Similarity-Aware Deep Attentive Model for Clickbait Detection

Tanmay Bhatt - 2020112017
Abhinav Siddharth - 2020112007
Anjali Singh - 2020102004
Pranav Manu - 2020112019

Index

- Motivation
- Objective
- Existing Work
- Understanding the Proposed Method
- Work Done So Far...
- Expected Results
- Datasets

Motivation

Clickbaits are a type of web links designed to entice users to enter specific web-pages or videos.

We come across clickbait text and images multiple times each day. Usually, such links will lead to non-informative and misleading articles which causes the reader to stray from the topic which they were actually looking for.

Hence, it becomes imperative to come up with methods for automated clickbait detection such that users can be warned against potential clickbait material.

Objective

The objective of our project is to build a deep similarity-aware attentive model for capturing the discriminative information from local and global similarities.

In other words, given the headline and the article body, the model should be able to classify the text as clickbait or not.

Some examples of clickbait text are

- You'll never believe what happened after …
- How to achieve unbelievable results with this 1 weird trick!

Existing Work

Clickbait detection has been a fairly sought after topic in the recent years and hence considerable amount of work has been put forward toward the same

The first few approaches extracted linguistic features from both the header and the body text and passed these features on traditional classifiers like logistic regression, Bayes classifier, random forest, etc.

More later works employ CNNs and RNNs to look for similarities between the header text and the body to make the classification.
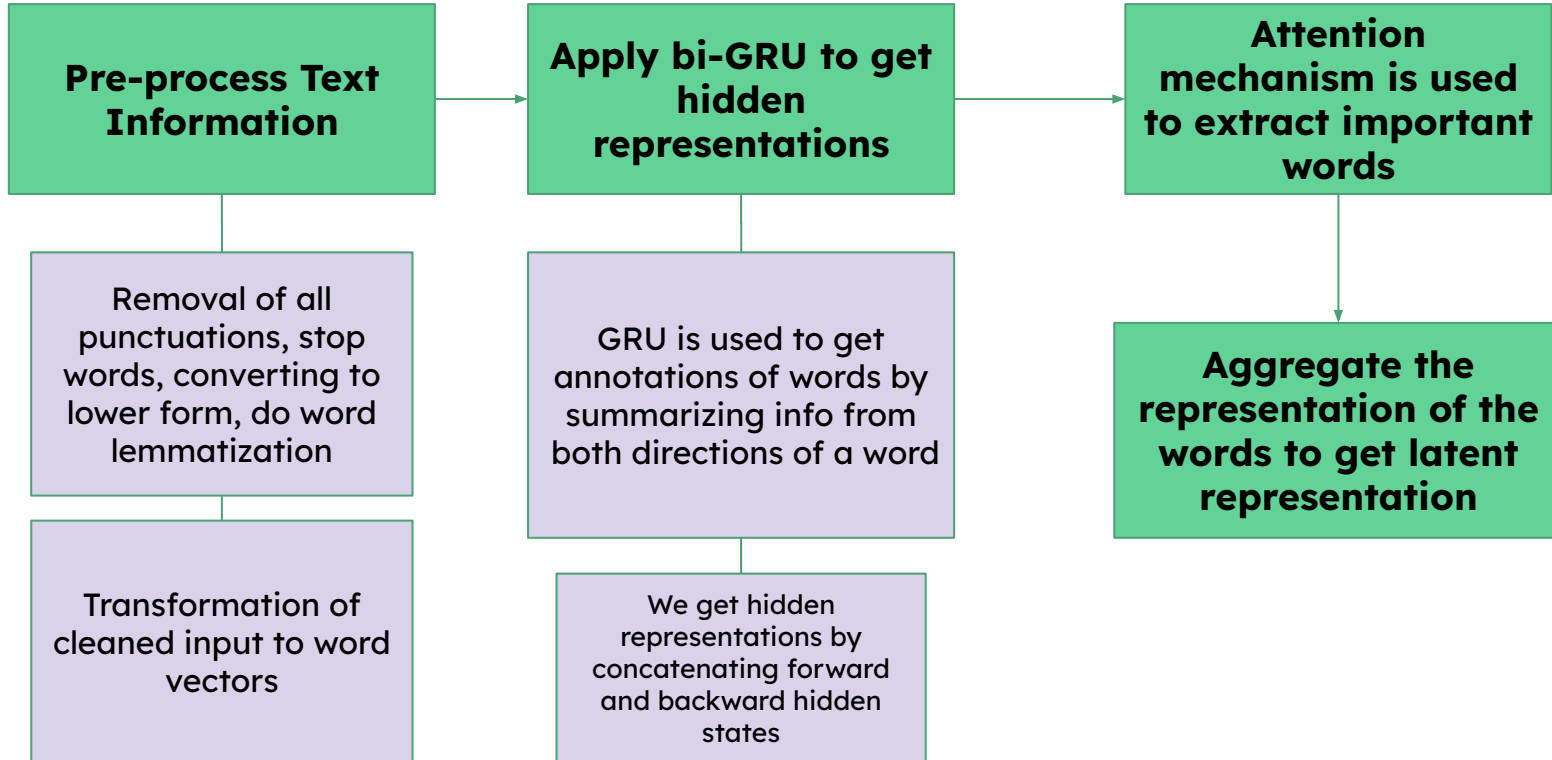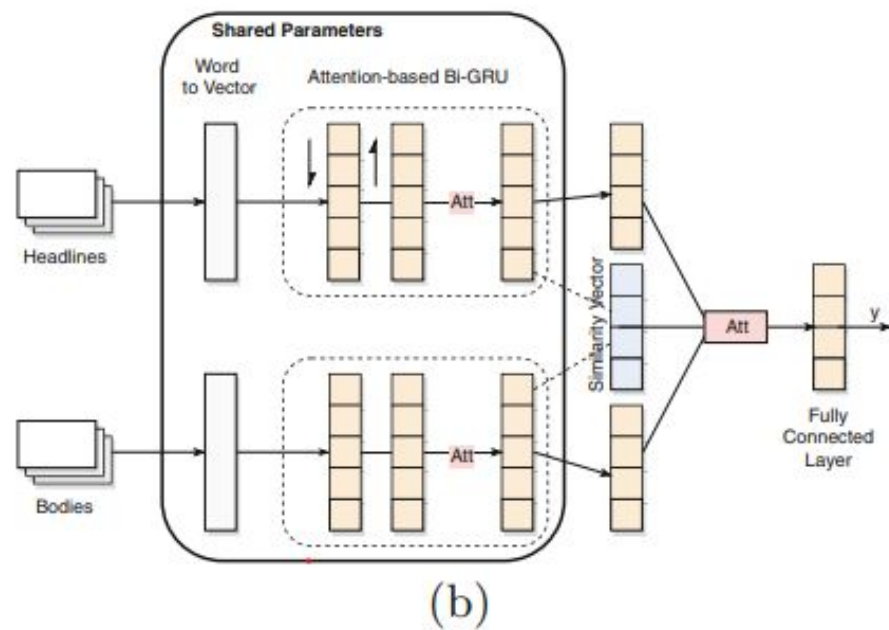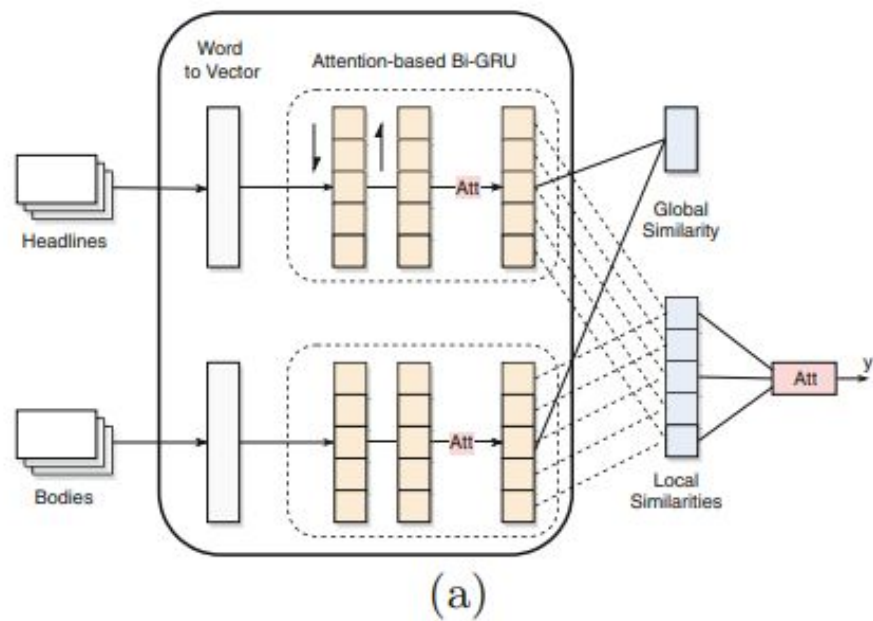
Understanding the Proposed Method

- We propose a model that will exploit both the local and global similarities between the header text and the body text. Moreover, the similarities are modelled as vectors so that they can be augmented with other features for future prediction easily.
- Given a set of titles, $H = \{h_1, h_2, ..., h_N\}$ and their bodies, $B = \{b_1, b_2, ..., b_N\}$ and the labels, $Y = \{y_1, y_2, ..., y_N\}$; $y_i = 1$ if the headline is a clickbait.

The framework in the paper includes: -

- Learning latent representations
- Learning the similarities
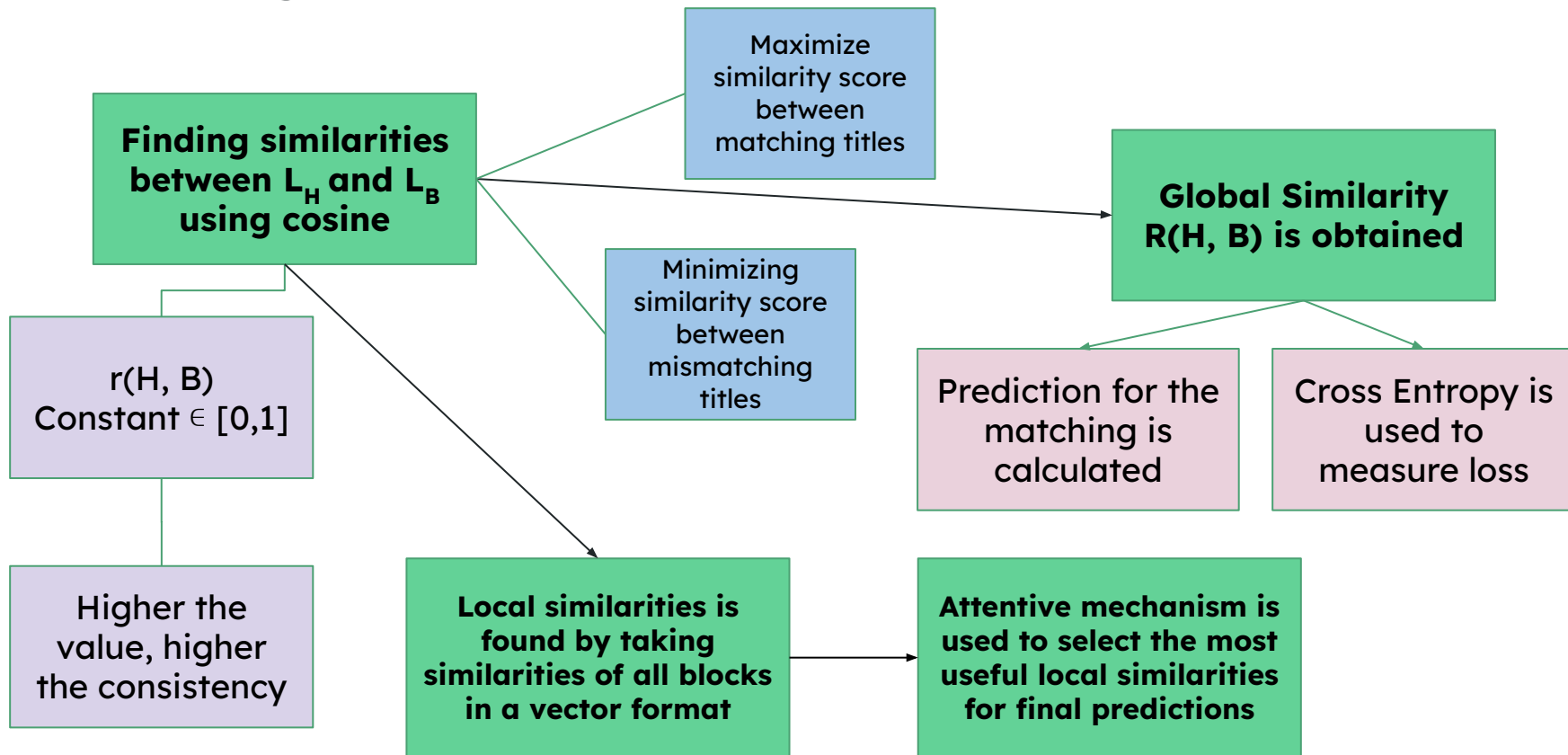- Using the similarities for further predictions

Learning the Latent Representation

Pre-process Text Information

Apply bi-GRU to get hidden representations

Attention mechanism is used to extract important words

Removal of all punctuations, stop words, converting to lower form, do word lemmatization

GRU is used to get annotations of words by summarizing info from both directions of a word

Aggregate the representation of the words to get latent representation

Transformation of cleaned input to word vectors

We get hidden representations by concatenating forward and backward hidden states

(a)

(b)

Learning the Similarities

Using the Similarities for further Predictions

Once we have the similarities, we can combine them with standard clickbait indicators like writing style and text quality. In order to combine these, we make use of the latent representations.

$$L'_H = f(W_H L_H + b_H)$$
$$L'_B = f(W_B L_B + b_B)$$

We then calculate self attention values which are representative of the writing style and text quality and combine it with the latent representations to get a final combination layer on which multilayer perceptron will be applied to get the prediction.

Work Done So Far..

- We have done the pre-processing work which includes the removal of all punctuations and stopwords, changed the sentences to their lower format.
- Performed Word Lemmatization on the dataset.
- Further on this, we have transformed the words into their word vector format.
- For the word, "thousands", this is the following vector form

```
[-0.9922755   2.062286   -4.143881    0.7422714  -2.623593    2.4011183
 -3.3908389   1.3311343   1.9180655    0.8354719  -0.10015057 -1.3349633
  2.345319    2.2079887  -0.74313396   2.1279888   4.4971514   0.91804177
 -3.5259938   0.13228197  0.50051516  -0.8743686  -0.81340885  1.5873406
 -1.4807873  -0.2890507  -2.6256404   -0.3273204  -0.74853665  1.4137839
 -0.11331296 -1.3351024  -1.9986244   -3.9590816  -0.83542895  0.1336828
  0.3800656   0.21332352  1.6442833   -1.2882215  -0.35732886 -1.4814194
  0.99194384  0.10475945 -0.48982957  -1.127037    0.99816555  2.974634
 -1.3218511  -1.0919155  -0.44962686   0.34218332 -0.11054708 -3.3773508
  1.803029   -0.857081   -1.6578285   -2.7807014  -1.5871106  -0.01479506
  2.652553   -1.7131612   0.79556453  -1.8487661   0.35171512  0.19546463
  2.3071728  -1.6871629  -0.59963423   2.7975104  -3.7773054  -0.4606304
 -2.1422515  -3.1007018   5.133338     3.435659   -2.1749642  -0.10913195
  0.0967379  -2.7284362   1.6017314   -0.05697238 -0.48255438 -1.5746205
 -2.701416    1.2413456   2.1664574    1.5282022  -0.21666825  3.2362263
  2.9088345  -0.04858487  1.2692846    0.7647497  -0.36702266 -1.2758273
  2.5526013   2.0355692  -1.7551188    0.14641859]
```

Expected Results

The model is expected to be able to predict whether a piece of text is clickbait or not with reasonable accuracy.

It should be able to perform clickbait analysis on the given datasets with comparable, if not better, accuracy as the existing techniques.

Datasets

- Clickbait Challenge - link
- FNC Dataset - link

Thank You :)