Predicting Violent Behavior using Language Agnostic Models

Yingjie Liu[1], Gregory Wert[1], Benjamin Greenawald[1], Mohammad Al Boni[2] and Donald E. Brown[1,2]
[1]*Data Science Institute, University of Virginia, U.S.A.*
[2]*Department of Systems and Information Engineering, University of Virginia, U.S.A.*

Keywords: Text Analysis, Natural Language Processing, Convolutional Neural Networks, Bidirectional Recurrent Neural Networks.

Abstract: Groups advocating violence have caused significant destruction to individuals and societies. To combat this, governmental and non-governmental organizations must quickly identify violent groups and limit their exposure. While some groups are well-known for their violence, smaller, less recognized groups are difficult to classify. However, using texts from these groups, we may be able to identify them. This paper applies text analysis techniques to differentiate violent and non-violent groups using discourses from various value-motivated groups. Significantly, the algorithms are constructed to be language-agnostic. The results show that deep learning models outperform traditional models. Our models achieve high accuracy when fairly trained only on data from other groups. Additionally, the results indicate that the models achieve better performance by removing groups with a large amount of documents that can bias the classification. This study shows promise in using scalable, language-independent techniques to effectively identify violent value-motivated groups.

1 INTRODUCTION

Due to the often vast linguistic and cultural differences, as well as the ever-evolving nature of value-motivated groups, it is challenging for governmental and non-governmental organizations to correctly classify the tendencies of these groups towards violence. As a result, a scalable and language agnostic solution for the detection of violent groups becomes imperative.

Based on the premise that the behavior of value-motivated groups can be inferred from their use of language, researchers in (Venuti et al., 2016) and (Green et al., 2017) developed text-mining algorithms that accurately evaluated important characteristics of language usage by religious and non-religious value-motivated groups. Greenawald et al. used these methods to predict violent groups from English text, and showed that language-dependent bag-of-words models achieved a higher performance than language-independent ones (Greenawald et al., 2018). However, this earlier work relied heavily on the semantics of the English language and the availability of Natural Language Processing (NLP) tools (e.g., stemming, part-of-speech tagging, sentiment analysis). Since value-motivated groups can produce text in many languages including English and some languages might have less developed NLP tools, language-dependent models might perform poorly or be inapplicable for the language of interest. In this work, we test bag-of-words models from (Greenawald et al., 2018) on a language with less mature NLP tools (i.e., Arabic.) The main contributions of this work include: 1) collecting a corpus of Arabic documents from violent and non-violent value-motivated groups[1]; 2) proposing two language independent deep learning models for violence prediction; and 3) comparing the proposed models to bag-of-words models from (Greenawald et al., 2018).

For this study, a **value-motivated group** is a group that operates under a common name, has a primary mission outside of making a profit and has a publicly available statement or set of values that generally reflect a worldview and historical narrative. It should be noted that under this definition, individuals can qualify as value-motivated groups. **Violence** is defined as the intentional use of physical force, threatened or actual, that has a high likelihood of causing human injury or death. A **violent group** is defined as a group whose members perform acts that fall under the above definition of violence, and the group must claim responsibility for that action.

In this study, text related to violent and non-

[1]Code and data can be accessed from: https://github.com/bgreenawald/Capstone

Liu, Y., Wert, G., Greenawald, B., Boni, M. and Brown, D.
Predicting Violent Behavior using Language Agnostic Models.
DOI: 10.5220/0006933701020109
In *Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2018) - Volume 1: KDIR*, pages 102-109
ISBN: 978-989-758-330-8

violent groups were collected from 20 groups with an even split of 10 per category. These groups were selected to contain a multitude of regional and ideological diversity. Types of language dependent and independent models in this study include 1) vector-space models, 2) convolutional neural networks, 3) recurrent neural networks, and 4) ensemble models.

2 RELATED WORK

Prior work in text classification has sought to classify intent and sentiment within language through computational methods. This research has yielded powerful tools and methods for NLP. For instance, tools now exist to classify the intent of a document without the creators explicitly stating its intent (Kröll and Strohmaier, 2009). Researchers have also been able to detect semantic change within publications and have been able to examine which topics tend to have the most change (Boussidan and Ploux, 2011). Venuti et al. (2016) and Green et al. (2017) used text as a medium to analyze ideological behaviors of value-motivated groups. They proposed a set of semantic and performative features to estimate the linguistic rigidity of religious and non-religious groups. They argued that linguistic rigidity can be used to infer the flexibility of groups which would help in policy making (i.e., initiating negotiations). These methods have shown potentials in inferring the purpose of a document.

The prior literature heavily concerns itself with predicting specific instances of violence. This is a problematic approach, however, because many incidences of violence are caused by specific environmental factors and are difficult to predict (Yang et al., 2010). Trying to ascertain violent intent in communications has also struggled. Automated attempts at detecting features such as anger have struggled because of the inability to classify unorthodox expressions of anger such as insults; this negatively affects prediction analysis done on traditional methods such as linguistic inquiry and word count. Topic modeling of violent communications, however, has managed to yield logically identifiable categories which imply violence (Glasgow and Schouten, 2014). Recent work has found some success using diachronic modeling to semi-accurately predict future incidences on violence by groups based of past incidences of violence (Kutuzov et al., 2017). In general, studies over time or at more aggregate levels have shown greater success. For instance, research has shown that longitudinal analyses can be performed on individuals to examine changes in the level of aggressiveness within their texts and thus over time (Hacker et al., 2013). Furthermore, Greenawald et al. (2018) showed that text can be predictive of violent groups. They compared the performance of language dependent and independent bag-of-words models. Their results suggested that language independent models were comparable alternatives although incorporating NLP tools yielded a boost in the performance. However, Greenawald et al. (2018) tested this hypothesis only on English text. In this paper, we analyze the robustness of language dependent models by testing them on Arabic text. Also, we implement deep learning models, which are language-independent in nature, and compare them to bag-of-words approaches.

Text analysis techniques have been extended to examine political discourse. Through techniques such framing analysis, these computational methods have been able to detect distinctions in the discourse of two groups focused on the same issues (Landrum et al., 2016). Other techniques such as latent semantic analysis have also proved useful, as they have been able to examine framing within political discourse (Hacker et al., 2013) Researchers have also been able to detect semantic change within publications and have been able to examine which topics tend to have the most change (Boussidan and Ploux, 2011). These techniques have shown limitations, however. Studies have shown that latent semantic analysis can grasp concepts but has difficulty with nuance; for example, it struggles in distinguishing between the desire to commit an action and the confession towards having committed said action (Cohen et al., 2005).

3 DATA COLLECTION AND PREPROCESSING

Much of today's text is digital, and in order to reflect that, the primary data source for this project is web-based content collected from the internet. If possible, data was collected from the official websites of each of these groups, but in some cases, digital archives of content published by the group were used. The subject and format of the content varied among the different groups and sources. The published content included newsletters, magazines, reports, profiles, speeches, and sermons among other publications.

In this study, discourse related to 10 violent and 10 non-violent groups were collected and labeled as such. For violent groups, documents were collected from international groups including Al-Qaeda in the Islamic Maghreb, Ansar Al-Sharia, Al-Shabaab, and ISIL; domestic insurrection groups including Azawad

Table 1: Text corpus collected from 20 violent and non-violent value-motivated groups.

	Group Type	Group Name	Number of Documents	Number of Words
Violent	International Groups	ISIS	55	676,615
		Ansar Al-Sharia	45	781,268
		Al-Shabaab	28	53,198
		Al-Qaeda in the Islamic Maghreb	6	2,353
	Domestic Insurrection	Hamas	2,181	2,632,273
		Hezbollah	678	433,406
		Houthis	285	147,577
		Syrian Democratic Forces	172	43,656
		Azawad Liberation Movement	6	2,741
	Cross Group	Al-Boraq forum	3,973	1,926,423
	Total		**7,429**	**6,699,510**
Non-Violent	News Organizations	Al Arabiya	3,896	2,465,732
		Al Jazeera	31	34,327
		CNN	24	5,398
	Political Organizations	GA on Islamic Affairs	2,224	1,311,662
		Socialist Union of Popular Forces	312	213,136
		Tunisian General Labor Union	68	26,915
		Movement of Society for Peace	47	14,481
	Islamic scholars	Salman Fahd Al-Ouda	663	538,051
		Rabee Al-Madkhali	134	581,907
		Mohamed Rateb Al-Nabusi	30	86,587
	Total		**7,429**	**5,278,196**

Liberation Movement, Hamas, Hezbollah, Houthis, and Syrian Democratic Forces; and a cross-group forum, Al-Boraq.[2] For non-violent groups, documents were collected from the op-ed sections of news organizations including Al Arabiya, Al Jazeera, and CNN; political organizations including General Assembly on Islamic Affairs, Socialist Union of Popular Forces, Tunisian General Labor Union, and Movement of Society for Peace; and Islamic scholars including Mohamed Rateb Al-Nabusi, Rabee al-Madkhali, and Salman Fahd al-Ouda. The groups were selected to reflect regional diversity with groups spanning across the Middle East and North Africa, as well as ideological diversity with religious, nationalist, economic and political groups. As researchers strove to collect data from an array of ideological backgrounds to reduce or eliminate bias. Thus, to address the bias issue, groups with more nationalist purposes were included along with those with more religious ones. There also was an effort to get groups with similar worldviews across the two classes; for instance, Salafi rhetoric was chosen for both the violent and nonviolent sources. Beyond that there was an effort at obtaining geographic diversity with groups selected from Morocco to Iraq. Figure 1 shows the geographical location of groups or individuals included in our study. Both violent and non-violent groups were obtained from countries such as Syria, Tunisia, Algeria, and Morocco.

In total, around 61,000 documents were collected. However, the vast majority of these came from one group, Al-Boraq, because the source for Al-Boraq documents was a large forum, where each forum post was counted as a document. Naturally, this led to a large number of documents. Upon running preliminary models, it became clear that when Al-Boraq was included in the training set, the model just learned these documents. Thus, Al-Boraq was downsampled to a random sample of approximately 4,000 documents, leaving us with a balanced split of violent and non-violent documents. Table 1 shows the value-motivated groups used in our analyses.

As for data preprocessing, we used two different approaches. For language-independent models, data preprocessing was kept to a minimum. A few basic operations were performed (e.g., removing any non-Arabic characters such as the noise generated from scraping web pages or PDFs). Numbers were replaced with a single token (NUM), and punctuation was removed. No stop words were removed. Note that removing stop words does not necessarily violate our goal of keeping the model language-agnostic. Given a large enough set of documents (which the model needs to work anyway) in most languages, a simple frequency analysis will tell what words show

[2]Al-Boraq [web forum], January 8, 2006 - May 17, 2012. AZSecure-data.org version. Accessed October, 2017. http://azsecure-forums-darkweb/Alboraq.zip

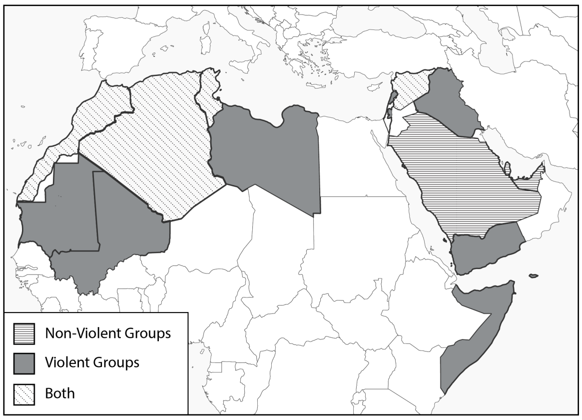

Figure 1: The locations of violent and non-violent value-motivated groups that we collected text documents from.

up most often, so removing stop words is permissible. However, we chose not to remove stop words for language independent models because the word frequency assumption might be invalid from some languages, and we desire our models to work across all languages. Finally, for the vector-space models, the Stanford NLP library was used to tokenize words.

4 MODELING APPROACH

We developed four different types of models: 1) vector-space models, 2) convolutional neural networks, 3) recurrent neural networks, and 4) ensemble models.

Vector-space models, also known as bag-of-words (BOW) models, map text documents into a multi-dimensional vector space such that each dimension represents a different concept (e.g., sports, politics, religion) and the weight for that dimension reflects the extent to which the document cover that concept. We chose to use unigrams and bigrams vector space with term-frequency inverse document frequency (TF-IDF) weighting scheme. We created a controlled vocabulary using two language-agnostic feature selection techniques: Chi-square and information gain (Yang and Pedersen, 1997). We used the intersection of the top 10,000 selected features from both methods as the final controlled vocabulary. Note, both the controlled vocabulary and the inverse document frequency were computed using only training documents. Finally, we represented each document using the controlled vocabulary and trained a binary logistic regression classifier.

We primarily focused on two different deep learning methods. The first methodology, convolutional neural networks (ConvNets), gained prominence in the field of image recognition; however, it translates quite naturally to the field of text classification, and has subsequently shown impressive results (Mikolov et al., 2013). The networks essentially involve sliding a filter across the input to find the important features. One can easily imagine sliding a window across a sentence, capturing consecutive words, and attempting to derive meaning from them. In this way, ConvNets capture the local context of a given word. To represent this idea as a numeric input that could be understood by a neural network, we used word embeddings. At their core, word embeddings map tokens from a vocabulary to a real-valued vector that can subsequently be fed into a neural network. The vectors try to numerically represent the context in which a given word appears and uses that as a proxy for word meaning. We chose a popular implementation developed at Google, word2vec, that uses a shallow neural network to achieve this mapping (Kim, 2014a). There are pre-trained word embeddings available but due to the niche quality of our dataset and the fact that we ultimately would like our pipeline to work on any relevant dataset, we trained the word embeddings on our data. Note, in our embedding encoding, we reserved two vectors to account for padding and unseen tokens. Our ConvNet architecture was heavily based on (Kim, 2014b), and consisted on the following layers: 1) input layer; 2) embedding layer with 8 dimensions; 3) dropout layer with 90% nodes kept; 4) two concurrent convolution layers with 250 filters of sizes of 3 and 4 respectively. Each layer had a stride of 1, and followed by a ReLU activation function (Nair and Hinton, 2010) and a max pooling of size of 2 and stride of 1; 5) the outputs of the two max pooling were concatenated and fed to a fully connected layer with 256 weights; a second dropout layer with 65% nodes kept; and 6) a single output layer with a sigmoid activation function. We trained the network on batches of size of 32, RMSProp optimizer (Tieleman and Hinton, 2012) using binary cross-entropy loss function, and regularized by the two dropout layers and an early stopping (20 training epochs).

Despite the many advantages, ConvNets have a major flaw. Namely, they attempt to learn the importance of local features, not global ones. Humans write in such a way that requires full context. Although the thesis of a document may be expressed in a sentence or two, the full bearing of a document necessitates understanding the document in its entirety. Thus, we require a model that can do the same. For this, we used long term-short memory (LSTM) architectures. LSTM networks are a form of recurrent neural networks, which work by not only using the word embedding for a given word but by also remembering features from earlier in a document, giving more context of a word. We chose to use a bidirectional LSTM model (BLSTM) to prevent biasing words at the end of a document and give words at the beginning and end equal amounts of information. Also, BLSTMs support building language independent models as in some language, authors write from right to left (e.g., Arabic), and therefore, regardless of the direction of the text, BLSTMs would be able to model the dependencies between sequences of tokens. Our BLSTM architecture consists of the following layer: 1) input layer; 2) embedding layer with 8 dimensions; 3) dropout layer with 90% nodes kept; 4) one BLSTM network with 128 output neurons; 5) one dropout layer with 25% nodes kept; and 6) a single output layer with a sigmoid activation function. We trained the network on batches of size of 256, ADAM optimizer (Kingma and Ba, 2014) using binary cross-entropy loss function, and similar to ConvNet, regularized by the dropout layer and an early stopping (20 training epochs).

Since the input layer is followed by an embedding layer in both ConvNet and BLSTM models, we need to fix the length of the input text. A common approach around this is to set the document length to the maximum length and pad shorter documents with a special token. However, if the distribution of document length is skewed to the right (i.e., few long documents and many more shorter ones), then padding to the maximum length would be impractical. To deal with such cases, another option is to set the maximum length to either the mean or median document length. However, documents longer than the fixed threshold will be cut and some informative content will be lost. To address the limitation of both approaches and increase the size of the data, we chose to perform data augmentation. First, we set the maximum length to the median document length, 300 in our dataset, then for each training document, we generated text patches of fixed length but with a random offset. The number of generated patches is given by,

$$B_{train}(d_i) = \left\lfloor \alpha * \frac{|d_i|}{\beta} \right\rfloor \tag{1}$$

where β is the fixed length threshold and α is an augmentation factor. We selected $\beta = 300$ and $\alpha = 2$. For example, if a training document has a length of 650 words, this method would generate 4 random patches with size of 300 words. At testing time, we generated overlapping patches with an offset of $\left\lfloor \frac{\beta}{\alpha} \right\rfloor$, and therefore, the number of generated patches is given by,

$$B_{test}(d_i) = \begin{cases} \left\lfloor \frac{|d_i|}{\frac{\beta}{\alpha}} \right\rfloor = B_{train}(d_i) & \text{if } |d_i| \% \beta = 0 \\ \left\lfloor \frac{|d_i|}{\frac{\beta}{\alpha}} \right\rfloor + 1 = B_{train}(d_i) + 1 & \text{otherwise} \end{cases} \tag{2}$$

For example, for a testing document of length 620 words, $\beta = 300$ and $\alpha = 2$, we generate 5 patches at offsets 0, 150, 300, 450, and 600. Note, the last patch in this case, as well as any training or testing documents of length less than β, will be padded. Next, we averaged the probabilities of all batches for the final output.

Finally, we fused results from LR, ConvNet and BLSTM models using an average model (Avg-EM), where for each testing document, we took the average probabilities from the included models.

5 EXPERIMENTS AND DISCUSSION

We performed empirical evaluations of the proposed language-agnostic models on a large collection of documents. First, we performed an exploratory analysis of our collected corpus. We trained document embeddings of 100 dimensions (Le and Mikolov, 2014) using doc2vec from gensim[3]. Next, we visualized the documents using T-SNE (Maaten and Hinton, 2008). Figure 2 shows Arabic documents from violent (blue) and non-violent (red) value-motivated groups. Although there are no clear and linearly separable clusters, we find that violent documents have a bi-modal distribution intermixed with a uni-modal distribution of non-violent ones. This supports our motivation for predicting violence from text yet highlighting the complexities of doing so. We used two experimental setups: 10-fold cross-validation (CV), and leave-one-group-out cross-validation (LOGO-CV). In both

[3]Gensim doc2vec models. https://radimrehurek.com/gensim/models/doc2vec.html.

Table 2: Unigram and bigram features with the largest shift in logistic regression weights.

Feature	Meaning in English	Classifier Weights W GAIAE	Classifier Weights W/O GAIAE
	organization	-0.2950	0.2050
-	for Islamic affairs	-0.4562	NA
	and endowments	-0.4154	NA
	Caucasus	NA	0.3268
	highness	-0.2560	0.02614

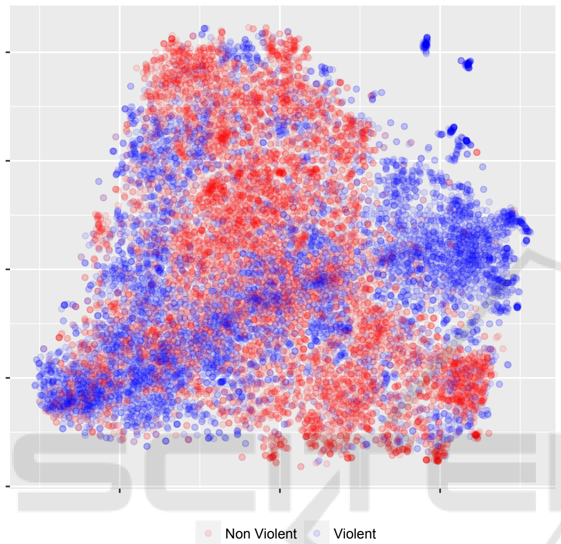

Figure 2: T-SNE visualization of violent (blue) and non-violent (red) documents.

setups, we computed classification accuracy and two F1-measures (positive and negative) at the document level. Using CV, LR and ConvNets classified almost all documents correctly (accuracy of 0.9896 for LR and 0.9882 for ConvNets.) We hypothesized that the reason for the high performance was that documents from a given group were divided into training and testing, and as a result, the models learned features that would distinguish groups and associated that with the violence label. To validate this hypothesis, we excluded training documents from a particular group and retrained the CV LR models. Then, we compared the controlled vocabulary, the classification performance, and the weights of the learned models. For this experiment, we chose the General Authority of Islamic Affairs and Endowments (GAIAE) as a test-bed. When removing documents from this group, the 10-CV accuracy dropped from 0.9960 to 0.1347. After comparing the controlled vocabulary for the different folds, we found that adding documents from GAIAE to training promoted about 2,100 features on average to be included in the top 10,000 controlled vocabulary. The significant change in performance was clearly caused by the big change in the vocabulary. To further explore the type of features that were included, we compared the coefficients of the trained CV models with and without GAIAE's training documents. Table 2 shows features with the biggest change after adding GAIAE's documents from the training. Features newly included in the controlled vocabulary such as "for Islamic affairs" and "and endowments" are clear indicators of the GAIAE, and since all training documents from GAIAE were labeled as non-violent, these features became indicators of non-violence. Other features such as "organization", which were included in the vocabulary in both cases, switched from being violence indicators to non-violence indicators.

Table 3: LOGO-CV performance of language agnostic models.

Model	Accuracy	Positive F1	Negative F1
LR	59.46%	0.9172*	0.4409
ConvNet	71.30%*	0.8633	**0.7135***
BLSTM	**71.99%***	0.9018	0.7032
Avg-EM	69.78%	**0.9227***	0.6384

* p-value< 0.05 with paired t-test compared to remaining models.

It is clear that such superficial features are less meaningful for prediction. Therefore, we ran deep learning models that would capture the context rather than individual key terms. Furthermore, the previous findings suggest that CV setup is not appropriate for this prediction problem. A better setup is the LOGO-CV where we exclude all documents from a given group from the training and test the classifier only on the held-out documents. LOGO-CV reflects the actual use of such models in real-world applications in which we predict the behavior of a new group whose violence is unknown. Table 3 shows the LOGO-CV macro-classification performance. The deep learning models have significantly outperformed BOW approaches. This indicates that the context is very important for predicting behavior from text.

Table 4: Groups with low accuracy.

Model	Group	Is Violent	Accuracy
LR	Mohamed Rateb Al-Nabusi	No	0.4666
	Al Jazeera	No	0.4516
	Socialist Union of Popular Forces	No	0.4006
	Al-Boraq	Yes	0.3838
	Movement of Society for Peace	No	0.3829
	Al Arabiya	No	0.2120
	Rabee Al-Madkhali	No	0.1417
	GA on Islamic Affairs	No	0.1344
	CNN	No	0.1250
	Salman Fahd Al-Ouda	No	0.0437
ConvNet	Al-Boraq	Yes	0.4375
	GA on Islamic Affairs	No	0.4290
	Syrian Democratic Forces	Yes	0.2733
	Rabee Al-Madkhali	No	0.2537
	Alarabiya	No	0.1450
BLSTM	GA on Islamic Affairs	No	0.4245
	Movement of Society for Peace	No	0.3830
	Rabee Al-Madkhali	No	0.1642
	Alarabiya	No	0.1345

Table 5: Comparison of language dependent and independent logistic regression.

Language Dependency	Accuracy	Positive F1	Negative F1
Independent	59.46%	0.9172	0.4409
Dependent	60.58%	0.9163	0.4645*

* p-value< 0.05 with paired t-test.

Also, ensembled models achieved the highest positive F1 score which they produced significantly lower negative F1 scores than ConvNet and BLSTM. This indicates that fusing models work well when classifiers have relatively close performance scores, and they are greatly affected by one weak classifier (e.g., LR on non-violent prediction). We further compared the performance at a group level and showed groups with less than 50% accuracy (See Table 4). It is clear that logistic regression models were biased by the large number of Al-Boraq documents since its documents were included in all models except the one where we evaluated on Al-Boraq. Deep learning models significantly predicted more groups than BOW.

Finally, we wanted to measure the boost in performance after using language dependent models. We measured this on the bag-of-words LR. We applied the same pre-processing steps as in Section 3, but we removed stop words [4], and applied Snowball stemming[5]. Table 5 shows the performance of two logistic regression classifiers with and without language-specific NLP tools. Although the accuracy scores are comparable, we observed a significant boost in the negative F1 score. This supports the findings from (Greenawald et al., 2018). However, even with language-specific information, ConvNet and BLSTM outperformed LR. This would suggest that either the NLP tools for Arabic are of low quality or, and most likely, the context, which unigram and bigram bag-of-words LR models do not capture, is very important for predicting violence from text.

[4]Arabic stop words list, https://github.com/ mohataher/ arabic-stop-words/blob/master/list.txt

[5]Arabic stemmer, http://arabicstemmer.com/

6 CONCLUSIONS

We sought to create a model that could differentiate between documents from violent and non-violent groups in a language-agnostic manner. We tested a variety of models using a leave one group out cross-validation (LOGO-CV). As expected, deep learning models generally outperformed traditional models in this task. Although logistic regression was the top performer in positive F1, these scores were close and the neural networks performed much better in other metrics. Also, incorporating language-specific NLP tools such as stemming improved the performance of bag-of-words logistic regression, yet it failed to outperform deep learning models.

Further, in the LOGO-CV setup, we observed that removing groups with large numbers of documents such as Al-Boraq or Alarabiya significantly boosted the predictive performance of the opposite class. However, assuming that we will not know the class label of the testing group, we cannot determine which groups to exclude from the training. We plan to extend this work to explore different ways to automatically select training data such as selecting the top k similar documents for every testing document or the top k groups with highest in-group similarity variance. We would also like to implement different data-driven ensemble models such as learning a new Logistic regression that take the predicted probabilities of the individual models as predictors.

REFERENCES

Boussidan, A. and Ploux, S. (2011). Using topic salience and connotational drifts to detect candidates to semantic change. In *Proceedings of the Ninth International Conference on Computational Semantics*, pages 315–319. Association for Computational Linguistics.

Cohen, T., Blatter, B., and Patel, V. (2005). Exploring dangerous neighborhoods: latent semantic analysis and computing beyond the bounds of the familiar. In *AMIA Annual Symposium Proceedings*, volume 2005, page 151. American Medical Informatics Association.

Glasgow, K. and Schouten, R. (2014). Assessing violence risk in threatening communications. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 38–45.

Green, S., Stiles, M., Harton, K., Garofalo, S., and Brown, D. E. (2017). Computational analysis of religious and ideological linguistic behavior. In *Systems and Information Engineering Design Symposium (SIEDS), 2017*, pages 359–364. IEEE.

Greenawald, B., Liu, Y., Wert, G., Al Boni, M., and Brown, D. E. (2018). A comparison of language dependent and language independent models for violence prediction. In *Systems and Information Engineering Design Symposium (SIEDS), In Press*. IEEE.

Hacker, K., Boje, D., Nisbett, V., Abdelali, A., and Henry, N. (2013). Interpreting iranian leaders' conflict framing by combining latent semantic analysis and pragmatist storytelling theory. In *Political Communication Division of the National Communication Association annual conference, Washington, DC*.

Kim, Y. (2014a). Convolutional neural networks for sentence classification. *CoRR*, abs/1408.5882.

Kim, Y. (2014b). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

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

Kröll, M. and Strohmaier, M. (2009). Analyzing human intentions in natural language text. In *Proceedings of the fifth international conference on Knowledge capture*, pages 197–198. ACM.

Kutuzov, A., Velldal, E., and Øvrelid, L. (2017). Temporal dynamics of semantic relations in word embeddings: an application to predicting armed conflict participants. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1824–1829.

Landrum, N. E., Tomaka, C., and McCarthy, J. (2016). Analyzing the religious war of words over climate change. *Journal of Macromarketing*, 36(4):471–482.

Le, Q. and Mikolov, T. (2014). Distributed representations of sentences and documents. In *International Conference on Machine Learning*, pages 1188–1196.

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

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

Tieleman, T. and Hinton, G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31.

Venuti, N., Sachtjen, B., McIntyre, H., Mishra, C., Hays, M., and Brown, D. E. (2016). Predicting the tolerance level of religious discourse through computational linguistics. In *Systems and Information Engineering Design Symposium (SIEDS), 2016 IEEE*, pages 309–314. IEEE.

Yang, M., Wong, S. C., and Coid, J. (2010). The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools. *Psychological bulletin*, 136(5):740.

Yang, Y. and Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. In *ICML*, volume 97, pages 412–420.