International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

Sentiment Analysis of Text and Audio Data

Dr. Munish Mehta
Department of Computer Applications
National Institute of Technology
Kurukshetra, India

Kanhav Gupta
Department of Computer Applications
National Institute of Technology
Kurukshetra, India

Shubhangi Tiwari
Department of Computer Applications
National Institute of Technology
Kurukshetra, India

Anamika
Department of Computer Applications
National Institute of Technology
Kurukshetra, India

***Abstract* – Sentiment analysis has shown a significant amount of growth in the past few years. Textual sentiment analysis has been quite common as well as popular amongst researchers. Here we have first studied the approaches proposed by various researchers regarding sentiment analysis of all modalities of data. And then we tried implementing some of those methods. In this paper we have not only focused on sentiment analysis of text but also talked about audio data. Audio sentiment analysis is an area that is still being explored by researchers and hence new techniques are being utilized to analyze audio data. Here we try utilizing deep learning in order to classify audio into various sentiments. For text, we have tested basic machine learning models and have tried to draw a comparison between the results.**

Keywords – Sentiment Analysis; Natural Language Processing; Feature Extraction; Machine Learning; SVM; CNN; Image Classification.

I. INTRODUCTION

Sentiment Analysis is the method of detecting the emotional tenor behind a sequence of words in order to get an understanding of the opinions and attitudes expressed within a piece of data. It is predicted that about 80 percent of the data in the world is unstructured or unorganized. Huge amount of data is produced every day that includes e-mails, social media chats, blog posts, reports, surveys and other online documents. But it is very difficult to examine, recognize, and understand this data and the process is expensive and time-consuming. Sentiment analysis, on the other hand, helps in making sense of all this formless data by automatically processing and classifying it into various emotional categories. Nowadays, with the widespread of network, people have transformed the way in which they express opinions. Now, it is done mostly through blogs, online forums, product analysis sites, social sites etc. So, billions of users are expressing their views online. These social platforms help people to interact with each other and share their views regarding any product or service. Social media is producing a huge amount of opinion enrich data as posts, comments, reviews, etc. Social media platforms also allow businesses and companies to remain in touch with their customers and keep getting their valuable feedback and suggestions. This enables them to make better decisions based upon consumer feedback. Nowadays, when a person plans to buy something he or she definitely prefers to read or view reviews online to get better understanding about the product. However, the volume of data generated online on a daily basis is way too large to be analysed manually. So here comes the role of sentiment analysis.

Opinion Mining or Sentiment Analysis is quite useful in monitoring of social sites as it gets a summary of the sentiments of the society on every topic. The ability to take out valuable information from social data is an exercise that is being extensively used by organizations all around the globe. Some popular sentiment analysis applications include monitoring of social sites, management of customer support and analysing customer response. Automatic sentiment analysis can be performed on any data source, to categorize survey responses and chats, Twitter and Facebook posts, or to scan emails and other documents. All this is significant information for the companies and can make them take decisions accordingly.

With its growing demand and advancement in sentiment analysis techniques, the analysis of sentiments is not only limited to textual data. Researchers are even exploring new possibilities in analysis of other modalities of data. Today with the increased utilisation of internet surfing, the enormous info produced is not only in textual form but more and more images and videos are being used to convey one's opinions. There has been significant amount of research on analyzing textual data but research related to other modalities of data including image, speech and video content has been limited. [1] Multi-modal emotion detection is a recent topic in the domain of sentiment analysis. In multi-modal sentiment analysis we also take care of the audio and visual context of the data. This can be employed in building virtual assistants, analysis of YouTube videos and depression monitoring.

II. RELATED WORK

In this section, we have briefly described some of the approaches used by researchers in sentiment analysis of text, image, audio data.

A. Classification of Textual Data

In [2] , the authors have performed sentiment classification on Twitter Dataset. Firstly, they used naive-bayes as a baseline algorithm on unprocessed data. Later they performed preprocessing of dataset followed by stop words removal in order to draw a comparison between the results. Then they used various machine learning approaches like SVM and Maximum Entropy. They experimented with unigrams, bi-grams, stop words removal finally achieving the best results when using SVM with unigram or bi-gram.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

In the research performed in [3], a rule based classifier called VADER was presented. VADER stands for Valence Aware Dictionary for Sentiment Reasoning. It is a rule-based classifier based on lexicon that can detect both polarity and intensity of the emotion. It was made to compare with seven well known lexicons at that time and it outperformed all of them in case of social media data. In [4], the authors have presented a way for combining the machine learning models as well as background knowledge of lexicons for effective sentiment classification in the form of pooling multinomial. They have presented a hybrid approach by using naive bayes algorithm that combines lexicon's knowledge and the training of the classification model. Datasets were extracted from IBM Lotus blog posts, political posts and movie reviews and it was found that Linear Pooling model performed well as compared to various other approaches.

B. Classification of Audio Data

In [5], the authors have suggested ways to use the acoustic features extracted from audio signal in order to detect the emotional status of the speaker. The speech signal was fed to the Voice Activity Detection System as input that recognizes and differentiates audio from speech signals. The audio was then fed to ASR model and speaker discrimination model for identifying the data and speaker-identity. ASR model then labeled the voices with different speaker-ids. The voices were then converted to text with the help of Automatic Speech Recognition System. Then the speaker Ids were further matched with the converted text. The text output produced from the ASR system respective to different speakers was a significant feature to predict the emotion expressed by different speakers. In [6] research, end to end ASR models were used to combine acoustic features with text for sentiment classification. Here, RNN-T model was utilized to perform end-to-end speech recognition, then the result was fed to sentiment decoder. Sentiment Decoder consisted of Bi-LSTM, attention model and softmax classifier. To reduce overfitting, spectrogram augmentation was implied. The authors were able to achieve In [7], the authors extracted text and acoustic features separately from the input. They have used HMM model for classifying acoustic features and Naive Bayes and SVM for text features. Final output is produced after combining the results obtained from both type of classifications. This study showed the importance of both text as well as acoustic features in sentiment identification of audio data.

III. TEXT SENTIMENT ANALYSIS

A. Process of text sentiment analysis:

1) *Collection of data:* The very first step is to collect the data on which we want to do sentiment classification.
2) *Preprocessing of data:* Before performing any operations on the data, the data needs to be cleaned and preprocessed. This includes removal of punctuations, stop words and other unnecessary features of the data that are not required for further processing.
3) *Feature Extraction:* The preprocessed data is then converted to feature vectors so that they can be fed to different classifiers. The feature vectors are the kind of numerical representation of text data. Feature vectors are much easier to work with if we are using machine learning models.
4) *Traing the model:* After the text has been converted to feature vectors the dataset is split into training and testing data. The models are trained with the help of labeled training data using suitable classification algorithms.

TABLE I. COMPARISON TABLE

Researchers' names and Year	Model used	Type of input data	Results
Kharde, Vishal, and Prof Sonawane [2]	Naive Bayes, SVM and Maximum Entropy	Text	It was found that SVM with bi-gram model gave the highest accuracy.
Hutto, Clayton, and Eric Gilbert [3]	VADER	Text	VADER outperformed the current state of the art in case of social data when compared to other lexicon based models.
Melville, Prem, Wojciech Gryc and Richard D. Lawrence [4]	Multinomial Naive Bayes	Text	It was found that Linear Pooling model performed well as compared to various other approaches
Maghilnan, S., and M. Rajesh Kumar, 2017 [5]	ASR models, Naive Bayes, linear SVM and VADER.	Audio data	Among ASR models, Bing Speech API gave the highest WRR. Among classification models, VADER was the most effective one.
Lu, Zhiyun, et al [6]	RNN-T model, Bi-LSTM, Attention model and SoftMax	Audio Data	RNN with attention and specAug proved to be way better when compared to other RNN pooling/attention models.
Murarka, Aishwarya, et al [7]	HMM, Naive Bayes and SVM.	Audio Data	It was found that better results can be obtained when text and acoustic features both are combined to produce the final output.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

5) *Evaluating the model:* Finally the trained models are tested and evaluated using the test dataset.

B. Dataset

Here we have used IMDB movie review dataset for sentiment analysis of textual data.
This dataset consists of 50000 movie reviews out of which 2500 are labeled as positive and 2500 as negative. We selected this dataset because there is an equal distribution of both kinds of sentiments.

C. Models

We have experimented with different machine learning algorithms such as SVM, Naive Bayes, Logistic Regression and KNN.

1) *Support Vector Machines:* It is a non probabilistic model that makes use of a representation of inputs as points in a multi-dimensional space. It draws a hyperplane in the multi-dimensional space for classification of input. Each category or class is assigned a separate region in that space. New input is classified based on its similarity with existing data points and regions.
2) *Naïve Bayes:* It makes use of Bayes Theorem to predict the possibility that an input belongs to particular class of labels. It assumes that the features are independent and the occurence of one feature does not affect the other. It is a simple probabilistic classifier and yet performs well in sentiment analysis tasks.
3) Logistic Regression: It wrks on the basis of a sigmoid function that always gives the value in between 0 and 1 irrespective of what the input is. This makes it suitable for classification problems.
4) KNN: It compares the input with its K nearest neighbours and assigns the label based upon the majority indicated by the K nearest neighbours.

IV. VISUAL SENTIMENT ANALYSIS

A. Process of Visual Sentiment Analysis :

1) *Preparing the dataset:* Our dataset contains greyscale face images of 48*48 pixels dimension. Each of the image depicts one of the given emotion class that is Angry, Surprise, Happy, Disgust, Fear, Neutral, Sad. The dataset file has two fields, namely “emotion” and “pixels”, where “emotion” field contains numbers from 0-6 as per the emotion depicted by the image and the “pixels” field contains a string with space-separated pixel value for each image in row major form. Our major job is to train the model so as to predict the “emotion” field.
2) *Pre-processing the given data*:Before proceeding forward the given data is pre-processed using the techniques such as resizing, reshaping, converting the image to greyscale and normalisation. Normalisation is done so as we can achieve convergence as fast as possible.
3) *Splitting the dataset:*The given data set is split into two parts the training set and the testing set. This is done so that we can know if the model is overfitted to the training set with the help of the testing set also known as the validation set.
4) *Building the model using Deep Neural Network:* The DNN consists of mainly 2 kinds of layers-

- The hidden layers / Feature Extraction Part
  - convolutions
  - pooling
- The classifier part
- Further the following layers are added:
- Convolution layer : A set of learnable filters are used in this layer. A filter can detect specific features or patterns present in the input image.
- Pooling layer : This layer is used to reduce the no. of parameters and computations. It also controls overfitting by reducing network’s spatial size.
- Batch normalization : In this the layer inputs are standardised so that we can speed up the learning process.
- Activation Layer : This layer uses an activation function that is responsible for deciding the final value of a neuron.
- Dropout Layer : This layer is basically responsible for preventing overfitting of model.
- Flatten Layer : In this layer the data is converted to a 1-d array (flattening) before passing it to the next layer.
- Dense layer : The dense layer is a fully connected layer which means all the neurons are connected to each other in the current and the next layer.

5) *Train the model:* To train the model means making it learn on a repeat. Now a few hyperparameters are defined like the no. of epochs, batch size, rate of learning, etc. Here we look for the best parameters only by trying them repeatedly by changing the values of these hyperparameters. For recording the performance of the model during the period the model is training, various types of callbacks are used. Some examples of callbacks that we use are:
   a. **EarlyStopping()** :is used to stop training when a monitored metric has stopped improving.
   b. **ModelCheckpoint()** : to save keras model or model weights at some frequency.
6) *The New model is evaluated*: We check for the performance of the model i.e. how well the model is learning patterns from the training dataset. In this we test that how the accuracy is changing upon increasing the number of epochs. For this purpose we use matplotlib to visualise the model with the help of a learning curve.
7) *The model is tested:* In testing, the model is used for prediction of some images. Here in testing also

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 12, December-2021

the test images are also pre-processed before they are passed to the model.

8) *The model is saved for further use:* For saving the model both the weights and the architecture are saved into .h5 file and .json file respectively. This is done so that we don't need to train the model every time we need to predict the emotion of the images. So now on we can load the model make the predictions.

B. *Dataset*

Here we are using the Kaggle_fer2013 dataset which is freely available on Kaggle. It includes 35587 images and we worked on the fer2013.csv file. This file contains 2 fields namely emotion and pixels. Another dataset that is used is the RAF_dataset which contains 15399 basic images and 3954 compound images.

V. RESULTS

A. *Text Sentiment Analysis*

The results that we obtained with three different classifiers namely Logistic Regression, SVM and Naïve Bayes in two categories (with and without stop words) are presented in Table 7.

TABLE 2: RESULTS ACHIEVED THROUGH VARIOUS MODELS

Model	Accuracy (with stop words)	Accuracy (without stop words)
Logistic Regression	89.4	89.49
SVM	89.8	89.99
Naïve Bayes	85.6	86.07

As evident from the table stop words removal have made a slight increase in accuracy of almost all the models. Support Vector Machine (SVM) performed the best as compared to Logistic Regression and Naive Bayes. However Naive Bayes shows the highest amount of increase after stop words removal.

B. *Image Sentiment Analysis*

TABLE 3: ACCURACY OF THE NETWORKS

Network	FER2013	
	Validation	Test
A	63%	50%
B	53%	46%
C	63%	60%
Final	66%	63%

VI. CONCLUSION

After going through the work of different authors we have come to know that all the experiments that were carried out yield either a positive or negative sentiment polarity or emotions such as happiness, love, anger, fear and sadness as an output. We also came to know that we can achieve better results by combining different modalities such as visual and textual as compared to using one at a time. The results that we have drawn from the study of these works are summarised in Table 1. For text classification, lexicon based approaches tend to execute faster as they do not require training of the dataset and are effective for small volumes of input data. However, these approaches rely completely on lexicons and if a particular word cannot be found in the dictionary then it cannot classify that word. On the other hand, in order to work with large volumes of data, machine learning models can be utilized and can perform better. This requires a lot of preprocessing on the data so that it can fit into the model. Many researchers have proposed hybrid approaches that use both machine learning models and the background knowledge of lexicons. These models have also shown good results as they combine the best of both the methods.

As evident from the table stop words removal have made a slight increase in accuracy of almost all the models. Support Vector Machine (SVM) performed the best as compared to Logistic Regression and Naive Bayes. However Naive Bayes shows the highest amount of increase after stop words removal.

REFERENCES

[1] Cai, Guoyong, et al. "Multi-level Deep Correlative Networks for Multi-modal Sentiment Analysis." *Chinese Journal of Electronics* 29.6 (2020): 1025-1038.

[2] Kharde, Vishal, and Prof Sonawane. "Sentiment analysis of twitter data: a survey of techniques." *arXiv preprint arXiv:1601.06971* (2016).

[3] Hutto, Clayton, and Eric Gilbert. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 8. No. 1. 2014.

[4] Melville, Prem, Wojciech Gryc, and Richard D. Lawrence. "Sentiment analysis of blogs by combining lexical knowledge with text classification." *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* 2009.

[5] Maghilnan, S., and M. Rajesh Kumar. "Sentiment analysis on speaker specific speech data." *2017 International Conference on Intelligent Computing and Control (I2C2)*. IEEE, 2017.

[6] Lu, Zhiyun, et al. "Speech sentiment analysis via pre-trained features from end-to-end asr models." *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020.

[7] Murarka, Aishwarya, et al. "Sentiment Analysis of Speech." *International Journal of Advanced Research in Computer and Communication Engineering* 6.11 (2017): 240-243.

[8] Mehta, Munish, Kanhav Gupta, and Shubhangi Tiwari. "A Review on Sentiment Analysis of Text, Image and Audio Data." *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2021.