IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105170

To Analyze Live Streaming Text data for Real Life Applications using Deep Learning Techniques

Mrs.Vanita raut,Mrs.Anita Mahajan,Ms.Sanjana karmalkar,Ms.Snehal sarade

G.H raisoni institute of engineering and technology, Pune,India

Abstract: Text conversion and categorization are fundamental machine learning operations in which an object is assigned to a subset of unnamed candidates. We propose a novel text classification approach based on Deep Learning in this research (DL). Our suggested method has a lot of appealing features: it gathers certain metadata from each item and builds the training set train first. The weight achieved under each class mark was used to classify each object. The Recurrent Neural Network (RNN) was employed in the proposed classification model. For training and testing, the streaming text was dispersed using cross-fold validation. The 70-30% of the data was used for training and research purposes, respectively. The consequences of the strategy in the chapter on outcomes are revealed through a partial implementation. The proposed method outperforms traditional approaches to text classification, according to the results of the experimental study.

Keywords: Text processing, Machine learning, deep learning, natural language processing, data analysis, artificial intelligence

INTRODUCTION

Text real data classification is very tedious task for classification algorithms, it consist an on web data is now becoming a potential research region of social data analytics environment. Basically to identify the emotions from some co-related textual features, as well as text real data processing technique. Many existing system has already introduces emotion detection using Machine learning techniques, but the traditional feature extraction strategies it's not able achieve required accuracy on large text data. This proposed work Text Real data Analysis using hybrid deep learning approach like machine learning and NLP for synthetic as well as real time voice text dataset. The system investigate the various strategies of image real data identification and classification using different machine learning or deep learning algorithms and predict the future work possibility for system enhancement.
In this research to design and develop a NLP based approach that analyze real time streaming data and provide automation, to perform the event on real life applications

LITERATURE SURVEY

According to Lubomir Stanchev et. Al. [1] was represented by Semantic text Clustering Using WordNet and DBPedia Knowledge. As long as they are on the same topic, this technique may group documents with no words in common. As a function of the semantic similarity between the words and phrases in the documents, we compute the similarity between two documents. As a probabilistic graph that can be used to calculate the similarity between two words, we model knowledge from WordNet and DBPedia. The Cosine Similarity (CS) algorithm has been used to generate the weight of similarity between two vectors, producing the classified outcomes until the ordering algorithm is applied to the sorted vector method. Aditi, Gupta, Jyoti Gautam et. A Survey on Methodologies used for Semantic Text Clustering was proposed by Al. [2].
The study of different research papers that have been studied explains this method and highlights each clustering algorithm's benefits and demerits. This will have a more focused way of directing the future study. Chalitha, Kulathunga et al. [3] suggested a methodology for financial documents focused on ontology and domain-specific clustering. The architecture of the Relational Data Framework (RDF) has been used to derive semantic information. In general, it is used to describe the correct definitions of the ambiguous terms in the articles.
Most of the proposed methodologies were evaluated on general text datasets. Most of the few domain-specific clustering studies available were restricted to particular domains where full domain ontologies are available. Although there are many domain ontologies in the economic domain, none is complete and sufficient for semantic text clustering. A Text Classification Model Using Convolution Neural Network and Recurrent Neural Network has been proposed by

This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105170

Radhika K, Bindu K.R.et. Al.[4] It suggested a model of text classification using CNN and RNN. Text classification is characterized as the text's classification into one of the categories in which the text belongs to Neural Networks. Document selection is tested and tested using neural networks. This paper uses the Convolution Neural Network and Recurrent Neural Network to construct a text classification model. With our dataset, we train and test both CNN and RNN models. The dataset that we used is a collection of essays. We conclude that RNN performs better than CNN for our essay dataset from the train and test accuracy collected.

Xipeng Qiu, Pengfei Liu, Xuanjing Huang et. Al.[5] has implemented a Recurrent Neural Network with Multi-Task Learning for Text Classification. We use the multitask learning method here to learn together through many associated tasks. We propose three different methods of exchanging knowledge based on recurrent neural networks to model text with taskspecific and shared layers. On all these activities, the entire network is jointly trained. Experiments on four tasks for benchmark text classification show that, with the aid of other similar tasks, our proposed models can boost a task's performance. The multitask learning system for jointly learning across multiple related tasks based on RNN is also defined. From an approach like this.

In the Age Groups Classification in Social Network Using Deep Learning, Guimaraes, Rita Georgina, et al. [6] provided. This study indicates that the age group is one of the most important parameters found in the user profile, showing that typical behaviours occur among users of the same age group, especially when those users write about the same topic. In order to decide the features are important, such as the use of punctuation, the number of characters, media sharing, topics, among others, a thorough study of 7000 sentences was carried out; and which ones can be disregarded for the classification of the age groups. For the classification of the adolescent and adult age group, various learning machine algorithms are tested. The Deep Convolutional Neural Network (DCNN) has the best results, achieving an accuracy of 0.95 in the validation tests. In the area of recommender systems and personalization, the inference of user real data can be very useful to make up for the lack of explicit user feedback on a provided service. In addition to machine learning, other methods, such as those based on the similarity of results, can be used for this purpose [7].

The sources of data for real data analysis (SA) are online social media, the users of which generate an ever-increasing amount of information. Thus, these types of data sources must be considered under the big data approach, given that additional issues must be dealt with to achieve efficient data storage, access, and processing, and to ensure the reliability of the obtained results In this context, Thai et al. suggested not to focus solely on the structure and correlations of data, but on a lifelong learning approach to dealing with data presentation, analysis, inference, visualization, search and navigation, and decision making in complex networks [8]. Li et al. [9] studied the impact of data quality on real data classification performance. They considered three criteria, namely informativeness, readability, and subjectivity, to assess the quality of online product reviews. The study highlighted two factors that a ect the level of accuracy of real data analysis—readability and length of the reviews. Higher readability and shorter text datasets yielded higher quality of real data classification. However, when the size or domain of the data varies, the reliability of the proposed method is questionable A deep neural network [10] is a neural network with more than two layers, some of which are hidden layers. Deep neural networks use sophisticated mathematical modeling to process data in many different ways. A neural network is an adjustable model of outputs as functions of inputs, which consists of several layers: an input layer, including input data; hidden layers, including processing nodes called neurons; and an output layer, including one or several neurons, whose outputs are the network outputs. Recurrent neural networks [11] are a class of neural networks whose connections between neurons form a directed cycle, which creates feedback loops within the RNN. The main function of RNN is the processing of sequential information on the basis of the internal memory captured by the directed cycles. Unlike traditional neural networks, RNN can remember the previous computation of information and can reuse it by applying it to the next element in the sequence of inputs.

PROPOSED SYSTEM ARCHITECTURE

The proposed system provides a real-time message and automatic reading system for communication-based applications. This system basically focuses on, automated text reading using NLP algorithms. Whenever any users received the text information, the system immediately asks the user for the next event. (e.g. user wants to read this text or not). Based on the user's input system automatically perform the next transaction that will provide voice information of the entire text.

Feature Extraction

It is the process of converting the text feature into feature vector. For the representation of text we are going to use the vector space model in our proposed system.

Feature Selection

Feature selection is used for dimensionality reduction of original feature set to get the more relevant feature space for classification.

This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105170

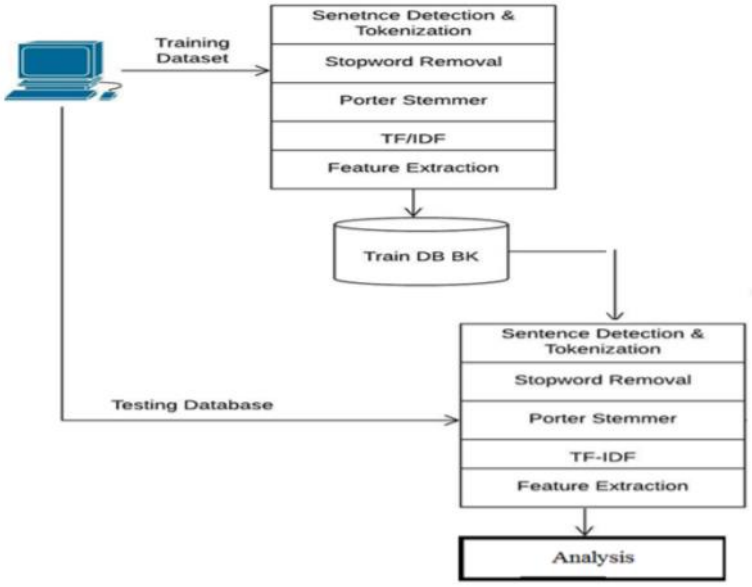

Figure 1 : Proposed system design

Similarity based techniques: Deep Learning (RNN)
The similarity between two documents which are considered as nodes can be calculated using cosine similarity. The cosine similarity is calculated using below formula.

$$\text{soft\_cosine}_1(a,b) = \frac{\sum_{i,j}^{N} s_{ij} a_i b_j}{\sqrt{\sum_{i,j}^{N} s_{ij} a_i a_j} \sqrt{\sum_{i,j}^{N} s_{ij} b_i b_j}},$$

Module Training: The sleeted features forward to module raining and based on that system generates some Background Knowledge (BK) which is basically used for classification

Module Testing: The remaining test data has validates in module testing, it will classify each instance with help of training rules. The similar algorithm will carry forward which is used for module training.
Analysis: It shows the classification results of entire system for testing data and demonstrates the graphs based on confusion matrix

Algorithms

1: Method to stop Word Elimination
Input: List of stop words L[], String Data D for deleting stop words.
Output: Checked D data with all stop words removed.

Phase 1: Initialize the S[] string for details.
Phase 2: a=0,k=0 is initialized
Stage 3: for each of the (read a to L)
If(a.equals(L[i]))) (L[i]))
Remove S[k] then
Ending for
Step 4: add S to D.
Phase 5: Process to End

2 Algorithm for Stemming.
Input: Phrase w
Output: w with past participles also being excluded.
Phase 1: The initialization of w

This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105170

Phase 2: Initialize all Porter Stemmer steps
Phase 3: for each of the (Char ch from w)
When(ch.count==w.length()) &&(ch.equals(e))
Drop ch from the list of (w)
Phase 4: if(ch.endswith(ed)))
Delete from 'Ed' (w)
Phase 5: k=w.length ()
If(k(char) .equals(tion)) to k-3.
Replace Te with W.
Phase 6: Termination of Procedure

3 TF-IDF 3 TF-IDF
Input: Each Vector Word as Term T, All Vectors V[i...n]
Output: weight of TF-IDF for each T

Phase 1: {c1, c2, c3....cn} vector =
Phase 2: Aspects that are available in any statement
D = {cmt1, cmt2, cmt3, cmtn} Stage 3:
Open comments and comments in each document
Calculate the score for Tf as
Phase 4: = tf (t,d) (t,d)
Relevant word t=
D= Precise text should be found in a word.
Phase 5: idf = t = sum (d)
Phase 6: Give back tf *idf

RESULTS AND DISCUSSION

For the system performance evaluation, calculate the matrices for accuracy. The system is executed on java 3-tier architecture framework with INTEL 2.7 GHz i3 processor and 4 GB RAM with deep learning approach. The below figure 2 shows the classification results for different test data size.
The study below is the classification graph of the method. The graphs demonstrate how the overall inputs are grouped into groups by the system. The proposed framework is applied to the RNN mix, which offers a satisfactory level for all outcomes. 112 documents for preparation and 30 documents for testing are issued for performance assessment. The method contrasts the suggested outcomes with two different current systems here.

Figure 2: Data Transmission

This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.105170

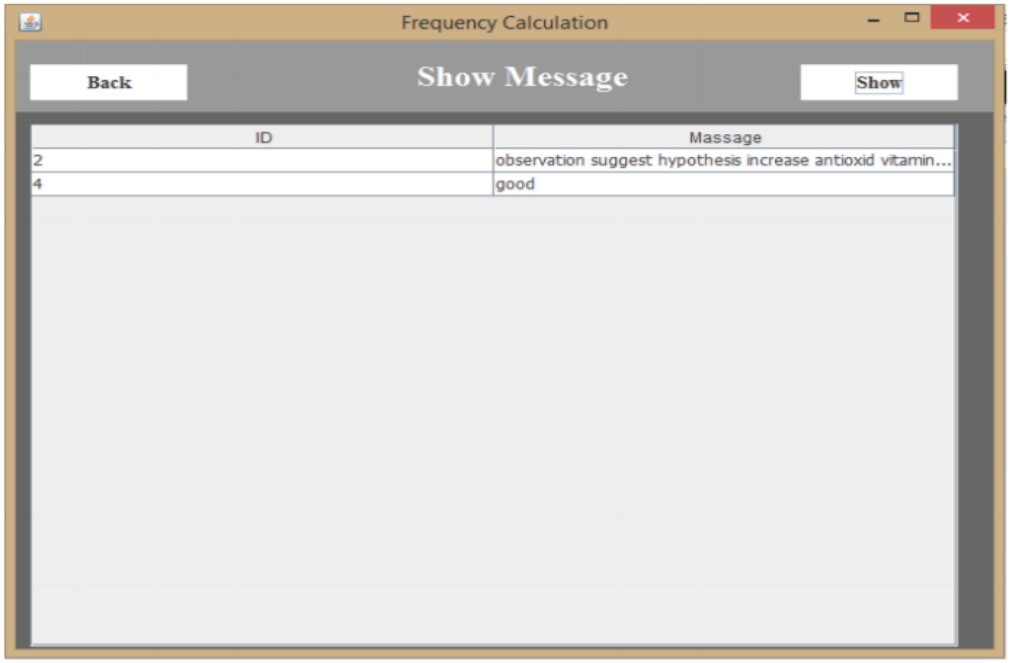

Figure 3 : TF-IDF NLP calculation on incoming data

CONCLUSION

A data preparation employing NLP and computer vision method was suggested in this paper. A text multi-label categorization by employing RNN approach was proposed in this work. The suggested learning system particularly models the associations by naming machine learning, which is concurrently configured along with identification. Consequently, the trained label association graph can meet the job of micro classification well, even when capturing the underlying superposition between labels correctly. In comparison, to capture the conversation inter-label related information, we have adopted neighborhood regularization. Using NN weight calculation and some classification approach, the future methodology may categorize the powerful label using a test example. Present measurements have illustrated the efficacy of our technique over several benchmark datasets.

- It works on real time text data and classifies the results accordingly.
- The systems works like automation system, which provides runtime results.
- It works any kind of text classification applications.
- System also works to classify real time voice sentiments.
- A text should be recognized by NLP algorithm and provides result accordingly.
- To apply this techniques on various real life applications and show the effectiveness of systems.

The system works like smart automation system like smart mirror etc.

FUTURE WORK

Often, we will concentrate on such issues with a precision device with a false positive ratio. When we deal with strong or big data, the second element is machine execution sophistication. For average time calculation or simultaneous distribution, the device may operate with the HDFS application. HDFS foundation framework with parallel neural networks can be executed for the performance was obtained.

REFERENCES

[1] Vanita G Kshirsagar et al. [R1], They worked on the facial expression based on that identify the sentiment and mood of the person. They work on the human mental condition using RNN method.

[2] Mayur Saptal et al. [R2] Worked to find the sentiment of the people using their facial expression or their voice or the songs they listen based on that it will identify the sentiment based on they will recommend the song to make a person happy. They apply LSTM technique.

[3]VANITA G. KSHIRSAGAR AL., STUDIED ON THE SENTIMENT ANALYSIS BASED ON THE NLP PROCESS. THE LITERATURE SURVEY IS BASED ON THE NLP AND ANN.

[4] Gupta, Aditi, Jyoti Gautam, and Ajay Kumar. ‖A survey on methodologies used for semantic text clustering.‖ 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS). IEEE, 2017.

[5] Kulathunga, Chalitha, and D. D. Karunaratne. ‖An ontologybased and domain specific clustering methodology for financial documents.‖ Advances in ICT for Emerging Regions (ICTer), 2017 Seventeenth International Conference on. IEEE, 2017.

[6] Nema, Puneet, and Vivek Sharma. ‖Multi-label text categorization based on feature optimization using ant colony optimization and relevance clustering technique.‖ Computers, Communications, and Systems (ICCCS), International Conference on. IEEE, 2015.

This work is licensed under a Creative Commons Attribution 4.0 International License

DOI 10.17148/IJARCCE.2021.105170

[7] Pratama, Timothy, and Ayu Purwarianti. ITopic classification and clustering on Indonesian complaint tweets for bandung government using supervised and unsupervised learning.I Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), 2017 International Conference on. IEEE, 2017.
[8] Guimaraes RG, Rosa RL, De Gaetano D, Rodriguez DZ, Bressan G. Age Groups Classification in Social Network Using Deep Learning. IEEE Access. 2017;5:10805-16.
[9] Kulathunga C, Karunaratne DD. An ontology-based and domain specific clustering methodology for financial documents. InAdvances in ICT for Emerging Regions (ICTer), 2017 Seventeenth International Conference on 2017 Sep 6 (pp. 1-8). IEEE.
[10] Pouli, V.; Kafetzoglou, S.; Tsiropoulou, E.E.; Dimitriou, A.; Papavassiliou, S. Personalized multimedia content retrieval through relevance feedback techniques for enhanced user experience. In Proceedings of the 2015 13th International Conference on Telecommunications (ConTEL), Graz, Austria, 13–15 July 2015; pp. 1–8.
[11] Thai, M.T.; Wu, W.; Xiong, H. Big Data in Complex and Social Networks; CRC Press: Boca Raton, FL, USA, 2016.
[12] Li, L.; Goh, T.-T.; Jin, D. How textual quality of online reviews a ect classification performance: A cas3611e of deep learning real data analysis. Neural Comput. Appl. 2018, 1–29.
[13] Aggarwal, C.C. Neural Networks and Deep Learning; Springer: Berlin, Germany, 2018
[14] Britz, D. Recurrent Neural Networks Tutorial, Part 1–Introduction to Rnns. Available online: http://www.wildml.com/2015/09/recurrent-neural-networkstutorial-part-1-introductionto-rnns/ (accessed on 12 March 2020).
[R1] Dr. Shweta Sharma Vanita G. Kshirsagar, "Early detection of depression using facial expression and sentiment analysis", GIS Science Journal UGC-CARE Approved-II, 2021.
[R2] Mayur Saptal Vanita Raut
, "Sentiment analysis in songs and advanced recommendation system",I
International Journal of Current Engineering and Technology,2021

[R3] Shweta Sharma Vanita G. Kshirsagar
, "Review Paper on Sentimental Analysis for Recommendation System"
Proceeding of First Doctoral Symposium on Natural Computing Research,Springer Link, 2021
[R4] Rutuja Sandbhor Vanita G. Kshirsagar , Nisha Kadam
, "Prevention of Suicidal Causes on Social Media Using Machine Learning",
International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET),2020
[R5] Sachin Dhage Prof. Vanita G. Kshirsagar, Adam Ramakant, Suraj Dalve, Ashutosh Rapatwar
, "Accident Prevention System using IR Sensors", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2020
[R6] Prof. Vanita Raut Mr. Nivas Patil
, "Ranking Based Recommendation using Online Social Users Data",
IJSRD - International Journal for Scientific Research & Development, 2019

This work is licensed under a Creative Commons Attribution 4.0 International License