Multiclass Sentimental Analysis for Movie Reviews

NAGA TEJA GUTTIKONDA
MSc Computer Science
Lakehead University
nguttiko@lakeheadu.ca

***Abstract*—: Among the machine learning approaches the deep learning has been trending recently with lot of attention in the domain of user behaviour analysis. The process of studying the user behaviour from the unstructured movie reviews is a challenging task. Multi class classification with deep learning approach is implemented to analyse the user behaviour.Word2Vec model is used for feature vector conversion .Keras library is used for the implementing convolutional neural network model.Dataset of 156060 records is used where 7:3 split ratio is used for training and testing the model.The implementation results reveal that proposed model can reach up to 65 percent accuracy , F1 score of 0.67 , Precision is 0.63 and Recall is 0.62 on multiclass based prediction in machine learning. The results can help the people movie industry to know the feedbacks of movie and further steps to be taken in future.**

***Index Terms*—Sentimental Analysis , Movie reviews , Neural networks , Word2Vec , Bag of words , TF-IDF , Sentiment**

I. INTRODUCTION

With its impressive success the emergence of deep learning has sparked a paradigm shift in Machine Learning. There have been various studies to enhance the accuracy of user behaviour mining from basic linear models to complex modern deep approaches.

Social networking has revolutionized the internet by turning users from passive information receivers into contributors and influencers of data. It has direct impact on various business products currently in the market. Negative reviews given by users may affect the rate of success and survival of any business in market. The sentiment analysis is formally defined as the task of identifying and analysing subjective information about people's opinions in social media sources. In recent times it has gathered lot of attention due its implications with business. The challenges in this model are 1)There are large number of varieties in expressions to indicate the range of sentiments. Therefore, how long the dictionary of sentiments is it cannot list the all possible ways in which people represent their feelings 2) Words may have different meanings depending on the context. For example, "short size "mostly indicates the negative opinion, whereas "short wait time "indicates a positive one 3) Users will not express their opinion directly they could be indirect and confusing 4)The dependencies between various constituents of sentence might be at the long distance. For example, the scope of negation and polarity of adjective in the given sentence cannot be determined.

Users might express the sentiments at various range of expressions with different degrees but as the part of task simplification sentimental analysis approaches classify the sentiments into either positive, negative or neutral

Sentiment analysis might be done at various levels

a)Document level: At document level, classification is called document-level sentiment classification of all opinion records that may have negative or positive sentiments. For example, this involves predicting negative and positive opinions on product reviews under all comments and feelings written or spoken by an opinion holder. Consequently, it can not be used to compare and analyse specific individuals. At this point the document is believed to represent an person or a commodity.

b)Sentence level: Classification is done at sentence level by analysing each sentence to determine whether the sentence consists of a neutral, negative or positive emotion. The sentence-level classification is performed in two steps. The first step is to classify sentences into objective or subjective categories; thus, subjective sentences are graded to a positive or negative feeling. Sentences may then take two forms: either factual or subjective. Therefore, subjectivity is not about emotion, but an objective statement may mean decision.

c)Aspect level: Aspect-base or feature-base levels are composed of either negative or positive emotions based on ideas and opinions. They result in a target of opinion that allows us to better recognise problems of sentimental research. Many consumers believe they can make better decisions when selecting and purchasing items based on reading others ' experiences in web reviews

An entity named sentiment lexicon is used in various projects for the sentiment classification These words show both the positive and negative meanings. Even though these are important in classification they might become problem sometimes under certain instances. There are three approaches to make the sentimental lexicon such as manual making, dictionary based, and corpus based. Manual making is very rarely used because it takes lot of time making the task tedious. Dictionary based is centered by small set of sentimental words and rely on online lexical resources such as WordNet. It is carried out in two steps. In first step we manually collect small set of sentimental words known as the polarity sentiment. Under second step the tree is grown further by setting up their synonyms and antonyms in the online dictionary. Corpus based approach initially use a set of words of sentiment with polarity sentiment and expand this by helping syntactic patterns to find the new words of sentiments in large corpus A mathematical approach is proposed for the measurement of sentiment score a method for the generation of vectors

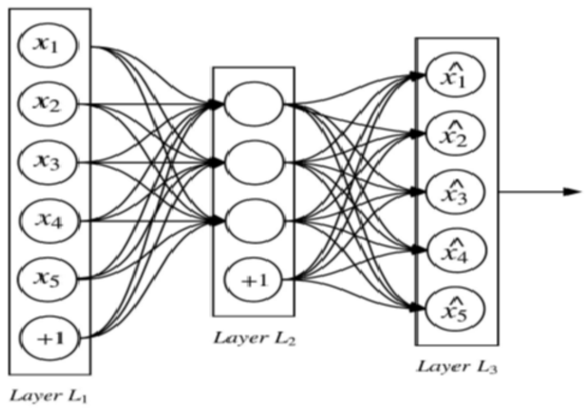

Fig. 1. Autoencoder Network
Source : Adapted from [5]

is provided for the categorisation of sentiment polarity. Two experiments in polarity categorization are conducted based on the sentence level and the test level respectively

Deep learning is the application of artificial neural networks (short neural networks) to learning tasks using multilayer networks. It can leverage much more neural network learning (representation) capacity, which was once deemed feasible only with one or two layers and a small amount of data.

We can use various types of neural networks for sentimental analysis namely Auto encoder network: It is a three-layer network which sets the target values equal to the input values its ability to learn non-linear representations make it more expressive than PCA (Principal component analysis)

We can use various types of neural networks for sentimental analysis given below:

1)Auto encoder network: It is a three-layer network which sets the target values equal to the input values its ability to learn non-linear representations make it more expressive than PCA (Principal component analysis

2) Convolutional Neural network : CNN is a special type of feedforward neural network which is inspired from the human visual cortex where lot of cells together are responsible for producing light in smaller areas known as receptive fields

Configuration of CNN is formed by using following parameters:

a) Number of hidden CNN and NLP Layers

b) Size of kernel in each layer

c) Factor of subsampling in each layer

d) The choice of selecting activation functions

Various terminologies used in CNN:

a) Convolution layer: It uses filter to perform convolution operations instead of matrix multiplication for scanning the input with respect to its dimensions

b) Pooling Layer: It is particularly applied after the layer of convolution to perform down sampling

c) Fully connected layer: This is the type of layer which receives the flattened input where each input is connected to all the neurons present.

Commonly used activation functions:

a) Rectified Linear Unit (reLU): It is used to introduce the non linearities into the network because the real-world problems are not at all linear in nature.

b) SoftMax: It is seen as generalized logistic function where vector of scores is taken as input to give the output in the range of 0-1

2)Auto encoder network: It is a three-layer network which sets the target values equal to the input values its ability to learn non-linear representations make it more expressive than PCA (Principal component analysis

II. Literature survey

(Mohammad ettia et al.,) A language independent model for multi class sentimental analysis using a simple neural network using 5 layers.It does not rely on language specific parameters such as ontologies or dictionories. In this paper we presented our multi-class sentiment classification systems and found that the deep neural network model can outperform traditional methods based on language-specific feature engineering.It performs well for ASTD dataset which has the 5 classes.Through the class imbalance can lead to decrease in system performance and finded that oversampling will be useful to deal this problem.And also through results it is finded that system performance could be impacted by the annotation quality. This CNN layer has 300 filters and a width of 7, which means that each filter is trained to detect a certain pattern in a 7-gram window of words

(Krishna kumar mohbey et al., 2019) For this Tweets relevant to India's general elections 2019 is used as the study's data corpus. For evaluating the user view, multi-class classification made with a novel deep learning approach is implemented. We used nine different groups here, which for the election agenda reflects larger issues in the country. In addition, a comparative study is provided between conventional approaches such as Naïve Bayes, SVM, decision tree, logistic regression and deep-learning approach employed. Experimental findings show that the proposed model can achieve accuracy of up to 98.70 in machine learning based on multi-class prediction. Each parameter is set to their values of default in which programmer can change it based on requirement. y. Now if the same model is applied with the same default value parameters to different datasets then a good result will not be produced uniformly for all cases. Therefore, we need to adjust the parameter in such a way that the mode can be used for each specified information collection. The developed approach to deep learning achieves 98.70 percent good accuracy results compared to conventional methodology. As a part of future work, we will apply user analysis approach in new domains of the industry

(Saman Ghili Et al., 2017) Two different datasets are taken one with binary labels and another with multiclass labels. For the purpose of binary classification bag of words and skipgram models followed by various classifiers are used. In the case of multi class we have implemented recursive neural tensor networks with the over expense of RTNN low rank RTNN is introduced. The challenge faced with first part of data is losing the order of words in sentences due to failure in aggregating word vectors. The accuracy is improved by 1.5 percent at the end using ensemble averaging technique

(Qin Li Et al., 2015) We proposed the recurrent model of the BiGRULA (bidirectional gated recurrent unit for sentiment classification) neural network with topic-enhanced word embedding and an attention mechanism for classifying sentiments. We applied our BiGRULA model to an overview of real-world hotel feedback and demonstrated its ability to extract important topics.We evaluated and demonstrated the advantage of using subject-enhanced word embedding based on the lda2vec model for document classification compared to other text representations. We assessed and compared BiGRULA's efficiency in the classification of sentiments with other neural network models. Our algorithm achieved 89.4 percent accuracy, with an increase of 3.0 percent over the highest of three reference algorithms. We applied our BiGRULA model to an overview of real-world hotel feedback and demonstrated its ability to extract important topics. Context vectors are word vectors used in our BiGRULA model that are topic / meaning-enhanced. They were determined as follows: first, given a focal word in the text corpus, five target words were selected in a moving window behind and after the pivot term. The cycle has been replicated in the entire corpus.Adding the sentence level attention is part of our future work

(Xing Fang Et al.,) A word token is a positive (negative) word and the part-speech tag. In total, we picked 11,478-word tokens with each of them occurring at least 30 times over the entire dataset. For phrase tokens, out of the 21,586 known sentiment phrases, 3,023 phrases have been selected and each of the 3,023 phrases also has an occurrence not less than 30.We obtain the F1 score of 0.85 on manually labelled sentences where with help of ROC curve we can see that all three models perform very well. The main limitation is since the scheme relies highly on sentiment tokens it may not work well for reviews which contain implicit sentiments. Future work includes testing our data with more datasets and finding the sentiments within the scope of individual product.

III. Implementation

Dataset : First we take the dataset which has 156060 rows and 4 columns namely PhraseId, SentenceId , Phrase , Sentiment and we will print the first 10 rows of data.

You get the following output:

TABLE I
First Ten Records of Dataset

PhraseId	SentenceId	Phrase	Sentence
1	1	A series of escapades demonstrating the adage ...	1
2	1	A series of escapades demonstrating the adage ...	2
3	1	A series	2
4	1	A	2
5	1	series	2
6	1	of escapades demonstrating the adage that what...	2
7	1	of	2
8	1	escapades demonstrating the adage that what is...	2
9	1	escapades	2
10	1	demonstrating the adage that what is good for ...	2

1)Preprocessing : First of all we should perform the preprocessing to clean the data

a) Stop words removal: In this step we extracted from each of the reviews all frequently occurring phrases. A predefined list of stop words has been used for removing stop words. Each review is contrasted with the list of available stop words and the corresponding words are omitted from that review. These terms don't contribute to model efficiency.

b) Punctuations removal: There is a lot of punctuation available in reviews that are meaningless, all the punctuation symbols like.,! ,? In this move, etc., are revoked

d) Lowercase conversion: All the letters in the review are converted to the lower-case format

e) URL removal: Any URL and hyperlinks are removed from the reviews

2)Feature vector conversion: To convert cleaned sequence of words to numeric feature vectors the following methods can be used : Bag of words: It is a simplest way of numerically representing a text. It is a way to determine significant words in each text. After learning the BOW vectors for every review in the labeled training set, we fit a classifier to the data.

Word2Vec: It is another way of numerically transform each word of a text to the vector. It is generally in independent of main objective and does not require labelled dataset. It is computationally efficient predictive model for learning word embeddings from the raw data

3)Analysis : The given data set is analyzed and represented in form of bar graph for better graph

3) Model :We create the neural networks model using Keras .Here we create the model with 3 convolution layers , 3 max pool layers and 1 dropout layer. Word2vec has been imported from genism package and used for the feature vector conversion.Use 7:3 ratio for training and testing using random state of 2003

4) Training : For training the model we use Adam as optimizer batch size 512 ,number of filters is 256 and trained for 30 epochs to obtain the optimal solution.

5)Testing : Data is tested with CUDA GPU and model is runned to calculate the accuracy and F1 score

```
Proposed network
class CnnRegressor(torch.nn.Module):
  filter_sizes = [1,2,3,5]#[1,2,3,5]
num_filters = 256 #224 best
drop = 0.5

#    print("Creating Model...")
inputs = Input(shape=(MAX_SEQUENCE_LENGTH,),
    dtype='int32')
```

```
embedding =
    Embedding(input_dim=len(train_word_index)
    + 1, output_dim=EMBEDDING_DIM,
    weights=[train_embedding_weights],
                 input_length=MAX_SEQUENCE_LENGTH,
                     trainable=False)(inputs)
reshape =
    Reshape((MAX_SEQUENCE_LENGTH,EMBEDDING_DIM,1)
conv_0 = Conv2D(num_filters,
    kernel_size=(filter_sizes[0],
    EMBEDDING_DIM), padding='valid',
    kernel_initializer='normal',
    activation='relu')(reshape)
conv_1 = Conv2D(num_filters,
    kernel_size=(filter_sizes[1],
    EMBEDDING_DIM), padding='valid',
    kernel_initializer='normal',
    activation='relu')(reshape)
conv_2 = Conv2D(num_filters,
    kernel_size=(filter_sizes[2],
    EMBEDDING_DIM), padding='valid',
    kernel_initializer='normal',
    activation='relu')(reshape)

maxpool_0 =
    MaxPool2D(pool_size=(MAX_SEQUENCE_LENGTH
    - filter_sizes[0] + 1, 1), strides=(1,1),
    padding='valid')(conv_0)
maxpool_1 =
    MaxPool2D(pool_size=(MAX_SEQUENCE_LENGTH
    - filter_sizes[1] + 1, 1), strides=(1,1),
    padding='valid')(conv_1)
maxpool_2 =
    MaxPool2D(pool_size=(MAX_SEQUENCE_LENGTH
    - filter_sizes[2] + 1, 1), strides=(1,1),
    padding='valid')(conv_2)

concatenated_tensor =
    Concatenate(axis=1)([maxpool_0,
    maxpool_1, maxpool_2])
flatten = Flatten()(concatenated_tensor)
dropout = Dropout(drop)(flatten)
preds = Dense(5,
    activation='softmax')(dropout)

# this creates a model that includes inputs
    and outputs
model = Model(inputs=inputs, outputs=preds)
```

The main goal of our application is to improve the accuracy which can be done by following ways 1) By adding the dropout layer

2) By trying various optimizers

3) By changing the structure of network

4) By performing data shuffling

5)By trying different batch size

6) By checking with different epochs during training and also by using various data cleaning techniques

IV. Results and discussions

Here the network is tested with various optimizers and feautre vector conversion techniques for improving the performance of the model the loss.The metric such as precision , Recall and F1 score are used to measure the performance of the model. By using Word2Vec technique we obtained the accuracy of 64 percent along F1 score of 0.67 precision of 0.62 and Recall of 0.63. The metrics table is given below

TABLE II
Metrics Table

Iteration	Precision	Recall	F1 score	Support
1	0.60	0.24	0.34	2151
2	0.54	0.43	0.48	8070
3	0.69	0.87	0.77	23987
4	0.59	0.47	0.53	9872
5	0.60	0.30	0.40	2738
Macro Average	0.60	0.46	0.50	46818
weighted avg	0.63	0.65	0.63	46818

V. Conclusion and Future work

By usage of Word2Vec model with basic CNN the Sentimental Analysis has been done.It can be made further efficient by changing the network structure and training with more parameters Future work includes developing the model with better accuracy by doing further data cleaning and modifying network structure

VI. References

References

[1] “A comparative study on sentiment analysis”, Alireza yousefpour,Roliana Ibrahim,Haza Nuzly abdull hamed and mohammad sadegh hajmohammadi, “ Advances in Environmental biology”, American-Eurasian network for scientific information aug 2014

[2] “Multi class approach for user behaviour prediction using deep learning framework on twitter election dataset”,Krishna kumar mohbey,”Springer Nature Switzerland AG 2019”

[3] “Multilingual Multiclass sentiment classification using convolutional neural networks”,Mohammad attia, Younes samiah, Ali Elkahky, Laura kallmeyer

[4] “Deep learning for sentiment analysis of movie reviews”,Hadi Pouransari, Saman Ghili”

[5] “Deep learning for Sentiment Analysis: A survey”,Lei Zhang, Shuai Wang, Bing Liu

[6] “Tourism Review sentiment classification using a Bidirectional recurrent neural network with an attention mechanism and topic enriched word vectors”,Qin li, Shaobo Li, Jie Hu, Seri Zhang, Jianjun Hu,”sustainability 2018”

[7] “Sentiment Analysis using product review data “,Xing Fang and justin zhan,”Journal of big data 2015”.