

Identifying Bias-Heavy Sentences in News Articles

Andy S. Chen, Nicholas P. Hirning, Shreya S. Shankar
Stanford University



Motivation

- Recent events surrounding 2016 presidential election have spurred investigation of media and news provider bias
- Characterizing news providers by articles' sentences and words

Problem Definition

- Goal: classify an article by its news provider and identify what words or phrases rationalize the classification
- Data: approximately 10,000 news articles from 5 major news providers in the past six months (CNN, Reuters, New York Times, Breitbart, Fox News)

Sample Data Points

Article Text	Provider
NEW ORLEANS - New Orleans police say a drive-by shooting killed two men in front of a high school during a basketball game. Local media report that Edna Karr High School's auditorium was at capacity and was locked down for more than an hour Tuesday night...	CNN
A website set up to shame people who took inappropriate photos and smiling selfies at Berlin's Holocaust Memorial has achieved its objective, according to the Israeli artist behind the 'Yolocaust' project. Berlin-based satirist and author Shahak Shapira combined the people's photos with horrific concentration camp footage, using the Yolocaust site to shine a spotlight on selfie culture...	Fox News

- Test and dev datasets had 2,000 articles each; train dataset had 6,000 articles
- News providers' articles were evenly distributed among the dataset
- Implemented bag-of-vectors model for a baseline classification algorithm
 - Condensed each article to a single vector using averages of GloVe word vectors
 - Used two-layer feed-forward neural network with three hyperparameters (learning rate **LR**, hidden dimensions **h**, training epochs **e**)

Hyperparameter Determination of Baseline

h	LR = 0.01	LR = 0.1	LR = 1	LR = 10
25	0.522	0.576	0.588	0.258
50	0.527	0.576	0.609	0.257
75	0.530	0.579	0.597	0.258
100	0.535	0.572	0.572	0.258
125	0.529	0.582	0.547	0.258
150	0.532	0.569	0.605	0.258

- With optimal configuration (**LR** = 1, **e** = 1600, **h** = 200) and GloVe vector dimension of 300, test set classification accuracy was **75.8%**
- Objective: design a system to improve classification accuracy and also identify sentences which characterize this classification

Neural Networks for Text Classification and Rationale

CNNs and *k*-max pooling

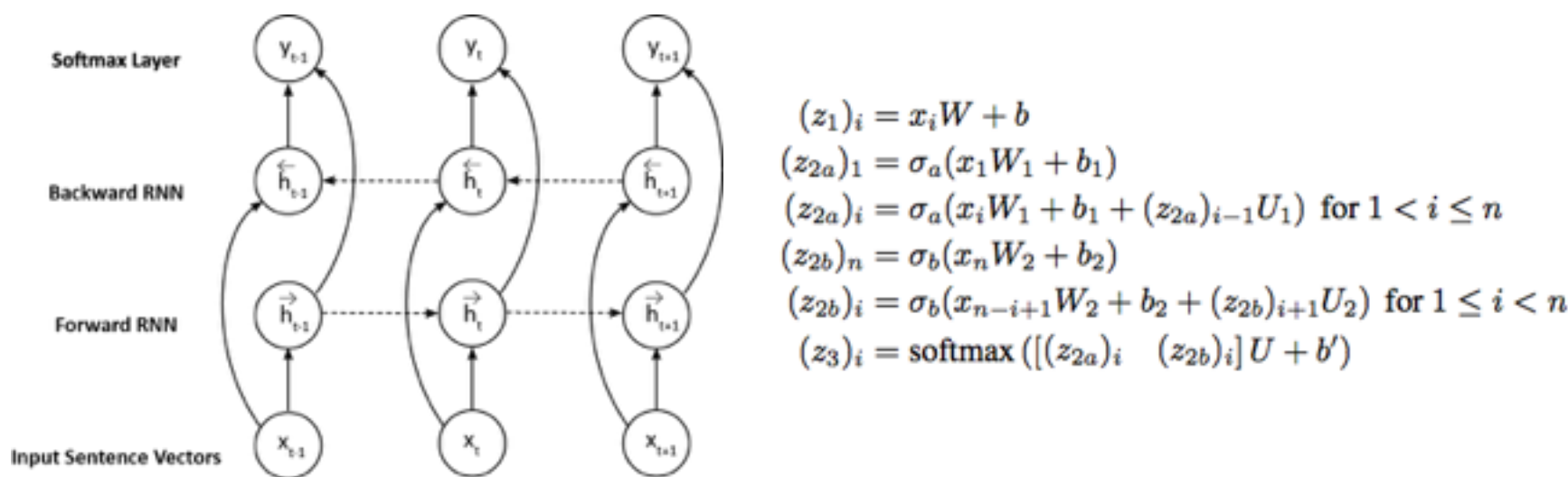
- Articles tokenized with Python Natural Language ToolKit into GloVe vectors
- Inputs were passed through a linear layer, then a repeated loop of applying a CNN followed by a *k*-max pooling operation, then a final hidden layer with dropout
- We let $\mathbf{x}_{i,j}$ denote the *j*th word vector in the *i*th sentence of the input article, d_i is the length of the *i*th sentence, we are applying a CNN filter of width **n** and *k*-max pooling, and we apply this operation a total of **t'** times

$$\begin{aligned}(z_1)_{i,j} &= x_{i,j}W + b_1 \\ \begin{cases} (z_{3,1})_{i,j} &= (z_1)_{i,j} \\ (z_{2,t})_{i,j} &= [(z_{3,t})_{i,j} \quad (z_{3,t})_{i,(j+1)} \quad \cdots \quad (z_{3,t})_{i,(j+n)}] F \text{ for } j+n \leq d_i \text{ and } 1 \leq t \leq t' \\ (z_{3,t})_{i,j,k} &= \max_{l=0,1,\dots,k'} (z_{2,t-1})_{i,j+l,k} \text{ for } 1 < t \leq t' \end{cases} \\ (z_4)_i &= \frac{1}{d_i - t'(n + k' - 2)} \sum_{j=1}^{d_i - t'(n + k' - 2)} (z_{3,t'})_{i,j} \\ z_5 &= \sum_i [(h_i \circ (z_4)_i)U + b_2]\end{aligned}$$

- Loss is computed by applying softmax to output vector **z**₅ and applying cross-entropy

Bidirectional RNNs for Rationale Identification

- For every article, we aim to rank sentences by their importance to the classification
- One RNN passes through the article in the forwards direction operating on every sentence, and another in the backwards direction
- The results are concatenated and passed through a hidden layer and a softmax function to generate **1x2** vectors for every sentence representing the "probability" of keeping/dropping a sentence
- The "keep probabilities" are normalized and used as masking weights for the words in each article when feeding into the classifier
- Loss was L2 difference between original classifier outputs and resulting outputs (regularized with variance of the masking weight distribution)
- We let \mathbf{x}_i be the *i*th sentence vector and **n** represent the number of sentences

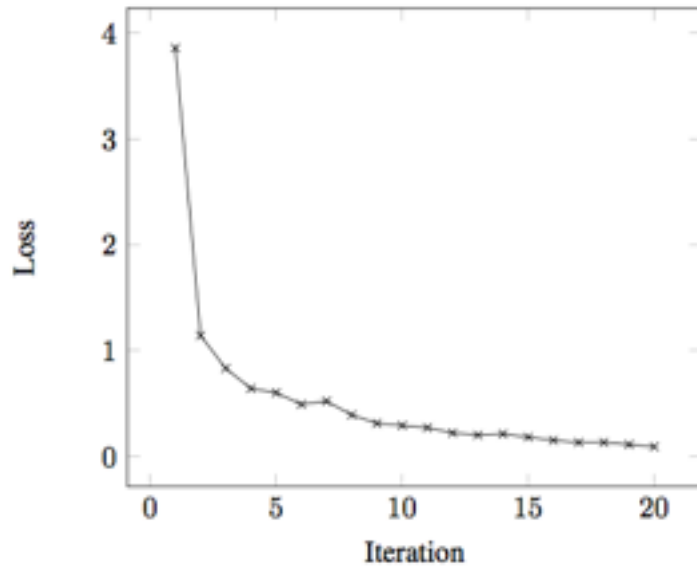


Data and Results

- For the CNN with *k*-max pooling, the optimal parameters were determined to be **k** = 1, **n** = 3, **t'** = 1, **LR** = 0.01, **e** = 60, **h** = 240 resulting in **84.0%** accuracy
- We calculate the accuracies per news provider and show the learning curve

Test Set Accuracies

News Provider	Accuracy
CNN	93.4%
Reuters	87.2%
New York Times	86.5%
Fox News	77.4%
Breitbart	55.8%



- For the bidirectional RNN, the optimal parameters were determined to be regularization constant of 1, **LR** = 0.01, **e** = 20, **h** = 240
- On the test set, the original article masked by the bidirectional RNN weights was consistent with the original classification **99%** of the time

Rationale Examples

Most Important Sentences	Provider
MANILA , Philippines - The Philippine military chief says troops have launched airstrikes and ground assaults against one of Southeast Asia's most-wanted terror suspects who is trying to establish a new base for an alliance of armed groups backing the Islamic State group .	CNN
Should places in this country that ignore the laws of this country when it comes to immigration receive federal money into their communities? White House Chief of Staff Priebus defended President Donald Trump 's move to block federal funds meant for so-called "sanctuary cities" that harbor immigrants who entered the U.S. illegally.	Fox News
But this year you may be aware of one nominee, the Iranian director Asghar Farhadi ,who is among the prominent artists to be affected by President Trump's ban on visas for travelers from predominantly Muslim countries.	NYT

Conclusions and Future Work

- Classification of articles by news providers and rationale identifications for these classifications are reasonably accurate (based on keywords or sentiments on certain topics)
- In tandem, these systems can be used to identify biased sentences at >80% accuracy
- Models can be improved upon and/or extended in several ways: more data; dynamic *k*-max pooling; different activation functions for bidirectional RNN; classification of news articles by other bias-inducing characteristics

Acknowledgements

- Arun Chagarthy, our project mentor
- Stanford Computational Linguistics Lab for guidance on obtaining data
- Chris Manning, Richard Socher, and CS224N staff for teaching the course and providing access to Microsoft Azure GPUs

References

- Kim, Yoon. "Convolutional Neural Networks for Sentence Classification." Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2014): n. pag. Web.
- Lei, Tao, Regina Barzilay, and Tommi Jaakkola. "Molding CNNs for Text: Non-linear, Non-consecutive Convolutions." Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (2015): n. pag. Web.
- Lei, Tao, Regina Barzilay, and Tommi Jaakkola. "Rationalizing Neural Predictions." Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (2016): n. pag. Web.
- Pavitra, R., and P. C. D. Kalaiyaani. "Weakly Supervised Sentiment Analysis Using Joint Sentiment Topic Detection with Bigrams." 2015 2nd International Conference on Electronics and Communication Systems (ICECS) (2015): n. pag. Baselines and Bigrams: Simple, Good Sentiment and Topic Classification. Stanford University. Web.
- Pennington, Jeffrey, Richard Socher, and Christopher Manning. "Glove: Global Vectors for Word Representation." Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2014): n. pag. Web.
- Recasens, Marta, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. "Linguistic Models for Analyzing and Detecting Biased Language." *Linguistic Models for Analyzing and Detecting Biased Language*. Stanford University, n.d. Web.