

SENTIMENT ANALYSIS USING RECURRENT NEURAL NETWORKS

Meenakshi¹, S.R.Akkash², S. Dinesh Kumar², D. Rajasekar², S. Sharan Prasad²

¹Assistant Professor, Department of ECE, Velammal Engineering College, Chennai

²UG Students, Department of ECE, Velammal Engineering College, Chennai

ABSTRACT: Sentimental Analysis is the process of identifying and classifying human emotions or sentiments expressed in source text. Social media nowadays is a collection of people interacting with each other, expressing multiple emotions. These emotions range anywhere between positive emotion and negative emotion. The main objective of this project is to survey public emotions and sentiments about a product or an incident. This Sentimental Analysis can be obtained by using Lexicon-based approach or by using Machine Learning/Deep Learning/Natural Language Processing algorithms. In this project, Sentimental Analysis is achieved using a Deep Learning algorithm known as Recurrent Neural Networks and a Natural Language Processing technique known as Word2Vec. Furthermore, by recording the user's audio we can analyse the sentiment of the user. Sentimental Analysis finds its application in Opinion Mining, Market Research, Brand Monitoring, Customer Feedback, Social Media Monitoring and Crowd Analysis.

General Terms

Sentiment_Analysis, Neural Networks, Sound_Sensor, Python, Tensorflow, Rnn, Raspberry, Word2vec, Pyaudio, Speechrecognition, Pickle.

Keywords

RNN, CSV, I2C, IDE, Analysis, pyaudio, Raspberry pi, tensorflow.

INTRODUCTION

In our modern world, emotions and sentiments can be used as information and data for multiple purposes. Social media is a great platform for people to express their feelings, thoughts and insights about particular events, incidents or products. Data as in emotions and sentiments are a valuable source of information for analysing and evaluating the opinions and sentiments of people to find the general consensus. Sentimental analysis is a way to find this out. The approach taken here is to find whether the given text is positive, neutral or negative. The emotion or the sentiment of one particular user is analysed and found at a time. Then we can find the emotions of multiple users or people and conclude the general emotion about a particular product

or an incident. Once the text is analysed, we can identify what people think about a particular product, incident or an happening.

Sentiment Analysis

Sentiment Analysis is the contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations. With the recent advances in deep learning, the ability of algorithms to analyse text has improved considerably. Creative use of advanced artificial intelligence techniques can be an effective tool for doing in-depth research. Sentiment analysis can be used in different websites such as Twitter, Facebook, Amazon, IMDb and many similar social media sites. Sentiment Analysis finds its use in social media monitoring, brand monitoring, customer feedback, opinion mining and crowd analysis.

Neural Network

A neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial neurons or nodes. Thus a neural network is either a biological neural network, made up of real biological neurons, or an artificial neural network, for solving artificial intelligence (AI) problems. The connections of the biological neuron are modelled as weights. A positive weight reflects an excitatory connection, while negative values mean inhibitory connections. All inputs are modified by a weight and summed. This activity is referred as a linear combination. Finally, an activation function controls the amplitude of the output. For example, an acceptable range of output is usually between 0 and 1, or it could be -1 and 1.

PROPOSED METHODOLOGY

Modules

This model consists of two modules, the sentiment analysis module and the hardware module. The sentiment analysis module is the main module which is used to analyse the emotion of the given text. Recurrent neural networks are used to make the analysis of the text possible. The other module is the hardware module

which is the hardware implementation of the Sentiment Analysis module for both text-based and audio-based analysis.

SENTIMENT ANALYSIS MODULE

The sentiment analysis module consists of a Recurrent Neural Network (RNN) and a word embedding module Word2Vec. A recurrent neural network is a classification of neural network where the output from the previous step is fed as input to the current step. Word2vec is a group of related models that are used to produce word embeddings. These models are shallow, two-layer neural networks that are trained to reconstruct linguistic contexts of words.

HARDWARE MODULE

The hardware module comprises of the Raspberry Pi 4 microcomputer and a microphone which is used to record the audio signals from the user. The raspberry Pi operates on Raspbian OS which has an integrated Python IDE on which the module runs on. The microphone is connected to the power supply. In order to operate the Raspberry Pi and navigate in it, the Raspberry Pi is connected to the display peripheral such as a monitor or a laptop display Title and Authors.

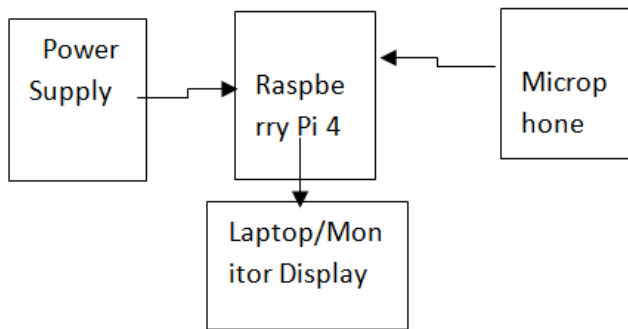


Fig1. Block Diagram of Hardware module.

Working Methodology

The Sentimental Analysis module works in a straightforward way. The input given will be either a comment or a review about an incident or a product respectively, which is basically a string of text. This string is pre-processed before it is given as input to the neural network. The first step in pre-processing is to make sure the string is completely in lower case and then special characters are removed. Then comes Stemming, Tokenization and Stop word removal. Tokenization is the process of chopping up of a sentence or a character into pieces known as tokens. Stemming is the process of reducing the word to its word stem that

affixes to prefixes or suffixes or to the roots of words known as lemma. Stop words are basically articles and some verbs which do not help us in finding the context of the sentence. Hence it is essential to remove the stop words in a sentence. Then word embedding and feature extraction are done using Word2Vec and TfidfVectorizer. The main purpose of word embedding is to map each word to a vector. Feature extraction is used for reducing the initial raw data into more manageable groups. Once word embedding and feature extraction are done, it is time for classification, where Recurrent Neural Networks (RNN) come into work. RNNs are a type of neural networks that remember their past and consider their past for future requirements. Once the neural network model is built, the string is entered to predict whether the given string is positive, negative or neutral. Sentiment thresholds are used to classify the sentiment. In addition, we can also use voice input and classify the audio's sentiment using a microphone and with the help of packages such as pyaudio and SpeechRecognition..

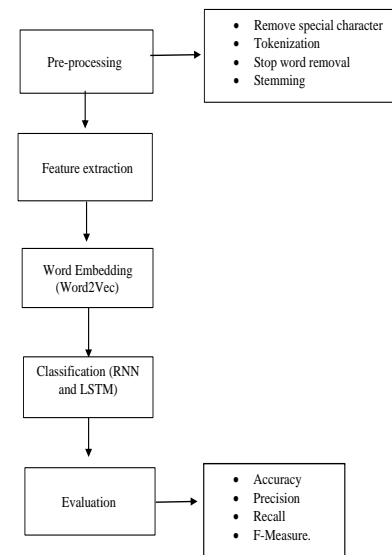


Fig2.Flow of sentiment analysis

HARDWARE

Raspberry Pi

This product's key features include a high-performance 64-bit quad-core processor, dual-display support at resolutions up to 4K via a pair of micro-HDMI ports, hardware video decode at up to 4Kp60, up to 4GB of RAM, dual-band 2.4/5.0 GHz wireless LAN, Bluetooth 5.0, Gigabit Ethernet, USB 3.0, and PoE capability

Sound Sensor

It uses a microphone which supplies the input to an amplifier, peak detector and buffer. When the sensor detects a sound, it processes an output signal voltage which is sent to a microcontroller then performs necessary processing. The working principle of this sensor is related to human ears. Because human eye includes a diaphragm and the main function of this diaphragm is, it uses the vibrations and changes into signals.

Software

TENSORFLOW

TensorFlow is the most famous deep learning library these recent years. A practitioner using TensorFlow can build any deep learning structure, like CNN, RNN or simple artificial neural network. Practitioners use TensorFlow because it is easy to deploy at scale. It is built to work in the cloud or on mobile devices like iOS and Android.

RNN

A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behaviour. Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition. Basic RNNs are a network of neuron-like nodes organized into successive "layers." Each node in a given layer is connected with a directed (one-way) connection to every other node in the next successive layer.

Word2Vec

Word2Vec can be obtained using two methods (both involving Neural Networks): Skip Gram and Common Bag of Words (CBOW). We use Skip Gram method in our model. The model outputs C probability distributions. For each context position, we get C probability distributions of V probabilities, one for each word. The network uses back-propagation to learn.

PyAudio

Pyaudio provides bindings for PortAudio, which is a cross-platform audio I/O library. Pyaudio is used to play and record audio on a variety of platforms.

SpeechRecognition

Speech Recognition is a library which is used for performing speech recognition, with support for several engines and APIs, online and offline.

Pickle

The pickle module implements binary protocols for serializing and de-serializing a Python object structure. "Pickling" is the process whereby a Python object hierarchy is converted into a byte stream, and "unpickling" is the inverse operation, whereby a byte stream (from a binary file or bytes-like object) is converted back into an object hierarchy. Pickling (and unpickling) is alternatively known as "serialization", "marshalling," 1 or "flattening"; however, to avoid confusion, the terms used here are "pickling" and "unpickling". The idea is that this character stream contains all the information necessary to reconstruct the object in another python script.

```
pickle.dump(data, file)
```

Fig 3 Dumping the code

```
data = pickle.load(file)
```

Fig 4 Loading the code

Experimental Result

The sentiment analysis module consists of a Recurrent Neural A recurrent neural network is a classification of neural network where the output from the previous step is fed as input to the current step. Word2vec is a group of related models that are used to produce word embeddings. These models are shallow, two-layer neural networks that are trained to reconstruct linguistic contexts of words.

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. It allows the visualization of the performance of an algorithm. The confusion matrix of the model is given in Fig 5.

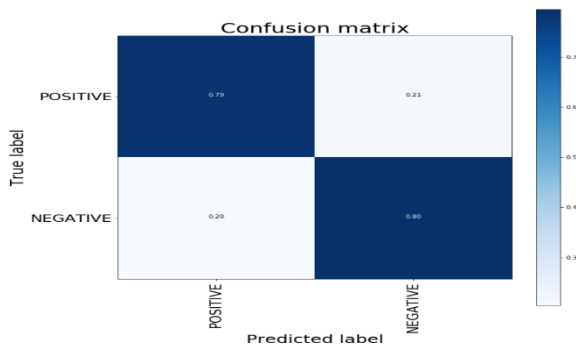


Fig 5. confusion matrix

CONCLUSION

Sentiment Analysis is the application which is used by many businesses to expand their growth. It uses machine learning and deep learning algorithms which are integrated with the text mining to bring meaningful data from unstructured data. This meaningful data brings lots of benefits to customers and businesses to get a better insight of the products and emotions and mood of the people on some particular topic like Elections or Movies. It is used in large extent with Social Networks and get the sentiment analysis from some popular celebrities and politician accounts, for example – “Follow the tweets of Current Indian Prime Minister Narendra Modi which is one of the most followed and active Twitter Politician around the world and get the tweets and comments over the tweets from Twitter API and perform a Sentiment Analysis to find the sentiments related with the tweets and comments.” Sentiment Analysis is also very popular application in E-commerce industry with the likes of Amazon, Flipkart, Snap deal etc.

The heading of a section should be in Times New Roman 12-point bold in all-capitals flush left with an additional 6-points of white space above the section head. Sections and subsequent sub- sections should be numbered and flush left. For a section head and a subsection head together (such as Section 3 and subsection 3.1), use no additional space above the subsection head.

FUTURE SCOPE

The future of sentiment analysis is going to continue to dig deeper, far past the surface of the number of likes, comments and shares, and aim to reach, and truly understand, the significance of social media interactions.

As a result of deeper and better understanding of the feelings, emotions and sentiments of a brand or organization’s key, high-value audiences, members of these audiences will increasingly receive experiences

and messages that are personalized and directly related to their wants and needs.

With the help of Convolutional Neural Networks (CNN), it is possible to analyse the emotion of people from their facial expression. With the help of this, it is feasible to completely analyse the emotion and sentiment of the user from their words and expressions.

Another use of Sentiment Analysis is Crowd Analysis. The sentiments of a group of people in a close environment can be captured and analysed, so that depending on what they need, the environment factors can be adjusted.

REFERENCES

- [1]. Cabanlit and K. J. Espinosa, Optimizing N-gram based text feature selection in sentiment analysis for commercial products in Twitter through polarity lexicons, in Proc. 5th Int. Conf. Information, Intelligence, Systems and Applications, Chania, Greece, 2014, pp. 94–97.
- [2]. Hodeghatta, Sentiment analysis of Hollywood movies on Twitter, in Proc. 2013 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining, Niagara Falls, Canada, 2013, pp. 1401–1404.
- [3].Jaspreet Kaur and Brahmaleen Kaur Sidhu (2018) “Sentiment Analysis Based on Deep Learning Approaches”, Proceedings of the Second International Conference on Intelligent Computing and Control Systems. IEEE Xplore Compliant Part Number: CFP18K74-ART;
- [4].Mondher Bouazizi and Tomoaki Ohtsuk (2019) “Multi-Class Sentiment Analysis on Twitter: Classification Performance and Challenges” BIGDATA MINING AND ANALYTICS ISSN 2096-0654 03/05 pp181–194 Volume 2, Number 3.
- [5].Vikas Goel, Amit Kr. Gupta, Narendra Kumar (2018) “Sentiment Analysis of Multilingual Twitter Data using Natural Language Processing”,8th International Conference on Communication Systems and Network Technologies.
- [6].Soler, F. Cuartero, and M. Roblizo, Twitter as a tool for predicting elections results, in Proc. 2012 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining, Istanbul, Turkey, 2012, pp. 1194–1200.
- [7]. Yaser Maher Wazery ,Hager Saleh Mohammed, Essam Halim Houssein (2018) “Twitter Sentiment Analysis using Deep Neural Network”.