

Is that headline Clickbait?

A Transformer-based News Clickbait Detector

*By: Sandeep Kumar
Charmi Chokshi*

Mentor: Xing Han Lu

What?

- A headline is Clickbaiting when it is,
- Arousing curiosity instead of providing informative summaries
- Mostly misleading to their reader
- Generated just to increase user engagements and 'clicks' in posts



Why?

- To promote responsible journalism
 - Recent example below of a celebrity calling out a publication for using his interview as clickbait
- To stop the spread of misleading information in today's information world
 - Recent example on the right showing an actual note VS how print and digital media take it over

Tyler aired the publication out for being thirsty for clicks. He said, "I guess people write these kinds of headlines for **clickbait**. It's sad. I didn't call out Janet and Jill, and they're not holding up another Why Did I Get Married." According to Tyler, he doesn't even have a script for the third installment, just an idea for a storyline.



CNET Your guide to a better future

Science

Giant Asteroid Hurling Toward Earth! How to Know If the Risk Is Real

The ones to worry about are rarely the ones in the alarmist headlines.

Eric Mack Oct 30, 2022 9:54 p.m. PT

4 min read

How?

1. Sequence-based models (RNN, LSTM, GRU) on headline's BERT Embeddings
2. Finding cosine similarity of headline with its respective paragraph(s) by using default BERT Embeddings
3. Fine-tune Transformer model on the task (BERT / RoBERTA)
4. Use Text Summarization on the article and then find similarity with the heading to see if it is related actually

Evaluation Metrics:

- Classification task: Accuracy or F1 Score
- Regression task: MSE

Dataset

- **Webis-Clickbait-17**: the Webis Clickbait Corpus 2017 comprises a total of 38,517 Twitter posts from 27 major US news publishers
- Each sample contains article headline, passages, keywords, captions, etc.
- All posts are annotated on a 4-point scale (not clickbaiting to heavily clickbaiting), and their class.

Thank You!

Questions?