Multi-class Sentiment Analysis using Deep Learning

JAY NAIMESH PATEL
Student Id: 1111407
MSc in Computer Science
Lakehead University
Email:jpatel48@lakeheadu.ca
Guided By: Dr. T.Akilan

***Abstract*—Some traditional approaches of sentiment analysis method ignores some contextual information. It's always a difficult task to achieve acceptable results in semantic realization. In the recent era deep learning has achieved a milestone by getting excellent results in the field of sentiment analysis using Convolution Neural Network (CNN) and Recurrent Neural Network (RNN). The literature depicts a detailed research on text-based movie reviews Multi-class Sentiment Analysis using CNN. To carry out the research a Rotten Tomatoes raw trained data were used to calculate the accuracy, loss, precision and recall.**

***Index Terms*—Sentiment analysis, Movie Reviews, Natural Language Processing, Machine Learning, Deep Learning, Neural Network, Pre-processing**

I. INTRODUCTION

In the era of Big data, the amount of data like, video, audio and text have been found rapidly increase. Among all types of data, the text data is largest, and studies related to text analysis have been performed on large scale.

Movie reviews are an essential way of evaluating a film's success. Although quantitatively providing a numerical rating to film teaches about the performance of film, a compilation of film reviews is what offers us a deeper qualitative perspective into different facets of the film. A text film review teaches us about the strength and weakness of the film and deeper analysis on movie reviews can tell us whether the film reaches the reviewer's standard. The study of getting semantic information from text is called sentiment analysis.

By definition, a sentiment analysis means use of the given text and analyzing those text to get the semantic understanding of the text. It is also known as opinion mining. The field of sentiment analysis is a part of natural language processing and text mining. The opinions are used for appropriate actions such as decision making in marketing, business expansion, and many more.

In order to comprehend the mining and analyzing of the massive data being exchange and produced on regular basis the regular machine learning techniques and Neural Networks were not sufficient, hence deep learning was the key.

Deep learning is sub section of machine learning. The normal neural network comprises of single network with an input, hidden and output layers. Whereas, Deep Neural Networks consist of multiple neural networks where the output of one network is feed as an input to another network and so on. Due to this type of architecture it resolves a limitation of number of hidden layers in Neural Networks which leads into working with such data more efficiently and feasibly. Deep learning networks learn features on its own and can extract features from textual, audio and video types of data. Deep learning networks have been used in textual sentimental analysis.

II. LITERATURE REVIEW

The proliferation of social media has provided unparalleled incentives for people to share their viewpoints freely, but has generated extreme bottlenecks when it comes to making sense of those beliefs. Around the same time, the importance of getting a real- awareness of citizens ' interests has grown: due to the widespread existence of social networking (where information is spread quite unevenly and rapidly) certain things easily and unpredictably become relevant by word of mouth. The World Wide Web has opened up several different avenues of human experiences.

Users will express their thoughts about various subjects and share viewpoints with other members, journals, forums; web groups are also areas where people will compose their feelings. Classical technology in the categorization of text pays particular attention to deciding if a text is relevant to a specific subject, such as schooling, economics, entertainment etc. Work continues on however, a small concern focuses on how to identify the contextual orientation of the language. Analysis of sentiment is one form of study of quantitative linguistic means of sentiment is the job of recognising positives and negative feelings, emotions and assessments.

III. PROPOSED MODEL

The raw trained data of Rotten Tomatoes were used to train and test the 1D Convolution neural network model. The Rotten Tomatoes movie review dataset is a corpus of reviews which is used for sentimental analysis. The dataset is splitted into training and testing for analysis purpose. The review contains Phrase Id, Sentence Id, Phrase and Sentiment. Each Phrase is associated with a Sentiment. The sentiment number ranging from 0 to 4 associating different sentiments with each number.

The libraries used to processed the dataset and to get better accuracy are:

1) NLTK: It helps to pre-processed, analyze and understand the text.

2) Keras: It is efficient and powerful open source package to develop and evaluate deep learning models.

3) Sklearn: It is a useful library for machine learning. It consists of many efficient tools for statistical modeling and machine learning such as classification, regression, clustering and dimensionality reduction.

4) Pandas: To manipulate the numerical data from dataset.

The method of translating data into something which a machine may comprehend is called pre-processing. Preprocessing steps needs to be carried out in order to overcome the fitting issue. Mainly, the preprocessing steps involves tokenization, stemming, vectorization. But before proceeding further, the data from dataset needs to be filter. The dataset comprises of sentences or maybe just a single word as a part of movie review. So, only the full sentence needs to be taken for our training and testing purpose. The dimensions of dataset is also reduced.

A. Pre-Processing Steps

1) Tokenization: Tokenization is the method of tokenizing a string or breaking it into a set of tokens. One may think of token as pieces like a word is a token in sentence, and a sentence in a paragraph is a token. Appendix VII.A showing the code to get tokens from phrases column of dataset and appending those tokens into a single list with their sentiments.

2) Stop Words Removal and Punctuation Removal: The method of removing most commonly used words such as "the, an, etc." is known as stop words removal. For purpose of classifying or analyzing these stop words don't add value to the document's meaning. Removing stop words is an important aspect in text classification and sentiment analysis as more focus can be given to those words which defines a meaning of text. The punctuation is removed in order to get those words which comprise any meaning related to that aspect.

3) Stemming: Stemming a process of removing the affixes to suffixes and prefixes to roots words globally known as lemma. A root form of any word is called as lemma. Stemming is required in sentiment analysis because, different words comprises of different sentiment values comes under same stem. An example from page is closeness and close is stemmed into the same term, where one term has a positive sense at the first place and the other has negative. There two methods on which analysis is done like Porter Stemmer and Lancaster Stemmer.

4) Embedding: Processing of any textual data or extracting information from text requires a technique to convert strings or given text into a real number (vector) this process of calculating real number from a word is known as Word Embedding or Word Vectorization. Pre-trained techniques are used to get vector forms of each words, the method is Term Frequency-Inverse Document Frequency (TFIDF). Appendix.

$$W_{i,j} = tf_{i,j} \times log(\frac{N}{df_i}) \quad (1)$$

Where,

$W_{i,j}$ = TF-IDF weight for token i in document j
$tf_{i,j}$ = Number of occurrences of token i in document j
df_i = Number of documents that has token i
N = Total number of documents in the training corpus

B. CNN Model

Defining a convolution model in order to train and test the movie review dataset. Keras is used to create 1D convolution neural network. Before making the model the data needs to be converted from to raw data to numpy array. After this the input shape is changed as such it can fit in our proposed model. Convolution layer, Maxpooling layer and flatten layer are used to processed the input. The appendix VII.C shows the code regarding the implementation of proposed model.

The batch size of 128 is used to define our model, moreover the best optimizer found from experimental analysis such as Adam is used and the suitable learning rate is chosen to have the appropriate accuracy for our model.

C. Saving a Model

The benefits of saving a model is to resume the model from where it is left off and avoid unnecessary training times. The code showing the saving of model can be found at appendix VII.D.

TABLE I
HYPER-PARAMETER VALUE

HYPER-PARAMETER	VALUE
Learning Rate	0.01
Epoch	50
Optimizer	Adadelta
Batch Size	64
Number of Layer	2
Kernel Size	5
Filters	64

IV. EXPERIMENTAL ANALYSIS

In this section, we compared some empirical performance using different aspects such as comparing the different batch size values, trying different optimizers for getting the best accuracy, furthermore even some impact was also seen when learning rate was changed.

A. Batch Size

The evaluation of error gradient from the number of trainable data is hyperparameter for learning algorithm is called "Batch Size". The larger the batch size the poor the generalization is. Table II shows the impacts on accuracy and loss for different batch sizes.

TABLE II
COMPARISON ON DIFFERENT BATCH SIZE

BATCH SIZE	LOSS	ACCURACY
256	1.0177	0.6112
128	1.0232	0.6094
64	1.010	0.6126

B. Optimizers

Optimizers are an algorithm or methods through which losses in a neural network can be reduced by changing some attributes such as learning rates. Table III shows the comparison on various optimizers.

TABLE III
COMPARISON ON DIFFERENT OPTIMIZERS

OPTIMIZER	ACCURACY	LEARNING RATE
Adam	0.5978	0.01
Adadelta	0.6126	0.01
SGD	0.60123	0.01

C. Evaluate Parameters

The recall is defined as relevant search found at an instance from dataset. Precision is a measure of result relevancy. The model got an appropriate precision and recall for the dataset.

V. APPLICATIONS

CNN models are used in search database extraction, search query extraction, estimation, selection, typical NLP functions, and so on. 1-D convolutions are often used by unsupervised models on time series inside the frequency domain to find anomalies in the time domain.

Sentimental analysis is an analysis of text data to extract features from it such emotions that are meant by that text. Using the sentiment analysis the business professionals can identify the customer's sentiment towards any brand or product which the business professional can use as a feedback through which betterment of product takes place.

The secret to operating a profitable enterprise with the data on emotions is the opportunity to allow use of unstructured data with actionable perspectives. However, deep learning is better choice than the conventional machine learning algorithms as deep learning provides automatic feature extraction from data.

VI. CONCLUSION

The proposed model was designed to predict the sentiments for the movie reviews corpus available from Rotten Tomatoes. Many comparisons were done on different tyoes of optimizers, batch sizes and vectorization techniques. And at the final stage of analysis i found a model that was able to provide best accuracy. The model uses a batch size of 256, and 2 convolution layers are used excluding the input layers. Relu and Softmax activation functions were used to get the best accuracy for proposed model. It uses Adadelta as a optimizer. Moreover, the model comes up with a suitable accuracy of 0.6126 having 1.010, 0.6565, 0.5892, 0.5360 values for loss value, precision, f1 measure and recall respectively.

REFERENCES

[1] https://machinelearningmastery.com/display-deep-learning-model-training-history-in-keras/
[2] http://sentiment.christopherpotts.net/stemming.htmlporter
[3] https://github.com/cacoderquan/Sentiment-Analysis-on-the-Rotten-Tomatoes-movie-review-dataset/blob/master/train.tsv
[4] https://towardsdatascience.com/choosing-the-right-encoding-method-label-vs-onehot-encoder-a4434493149b
[5] https://en.wikipedia.org/wiki/Tf
[6] P.D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02), pp. 417-424, 2002

VII. APPENDIX

A. Pre-processing

```
for l in range(len(documents)):
 label = documents[l][1]
 tmpReview = []
 for w in documents[l][0]:
  newWord = w
  if remove_stopwords and (w in stopwords_en):
   continue
  if removePuncs and (w in punctuations):
      continue
  if useStemming:

   newWord = lancaster.stem(newWord)
  if useLemma:
   newWord = wordnet_lemmatizer.lemmatize(newWord)
    tmpReview.append(newWord)
  documents[l] = (tmpReview, label)
  documents[l] = (' '.join(tmpReview), label)
```

Listing 1. Pre-Processing Steps

B. Embedding

```
from sklearn.feature_extraction.text import
    CountVectorizer ,TfidfVectorizer
from keras.utils import to_categorical

vectorizer = TfidfVectorizer(max_features = 2000,
    ngram_range=(2, 2))
X = vectorizer.fit_transform(df["text"])
Y = df['sentiment']

X_train = vectorizer.transform(X_train).toarray()
Y_train = Y_train
X_test = vectorizer.transform(X_test).toarray()
Y_test = Y_test
```

Listing 2. Vectorization

C. Model

```
model = Sequential()
model.add(Conv1D(filters=64, kernel_size=5,
                 activation='relu',
                 input_shape=(2500,1)))
model.add(Conv1D(128, kernel_size=5, activation='
    relu'))
model.add(MaxPooling1D(pool_size=1))

model.add(Flatten())
model.add(Dense(num_classes, activation='softmax'))
```

Listing 3. Convolution Neural Network Model

D. Saving The Model

```
model.save('/content/drive/My Drive/1111407.h5')
```

Listing 4. Model Save