CB;DR
Clickbait Detection using Parallel Neural Networks

Peter Adelson, Sho Arora, and Jeff Hara
padelson@stanford.edu, shoarora@stanford.edu, jhara18@stanford.edu

PROBLEM

As attention has become an increasingly scarce resource, clickbaiting titles have become highly prevalent with underlying content often fails to deliver [1]. Clickbait scoring provides immediate application in providing automatic flagging of clickbaiting articles. We explore the performance of a parallel neural network architecture on the task, using article, post, and title text as input to generate a clickbait score. We are able to achieve a MSE of 0.067 and F1 score of 0.648.

TF-IDF

As a baseline, we used Term Frequency, Inverse Document Frequency as a feature meant to capture article semantics. TF-IDF is given by the following equation:

$$f_{t,d} * log \frac{N}{n_t}$$

Relating the frequency of a term t in document d to the total occurrence of the word. To create a vocabulary size that was tractable, we had a threshold number of documents a word had to appear in. We limited TF-IDF usage to the article text.

GloVe

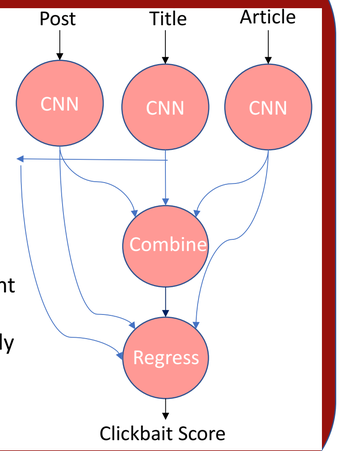
Global Vectors, or GloVe, are a method of representing words as vectors of numbers. Vectors are created through unsupervised learning, focusing on word-word co-occurrences. This generates vectors where words that co-occur have a low Euclidean distance in their vector space [3]. To process text in our model, we converted it into pre-trained GloVe vectors available from GloVe6b trained on gigaword and wikipedia.

DATA

We used the clickbait dataset provided by Bauhaus-Universitat (clickbait-challenge.org).
Features include article text, post text, title text, associated post metadata, and human-generated clickbait scores ranging between 0 and 1 with higher scores indicating clickbait. The dataset had 20,000 unique points. We used a 60/20/20 train/dev/test split. It's balance is about 0.25 entries with median scores greater than 0.5.

MODEL ARCHITECTURE

The model is based around three CNNs, operating in parallel. Their output is then joined in an activation layer. This layer then outputs a result that is joined with the output of the three neural networks to output a final clickbait score. Each neural network focuses on a different field from post, title, and article. We have affectionately called our model 'PNet', for Parallel Network.



FUTURE WORK

The model architecture provides interesting opportunities for extensions with potential for better results. Different models such as the QRNN could be attempted. Different activation functions could be used in the final layers of the model. Finally, investigating the associated post media image could improve results.

BASELINES

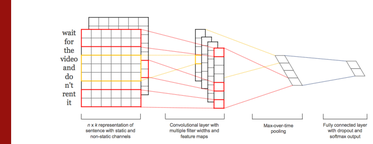
In order to judge the performance of our model, we generated two baselines. The first used TF-IDF features and an SVM. The second baseline used 60 features, including Flesch-Kinkaid scores of the article, number of proper nouns, and average word length [4]. We ran both on SVMs using hinge loss. We also attempted both baselines with logistic regression. We also used a solo CNN.

SVM Equation [5]

$$\begin{aligned} \min_{\gamma,w,b} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i, \;\; i = 1, \ldots, m \\ & \xi_i \geq 0, \;\; i = 1, \ldots, m. \end{aligned}$$

CNN

We use convolutional neural networks. Briefly, convolution is sliding a filter over multi-dimensional input. As input we use the multidimensional word vectors over the text. Image credit [6].



RESULTS

Model	Train MSE	Train F1	Test MSE	Test F1
TF-IDF (LR)	101.189	0.908	43.008	0.445
Features (LR)	5.484	0.482	5.382	0.501
Features (SVM)	0.073	0.522	0.075	0.517
CNN (Text)	0.019	0.909	0.094	0.299
CNN (Title)	0.058	0.530	0.088	0.385
CNN (Post)	0.017	0.913	0.062	0.650
PNet 2 (Title/Post)	0.039	0.748	0.061	0.646
PNet 3	0.014	0.896	0.067	0.648

DISCUSSION

The parallel network architecture demonstrates itself capable of outperforming the non-neural network baselines. Most of the signal on clickbait appears to be concentrated in the post associated with the article. As the content is human-scored, this indicates that people judge clickbait primarily based off how it is presented in the post. Because of this, a CNN run on just the post text has as much signal and less interference and is capable of edging out the PNet. The additional features of text and title allow a great degree of overfitting for the PNet, a phenomenon also seen with CNN text.

REFERENCES

[1] Potthast et al. 'Clickbait Detection'. 2016. http://www.uni-weimar.de/medien/webis/publications/papers/stein_2016b.pdf
[2] J. Ramos. Using TF-IDF to Determine Word Relevance in Document Queries. Technical report, Department of Computer Science, Rutgers University, 2003.
[3] Pennington, Jeffrey, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014) 12, 2014.
[4] Cao et Al. Machine Learning Based Detection of Clickbait Posts in Social Media. 2017.
[5] Andrew Ng. CS 229 Lecture Notes. Support Vector Machines.
[6] Yoon Kim. Convolutional Neural Networks for Sentence Classification. 2014