BaitBuster: Destined to Save You Some Clicks

Md Main Uddin Rony¹, Naeemul Hassan¹, Mohammad Yousuf²
¹The University of Mississippi, ²The University of Oklahoma

Data Collection

- 32,000 manually labeled news headlines. [1]
- About 1.7 million Facebook posts.
- From 68 mainstream and 85 unreliable media.
- Considered headlines, bodies, and the statuses.

Overview

- Powered by clickbait detection model with the accuracy of 98.3%.
- Provides brief explanation behind clickbait decision.
- Allows users to read the brief summary of the content without leaving the current page.
- Collects user feedback on clickbait decision.
- Social bot regularly publishes automatically generated report about contemporary clickbait contents.

Clickbait Detection

- Distributed subword embedding was used to create sentence vectors.
- Pre-trained embeddings using 477236 words of the Media Corpus.
- Used Softmax function as the classifier.

Clickbait Detection API

- Article Body Scraper
- Clickbait Detection Model
- RESTful API Services
- Bot API
- Facebook Page

Motivation

- Clickbaiting has become widespread
  - All types of media practice clickbait.
  - It has become a source of easy revenue.
- Negative impact on media ecosystem
  - Damages media-user relationship.
  - Depletes media brand value.
- Lack of comprehensive solution
  - A little but inefficient approaches to fight against this ill practice.
  - Solutions depending on checking the headline only are not sufficient.

Ground Truth

- 32,000 manually labeled news headlines. [1]

Clickbait

- Eek! What’s Lurking in the Shadows?! I Have to Know!
- I Left My Daughter And THIS Happened!
- Non-Clickbait
- America’s democracy has become illiberal
- Twitter Expands Fight Against Abuse

Problem Formulation

- Two categories of headlines
  - Clickbait
  - Non-Clickbait

Printed in the USA by

CDP Media

Copyright ©2017

References