

AUTOMATIC HEADLINE GENERATION (AHG) FROM SINGLE BENGALI DOCUMENT USING LSTM

*Abdur Rahman, Satyajit Sen, Mohammed Armanuzzaman Chowdhury
Department of Computer Science and Engineering,
Port City International University

ABSTRACT

A headline is viewed as a dense synopsis of a record. Rather than perusing each record, the feature can be utilized to choose which ones contain significant and applicable data. The aim of this research is to generate Headline of Bengali News from any news article and objective of our work is to save time (as headline quickly and briefly draw attention to the story). With the help of Long Short-Term Memory (LSTM), Bidirectional encoder and decoder with attention mechanism, it is possible to predict news headlines like more humanly. In our work, we emphasize on Bengali News Headline generation from single Bengali news document, where a dataset of 1565 Bengali News Documents is used. For generating news, headline like more humanly, Long Short-Term Memory (LSTM) Bidirectional encoder and decoder with attention mechanism is used as a model. For testing the performance of generated headline, ROUGE matrix is used. The performance of the model is evaluated on the basis of Precision, Recall, and F1 score. For finding similarity between human generated headline and model generated headline, F1 score above 0.50 is considered as a threshold value for valid headline. Hence, obtained rate of accuracy is 62.61%.

Key words: LSTM, ROUGE, Bengali News, Single Document, Automatic Headline Generation.

INTRODUCTION

Language is one of the archaic original media or tools of civilization, which is being enriched day by day in the course of evolution. Just as the previous language has been enriched, it has enriched the standard of living. There is an ocean of data available online and increasing rapidly. People are searching with search engines for their coveted information. In some case they are getting what they absolutely require and some of the time they are overpowered by comparable kinds of copy data. For a seeker of selective information, it's really hard and time-consuming to experience all the related documents available on the web. Automatic Headline Generation (AHG) may be very efficient for these types of circumstances. Anyone can go through a headline of a large document and ensure whether that is effective for him or not. Lately, it has ended in a part and parcel for opinion monitoring, indexing efficiency, recommendation of news and blogs.

There are various types of headlines which are prefaced based on the operation. Headline creation can be of two types depending on the input document. When the headline is generated from a single document it can be described as a single document text headline adversely if the input is similar types of two or more documents it is named as multi-document text headline. A headline can also be described by its nature such as an extractive headline is created by drawing out the most momentous points from a document without modification. On the contrary, abstract headline represents the main gist of an article by reproducing significant parts from that article (Alexander M. Rush, 2015). An extractive headline does not include any text that is not part of the original article but an abstractive headline can add new words or sentences into the headline relevant to the source text. Thus, the greater part of the cases implementing extractive headline generation method is easier than abstractive summarization. Various researchers work with many Bengali News articles connected dataset for

*Corresponding Author: Abdur Rahman; csetanim11@gmail.com