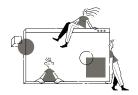
einigen & s SUMMARIZING REDUCES PERUSING TIME Spaltungen in zwei wichtigen #zirken werden sich wohl auf das hat Erissilakou: Grun wahlt man aus celten: For diverzeugung: weil man ein starviogars Team braucht, wenn der Park Kickl Eck verschwindet oder man at für Fluglärm nicht schlafen Dake vie für Verhindeusionen? schor TEXT SUMMARIZATION WRITEUP aftsführer her setz gen ell man etnun möchte: Frühjahr hat einen genung geme a der ganzen CO:produkt (B). Wir wollen zeigen, das Vergleich zum viel machen." Dass es bel Das Wachstum Brandlands nicht immer nur um n einem sehr niedrige. Toyota ruft Autos zur Der japanische Autoherste yota muss wieder Autos zu ammergut. ertel bend 35,000
and dex see ja übeljemen und laufend
sinde dex se ja übeljemen und laufend fen. In den USA müssen Autos der Modelle Corol pylertel beund 35,000 PREPARED Sollen wir das aus dieseisordnungen von 15 Rania Almneie dass sie das, was sie sind, en, dass Fußgäng Hanadi Alshahrani dass sie das, was sie sind.
werden, indem man sie berung inkompati
sie sein können oder sein se denstecht
sie das nicht mehr, weil sie denstecht
Scherben, in Splittern) an ihr Endruge Najd Alqahtani ST. US-Verkehrssicherheitsbel

Abstract

When you open news sites, do you just start reading every news article? Probably not. We typically glance the short news summary and then read more details if interested. Short, informative summaries of the news is now everywhere like magazines, news aggregator apps, research sites. So we built a model to summarize the text.

We get the data from the Kaggle website, then apply the EDA to it, and prepare it for modeling by doing things like expanding contractions, removing punctuation, removing stopwords from text, and so on. For more information, see the following components:

Design



As a group, we wanted to analyze the news text data in order to build the model that summarized it. Instead of manually summarizing the text of the news, you can save time and effort by automating the process.

Algorithms



The first step toward achieving our goal was EDA and data preparation, which included removing accented and unnecessary characters, stopwords, numbers, and punctuation from text, among other things. Following that, we created Seq2Seq models using only LSTMs, Bidirectional LSTMs, and hybrid architecture.

Data

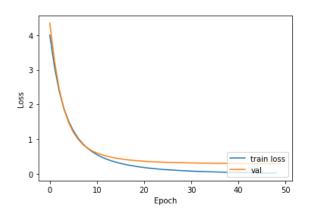


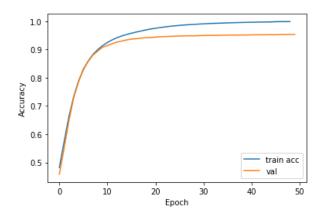
We used two datasets downloaded from the Kaggle website and combined them to create a single dataset with 100132 observations. The data has two features:

- Text: news text
- · Summary: the summary of the text

Communication







We have built 3 models (Seq2Seq with LSTMs,Seq2Seq with Bidirectional LSTMs,Seq2Seq with hybrid architecture), and the best model is Seq2Seq with Bidirectional LSTMs. As shown in the figure, the accuracy of the validation data is close to one and the loss is very close to zero.

Tools

