Multiclass Sentimental Analysis on Movie Reviews using CNN

Srushti Wadekar
Computer Science
Lakehead University
swadekar@lakeheadu.ca

***Abstract*—This paper consists of the detailed and elaborated steps required to generate the classification model for sentiment analysis of movie review dataset. Sentiment analysis are descriptive text mining, which defines and removes intangible knowledge from source content and lets an organisation perceive their name, commodity or service's social feeling by analysing online communication. This paper focuses on the Convolution based neural network method to implement a multi-class classification model. Multi-class classification model is a model where classification task is considered by considering more than two classes. Convolution neural network is a type of deep learning technique which employs the mathematical technique of convolution. After selecting the optimum hyperparameters, accuracy of 68.13 percent and F1 score of 66.79 percent was achieved.**

***Index Terms*—Sentiment, Convolution, Multiclass, Classification**

I. Introduction

Sentiment Analysis is the most common text classification tool that analyses an incoming message and tells whether the underlying sentiment is positive, negative our neutral. A sentiment analysis framework incorporates natural language processing (NLP), with machine-learning techniques, to allocate weighted emotion scores within a paragraph or phrase to persons, subjects, topics and groups. Sentiment analysis allows data analysts in big corporations to gauge consumer sentiment, perform complex market studies, track the credibility of company and brands and consider the perspective of consumers. It represents a large problem space. There are also many names and slightly different tasks, e.g., sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc. It has a large range of applications in nearly all areas. Owing to the explosion of business devices, the industry surrounding sentiment analysis has also excelled. This provides a good research drive. Sentiment/Opinion is fundamental to virtually all human behaviour, because it is essential to our actions. We like to learn the views of others as we must make a choice. Businesses and companies of the modern world often seek customer or general feedback about their goods and services. With the explosive growth of social media (e.g., reviews, forum discussions, blogs, micro-blogs, Twitter, comments, and postings in social network sites) on the Web, individuals and organizations are increasingly day by day, sentiment analysis of these opinions is very crucial.

Reviews influence a lot of people in the growing world of social media and online shopping. Reviews not only have the power to influence consumer decisions but can strengthen a company's credibility. Reviews have the power to gain customer trust, and they encourage people to interact with the company. Hence, getting the sentiment from the reviews if of much importance nowadays. Sentiment from reviews is essential in data analytics field. Multiclass sentiment analysis is when there are more than two class labels apart from positive and negative. In this paper, a multiclass sentiment analysis on movie reviews is performed by considering five class labels or five different sentiments.

II. Literature Review

Sentiment analysis may be rendered on three levels: text stage, sentence and attribute stage (or aspect) [2]. The purpose at document level is to determine the course of view of the whole text. Therefore, the job is to assign each text into one of the positive or negative groups. On the level of the sentence, the goal is to classify the views of the opinion sentences. First, it is normal to define the subjective phrases and then decide each sentence's feeling. The features of the item to which the subject has responded are defined first and then the meaning of the expression is discovered in the aspect sentiment analysis. A procedure used for collecting the attributes of the item and the measures of prepossessing in the planning of the clustering sentences. To define a user feedback, we use the word "analysis" is described in [1]. The purpose behind the usage of clustering strategies is to identify the attributes of the topic which users in the reviews have shared their views. In other terms, because we have links to all phrases in all comments and still consider specific phrases, specific phrases are usually linked to similar things. When and sentence is described by a vector, the sentences in each cluster are identical sentence, and presumably discuss the same dimension of the object, because we use the clustering method for such function vectors.

In recent years Twitter data mining has been a hot research subject. The essence of the data gathered differs considerably according to the purpose and intended outcome. The methods for manipulating data and collecting the information needed are also different. A method to m easure public opinion transition over time and to classify the news that contributed to public opinion breakdowns was proposed in [2]. In a related context, Sriram et al. [3] proposed a method to classify tweets depending on their natures into a set of classes including private messages, opinions and event, etc. Unigrams, bigrams and adjectives in different ways to classify a set of movie reviews into positive or negative. Recently, new models have been built Gao and Sebastiani [4] proposed a new approach based on the distribution or frequency of the types of sentiment they examine. The writers have found that a quantifier algorithm is a better approximation of the frequency than standard classification-driven algorithms from classification to quantification.

A framework for multi-class sentiment classification is proposed, which includes two parts: 1) selecting important features of texts using the feature selection algorithm, and 2) training multi-class sentiment classifier using the machine learning algorithm is proposed in [5]. 10-fold cross validation was used to achieve the classification accuracy concerning each combination of attribute selection algorithm, machine learning algorithm, feature set size and data subset. Based on the obtained 3600 classification accuracies (4 feature selection algorithms ×5 machine learning algorithms ×15 feature set sizes ×12 data subsets), the average classification accuracy of each algorithm is calculated, and the Wilcoxon test is used to verify the existence of significant difference between different algorithms in multi-class sentiment classification.

A Deep learning technique for sentiment analysis of movie reviews was implemented with the help of binary classification method and recursive neural networks in [6]. Binary classification methods have provided fair accuracies but have loosened the order of terms in a sentence, so critical semantics can not be obtained from the reviews of inputs. On the other hand, recursive neural networks allow one to take the word order into account in a sentence. Although RNNs can attain reasonably high accuracy in classification, RNNs may still not communicate relationships, whether negated positives or negated negatives. In RNTN each node in the parse tree is represented by a hidden vector of length d (like word2vec). The hidden vectors at the leaf level (i.e., vectors corresponding to single words) are in fact the word-vectors. For a non-leaf node, the hidden vector is obtained from the hidden vectors of its left and right children. The activation function of the neurons is assumed to be tangent hyperbolic tanh in this case. Using the tanh activation functions helps to overcome the issue of negated phrases. To overcome the overfitting issue of tensor, low-rank approximations of matrices is been taken into consideration.

III. Proposed Model

Artificial neural networks (ANN) compose of simple elements called neurons, which can make simple judgments in the area of mathematics. The neurons can together evaluate complicated issues, imitate almost any feature, including very complex, and provide detailed answers. There are three levels of neurons on a superficial neural network: an input layer, a hidden layer, and an output layer. A Deep Neural Network (DNN) includes more than one hidden layer that increases model flexibility and enhances predictive power dramatically. A neural network can perform as a classification model. Artificial neural networks are relatively crude electronic networks of neurons based on the neural structure of the brain. They process records one at a time and learn by comparing their classification of the record (i.e., largely arbitrary) with the known actual classification of the record. The errors from the initial classification of the first record is fed back into the network and used to modify the networks algorithm for further iterations. In this paper, the classification model is generated using a convolution neural network for multi level classification of movie reviews dataset.

A CNN is a multi-layer neural network used for images processing for image recognition, segmentation or object detection. CNNs work to reduce an item to its key characteristics, and to classify the type using the cumulative likelihood of defined characteristics. The value of CNNs is that they require fewer hyperparameters and less testing relative to other classified algorithms. Given a fixed-dimensional input from the lower layer, the classification layer affine transforms it followed by a relu activation function. Convolution neural networks usually consists of three layers i.e. convolution layer, pooling layer and fully connected layer.

The dataset used for this multi level classification problem is the rotten tomatoes movie review dataset. The two main features taken into consideration while training the model where the phrase or the review and the sentiment of the review. The datatypes and the description of these features is given below in Table.1.

TABLE I
DETAIL ABOUT ATTRIBUTE

Attribute Name	Data type	Description
Phrase	object	The review given to movies
Sentiment	int64	A categorical variable of the sentiment/emotion behind the respective review.

The first layer in the proposed model starts with a convolution neural network layer which takes (2000,1) input features. The kernel size defined in the layer is set to 3. Kernel size is the filter size which is used as a scanning window. In the convolution layers, an input is analyzed by a set of filters that output a feature map. This output is then sent to a pooling layer, which reduces the size of the feature map. This helps reduce the processing time by condensing the map to it's most essential information. The output if CNN layer is passed further to the dropout layer and dense layer.

The first two convolution layer in the model uses relu as the activation function. This enables the generalisation or modification of the model with various data and the separating between the output. The batch normalized output is passed to the max pooling layer in the training model. The purpose of a pooling layer is to gradually reduce representation spatial size to limit parameters and network computation. Pooling layer operates on each feature map independently. The pooling layer summarizes the features by considering the feature map. It operates upon each feature map separately to create a new set of the same number of pooled feature maps. Convolution and pooling procedures are repeated several times and depending on the network, the amount of iterations is then sent to the dropout layer. A dropout layer is used after pooling layer to minimize the overfitting issue of the model as the dataset has class imbalance issue which can lead to overfitting. A flatten layer is used to convert the format of the convolution layer outputs into a linear form that can be passed to the dense layer. Finally, softmax activation function is used on the final output as it gives out probability for every class label and it suits the best for a multi-classification problem.

This modelling allows CNNs to learn the position and scale of features by making them especially good at the classification of hierarchical or spatial data and the extraction of unlabelled features. Keras and panda library is used in the proposed model for data preprocessing and model generation. The data is preprocessed and the first five rows of the preprocessed data is shown in Table.2.

TABLE II
FIRST 5 ROWS OF DATASET

Phrase Id	Sentence ID	Phrase	Sentiment
1	1	A series of escapades demonstrating...	1
2	1	This quiet , introspective and entertaining...	2
3	1	Even fans of Ismail Merchant 's work , I suspect...	2
4	1	A positively thrilling combination of...	2
5	1	Aggressive self-glorification and a manipulating...	2

IV. Experimental Analysis

The training model was implemented on the rotten tomatoes movie review dataset. The dataset gives information about various movie reviews and it's possible corresponding sentiment. It was observed that there was a class imbalance issue in the dataset where many phrases where biased to sentiment 0. Hence, it was neccessary to remove this class imbalance to avoid the model to behave biased and also to avoid the overfitting issue. Plots were generated to visualize the count of phrases in every class label on the raw dataset and the preprocessed dataset. Figure.1 demonstrate the class imbalance issue in the raw dataset of movie reviews. Figure.2 demonstrates the count od phrases in every class labels after preprocessing the dataset.

The dataset is split into the ratio of 70:30 for training and testing with the help of sklearn library. The CNN model takes input in array format, hence the dataset is converted into numpy array. The preprocessing of dataset is done in which the stopwords and punctuation are removed from the phrases or reviews. Also, stemming and lemmatization is performed on the dataset. TF-IDF vectorization is performed on the dataset with maximum features as 2000 to covert the words into vector. Finally the training and testing sets are converted into numpy arrays to work with the keras model.

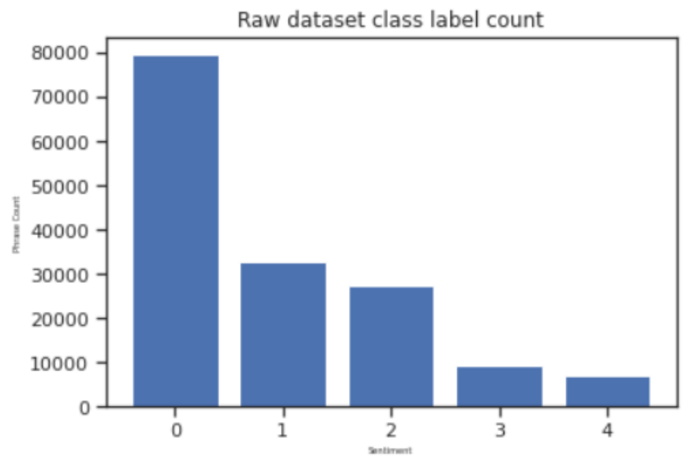

Fig. 1. Plot for the class imbalance issue

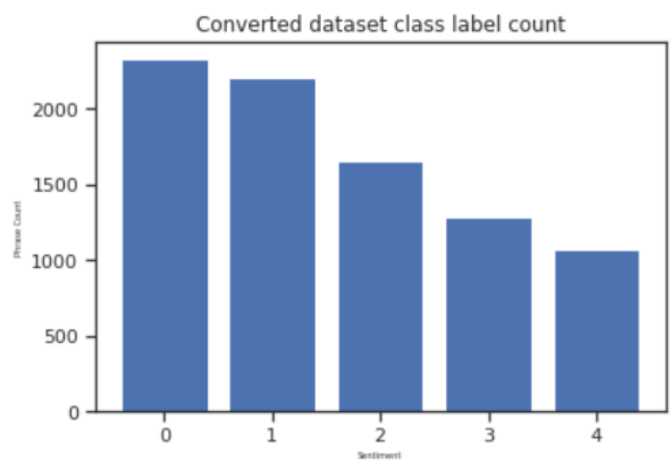

Fig. 2. Plot after converting into full sentences

The classification model was trained using one-dimensional convolution neural network. Batch normalization layer was added after convolution layer to fasten the model training process. It was observed that applying batch normalization on the output neurons from convolution layer increased the R2score which lead to a better model performance. Hence, batch normalization was applied on both the convolution layers in the model. Finally, for the pooling layer, MaxPooling and AvgPooling layer was included in the model. MaxPooling was taken into final consideration due to it's high performance. As the data is highly imbalance, dropout layer was added to avoid any kind of overfitting issue. This classification model resulted in accuracy of 68.13 percent and F1 score of 66.79 percent. Figure.3. shows the training progress plotted for 100 epochs.

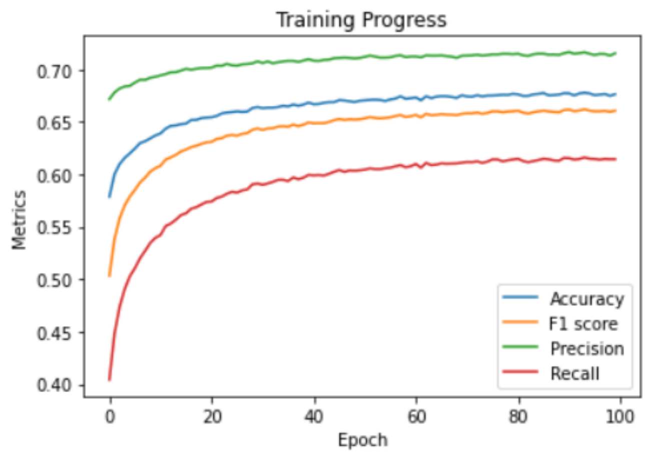

Fig. 3. Training Progress for 100 epochs

A. Training parameters

There were several training parameters such as batch size, kernel size, no of epochs, learning rate, padding and stride rate that were considered while building the prediction model. Batch size of 64 and 128 was considering during training the model where batch size of 64 trained the model in a more accurate way. Number of epochs tried while compiling the model were 10,50,100 and it was observed that increasing the number of epochs gave more accurate results. Filter size of 64 and kernal size of 3 suited the best for training the model. Table.3 shows the optimum value of the hyperparameters that suited the best for the model.

TABLE III
OPTIMUM HYPERPARAMETER

Hyperparameter	Value
Optimizer	Adam
No of Epochs	100
Batch Size	64
No of Layers	5
Loss Function	Cross Entropy

B. Activation function

Softmax activation function was used in the output of final dense layer as it returns the probability of the class labels. Hence, softmax is has significance in training a multi classification model for movie review dataset. Relu activation function is used in the intermediate output from convolution layer to maintain the non-linearity of the solution.

C. Loss function

Loss function for a model is necessary to evaluate the performance and back propagate to adjust the parameters to generate better performance. Various optimizer were used in the loss function to optimize the model at every step. Optimizer such as SGD, AdamW, Adamax, Adam were taken into consideration in the loss functionality. Table.3. shows the performance with respect to accuracy, f1 score, precision, recall and L1 loss by considering different optimizer. Loss function that is used to evaluate the loss is categorical cross entropy. It was observed that evaluation based on cross entropy worked better compared to the L1 loss measure to evaluate the loss of the model. Figure.4 shows the plot for loss generated over 100 epochs.

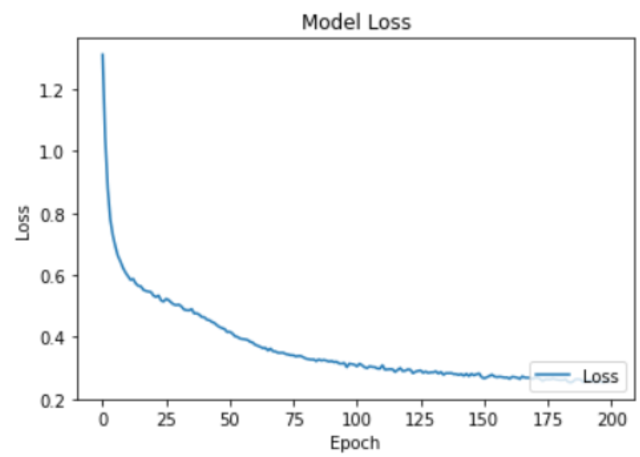

Fig. 4. Plot for loss over 100 epochs

TABLE IV
EVALUATION BASED ON DIFFERENT OPTIMIZER

Optimizer	Accuracy	F1 Score	Precision	Recall	Loss
Adam	68.13	66.79	70.16	63.24	0.76
AdamW	64.76	63.45	67.11	59.23	0.89
Adamax	66.12	65.02	68.79	61.77	0.83
SGD	63.13	61.68	65.87	58.11	0.96

These results lead in selection of Adam as an optimizer while training the model.

V. CONCLUSION

In this paper, rotten tomatoes movie review dataset was used for multiclass sentiment analysis. CNN model was used for building the mode as it uses the information given at a particular instant for analysis. Another reason for CNN performing better is due to the fact that CNN does not depend on historic data for classification. As a result, the model can better estimate the dynamic changes in the training phase which leads to a better classification accuracy. Thus, on analysis and experimentation, a one-layered convolution neural network by application of MaxPooling and dropout layer resulted in a maximum accuracy of 68.13 percent and F1 score of 66.79 percent.

VI. Appendix

The dataset used to train the model and the source code of trained model is present in the given Github link. Link: https://github.com/Srushti10/SentimentAnalysis

References

[1] M. Bouazizi and T. Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter," IEEE Access, vol. 5, pp. 20617-20639, 2017.

[2] Mohsen Farhadloo, and Erik Rolland, "Multi-task CNN Model for Attribute Prediction,"Conference Paper, December 2013.

[3] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, M. Demirbas, "Short text classification in Twitter to improve information filtering," Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., pp. 841-842, Jul. 2010.

[4] Lei Ren,YAQIANG SUN,HAO WANG, and LIN ZHANG, "Prediction of Bearing Remaining Useful Life With Deep Convolution Neural Network,"IEEE special section on Cyber-Physical-Social Computing and Networking, February 2018.

[5] Lei Ren,YAQIANG SUN,HAO WANG, and LIN ZHANG, "Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms,"Expert Systems with Applications Volume 80, September 2017.

[6] Yang Liu, Jian-Wu Bi and Zhi-Ping Fan, "Deep learning for sentiment analysis of movie reviews," IEEE 14th International journal, February 2018.