CARE: Clickbait Fighter Using Support Vector Machine

Changyuan Qiu* Chenshu Zhu* Haoxuan Shan*
Department of EECS, University of Michigan
{peterqiu, jupiterz, shanhx} @umich.edu

1. Introduction

Clickbait is a rampant problem hauting online news readers. Motivation behind clickbaiting is to boost site traffic (and therefore, advertisement revenue) by exploiting the curious nature of human readers. The clickbait technique works by dangling a hyperlink with enticing headlines to lure people into clicking; and then redirect them to the publishers' own websites. This poses a problem that, as there exist considerable discrepancies between the headline itself and the destination content it points to, news readers would waste a significant amount of their time on these advertisements.

To address this problem, we introduces the model CARE: (**C**lickb**A**it fighter using support vecto**R** machin**E**), which could effectively identify clickbaits utilizing feacture extraction methods and support vector machine. We approach this issue as a binary classification problem, with a dataset of news articles to be labeled either clickbait or non-clickbait. We expect our model to be capable of capturing clickbait patterns in the headline, measures the correlation between the headlines and its contents (if applicable), and then classify the news article into the correct category.

2. Related work

Attempts to eliminating clickbaits have been made by many news agencies. Facebook, for instance, launched a war against clickbaits in 2017. Their main strategy was to utilize users' feedbacks to identify and block domains that are notorious for producing clickbaits.

In the academic world, Chakraborty et al. [1] took the lead to explore machine learning models in clickbaits detection in 2016 and create a browser extension 'Stop Clickbait' for deterring clickbaits, yet they called for future work devoted to improving the classification and blocking performances.

As a result, we plan to examine the effectiveness of the features (sentence structure, word pattern, clickbait language and N-gram features) used in Chakraborty's paper [1], try to extract new feactures and make further improvements.

3. Datasets

At the time of this writing, we found 2 datasets available for model evalutaion. The first one is the Webis Clickbait Corpus 2016, which comprises 2992 Twitter tweets sampled from top 20 news publishers as per retweets in 2014 [2]. This comprehensive dataset is the first of its kind and gives rise to the development of technology to tackle clickbait. All data in this corpus contains both the headline and the content of each tweet. While for the second dataset (accessible from Chakraborty's work) [1], it contains only headlines, with a total amount of 18513 "Non-clickbait" headlines and 7623 "Clickbait" ones. We plan to first train and cross-validate our model on the second dataset to enhance feature extraction, and then move on to the first dataset for systematic analysis and further exploration (like we may also add correlation between headline and content as a new feature and see how the new feature would affect performance).

4. Methodology

We plan to use Support Vector Machine (SVM) to serve as the backbone. We will first utilize feature extraction methods to preprocess the training data and collect useful features from the dataset. Then we will perform cross-validation model selection on our SVM classifier to find the optimal hyperparameters. Finally we will train the classifier and evaluate its performance on the test set. As time permits, we also hope to explore a neural network model in addressing the problem.

For feature extraction parts, besides the features discussed in Chakraborty's paper [1], we will also explore other features and perform corresponding evaluation. For instance, we consider word sentiment to be a viable choice since clickbait headlines generally uses more sentimental words to catch reader's attention.

5. Evaluation

As a binary classification problem, we will prioritize accuracy among all performance metrics in terms of evaluation. Then we will consider precision next to maximize the percentage of click-bait that get correctly classified. We will also evaluate our model and report our model's performance on recall rate, f1-score, ROC curve and other metrics.

6. Project Management & Milestones

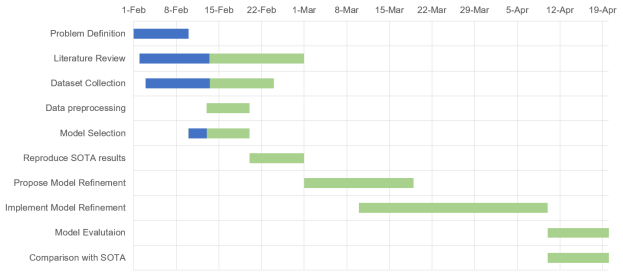

Figure 1. Project Management Gantt Chart (Tentative)

References

[1] Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. Stop clickbait: Detecting and preventing clickbaits in online news media. CoRR, abs/1610.09786, 2016. 1

[2] Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen. Clickbait Detection. In Nicola Ferro, Fabio Crestani, Marie-Francine Moens, Josiane Mothe, Fabrizio Silvestri, Giorgio Maria Di Nunzio, Claudia Hauff, and Gianmaria Silvello, editors, Advances in Information Retrieval. 38th European Conference on IR Research (ECIR 2016), volume 9626 of Lecture Notes in Computer Science, pages 810–817, Berlin Heidelberg New York, Mar. 2016. Springer. 1