AirLine Tweets Sentiment Analysis using RNN

Aarcha Jayachandran Prasanna kumari JYCRCH97S517222F@STUDIUM.UNICT.IT
Monci Mamachan[1] MMCMNC96R60Z222S@STUDIUM.UNICT.IT

1. Model Description

Social media such as Twitter created since 2006 is increasing day by day with a wide-spread of customers expressing their feedbacks about goods and services. RNN is a class of neural nets that remember its inputs due to an internal memory that involves sequences of data, such as, text, genomes, or numerical time series data. RNN is a class of neural nets that remember its inputs due to an internal memory that involves sequences of data, such as, text, genomes, or numerical time series data.

2. Dataset

We applied sentiment analysis for airline customers and classified the reviews as positive or negative. The dataset is taken from a Kaggle repository A sentiment analysis job about the problems of each major U.S. airline. Twitter data was scraped from February of 2015 and contributors were asked to first classify positive, negative, and neutral tweets, followed by categorizing negative reasons (such as "late flight" or "rude service").Twitter has a basic length of the tweets as 140 characters. Tweets have positive, negative, neutral (neither positive nor negative) which helps customers to choose better service airlines. In this work, 14640 tweets of six different US airlines such as Virgin America, United, US Airways, Delta and Southwest dataset taken from Kaggle released by CrowdFlower in CSV form. The percentage of tweets is higher in the negative of 91.78%̇The unstructured information (assessments) on a specific site is seen by the clients and consequently making a picture about the items or administrations and thus at long last producing a specific judgment.The reason for extracting information from tweets is that the success or failure of the particular service or product is analyzed by the way how people responded to that service or product.

3. Training procedure

The dataset contains 14,640 tweets. It contains other features also. But we consider only text tweets for our analysis. Based on the given tweet, we need to classify it as a positive or negative review First, we preprocessed the dataset using various natural language processing techniques.Later we divided the dataset into training and testing sets.Next,we prepared word embedding.We applied SimpleRNN on the training set and applied the model on testing data.Each hidden layer has its weights and sigmoid activation function. Recurrent neural networks deal only with short-term dependencies, whereas it cannot handle long dependencies by vanishing gradient problems during back propagation through time. ht=f0(Whhht1+WxhXt)

f0 is the tanh function for ReLu (Rectified Linear Unit) function. ht1 is the preceding state of the word. The input for the current state at a time (t) is Xt

We added one embedding layer, one SimpleRNN layer, and the output layer. The error function used in the model id "binary crossentropy". As this problem is a binary classification problem, we applied "sigmoid" as an output function.

4. Experimental Results

The dataset contains14,640 tweets. It contains other features also. But we consider only text tweets for our analysis. Based on the given tweet, we need to classify it as a positive.negative review First, we preprocessed the dataset using various natural language processing techniques.Later we divided the dataset into training and testing sets.In this paper, we applied deep learning methods for airline tweets classification As the dataset contains text data, first we applied various natural language processing methods and prepared a preprocessed data, we applied simple rnn model and achieved accuracy of 89%.