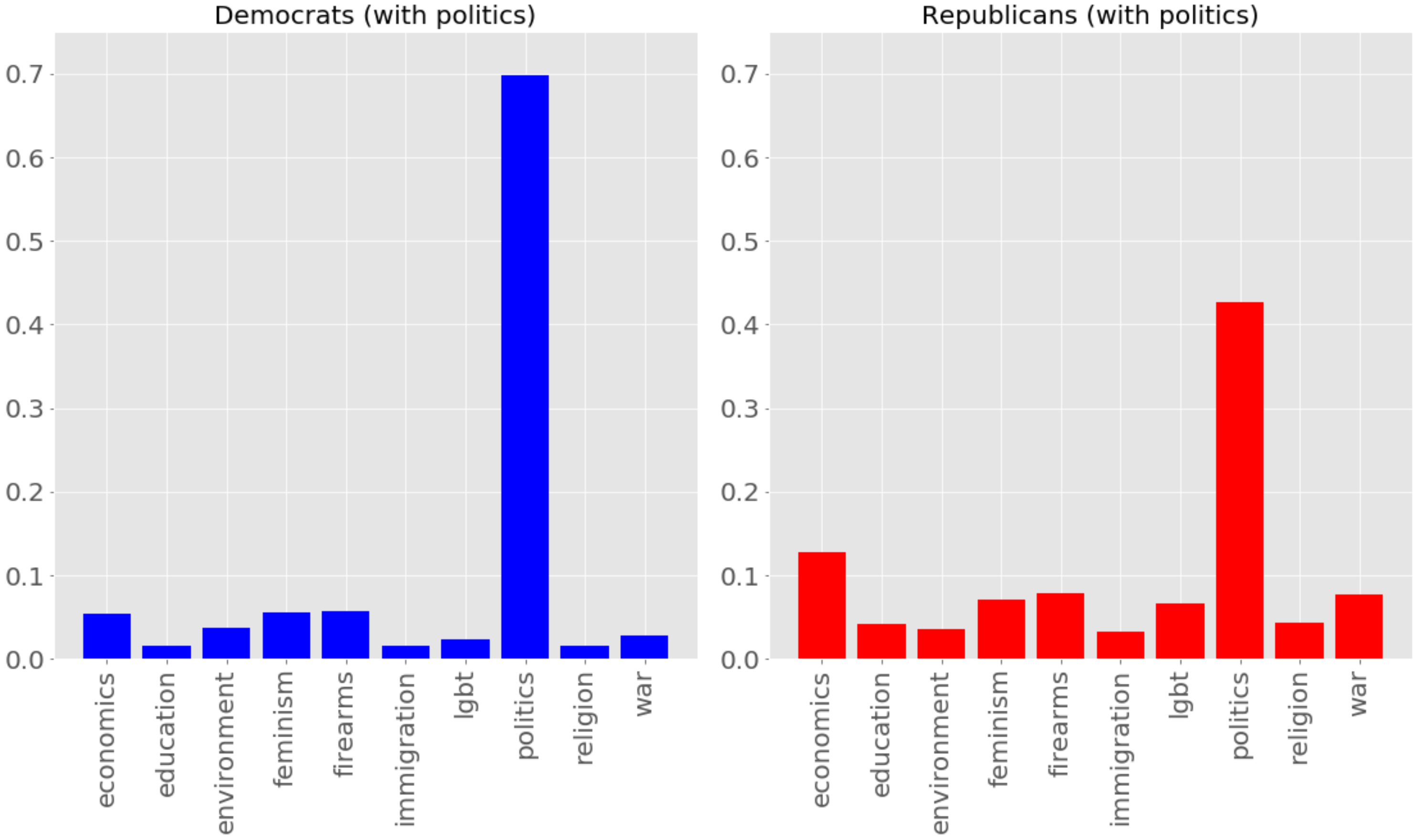

Figure 6: Topic distribution per party affiliated subreddit

From Figure 6 we see that the topic with the highest share is 'politics'. This may be due to the overlap between this topic and most of the other topics. This adds unnecessary noise to the entire distribution of topics, which we want to look into. Despite the noise, it still brings us some valuable insight into the two political subreddits. The Democratic subreddit has a higher share of topics classified as ’politics’ compared to the Republican subreddit. The simple conclusion here would be that the Democratic subreddit is used to discuss politics related to candidates and primary elections more than the Republican subreddit. However, when considering how earlier work has determined that language changes based on affiliation, these results may indicate that the overlap of people posting on both the ’politics’ and ’democrats’ subreddits is higher than those who post on ’democrats’ and the rest of the topics.

In order to make a clearer comparative analysis, we choose to remove sentences associated with the ’politics’ topic in the following histograms:

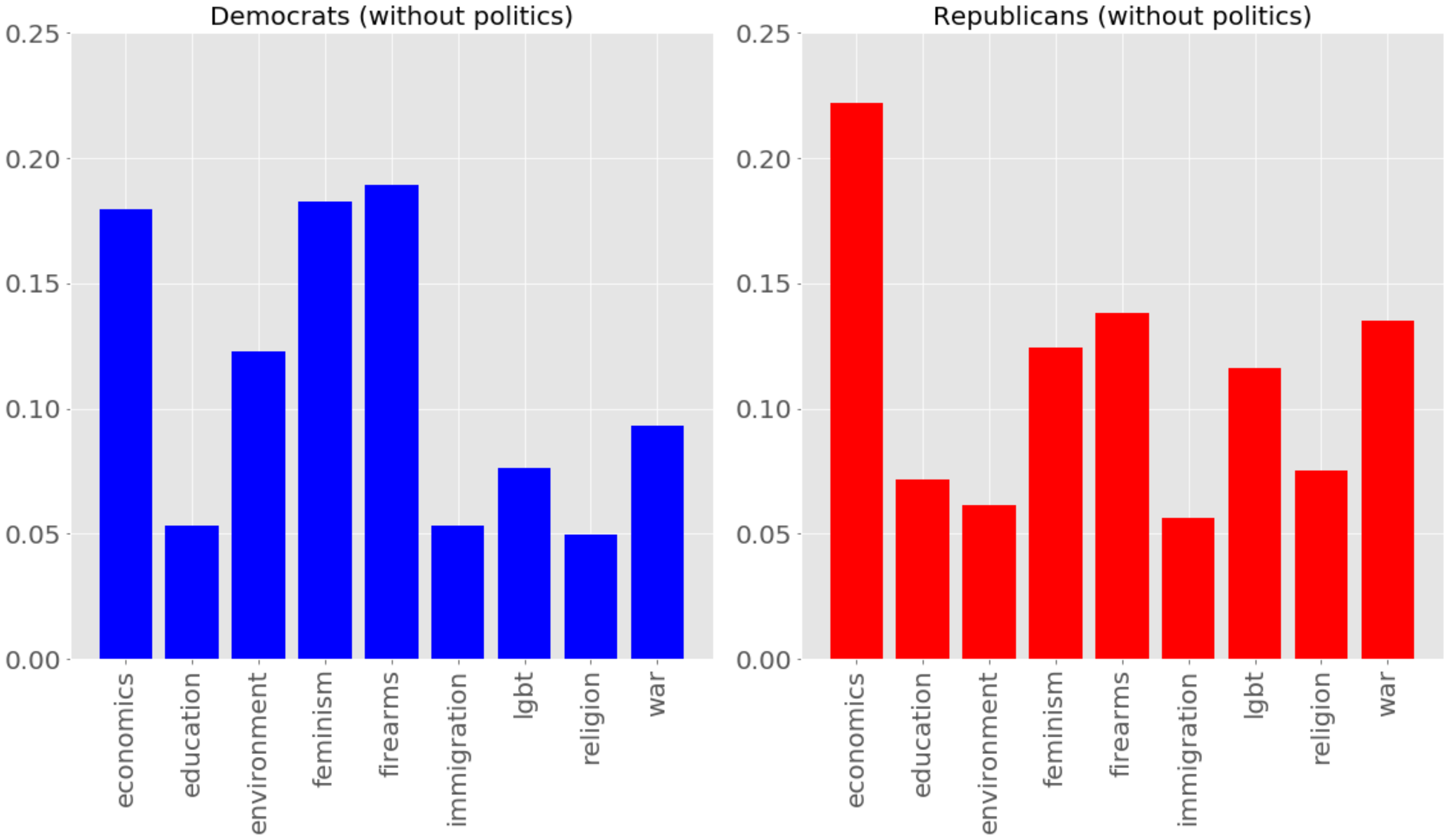

Figure 7: Topic distribution per party affiliated subreddit, excluding ’politics’

Now we see a clear difference in share for some of the topics. Besides ’politics’, Democrats also have a larger share of topics regarding ’environment’, ’feminism’ and ’firearms’. Republicans on the other hand have a higher share of topics like ’economics’, ’lgbt’, ’religion’ and ’war’. The topic ’environment’ seems to have a large share of discussions in the Democratic subreddit compared to the Republican, which was expected from the literature.

As we do not know how accurate the model is in predicting the topic, we provide some statistics on what topics the model has the highest average confidence score. The confidence score is the probability that a sentence is classified as a given topic computed by the *softmax* function of the output layer. We compute the confidence scores on each sentence's classified topic and average over the topics for each party-affiliated subreddit:

Table 1: Average confidence score of model per topic per political community

	republican	**republican counts**	**democrat**	**democrat counts**
economics	0.783	130	0.742	54
education	0.786	42	0.870	16
environment	0.856	36	0.854	37
feminism	0.640	73	0.778	55
firearms	0.668	81	0.678	57
immigration	0.745	33	0.727	16
lgbt	0.522	68	0.603	23
religion	0.748	44	0.755	15
war	0.699	79	0.703	28

From Table 1 we see that the democratic topics have the highest average confidence score in six out of the nine topics. This could indicate a similar pattern as with the 'politics' topic, that the language used in the Democratic subreddit is more similar to that used in these topics, than the language used in the Republican subreddit.

It is worth noting that the topic with the lowest average score is that of 'lgbt'. We also see that for the topics 'lgbt' and 'firearms' both parties have a very low average confidence score. As we do not consider the topics to be closely related, given that our tSNE plots did not put these close to each other, we want to evaluate this qualitatively in the next section.

6.1.1 Firearms and LGBT Confusion

Table 2: Sample of predictions with lowest confidence scores classified as 'lgbt' and 'firearms'

Text	Party	Predicted	Confidence
How Bernie's Brovolution Rigs Its Numbers \n	democrat	lgbt	0.181
		firearms	0.169
Where Are They Now? \n	republican	lgbt	0.173
		immigration	0.141
I found this amusing. \n	republican	firearms	0.141
		lgbt	0.136
October 2, 2012\n	republican	firearms	0.174
		lgbt	0.151

As can be seen in Table 2 the low score texts from the dataset of republicans-democrats classified as 'lgbt' and 'firearms' contain very few words, without much information. They are written as if they refer to something else containing the relevant information for the post, possibly an attached image or a news item that the submission links to.

In the discussion about the confusion matrix of the model, we observed the tendency to confuse 'firearms' for 'lgbt'. We look at examples of sentences from the dataset, where 'firearms' subreddits were mislabeled as 'lgbt'. We notice that most of these do not have words directly associated with guns or firearms in them. It is most likely these were just posts of images of guns. The model might have confused it with 'lgbt' topic due to the adjectives and terms used as epithets for the guns: 'proud', 'beautiful' or 'cute'; or terms associated with family: 'childhood', 'dad', 'friends' or 'family'. It is also possible that both of these subreddits have a large amount of short titles followed by photos (without easily identifiable text).

Table 3: 'Firearms' submissions mislabeled as 'lgbt'

text
yippee ki yay mother fuckers!
Reddit Mods hate this image.
Proud and Beautiful
It's so cute . I want one for the novelty
A childhood dream of mine has been realized. I am now the proud owner
Dad's 50th Bday Present
Whether we are celebrating Christmas with friends and family, or depressingly alone, let us not forget the real reason for the seaaon
Excitement got the best of me! Here she is!!!! July of 1944!!!!!Pe ee ee ee ee
I think we all know what that would be used for if it were in our homes.

6.1.2 General Evaluation

In Table 4 we look at other sentences with low confidence scores. We have chosen six examples that illustrate the challenges that our model faces. The examples are from different topics, but they all have relatively low confidence scores

Table 4: Sample of labeled sentences with low confidence scores

Text	Party	Predicted	Conf.
What Martin Luther King, Jr. can teach us about the Panama Papers: With the corruption of the global elite laid bare for all to see, King's Marxist doctrine feels as urgent as ever \n	democrat	religion	0.336
		economics	.304
Sorry Folks, Independence Day Is Canceled \n	republican	economics	0.181
		war	0.161
Looking for moderator volunteers to help clean up this subreddit. Post a comment in this thread to apply \n	republican	environment	0.420
		religion	0.253
CONSERVATISM IS CALLING - YouTube \n	republican	feminism	0.193
		politics	0.190
The Only Presidential Candidate that Makes Indian Country a Priority:Native People Feeling the Bern \n	democrat	immigration	0.344
		politics	0.299
President Bush's Scottish terrier dog Barney dies at age 12 \n	republican	war	0.280
		religion	0.236

If we look at the first and the fifth text, we get an indication, that part of the classification of these posts is based on word association. 'Martin Luther' is a prominent name in Protestantism and 'Indian Country' and 'People' could be associated with India and migration of workers. However, we see that the model's second guess is very likely to be the accurate classification i.e. 'Panama papers' and 'corruption' is associated with 'economics' and 'Presidential Candidate' with 'politics'.

Other posts are either related to pictures or not specifically covered by any of the chosen topics, and would therefore be hard to classify correctly, which again would give them a low confidence score. The second, third and last post are examples of this.

And finally, there are some sentences that are clearly misclassified, which is illustrated by the fourth post. To us it is highly political due to the word 'conservatism'. Why it is classified as 'feminism' might be explained by the type of posts that are found within the subreddit.

In posts with high confidence scores, we would expect classification to be correct (e.g. post about war is classified as 'war'). And even though many of the samples that we look at are classified correctly, we also find some posts with high confidence scores that are not.

Table 5: Sample of labeled sentences with high, but misleading confidence scores

Text	Party	Predicted	Confidence
The EU, after a 3 year study by 21 scientists, says water isn't healthy and cannot be used to protect against dehydration \n	republican	environment	0.990
		politics	0.007
Wisconsin: Left's war on democracy \n	republican	war	0.998
		religion	0.001
Bernie's Army of Coders - Inside the DIY volunteer tech movement helping drive the insurgent campaign \n	democrat	war	0.987
		politics	0.007

In Table 5 we have found three examples where the model did not classify accurately, but still had high confidence score. As we look closer at the posts, we can guess why the model misclassified some of the posts. Buzz words like 'water' and 'protect' are similar to words that could be found in 'environment', but the actual post would be more accurately placed in the category of 'politics' or 'health' if we have chosen to include such a topic.

In the second and third post in Table 5, we notice the word 'war', which could explain why the model so confidently classifies the posts as 'war', whereas the sentences probably should have been classified as 'politics', as they refer to a party, a campaign and a political figure.

Still, as we manually go through some of the posts with high confidence, we find that most are classified correctly.

6.2 Coherence Analysis

We use a coherence analysis model to obtain an objective assessment of the LSTM model, in comparison with the topic modeling systems. To do this we employ the coherence model evaluation from (Röder, Both, & Hinneburg, 2015), as implemented by the *gensim*. This method works by segmenting the words associated with a topic into different clusters, as done by LDA, HDP, LSI, and our LSTM, then measuring the probabilities of the vocabularies in those clusters, and finally outputting a metric. The implementation expects the outputs from one of the three models. In order to fit our LSTM model to this, we extract the words within each of the 10 predicted classes.

In figure 8 we observe that the HDP model achieves a higher coherence than the LSTM model. It is difficult to ascertain why. From the subjective analysis we have done of the topics in appendix **C**, the HDP model does not seem to produce meaningful results. We do observe that the LSTM model is in second place, and with higher a coherence value.

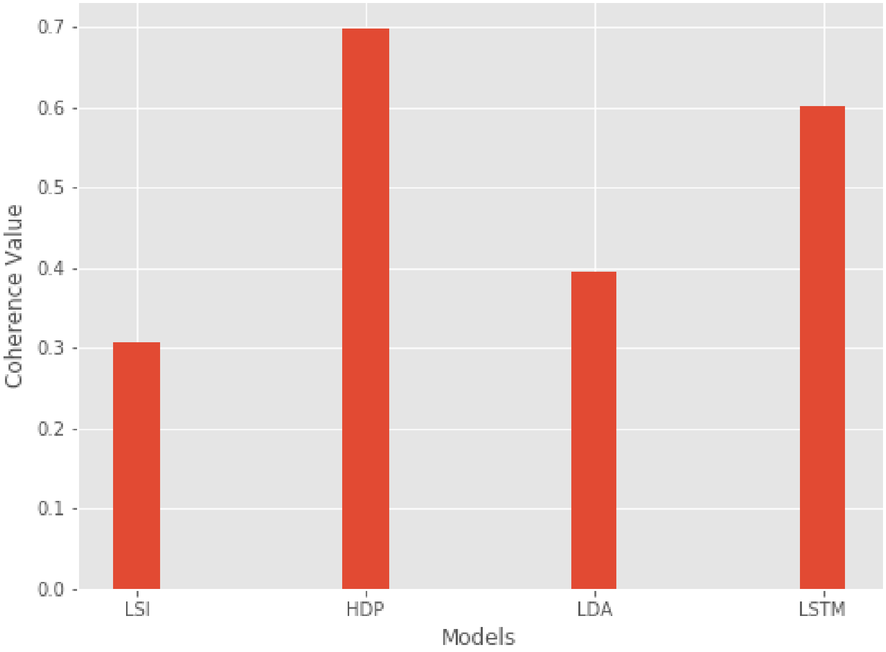

Figure 8: Coherence of models

7 Conclusion

The difference in distribution of topics between the two political fora indicate that some topics do gain more attention based on the political beliefs. Some differences might be caused by the choice of words, depending on the affiliation, but as we were able to produce results that align with former literature, the results do not seem to be badly categorized based on politically affected language.

As Reddit provides labels for discussion by the nature of its structure, using this site as a tool for classification provides quality results. However, our model's shortcomings are highlighted with posts that are lacking in information, due to a short title. Some of the results might have been misclassified only due to the lack of topics within our data.

Some subreddits might be dominated by either left- or right-leaning users, which may cause the language to be easily recognized when matching it with the party-associated subreddit. This may cause additional inaccuracies in the distribution of topics due to the artificially increased confidence scores. We found some evidence that the choice of words affects classification, since topics such as 'lgbt' consistently were confused with 'firearms' due to both forums expressing a high amount of words related to 'love'.

7.1 Further research

Since the use of a Dirichlet prior has been shown to introduce bias in the results, Gerlach, Peixoto, and Altmann (2018) has introduced an alternative called Hierarchical Stochastic Block Modelling (hSBM), which can be used for topic modeling. The method is based on a fully Bayesian framework (Gerlach et al., 2018).

Further improvements to the model could be made by parsing the text from the links to any news items. This could increase the average sentence length and avoid any confusion due to insufficient information.

We could also further expand on the list of topics, either by including more topics (e.g. health) or by splitting topics into subtopics (economics could be split into business, taxation, etc.). This would allow the model to better handle special combinations of topics.

Appendices

A User size of Reddit communities

We observe the number of members of each of the subreddits covered. This informs us about the significance of each of the communities. Larger subreddits means more relevant statistics about controversiality, as we have a larger sample of the real world distribution.

In the first dataset we have $93.8k$ members on the democrat subreddit and $82.1k$ members on the republican subreddit. These numbers are very close, so we can say that the submissions on these communities are approximately equally significant in terms of representing each of the political spectrums.

The second dataset we have multiple subreddits. They are as follows:

`https://www.reddit.com/r/environment`	561k
`https://www.reddit.com/r/economics`	623k
`https://www.reddit.com/r/business`	480k
`https://www.reddit.com/r/religion`	37.8k
`https://www.reddit.com/r/education`	83.6k
`https://www.reddit.com/r/immigration`	15.0k
`https://www.reddit.com/r/lgbt`	318k
`https://www.reddit.com/r/gay`	112k
`https://www.reddit.com/r/war`	12.9k
`https://www.reddit.com/r/feminism`	136k
`https://www.reddit.com/r/firearms`	106k
`https://www.reddit.com/r/politics`	5.1m

B LSI topics

```
[(0,
  '0.269*"people"␣+␣0.218*"need"␣+␣0.188*"support"␣+␣0.136*"right"
     ␣+␣0.123*"good"␣+␣0.121*"new"␣+␣0.115*"election"␣+␣0.115*"
     state"␣+␣0.114*"high"␣+␣0.113*"change"'),
 (1,
  '0.336*"people"␣+␣0.178*"kid"␣+␣0.162*"gay"␣+␣0.159*"
     sex_marriage"␣+␣0.141*"property"␣+␣0.138*"movement"␣+␣0.136*"
     gay_right"␣+␣0.136*"gay_sex"␣+␣-0.133*"israel"␣+␣-0.130*"need
     "'),
 (2,
  '-0.613*"state"␣+␣-0.171*"split"␣+␣-0.146*"unemployment_rate"␣+␣
     -0.130*"national"␣+␣-0.128*"unemployment"␣+␣0.127*"people"␣+␣
     -0.119*"average"␣+␣-0.102*"large"␣+␣-0.097*"unicameral"␣+␣
     -0.095*"tax"'),
 (3,
  '0.542*"state"␣+␣0.171*"split"␣+␣0.145*"unemployment_rate"␣+␣
     -0.134*"money"␣+␣-0.129*"year"␣+␣-0.121*"supporter"␣+␣
     -0.121*"favorability"␣+␣-0.121*"general_election"␣+␣0.121*"
     unemployment"␣+␣-0.113*"tax"'),
 (4,
  '-0.220*"government"␣+␣-0.202*"right"␣+␣-0.200*"tax"␣+␣-0.154*"
     spouse"␣+␣-0.147*"federal"␣+␣-0.140*"work"␣+␣-0.114*"post"␣+␣
     0.111*"general_election"␣+␣0.111*"favorability"␣+␣-0.110*"
     country"'),
 (5,
  '0.334*"right"␣+␣0.307*"spouse"␣+␣0.197*"support"␣+␣-0.158*"
     people"␣+␣0.157*"federal"␣+␣0.155*"deny"␣+␣0.154*"homosexual"
     ␣+␣0.154*"property"␣+␣0.154*"joint"␣+␣0.124*"tax"'),
 (6,
  '-0.238*"talk"␣+␣-0.194*"fact"␣+␣-0.176*"feel"␣+␣-0.159*"medium"
     ␣+␣-0.145*"subject"␣+␣-0.142*"find"␣+␣-0.137*"topic"␣+␣
     -0.126*"outside"␣+␣-0.124*"conversation"␣+␣-0.124*"reality"')
     ,
 (7,
  '-0.234*"win"␣+␣0.194*"tax"␣+␣-0.178*"point"␣+␣-0.151*"turn"␣+␣
     -0.148*"election"␣+␣-0.144*"right"␣+␣-0.134*"gop"␣+␣-0.114*"
     war"␣+␣-0.110*"work"␣+␣-0.110*"way"'),
 (8,
  '0.282*"war"␣+␣0.235*"fight"␣+␣0.152*"resign"␣+␣0.135*"way"␣+␣
     -0.134*"win"␣+␣0.133*"force"␣+␣-0.126*"people"␣+␣0.112*"tea"␣
     +␣0.111*"email"␣+␣0.107*"picture"'),
 (9,
  '−0.237*"win"␣+␣−0.192*"supporter"␣+␣0.163*"war"␣+␣0.150*"
     investment"␣+␣0.146*"wealthy"␣+␣0.141*"feel"␣+␣−0.135*"time"␣
     +␣0.133*"thank"␣+␣0.132*"lot"␣+␣0.131*"low"')]
```

C HDP topics

```
[(0,
  '0.001*sandy␣+␣0.001*negate␣+␣0.001*shove␣+␣0.001*share␣+␣0.001*
    testimony␣+␣0.001*need␣+␣0.001*man␣+␣0.001*upvot␣+␣0.001*
    slice␣+␣0.001*west␣+␣0.001*running␣+␣0.001*grandchild␣+␣
    0.001*worship␣+␣0.001*respect␣+␣0.001*quinnipiac␣+␣0.001*
    adviser␣+␣0.001*past␣+␣0.001*reference␣+␣0.001*win␣+␣0.001*
    fil'),
 (1,
  '0.002*people␣+␣0.001*forfend␣+␣0.001*wambach␣+␣0.001*gay_right␣
    +␣0.001*tucker␣+␣0.001*reagan␣+␣0.001*save␣+␣0.001*shapiro␣+␣
    0.001*ban␣+␣0.001*regardless␣+␣0.001*fdr␣+␣0.001*cousin␣+␣
    0.001*native␣+␣0.001*institution␣+␣0.001*post␣+␣0.001*
    eachother␣+␣0.001*flavor␣+␣0.001*sexism␣+␣0.001*relationship␣
    +␣0.001*rule'),
 (2,
  '0.002*upset␣+␣0.002*shop␣+␣0.002*begin␣+␣0.001*andrew␣+␣0.001*
    bernard␣+␣0.001*jeanne␣+␣0.001*few␣+␣0.001*art␣+␣0.001*thin␣+
    ␣0.001*disability␣+␣0.001*lapd␣+␣0.001*explain␣+␣0.001*
    journalist␣+␣0.001*paycheck␣+␣0.001*bialek␣+␣0.001*stick␣+␣
    0.001*away␣+␣0.001*hobble␣+␣0.001*absence␣+␣0.001*shell'),
 (3,
  '0.002*missing␣+␣0.002*community␣+␣0.001*supporter␣+␣0.001*
    promotion␣+␣0.001*upvot␣+␣0.001*enlightenment␣+␣0.001*
    nationally␣+␣0.001*overt␣+␣0.001*scared␣+␣0.001*compensation␣
    +␣0.001*weekend␣+␣0.001*necessary␣+␣0.001*pakistani␣+␣0.001*
    sector␣+␣0.001*smithereen␣+␣0.001*denial␣+␣0.001*shelve␣+␣
    0.001*gentleman␣+␣0.001*wishy␣+␣0.001*brother'),
 (4,
  '0.003*state␣+␣0.002*reckless␣+␣0.002*whoop␣+␣0.001*aoc␣+␣0.001*
    degree␣+␣0.001*liberty␣+␣0.001*wish␣+␣0.001*aldous␣+␣0.001*
    quality␣+␣0.001*texan␣+␣0.001*orc␣+␣0.001*demand␣+␣0.001*
    money␣+␣0.001*huckabee␣+␣0.001*bigot␣+␣0.001*hostile␣+␣0.001*
    cheap␣+␣0.001*lady␣+␣0.001*horrible␣+␣0.001*false'),
 (5,
  '0.002*entitled␣+␣0.002*buster␣+␣0.002*warp␣+␣0.002*restart␣+␣
    0.001*bereavement␣+␣0.001*commitment␣+␣0.001*massachusett␣+␣
    0.001*colleague␣+␣0.001*rake␣+␣0.001*tipping␣+␣0.001*
    stalinist␣+␣0.001*daniel␣+␣0.001*varied␣+␣0.001*korea␣+␣
    0.001*custodial␣+␣0.001*sue␣+␣0.001*chairman␣+␣0.001*
    extremist␣+␣0.001*enlightenment␣+␣0.001*retreat'),
 (6,
  '0.002*tolerant␣+␣0.001*medal␣+␣0.001*twin␣+␣0.001*tax␣+␣0.001*
    painful␣+␣0.001*powerful␣+␣0.001*popular␣+␣0.001*whoopi␣+␣
    0.001*stasi␣+␣0.001*paycheck␣+␣0.001*relevant␣+␣0.001*divide␣
    +␣0.001*multifacet␣+␣0.001*debt␣+␣0.001*flip␣+␣0.001*prepare␣
    +␣0.001*islamophobe␣+␣0.001*overconfidence␣+␣0.001*pro_life␣+
    ␣0.001*plea'),
 (7,
  '0.002*shall␣+␣0.002*indoctrinate␣+␣0.002*jury␣+␣0.001*care␣+␣
    0.001*discuss␣+␣0.001*detrimental␣+␣0.001*stormy␣+␣0.001*
    voting␣+␣0.001*massive␣+␣0.001*gesture␣+␣0.001*embrace␣+␣
    0.001*flexibility␣+␣0.001*language␣+␣0.001*unfavorable␣+␣
    0.001*jon␣+␣0.001*unamerican␣+␣0.001*year␣+␣0.001*lavish␣+␣
    0.001*counter␣+␣0.001*deal'),
 (8,
  '0.003*fume␣+␣0.002*wellstone␣+␣0.001*moment␣+␣0.001*betray␣+␣
    0.001*universal␣+␣0.001*alienate␣+␣0.001*bachmann␣+␣0.001*
    tracking␣+␣0.001*grumpy␣+␣0.001*sink␣+␣0.001*ideological␣+␣
    0.001*unamerican␣+␣0.001*forget␣+␣0.001*november␣+␣0.001*jong
    ␣+␣0.001*august␣+␣0.001*lessig␣+␣0.001*push␣+␣0.001*killer␣+␣
    0.001*monolithic'),
 (9,
  '0.002*government␣+␣0.002*ruin␣+␣0.002*wrong␣+␣0.002*ban␣+␣
    0.001*msm␣+␣0.001*regard␣+␣0.001*backstory␣+␣0.001*poor␣+␣
    0.001*shitstorm␣+␣0.001*welp␣+␣0.001*firewall␣+␣0.001*food␣+␣
    0.001*mrw␣+␣0.001*thief␣+␣0.001*manhattan␣+␣0.001*gingrich␣+␣
    0.001*major␣+␣0.001*demographic␣+␣0.001*matter␣+␣0.001*dolore
    ')]
```

D LDA topics

```
[(0,
  '0.025*"support"␣+␣0.023*"defeat"␣+␣0.022*"predict"␣+␣0.022*"
    whoop"␣+␣0.022*"foreign"␣+␣0.022*"tactic"␣+␣0.022*"hopeful"␣+
    ␣0.019*"opponent"␣+␣0.018*"belief"␣+␣0.017*"lead"'),
 (1,
  '0.028*"fund"␣+␣0.021*"dollar"␣+␣0.020*"ideal"␣+␣0.020*"sach"␣+␣
    0.020*"hedge"␣+␣0.020*"man"␣+␣0.019*"dnc"␣+␣0.019*"ban"␣+␣
    0.018*"conservative"␣+␣0.018*"rich"'),
 (2,
  '0.020*"key"␣+␣0.016*"know"␣+␣0.012*"need"␣+␣0.011*"win"␣+␣
    0.011*"run"␣+␣0.011*"left"␣+␣0.011*"national"␣+␣0.010*"
    wall_street"␣+␣0.010*"play"␣+␣0.010*"chair"'),
 (3,
  '0.024*"point"␣+␣0.022*"half"␣+␣0.022*"awkward"␣+␣0.022*"breach"
    ␣+␣0.022*"increasingly"␣+␣0.022*"frontrunner"␣+␣0.022*"
    husband"␣+␣0.022*"prove"␣+␣0.020*"few"␣+␣0.015*"finally"'),
 (4,
  '0.029*"trade"␣+␣0.022*"fire"␣+␣0.017*"support"␣+␣0.015*"plan"␣+
    ␣0.015*"deal"␣+␣0.015*"release"␣+␣0.015*"goal"␣+␣0.015*"day"␣
    +␣0.015*"policy"␣+␣0.015*"cost"'),
 (5,
  '0.057*"wall_street"␣+␣0.019*"stand"␣+␣0.019*"run"␣+␣0.019*"
    stupid"␣+␣0.019*"joe_biden"␣+␣0.019*"officially"␣+␣0.019*"
    hall"␣+␣0.019*"town"␣+␣0.019*"announce"␣+␣0.019*"
    new_hampshire"'),
 (6,
  '0.027*"plan"␣+␣0.027*"gun_control"␣+␣0.026*"actually"␣+␣0.026*"
    attack"␣+␣0.026*"wall_street"␣+␣0.025*"rahm"␣+␣0.025*"corrupt
    "␣+␣0.025*"chicago"␣+␣0.025*"clear"␣+␣0.019*"catch"'),
 (7,
  '0.033*"public"␣+␣0.032*"back"␣+␣0.029*"union"␣+␣0.025*"employee
    "␣+␣0.020*"win"␣+␣0.017*"need"␣+␣0.017*"comment"␣+␣0.017*"
    endorsement"␣+␣0.016*"fivethirtyeight"␣+␣0.016*"financial"'),
 (8,
  '0.028*"time"␣+␣0.026*"kasich"␣+␣0.026*"endorse"␣+␣0.025*"john"␣
    +␣0.025*"new_york"␣+␣0.025*"wedding"␣+␣0.020*"claim"␣+␣
    0.019*"victory"␣+␣0.016*"election"␣+␣0.005*"win"'),
 (9,
  '0.025*"war"␣+␣0.016*"gun"␣+␣0.016*"ask"␣+␣0.016*"issue"␣+␣
    0.016*"reason"␣+␣0.016*"explain"␣+␣0.016*"wall_street"␣+␣
    0.016*"tie"␣+␣0.015*"stop"␣+␣0.015*"voting"')]
```

E List of topics

Based on literature (Balasubramanyan et al., 2012; Tsur et al., 2015; Yano et al., 2009; Zafar et al., 2016; Field et al., 2018; Reddit, 2019a)

- Energy and environment
- Economy, taxes and social security

- Republican primaries
- Security
- Religion
- Education
- Union rights and women's rights
- Senate procedures
- Mid-term elections
- Health
- Iraq war
- Immigrants
- Government shutdown of 2013
- Sandy Hook school shooting
- Zimmerman trial 2014
- Immigration reform 2014
- Same sex marriage
- Crime
- Business
- Law
- LGBT
- Racism
- Feminism and gender equality
- Human rights
- War
- Conspiracy

F Confident and Reasonable Predictions

Table 6: Selected sample of sentences with high confidence scores that have proper human interpretation

Text	Party	Predicted	Confidence
Treasury report says tax-cut plan will more than pay for itself, add $300 billion in revenue \n	republican	economics	0.998
		politics	0.001
If the government schools were really about education rather than indoctrination, today's required reading would be quite different: Required reading today and yesteryear. \n	republican	education	0.998
		firearms	0.001
Greenpeace, Sanders Hold Ground Against Clinton in Fossil Fuel Feud \n	democrat	environment	0.943
		politics	0.033
Unfriendly to women? The assertion is baseless. Having served 19 years in the Senate, and as a lifelong Republican, I have some perspective. The recent debate has focused on a narrow slice of what constitutes women's issues and how gender should direct women's views. \n	republican	feminism	0.967
		religion	0.016
Maryland: Your Governor has BIG plans for Gun Control | Heels and Handguns \n	republican	firearms	0.988
		education	0.003
Illegal Immigrants 'Seek Asylum' at Pelosi's Home: She Has Police Remove Them \n	republican	immigration	0.993
		politics	0.006
Here's What Bernie Sanders Didn't Tell His Crowd About Gay Marriage \n	democrat	lgbt	0.833
		feminism	0.063
GOP lawmakers tell Sessions to probe Clinton-Comey or resign \n	republican	politics	0.991
		religion	0.003
Santorum's sanctimonious hypocrisy knows no limits. He has been using god and religion to hide something he should not have to hide. \n	democrat	religion	0.993
		lgbt	0.003
Aircraft carriers need smaller ships to protect them, lest they be sunk. The military has many more bayonets now than in 1916. Special Forces soldiers on horseback were critical to ousting the Taliban. \n	republican	war	0.986
		politics	0.009

References

Balasubramanyan, R., Cohen, W., Pierce, D., & Redlawsk, D. (2012). Modeling polarizing topics: When do different political communities respond differently to the same news? Retrieved from `https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4525/4962`

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). Enriching word vectors with subword information. *CoRR*, *abs/1607.04606*. Retrieved from `http://arxiv.org/abs/1607.04606`

Choi, K., Fazekas, G., & Sandler, M. (2016). Text-based lstm networks for automatic music composition. *arXiv preprint arXiv:1604.05358*.

Chollet, F., et al. (2015). Keras: Deep learning library for theano and tensorflow. *URL: https://keras. io/k*, *7*(8), T1.

Chong, D., & Druckman, J. N. (2007). Framing theory. *Annu. Rev. Polit. Sci.*, *10*, 103–126.

Chow, A., & Hong, J. (2016). Topical classification and divergence on reddit.

Collins. (2019). *Collins enlish dictionary.* Retrieved 21-05-2019, from `https://www.collinsdictionary.com`

Dan Jurafsky, J. M. (2018). *Speech and language processing: Chapter 6: Vector semantics, part ii. online slides.* Retrieved from `https://web.stanford.edu/~jurafsky/slp3/slides/vector2.pdf`

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, *41*(6), 391–407.

Demszky, D., Garg, N., Voigt, R., Zou, J., Gentzkow, M., Shapiro, J., & Jurafsky, D. (2019). Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. In *Naacl '09 proceedings of human language technologies: The 2009 annual conference of the north american chapter of the association for computational linguistics* (pp. 477–485).

Dos Santos, C., & Gatti, M. (2014). Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of coling 2014, the 25th international conference on computational linguistics: Technical papers* (pp. 69–78).

Dunlap, R. E., Xiao, C., & McCright, A. M. (2001). Politics and environment in america: Partisan and ideological cleavages and public support for environmentalism. *Environmental Politics*, *10:4*, 23–48.

Field, A., Kliger, D., Wintner, S., Pan, J., Jurafsky, D., & Tsvetkov, Y. (2018). Framing and agenda-setting in russian news: a computational analysis of intricate political strategies. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 3570–3580).

Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). *A network approach to topic models.* Retrieved 20-05-2019, from `https://www-ncbi-nlm-nih-gov.ep.fjernadgang.kb.dk/`

pmc/articles/PMC6051742/

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). Fasttext. zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*.

Karpathy, A. (2017). *A peek at trends in machine learning.* Retrieved 29-04-2019, from https://medium.com/@karpathy/a-peek-at-trends-in-machine-learning-ab8a1085a106

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lin, W.-H., Xing, E., & Hauptman, A. (2008). A joint topic and perspective model for ideological discourse. In *Machine learning and knowledge discovery in databases* (Vol. 2, pp. 17–32). Springer.

Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, *9*(Nov), 2579–2605.

Mohtarami, M., Baly, R., Glass, J. R., Nakov, P., Màrquez, L., & Moschitti, A. (2018). Automatic stance detection using end-to-end memory networks. *CoRR*, *abs/1804.07581*. Retrieved from http://arxiv.org/abs/1804.07581

Pedersen, T. (2015). Screening twitter users for depression and ptsd with lexical decision lists. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality* (pp. 46–53).

Reddit. (2019a). Retrieved 20-05-2019, from https://www.reddit.com/r/redditlists/comments/josdr/list_of_political_subreddits/

Reddit. (2019b). *Rediquette.* Retrieved 20-05-2019, from https://www.reddithelp.com/en/categories/reddit-101/reddit-basics/reddiquette

Reddit. (2019c). *What are communities or subreddits?* Retrieved 20-05-2019, from https://www.reddithelp.com/en/categories/reddit-101/communities/what-are-communities-or-subreddits

Řehůřek, R., & Sojka, P. (2010, May 22). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50). Valletta, Malta: ELRA. (http://is.muni.cz/publication/884893/en)

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth acm international conference on web search and data mining* (pp. 399–408).

Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton project para.* Cornell Aeronautical Laboratory.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al. (1988). Learning representations by back-propagating errors. *Cognitive modeling*, *5*(3), 1.

Shen, J. H., & Rudzicz, F. (2017). Detecting anxiety through reddit. In *Proceedings of the fourth workshop on computational linguistics and clinical psychology—from linguistic signal to clinical reality* (pp. 58–65).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2005). *Hierarchical dirichlet processes.* Retrieved 20-05-2019, from `https://people.eecs.berkeley.edu/~jordan/papers/hdp.pdf`

Tsur, O., Calacci, D., & Lazer, D. (2015). A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)* (Vol. 1, pp. 1629–1638).

Wang, Y., Huang, M., Zhao, L., et al. (2016). Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 606–615).

Yano, T., Cohen, W. W., & Smith, N. A. (2009). Predicting response to political blog posts with topic models. In *Proceedings of human language technologies: The 2009 annual conference of the north american chapter of the association for computational linguistics* (pp. 477–485). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from `http://dl.acm.org/citation.cfm?id=1620754.1620824`

Zafar, M. B., Gummadi, K., & Danescu-Niculescu-Mizil, C. (2016). Message impartiality in social media discussions. Retrieved from `https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13154/12766`