

Abstrak

Clickbait Text Style Transfer Based on Encoder Decoder Model for Indonesian News Headlines Using Delete, Retrieve, Generate Approach

oleh:

Nurrizky Imani
18/429294/PA/18685

Perkembangan media digital telah menghasilkan peningkatan konten online di internet, yang mendorong industri berita untuk menarik perhatian pembaca dengan membuat konten yang menarik. Salah satu pendekatan untuk mencapainya adalah dengan memasukkan elemen clickbait ke dalam judul headline. Saat ini, menghasilkan judul yang menarik dilakukan dengan menggunakan teknik yang mengandalkan keterampilan manual manusia, yang dapat menjadi tidak efisien dan memakan waktu. Sementara itu, kemajuan dalam pemrosesan bahasa alami telah mengarah pada peluang untuk mengotomatisasi teknik pembuatan teks yang dapat mentransfer gaya teks.

Dalam penelitian ini, kami menyajikan implementasi untuk mentransfer kalimat non-clickbait ke kalimat clickbait untuk headline berita Indonesia. Penelitian ini menggunakan metode prototype editing, yang mendemonstrasikan penerapan pendekatan Delete, Retrieve, Generate berdasarkan jaringan syaraf tiruan untuk bahasa Indonesia. Dengan menggunakan kumpulan data headline berita Indonesia dengan 15.000 headline beranotasi yang diberi label sebagai clickbait dan non-clickbait, penelitian ini menerapkan BLEU dan metrik kerugian untuk menilai hasil akhir transfer gaya teks.

Hasil penelitian ini menunjukkan potensi pendekatan Delete, Retrieve, Generate dalam transfer gaya judul artikel clickbait berdasarkan model DeleteOnly dengan hasil terbaik yang dicapai dalam validasi kedua metrik BLEU yaitu 0,1628 dan loss 4,5428. Namun, penelitian ini juga menunjukkan kemampuan terbatas saat menggunakan pendekatan DeleteRetrieve dalam hal mengambil atribut clickbait, dengan hasil yang hanya mencapai metrik BLEU 0,0022 dan loss 7,976 pada validasi. Kedua model menunjukkan hasil kinerja yang berbeda antara data latih dan validasi, yang menyoroti tantangan yang ditimbulkan oleh jumlah data yang terbatas.

Keywords: Clickbait, Text Style Transfer, Encoder Decoder Model, News datasets

Abstract

Clickbait Text Style Transfer Based on Encoder Decoder Model for Indonesian News Headlines Using Delete, Retrieve, Generate Approach

by:

Nurrizky Imani
18/429294/PA/18685

The growth of digital media has led to an increase in online content on the internet, which has challenged the news industry to grab readers' attention by creating engaging content. One approach to achieving this is to incorporate clickbait elements within headline titles. Currently, generating engaging titles is done using techniques that rely on manual human skills, which can be inefficient and time-consuming. Meanwhile, advancements in natural language processing have led to the opportunity to automate the text generation techniques that can transfer text style.

In this research, we present an implementation for transferring non-clickbait sentences to clickbait sentences for Indonesian news headlines. The research uses a prototype editing method, which demonstrates the application of the Delete, Retrieve, Generate approach based on encoder decoder models for the Indonesian language. Using an Indonesian headline news datasets with 15,000 annotated headlines labeled as clickbait and non-clickbait, the research applies BLEU and loss metrics to assess the end results of text style transfers.

The results of this research show the potential of the Delete, Retrieve, Generate approach in clickbait headline style transfer based on the DeleteOnly model with the best result achieved in validation both BLEU metrics 0.1628 and 4.5428 loss. However, it also shows limited ability when using the DeleteRetrieve approach in terms of retrieving clickbait attributes with result only achieve BLEU metrics 0.0022 and 7.976 loss both validation. Both models show contrasting performance results in training and validation, highlighting the challenges posed by the limited amount of data available.

Keywords: Clickbait, Text Style Transfer, Encoder Decoder Model, News datasets