The Thirty-Third International
FLAIRS Conference (FLAIRS-33)

Narrative Origin Classification of Israeli-Palestinian Conflict Texts

Jason Wei, Eugene Santos Jr.
Thayer School of Engineering
Dartmouth College
Hanover, NH, USA
{jason.20, eugene.santos.jr}@dartmouth.edu

Abstract

The Israeli-Palestinian conflict is one of the most controversial in history. Not surprisingly, historic and contemporary literature on the topic tends to be polarized. Drawing inspiration from work in political ideology and hyperpartisan news detection, we collect two new datasets of history book excerpts and newspaper articles regarding the Israeli-Palestinian conflict and train sequence classifiers to predict whether a text is written by an Israeli or Palestinian source. Moreover, we find that data augmentation techniques improve performance, allowing our best model to detect narrative origin with an F1 score of 85.1% for history book excerpts and 91.9% for newspaper articles. Analysis of indicative phrases discovered by our models corroborate historian insight regarding the conflict.

Introduction

The Israeli-Palestinian conflict has been called "the most intractable conflict of our time" (Munayer and Loden 2014). Since the early 1900s, two peoples—the Jewish Zionists and the Palestinian nationalists—have been fighting over the same territory in the southeastern Mediterranean. The levels of violence, coupled with the religious and cultural significance of the contended territories, has created a deep divide between Israelis and Palestinians (Kressel 1987).

Understandably, historic literature and media coverage of the conflict are often biased (Hollander 2017; Sternthal 2001; Doumato and Starrett 2007). Case examples of text bias on the conflict are shown in Table 1. As a first step to bridging the vast gulf between the two societies, there has been a call for more comprehensive studies of the different viewpoints presented (Caplan 2011; Adwan 2012). A data-driven model for detecting narrative origin could provide insight into this fascinating history by exposing where and how the stories told by these two peoples differ.

In this paper, we use natural language processing (NLP) to analyze the narrative origin of a series of texts written on the Israeli-Palestinian conflict. Our contributions are the following:

Origin	Example
Israeli	The Palestinians carried out acts of terror against Jews in the cities, such as the shooting attack on Jews in Jaffa in April 1936, in which nine Jews were killed and more than fifty wounded.
Palestinian	While the Arabs suffered unemployment and high taxation and lived in stifling economic conditions, the Jews obtained a lot of economic privileges which raised their standard of living in all aspects.

Table 1: Examples of biased sentences regarding the Israeli-Palestinian conflict.

- We collect two datasets of writings on the Israeli-Palestinian conflict from both Israeli and Palestinian authors: one corpus of history texts and one of newspaper articles.
- We train and evaluate several text classifiers on these datasets, using two data augmentation techniques to improve performance. Analysis by a Middle East historian of biased phrases discovered computationally show that they are generally on target.

Data and code is publicly available at https://github.com/jasonwei20/isr-pal.

Related Work

Narrative origin detection aims to detect hidden attributes in language and is thus similar to a number of tasks in stylometric analysis including gender attribution, native-language classification, and political bias detection. In gender attribution, methodology typically aims to identify gender-indicative phrases with techniques such as hand-selected feature engineering (Koppel, Argamon, and Shimoni 2002), part-of-speech (POS) sequence mining (Mukherjee and Liu 2010), and character-level pattern discovery (Sarawgi, Gajulapalli, and Choi 2011), and has been applied to blog posts, scientific papers, and British National Corpus documents. Native language classification similarly identifies indicative key words using approaches such as n-gram based (Tetreault et al. 2012) and stylistic/syntactic feature based classification (Bergsma, Post, and Yarowsky 2012). Finally, political bias detection is the most similar to our classification task, with Iyyer et al. (2014) introducing recursive neural networks for detecting political ideology in books and magazine articles, and Potthast et al. (2018) taking a meta-learning approach to identifying hyperpartisan news.

For the Israeli-Palestinian conflict specifically, two empirical works are relevant. Lin et al. (2006) first introduced perspective classification by applying statistical models to a focused selection of website articles written on `BitterLemons.org` from 2001-2005. Later, Ellen and Parameswaran (2011) classified inferred biases in translated forum posts by extracting hand-engineered term frequency features and applying classical machine learning. While these datasets cover a relatively narrow timeline of events (Lin et al. 2006) or only include polarized narratives (Lin et al. 2006; Ellen and Parameswaran 2011), we collect two new datasets to mitigate these limitations respectively: **(1)** book excerpts that span almost the entire history of the conflict, and **(2)** newspaper articles from various Israeli and Palestinian sources that were not chosen based on detected bias.

Our paper builds on related work in several ways. Foremost, we collect and analyze two new text datasets on the Israeli-Palestinian conflict. Second, while related works used hand-engineered feature selection (Koppel, Argamon, and Shimoni 2002; Ellen and Parameswaran 2011) or bag-of-words representations (Wang and McKeown 2010; Sarawgi, Gajulapalli, and Choi 2011; Feng, Banerjee, and Choi 2012), we use distributed word embeddings in the context of sequence learning, which are more computationally efficient and have achieved superior performance in various NLP tasks (Mikolov et al. 2013; Kim 2014; Glorot, Bordes, and Bengio 2011). Finally, while previous works on the conflict focus on methodology (Lin et al. 2006) and performance (Ellen and Parameswaran 2011), we collaborated with a Middle East historian to analyze indicative phrases discovered by our model, with the aim of helping introduce NLP as a tool of analysis in the field of digital humanities (Tanasescu, Kesarwani, and Inkpen 2018).

Data Collection

Side by Side Dataset (SBS)

Our first dataset is a collection of history texts compiled in the book *Side by Side* (SBS) (Adwan 2012). In this book, historical accounts of major events in the conflict from 1917-2000 are told by an Israeli historian and by a Palestinian historian and placed "side by side" on the left and right pages of the book respectively. To convert this book into text-as-data, a physical copy was scanned and converted to electronic format by a third-party company. Then, we split the text into excerpts of 45 words in order to obtain approximately 1,500 text samples, each labeled with an Israeli or Palestinian origin based on the original classification in the book. In this dataset, topic selection was a control since events were reported on evenly by both sides of the conflict due to the nature of the book. Furthermore, all texts were compiled by the same editor, so stylistic differences were unlikely to be a confounding variable.

Descriptor		SBS	IP-News
Control	Style	Yes	Yes
	Time	Yes	Yes
	Topic	Yes	No
N_{train}	Israeli	425	197
	Palestinian	485	202
	Combined	910	399
N_{val}	Israeli	150	54
	Palestinian	150	55
	Combined	300	109
N_{test}	Israeli	150	108
	Palestinian	150	110
	Combined	300	228
l	Israeli	45	480
	Palestinian	45	755
	Combined	45	637
$\|V\|$	Israeli	5,643	15,788
	Palestinian	5,478	17,907
	Combined	8,059	23,711
w	Israeli	32,630	244,824
	Palestinian	35,203	296,554
	Combined	67,833	541,378

Table 2: Summary statistics for collected datasets of writings from Israeli and Palestinian authors. SBS: history text dataset. IP-News: newspaper article dataset. N: number of samples in the training (N_{train}), validation (N_{val}), and test (N_{test}) set. l: average number of words per sample (excerpt for SBS, article for IP-News). $|V|$: vocabulary size. w: total word count.

Newspaper Article Dataset (IP-News)

We also collected a larger and more colloquial dataset called IP-News, which comprises news articles by various Israeli and Palestinian authors. We searched for English news articles by Israeli and Palestinian writers between 2010 and 2017 containing the key words "Israeli" and "Palestinian." From the Nexis-Lexis database, we found a large number of *The Jerusalem Post* articles; the 248 top articles by search results were downloaded and labeled with a narrative origin of Israeli. We also found 128 articles on Nexis-Lexis from *New York Times* labeled with datelines; 111 articles corresponding to six Israeli cities were labeled Israeli, and 17 articles corresponding to five Palestinian cities were labeled Palestinian. No Palestinian newspapers had substantial English publications in Lexis-Nexis, so we scraped the website of *The Palestinian Chronicle*, a widely-read Palestinian-biased English online newspaper. 350 articles spanning each year from 2010 and 2017, were retrieved to match the class

	SBS Dataset				IP-News Dataset			
Model	Precision	Recall	F1 Score	Aug Gain	Precision	Recall	F1 Score	Aug Gain
LR	62.6	61.3	62.0	-	86.5	58.2	69.6	-
+SR	62.0	77.3	68.8	+6.8	85.1	72.7	78.4	+8.8
+SW	65.9	81.3	72.8	+10.8	72.6	89.1	80.0	+10.4
+SR,SW	74.7	72.7	**73.6**	+11.6	93.6	80.0	**86.3**	+16.7
RNN	77.5	73.3	75.3	-	76.4	88.2	81.9	-
+SR	81.8	78.0	79.9	+4.6	93.5	78.2	85.1	+3.2
+SW	82.4	74.7	78.3	+3.0	80.2	95.5	87.1	+5.2
+SR,SW	86.3	84.0	**85.1**	+9.8	93.9	84.5	**89.0**	+7.1
CNN	81.6	86.0	83.8	-	85.7	87.3	86.5	-
+SR	81.2	89.3	**85.1**	+1.3	94.1	87.3	90.6	+4.1
+SW	83.9	83.3	83.6	-0.2	85.8	93.6	89.6	+3.1
+SR,SW	82.8	86.7	84.7	+0.9	91.1	92.7	**91.9**	+5.4

Table 3: Evaluation (%) of various text classifiers in classifying narrative origin. LR: logistic regression, RNN: recurrent neural network, CNN: convolutional neural network. SR: synonym replacement augmentation, SW: sliding window augmentation. Aug Gain: improvement in F1 score compared to baseline with no data augmentation.

and time distribution of the Israeli articles. To best account for style guidelines differences between newspapers, we removed all metadata (e.g., titles, dates, authors, headers, footers, etc) and other style indicators (punctuation, line breaks, and paragraph breaks). Table 2 shows data splits and descriptive statistics for both datasets.

Experiments

Experimental Setup

We train the following models to classify narrative origin:

1. Logistic regression (LR) is the simplest neural network to satisfies the universal approximation theorem (Cybenko 1989) and serves as a good baseline.
2. Recurrent neural networks (RNNs) are suitable for text classification because of their ability to process sequential data; we implement a three-layer bidirectional LSTM-RNN (Liu, Qiu, and Huang 2016).
3. Convolutional neural networks (CNNs) are used in vision and have also achieved high performance in text classification; we implement “CNN-static” (Kim 2014).

For training and testing, we converted text inputs into numerical representations using 300-dimensional distributed embeddings pre-trained on the Common Crawl database with the GloVe method (Pennington, Socher, and Manning 2014). Because texts from IP-News were variable-length, long articles were split into multiple samples of 50 words each with the last sample zero-padded. All models were optimized over the binary cross-entropy loss function and were trained until the validation loss did not decrease after three epochs (early stopping).

Data Augmentation Techniques

Because our datasets are highly specific and thus relatively small, we implement two data augmentation techniques to alleviate potential overfitting.

- **Synonym replacement (SR).** Synonym replacement for data augmentation has been previously used in NLP with both positive (Wang and Yang 2015) and mixed results (Kolomiyets, Bethard, and Moens 2011; Zhang, Zhao, and LeCun 2015). For both datasets, we generate four augmented samples per training sample by randomly replacing three words (not including stop words) with synonyms identified by WordNet (Miller 1995).
- **Sliding Window (SW).** This augmentation technique is commonly used in data-scarce fields such as biomedical data science (Ortiz Laguna, Olaya, and Borrajo 2011; Um et al. 2017) but is not frequently used in NLP. However, we hypothesize that it would be helpful for our datasets because of its ability to generate a large number of data samples without introducing unnecessary noise or complexity. For both datasets, we concatenated all samples in the training set of the same narrative origin and obtained fixed-length inputs for training by sliding a window of size w=50 across the entire text with stride s=5.

Narrative Origin Detection

Here, we train and evaluate the LR, RNN, and CNN models. Each model was trained and evaluated with no augmentation, SR, SW, and both augmentation methods combined (Table 3).

As expected, both the CNN and RNN outperformed the logistic regression baseline due to their abilities to capture sequential data. The CNN had the strongest performance overall, with F1 scores of 85.1% for the SBS dataset and 91.9% for the IP-News dataset. Regarding data augmentation, both synonym replacement and sliding window augmentation generated substantial improvements in performance, especially for the baseline logistic regression model. Of note, F1-score on the best classifier improved 1.3% for the SBS dataset and 5.4% for the IP-News dataset, indicating that these techniques may be useful and worth exploring in other NLP applications.

Origin	Dataset	Top 1,2,3,4-grams
Israeli	SBS	**yishuvs**, rabbis, synagogues; **idf soldiers**, the yishuvs, yishuv forces; **yom kippur prayers**, jewish yishuv in, the yishuv forces; idf forces idf was, ethics which idf soldiers, the yom kippur prayers
	IP-News	**intifada**, iraqis, palestinians; **terrorist palestinians**, idf palestinians, intifada palestinians; intifada palestinians and, poll israelis palestinians, them palestinian journalists; the intifada palestinians and, should support palestinian journalists, protecting them palestinian journalists
Palestinian	SBS	**ali**, all, **townships**; mohammed ali, al din, **expelled all**; of mohammed ali, izz al din, salah al din; **sheikh izz al din**, al aqsa mosque did, that izz al din
	IP-News	**detentions**, **blockades**, blockaded; detention all, cleansing all, **apartheid by**; apartheid south african, lieberman settler rabbis, blockade imposed by; **israeli military detention centres**, racial apartheid mandela was, military blockade imposed by

Table 4: n-grams discovered by our RNN model to be most indicative of narrative origin.

Bias Analysis by a Middle East Historian

To explore the phrases most indicative of narrative origin, we retrieve all n-grams in the data for $n{=}\{1,2,3,4\}$ and rerun them through the RNN model trained with SR and SW augmentation. Table 4 shows the highest scoring n-grams.

A subjective consultation with a Middle East historian regarding these n-grams not only confirm that classifications are generally on target, but also shed light on terminology and content disparities in retellings of the Israeli-Palestinian conflict. In the SBS excerpts, the Israeli narrative commonly used religiously charged words such as *Yishuv* (Jews in Israel) and *Yom Kippur* (the holiest day of the year in Judaism). On the other hand, texts of Palestinian origin focused on the stories of famous Arabs: *Ali* is a common Arab name, *Ra'fat Ali* was a martyr who was killed in 1976, and *Shiekh 'Izz Al-Din Al-Qassam* was an Arab fighter of Syrian origin. For the IP-News dataset, Israeli journalists often referenced the role of Palestinian terrorists in the Second Intifada (the Palestinian uprising in the early 2000s), and Palestinian journalists frequently wrote about Arab prisoners held in detention camps and compared the segregation policies imposed on them to the *apartheid*. These results corroborate historian insight that the Israeli narrative tends to lay a religious and community-based claim to Israel, portraying the Palestinians as violent invaders, while the Palestinian narrative is often grounded in ethos and emphasizes the sacrifices made and suffering endured by their people.

Conclusions

In closing, we explored narrative origin classification on new Israeli-Palestinian conflict text datasets. Sequence classifiers achieved commendable performance that improved with data augmentation and can potentially aid historians in text analysis. A limitation of our work is that the SBS dataset is small, and the IP-News texts might have confounding variables since they are not aligned by event. Future work includes collecting more data and controlling for events in the IP-News texts through manual annotation, as well as obtaining better interpretations from the classifiers. We hope to start the discussion on applying deep learning to the field of digital humanities.

Frequently Asked Questions

Is the concept of bias oversimplified? For topics as complex and controversial as the Israeli-Palestinian conflict, there might not be a perfect solution for how to describe and study its history. To mitigate this, we used the more technical term "narrative origin" as much as possible, since Israeli sources might not always be Israeli-biased; the word "bias" was used sparingly and intentionally. Although our solution is not perfect, our hope is that releasing a dataset and analyzing the results of well-known, baseline models will open the door for further discussion on using computational models in the digital humanities.

What about self-critical sources (e.g., an Israeli source critical of its own government/country)? It is definitely possible that some Israelis write critically towards their own country. This could be removed through manual annotation, but based on our n-gram analysis, we found that despite potential noise from writers who are hostile to their own countries, the phrases learned by our model generally corroborate Middle East historian insight.

Why didn't data augmentation improve the CNN that much? We hypothesize that sliding window augmentation did not help much because CNNs inherently use a sliding convolutional window, and that improvement was not as pronounced because the baseline was already relatively high.

Does the small size of the dataset lead to models that overfit? Our dataset is indeed small, as collecting high-quality data that controls for time, style, and topic has been non-trivial. In terms of total number of words, however, our dataset sizes are comparable to other benchmark classification datasets. Our SBS dataset (~60k words) is similar in size to established benchmark datasets such as the Question Type Dataset (Li and Roth, 2002) and the Customer Review Dataset (Hu and Liu, 2004). Our larger IP-News dataset (~500k words) is similar in size to the Movie Review Dataset (Pang and Lee, 2005), the Stanford Sentiment Treebank (Socher et al, 2013), and the

Subjectivity Dataset (Pang and Lee, 2004). Overfitting is indeed possible, but we think it can be mitigated with data augmentation and standard regularization techniques, and we are currently working on collecting more data to address the issue moving forward.

Are presented models state-of-the-art? Nope. For this paper, however, the primary contribution is not a methodological innovation but rather (1) our rigorously collected dataset and (2) analysis that ideas discovered by a model reflect Middle East historian insight. We hope to open the door for more empirical methods to be used for historical analysis in the digital humanities.

How is data from synonym replacement used? We use WordNet (Miller 1995) to find synonyms.

Acknowledgement

This work was supported in part by Office of Naval Research Grant No. N00014-19-1-2211, Air Force Office of Scientific Research Grant No. FA9550-20-1-0032, and the Neukom Institute for Computational Science at Dartmouth College. Special thanks to Andrew G. Simon and Dale F. Eickleman for their advice and inspiration.

References

Adwan, S. 2012. *Side by Side*. New York, USA: The New Press.

Bergsma, S.; Post, M.; and Yarowsky, D. 2012. Stylometric analysis of scientific articles. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL HLT '12, 327–337. Stroudsburg, PA, USA: Association for Computational Linguistics.

Caplan, N. 2011. *The Israel-Palestine Conflict: Contested Histories*. Contesting the Past. Wiley.

Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. *MCSS* 2:303–314.

Doumato, E., and Starrett, G. 2007. *Teaching Islam: Textbooks and Religion in the Middle East*. Lynne Rienner Publishers.

Ellen, J., and Parameswaran, S. 2011. Machine learning for author affiliation within web forums – using statistical techniques on nlp features for online group identification. In *2011 10th International Conference on Machine Learning and Applications and Workshops*, volume 1, 100–105.

Feng, S.; Banerjee, R.; and Choi, Y. 2012. Syntactic stylometry for deception detection. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2*, ACL '12, 171–175. Stroudsburg, PA, USA: Association for Computational Linguistics.

Glorot, X.; Bordes, A.; and Bengio, Y. 2011. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, 513–520. USA: Omnipress.

Hollander, R. 2017. How the nyt continues to skew the israeli-palestinian conflict. *Committee for Middle East Reporting in America*.

Iyyer, M.; Enns, P.; Boyd-Graber, J.; and Resnik, P. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1113–1122. Association for Computational Linguistics.

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

Kolomiyets, O.; Bethard, S.; and Moens, M.-F. 2011. Model-portability experiments for textual temporal analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, HLT '11, 271–276. Stroudsburg, PA, USA: Association for Computational Linguistics.

Koppel, M.; Argamon, S.; and Shimoni, A. R. 2002. Automatically categorizing written texts by author gender. *LLC* 17:401–412.

Kressel, N. J. 1987. Biased judgments of media bias: A case study of the arab-israeli dispute. *Political Psychology* 8(2):211–227.

Lin, W.-H.; Wilson, T.; Wiebe, J.; and Hauptmann, A. 2006. Which side are you on?: Identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, CoNLL-X '06, 109–116. Stroudsburg, PA, USA: Association for Computational Linguistics.

Liu, P.; Qiu, X.; and Huang, X. 2016. Recurrent neural network for text classification with multi-task learning. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, 2873–2879. AAAI Press.

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

Miller, G. A. 1995. Wordnet: A lexical database for english. *Commun. ACM* 38(11):39–41.

Mukherjee, A., and Liu, B. 2010. Improving gender classification of blog authors. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, 207–217. Stroudsburg, PA, USA: Association for Computational Linguistics.

Munayer, S., and Loden, L. 2014. *Through My Enemy's Eyes: Envisioning Reconciliation in Israel-Palestine*. Authentic Media.

Ortiz Laguna, J.; Olaya, A. G.; and Borrajo, D. 2011. A dynamic sliding window approach for activity recognition. In Konstan, J. A.; Conejo, R.; Marzo, J. L.; and Oliver, N., eds., *User Modeling, Adaption and Personalization*, 219–230. Berlin, Heidelberg: Springer Berlin Heidelberg.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Potthast, M.; Kiesel, J.; Reinartz, K.; Bevendorff, J.; and Stein, B. 2018. A stylometric inquiry into hyperpartisan and fake news. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 231–240. Association for Computational Linguistics.

Sarawgi, R.; Gajulapalli, K.; and Choi, Y. 2011. Gender attribution: Tracing stylometric evidence beyond topic and genre. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning*, CoNLL '11, 78–86. Stroudsburg, PA, USA: Association for Computational Linguistics.

Sternthal, T. 2001. Critical thinking: Can you trust everything you read? *Committee for Middle East Reporting in America*.

Tanasescu, C.; Kesarwani, V.; and Inkpen, D. 2018. Metaphor detection by deep learning and the place of poetic metaphor in digital humanities. Florida Artificial Intelligence Research Society Conference.

Tetreault, J.; Blanchard, D.; Cahill, A.; and Chodorow, M. 2012. Native tongues, lost and found: Resources and empirical evaluations in native language identification. In *Proceedings of COLING 2012*, 2585–2602. The COLING 2012 Organizing Committee.

Um, T. T.; Pfister, F. M. J.; Pichler, D.; Endo, S.; Lang, M.; Hirche, S.; Fietzek, U.; and Kulic, D. 2017. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. *CoRR* abs/1706.00527.

Wang, W. Y., and McKeown, K. 2010. "got you!": Automatic vandalism detection in wikipedia with web-based shallow syntactic-semantic modeling. In *COLING*.

Wang, W. Y., and Yang, D. 2015. That's so annoying!!!: A lexical and frame-semantic embedding based data augmentation approach to automatic categorization of annoying behaviors using #petpeeve tweets. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2557–2563. Association for Computational Linguistics.

Wei, J. W., and Zou, K. 2019. EDA: easy data augmentation techniques for boosting performance on text classification tasks. *CoRR* abs/1901.11196.

Zhang, X.; Zhao, J.; and LeCun, Y. 2015. Character-level convolutional networks for text classification. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, 649–657. Cambridge, MA, USA: MIT Press.

Supplementary Material

This section contains implementation details not included in the main text, as well as a frequently asked questions section.

Implementation Details

We use the same CNN and RNN setup as in (Wei and Zou 2019).

Word embeddings. We use 300 dimensional word embeddings trained using GloVe (Pennington, Socher, and Manning 2014).

CNN. We use the following architecture: input layer, 1D convolutional layer of 128 filters of size 5, global 1D max pool layer, dense layer of 20 hidden units with ReLU activation function, softmax output layer. We initialize this network with random normal weights and train against the categorical cross-entropy loss function with the adam optimizer. We use early stopping with a patience of 3 epochs.

RNN. The architecture used in this paper is as follows: input layer, bi-directional hidden layer with 64 LSTM cells, dropout layer with p=0.5, bi-directional layer of 32 LSTM cells, dropout layer with p=0.5, dense layer of 20 hidden units with ReLU activation, softmax output layer. We initialize this network with random normal weights and train against the categorical cross-entropy loss function with the adam optimizer. We use early stopping with a patience of 3 epochs.