Sentiment Analysis on Current Event Opinions

Meltem Tutar (mtutar@stanford.edu)
Stanford University, CS230 Project

Predicting

We are performing sentiment analysis on comments that are in the political/news domain. Sentiment analysis is the computational study of people's opinions towards entities such as products, services, organizations, individuals, issues, events or topics

Our objective is to build predictive models that estimate positive, negative, and neutral sentiment given a text corpus.

Data

The dataset is from twitter comments on the 2016 GOP debate. Contributors labeled if the tweet was relevant, which candidate was mentioned, what subject was mentioned, and what the sentiment was for a given tweet. We have 13,874 tweets.

Example Text Representation

Text
"RT @NancyLeeGrahn: I couldn't understand"

Tokens
"I","could","n't","understand"

Embedding

0.1439	0.0862	0.2040	0.0572
0.4674	0.0716	-0.0171	-1.2526
-0.7179	0.2615	-0.3102	0.4974
-0.1276	-0.0382	-0.4162	0.3797
-0.3080	-0.4153	0.5791	-0.3694
I	could	n't	understand

Features

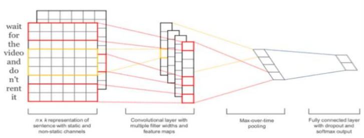
We kept only the [sentiment] and [text] features. We followed the following pipeline

- Cleaned data by removing entities such as hashtags and white space
- Tokenized ensuring the vocabulary size is roughly 10,000
- Used pretrained GloVe word Embeddings

Models

CNN with Max Pooling Over Time

Parameters include the filter sizes and number of filters. We utilized 100 Filters of size 2,3, and 5. We padded examples to have a minimum of 5 words. The size of the filter dictates the window of words that it analyzes and can be equivalent of learning from n-grams.



We applied a max over time pooling layer to ensure consistent size outputs. Linear and SoftMax layers were applied.

Models

Multinomial Naïve Bayes

We used TF-IDF to represent the text in vectors as weighted occurrences.

Multinomial naïve Bayes estimates and maximizes the joint probability distribution

$$p(y, x_1 \dots x_d) = q(y) \prod_{j=1}^{d} q_j(x_j|y)$$

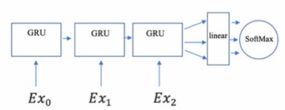
$$P(X_j = x \mid Y = y) = \frac{N_{ct} + \alpha}{N_y + \alpha n} \qquad P(Y = y) = \frac{NC}{n}$$

Using these estimates, predictions on test set can be made

$$\arg\max_{y \in \{1 \dots k\}} p(y, x_1 \dots x_d) = \arg\max_{y \in \{1 \dots k\}} \left(q(y) \prod_{j=1}^{d} q_j(x_j|y) \right)$$

GRU

Recurrent neural networks are able to capture relationships between sequential data. A GRU utilized update and reset gates to allow for learning across longer sentences.



There are 2 layers, with bidirectionality, dropout of 0.5, hidden dimension size of 256, and embedding size of 100.

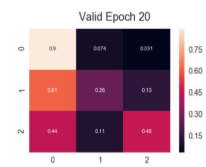
$$CE\ Loss = -\sum_{c=1}^{3} y_{o,c} \log(p_{o,c})$$

Results

The GRU loss and accuracy curves over 30 epochs

Model	Train Accuracy	Validation Accuracy
Multi NB	0.643	0.689
GRU	0.798	0.651
CNN	0.773	0.681

The confusion matrix on the GRU after 20 epochs



Discussion

CNN model had a loss of 0.785 and accuracy of 0.68 on the test set. We weren't able to make improvement from the baseline on unseen data, though we did have higher accuracy on the training set. It seems that the CNN/GRU modes are not generalizing well. The CNN loss curve is similar to the GRU curve. The large decrease in training error and eventual increase in validation error may indicate there is an overfitting problem. Dropout is applied, and various other regularization methods were also experimented with, without much change in the situation. The It is apparent that most of the errors are being made on examples from class 1 and 2.

Future

We believe that addressing the imbalances of classes may allow the CNN/GRU models to generalize better. We believe we need to fundamentally alter the models by experimenting with different objective functions to optimize, introducing new features such as POS tagging, and possibly learning from other training sets.

References

- Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and Trends® in Information Retrieval2.1–2 (2008): 1-135.
- Turney, Peter D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002.
- Zhang, Lei, Shuai Wang, and Bing Liu. "Deep learning for sentiment analysis: A survey." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8.4 (2018): e1253.
- Berardi, Giacomo, Andrea Esuli, and Diego Marcheggiani. "Word Embeddings Go to Italy: A Comparison of Models and Training Datasets." IIR. 2015.
- Kim, Yoon. "Convolutional neural networks for sentence classification." arXiv preprint arXiv:1408.5882 (2014).