IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 36, No. 3, June, 2022, pp. 381-386
Journal homepage: http://iieta.org/journals/ria

An Integrated Single Framework for Text, Image and Voice for Sentiment Mining of Social Media Posts

Gubbala Kumari*, A. Mary Sowjanya

Department of Computer Science and Systems Engineering, Andhra University College of Engineering(A), Visakhapatnam, AP 530003, India

Corresponding Author Email: gubbala.kumari@yahoo.com

https://doi.org/10.18280/ria.360305

Received: 10 April 2021
Accepted: 17 August 2021

Keywords:
sentiment analysis, LSTM, generative model, prediction, machine learning, deep learning, social media posts

ABSTRACT

The wide spread pandemic COVID-19 has propelled the entire world to rely on social media interaction digitally. Social media is thus a platform to express numerous kinds of direct and indirect sentiments by human beings. Psychologically, a person tends to share his/her feelings in terms of sentiments more openly over the social media. These sentiments, when intense may polarize oneself to commit severe mis-deeds. Here arises the role of the researchers to perform a real time identification of sentiments and classify them so that a prospective mishap can be averted. In this work, an integrated framework is proposed that does an early recognition of sentiments over social media in the digital domain. Along with sentiment categorization, another module has been integrated to the framework to perform a post-predictive analysis of the same. The proposed integrated framework involves combination of two distinct mechanisms. First, the proposed work channelizes the input data in line with its characteristics text, image, and voice. The text input is directly fed to our proposed 'Lexicon based LSTM with sentiment word mapping' mechanism. From the input image, both text and semantics are extracted through two different blocks. One block converts image-to-text and redirects the output to the above proposed model. We proposed a new generative model (GM) to extract the semantics of the image and the second block utilizes our generative model and redirects the outcome straight to the final output buffer of the framework. The voice-to-text module has been used for transforming voice input data to text data which is redirected to our proposed Lexicon based LSTM for further processing. A comparison of the proposed work has been made with state-of-the-art techniques. Our results indicate that the overall rate of accuracy of this framework is superior to the existing methods.

1. INTRODUCTION

The demographic ecosystem has now become virtual and interactive in digital domain. With the advent of social media platforms, humans are increasingly reciprocating their thought processes over the media. This phenomenon has led to a complete paradigm shift in the domain of digital data analysis. The abundance in social media posts has attracted many researchers towards analyzing this data thoroughly to generate meaningful outcomes. Prediction of inherent meaning and thought process of user is utmost essential now days due to the raising online activities of social media users. A few methods relating to the three types of classification tasks: text sentiments, image semantics and voice sentiments served as motivation for the proposed framework.

2. LITERATURE SURVEY

Bello et al. [1] reported that the most widely used social network platform is Twitter. More than five hundred millions of posts are being posted to Twitter on a daily basis. The number of users for the same has now reached as many as 350 million. Analysis of data over such a platform with respect to service, marketing and mainly for rescuing lives as per present scenario is challenging. Sentiment is the buzz word that seeks deep understanding through computing models. Economy of a region has now become dependent on such analysis. Analysis of tweet data may also lead to saving one's life in case the person is suffering from depression and/or other traumatic thoughts. Asghar et al. [2] used a combination of multiple number of features for performing sentiment analysis with accuracy. Along with variety of classifiers, these investigators have also used deep learning tools for suitable analysis and comparison of social media posts. Al-Twairesh and Al-Negheimish [3] proposed ensemble of surface and deep features for sentiment classification of Arabic tweets and also experimented with BERT model. Their results showed improvement on surface features only.

Word embeddings have been manually calculated by Zhao et al. [4] to generate ensemble feature sets for sentiment analysis. Ye et al. [5] translated input texts into real numbered vectors, preserving their regular meaning, background sentiment values and treated surface vectors as the base for classification. These vectors generated mainly lead to human errors and cause overload in computational time.

The studies of Araque et al. [6] and Giatsoglou et al. [7] indicated that with sufficiently large datasets and training samples, the word2vec tool performs effectively. Tang et al. [8], proposed learning continuous word representations as

features for Twitter sentiment classification. Due to the elimination of polarity words, the accuracy of the tool has been found to reduce drastically as per the findings of Hassan and Mahmood [9]. But, manually mapping the polarity terms helped in improving the accuracy. Improvements focusing on variations of long short term memory unit (LSTM) and combination of recurrent units with convolutional neural networks (CNN) have also been extensively used for sentiment analysis. Different deep learning techniques have been utilized for sentiment label prediction. A hybrid -LSTM model reported by Rehman et al. [10] incorporates the concepts of both LSTM and CNN. The pooling strategy in CNN has been customized to maximize key feature selection and word embedding was also utilized.

Rezaeinia et al. [11] proposed pre-trained word embeddings for sentiment analysis using some models like lexicon-based, POS tagging, word position algorithm and Word2Vec/GloVe for improved accuracy. Study of Narayanan et al. [12], showed that the opinions expressed on different topics in a conditional sentence are determined as positive, negative or neutral. Farias et al. [13] studied emotions and psycholinguistic analysis is by using word-sentiment lexicons. Ganapathibhotla and Liu [14] proposed an effective approach on comparative sentences for mining opinions on different datasets. Phan et al. [15] proposed a method on fuzzy sentiment for the better performance of sentiment analysis of tweets using feature ensemble models.

Feature combination scheme is reportedly utilized in many research works in sentiment analysis. The meaning of a word is given due weight age compared to the sentimental meaning of the same. The tokens, sentimental words, regular words, index of the terms, etc. need to be carefully chosen for the feature combination. Many of the proposed techniques miss such aspects which in fact prompted the present work. It is also noticed from the literature that, there is no single framework that can accept multiple variety of social media inputs in the form of text, voice, image, textual image, or a combination among these. In view of the above, we report an integrated framework for text, image and voice for sentiment mining of social media which can process the inputs more efficiently and generate accurate prediction labels.

The following section illustrates the proposed frame work in a phased manner. This is followed by experimental analysis on suitable datasets and comparison with state-of-the-art techniques followed by conclusions.

3. PROPOSED WORK

Conceptual representation of the proposed model has been depicted in Figure 1. The framework consists of five distinct phases described below.

Pre-processing: The raw input is prepared in accordance with the operational requirements.

Text categorization: Our proposed **'Lexicon based LSTM with sentiment word mapping'** mechanism is utilized for classification of texts.

Image data categorization: The image posts are considered for distinct and proper categorization as follows:

i) Textual images are converted to text through suitable image-to-text conversion algorithms.

ii) Sentimental symbolic images are subjected to our proposed generative model for extracting their semantics.

Voice categorization: The voice input if any, is passed through a suitable voice-to-text algorithm and fed to the text categorization module for further classification.

Overall categorization: This is the final stage which categorizes overall processed output buffer data and gives the sentiment risk and important weights as the final output.

All these phases are discussed below in subsequent sections.

3.1 Pre-processing

Pre-processing of the input raw social posts (tweets) needs to be performed carefully. Proper pre processing leads to efficient output prediction. Since the post may contain text, image, voice separation has to be carried out for channelizing individual formats of data through various processing stages. Once, the overall data is transformed into text format, we need to include additional information like time stamp, actual origin, and associated attachments. Special symbols and stop-words removed from the main text. Then, tokenization is carried out which leads to extraction of meaningful words and phrases. Finally, stemming is performed to derive basic word forms.

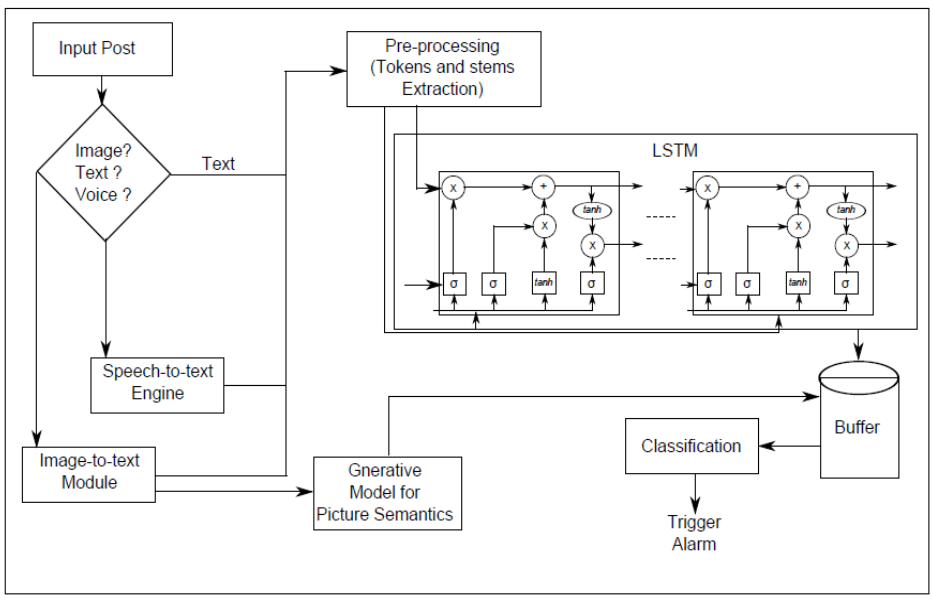

Figure 1. General overview of the proposed integrated framework

3.2 Text categorization

Long short-term memory networks (LSTM) have become the ultimate choice by researchers for performing the task of sentiment analysis. A meaningful label generation can be achieved through mapping of coherent words. However, the focus of this work is to do sentiment mining over a social media post. In the proposed framework a new **'Lexicon based LSTM with sentiment word mapping'** model is developed. The lexicons which are generally sentiment representatives are considered here as an assistive set of symbols. The other set of words are mapped with these sentiment lexicons and a universal label is generated. This task is executed in two steps:

First, the implementation of proposed **'Lexicon based LSTM with sentiment word mapping'** model. This can be boosted by pre-assumed sentiment terms.

Second, customized enhancement is implemented to improve the consistency of the supplementary vectors for usual sentiment analysis.

Consider a word mapping *M* that is derived from the referred vocabulary and a vector s. A mapping can be obtained as α_i = Mxs. A sentiment predictor can be trained in apriori with the sentiment words as the input. The number of layers pertaining to this predictor model can be decided based on the strength of the sentiment. We perform pilot executions to fix the number of layers. We decide the number of categories of the output labels of prediction in the similar fashion. A representative layer is formulated using the Eq. (1), where our model becomes a sentiment word mapping model and the actual prediction can be performed during final training.

$$L_{in} = S_n \oplus \alpha_i \tag{1}$$

where, S_n=sentiment terms and α_i =natural terms.

Both of these are combined and the resultant is given as input to the LSTM setup. Apart from this modification which is reflected as LSTM input, the remaining LSTM architecture is the same as the original one [10]. The predictor can be modeled mathematically as per the Eqns. (2), (3), and (4).

$$S_n = f(W^{\{s_n\}} \times \alpha_i + bias^{\{s_n\}}) \tag{2}$$

$$Label^{s_n} = Classify\ (S_n^{\{softmax\}}) \tag{3}$$

$$Level^{sentiment} = -\sum_i s_n^i log \alpha_i \tag{4}$$

The architecture of our proposed Lexicon based LSTM is shown in the following Eqns. (5), (6), (7), (8), (9), and (10).

$$in_n = \delta(W^{(n)}\alpha_i + W^{(n)}S_n + U^{(n)}h_{n-1} + bias^{(n)}) \tag{5}$$

$$g_n = \delta(W^{(g)}\alpha_i + W^{(g)}S_n + U^{(g)}h_{n-1} + bias^{(g)}) \tag{6}$$

$$out_n = \delta(W^{(out)}\alpha_i + W^{(out)}S_n + U^{(out)}h_{n-1} + bias^{(out)}) \tag{7}$$

$$U_n = tan^{-1}(W^{(U)}\alpha_i + W^{(U)}S_n + U^{(U)}h_{n-1} + bias^{(uU)}) \tag{8}$$

$$a_n = i_n \odot U_n + g_n \odot a_{n-1} \tag{9}$$

$$h_n = out_n \odot tan^{-1}(a_n) \tag{10}$$

The output of our proposed Lexicon based LSTM is (h_1, h_2, ..., h_{size}), where *size* is the length of the input words. We also introduced the concept of strengthening the model which is not present in the original architecture by emphasizing the sentiment word mapping with regular words. The proposed emphasizing vector is given in Eq. (11).

$$E = \sum_{j=1}^{size} weight_j \times portion_j \tag{11}$$

where, $weight_j$=weight of the sentimental word calculated manually using an Apriori table.

Similarly, $portion_j$=strength of the word that indicating likeness to that of the target sentiment label.

3.3 Image data categorization

Automated image semantic description has always been a challenging task in the field of artificial intelligence. We have made to an attempt tackle this challenge. If the input image is just a textual image then we transform the image into its equivalent text and then label it. However, the challenge is if the input image is an image other than textual image. In this context we proposed a generative model (GM) to extract the inhibit semantics of such images and label their sensitive meaning. For this purpose, we have adapted a deep computer vision model and designed an encoder based recurrent neural networks (RNN). The output vectors of equal length are utilized to generate the input image description by optimizing the probability values towards correct representation using the core formula as given below in Eq. (12).

$$\gamma^* = argmax_\gamma \sum_{Img,R} log\ (prob(R|(Img,\gamma))) \tag{12}$$

where, γ is being the key argument of the model, *Img* is the input image post and *R* is the suitable representation of the input.

The joint-probability of the set of representations can be given as Eq. (13):

$$log\ (prob(R|(Img)) = \sum_{n=0}^{N} logprob\ (R_n | Img, R_0, R_1, \ldots, R_{n-1}) \tag{13}$$

For simplicity γ has been ignored in the key argument. Our objective is to optimize the summation term, given the training data (R, *Img)* and model. The probability value into RNN with the hidden buffer (h_n) as Eq. (14):

$$h_n = function\ (h_{n-1}, Image_n) \quad (14)$$

This function is implemented using a regular LSTM where as for implementing the image input mechanism, a CNN (convolutional neural network) is used for implementing the image input. The overall summary of generative model is given in the following Eqns. (15), (16), and (17).

$$Image\_vector_{n-1} = CNN(Image) \quad (15)$$

$$Image\_vector_n = W_{gate} \times R_n \quad (16)$$

$$Prediction\_Label = LSTM(Image\_vector_n) \quad (17)$$

3.4 Voice categorization

The input voice from a post is characterized by converting it to text using deep learning based on the work proposed by Abdel-Hamid et al. [16]. An overview of voice to text transformation and its implementation using deep learning are shown in Figure 2 and Figure 3 respectively.

A: Voice input;
O: Feature vector;
Q: Set of intermediate states;
L: Phoneme;
W: Resultant words;
GMM: Gaussian mixture model;
HMM: Hidden Markov model.

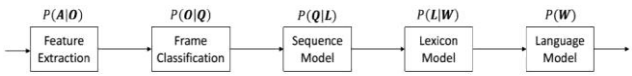

Figure 2. An overview of the voice-to-text transformation

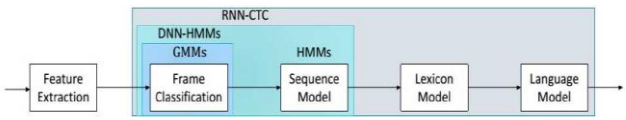

Figure 3. An overview of the specific implementation of voice-to-text using deep learning [16]

3.5 Overall categorization

The outputs from the above text, image, voice categorization is redirected to an intermediate buffer. In the buffer redundant, inconsistent labels, anomalies are removed. The label (for example happy, sad, good, bad, danger, comic, blame etc.) for a specific post from a specific user at a specific instant of time is generated as the final output in the proposed integrated framework.

The final buffer data is directed to a standard support vector model (SVM) that is trained with two different support vector machines (SVMs). One is meant for the image resultant output and the second one is meant for the text part. Since social media post is presently considered as an input from a single user at any instance of time, the output labels from both of these SVMs are combined to get a single semantic of the sentiment. In case, the two outputs are deviating from each other's semantics, then, retraining of either of these SVM are performed until they converge to an identical meaning. Thus, the support vector model shows dynamic learning capability in the context.

4. EXPERIMENTAL ANALYSIS

Different standard datasets are chosen for the implementation and validation of the three separate modules of the proposed work. For the text classification, we chose the dataset proposed by Go et al. [17] with 4,000 samples is chosen. For image data classification the Flickr8k [18] and COCO [19] datasets with 150 samples is used. For voice classification, the RAVDESS dataset with 1440 samples [20] is selected. The k-fold cross-validation method is used to measure overall rates of accuracy for each of the categorization tasks and the overall categorization as well [21-27]. The k-fold cross-validation method is used to measure overall rates of accuracy for each of the categorization tasks and the overall categorization as well. Accuracies are observed for each categories of text, visual and voice data respectively. Overall rate of accuracy is observed to be 87.66% for the proposed integrated framework. Comparison plots for sample observations are generated and shown in Figures 4, 5, and 6 for the three categories respectively. From these graphs, it can clearly observe that for all the types of categorizations (text, image semantics, overall), the proposed model is superior to other reported methods.

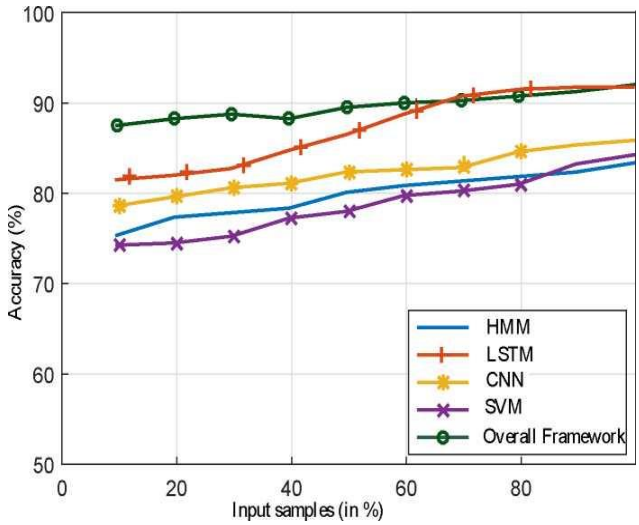

Figure 4. Comparison of proposed integrated framework with the other methods for text input

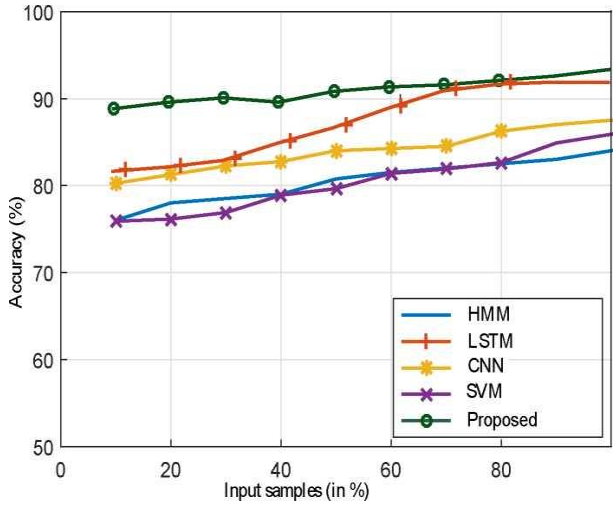

Figure 5. Comparison of proposed integrated framework with the other methods for visual input

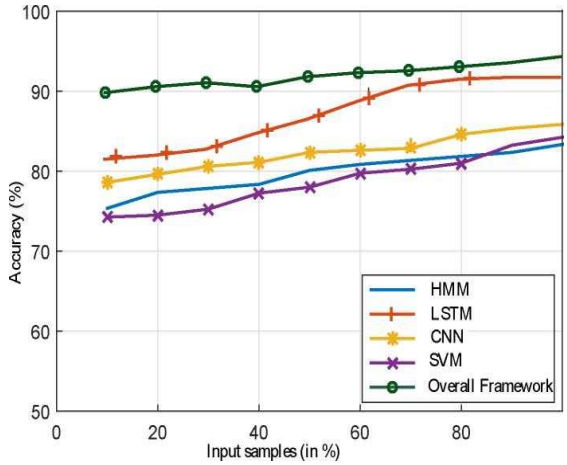

Figure 6. Comparison of overall performance of proposed integrated framework with other methods for complete framework

5. CONCLUSIONS

A new integrated framework has been developed to include text, image and voice sentiment mining from social media posts. The proposed integrated framework is successfully validated in the context of sentiment analysis. The framework is complete since it is able to categorize all types of inputs with distinctly proposed models. This work utilizes **'Lexicon based LSTM with sentiment word mapping'** for sentiment text classification along with a generative model to include image semantics. Validation of individual models are performed over benchmark datasets through a k-fold cross-validation technique. Overall performance of the integrated framework is found to be 87% for text, visual and voice data inputs respectively. Obtained results indicate that the performing of the proposed framework work is superior to all other methods.

ACKNOWLEDGMENT

This work is supported by the Ministry of Electronics and information Technology, Government of India under the Visvesvaraya PhD scheme for Young Faculty Research Fellows implemented by digital India Corporation, sanctioned to one of the authors (Dr. A.M. Sowjanya).

REFERENCES

[1] Bello, G., Menéndez, H., Okazaki, S., Camacho, D. (2013). Extracting collective trends from twitter using social-based data mining. In International Conference on Computational Collective Intelligence, pp. 622-630. https://doi.org/10.1007/978-3-642-40495-5_62

[2] Asghar, M.Z., Khan, A., Ahmad, S., Qasim, M., Khan, I.A. (2017). Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PloS One, 12(2): e0171649. https://doi.org/10.1371/journal.pone.0171649

[3] Al-Twairesh, N., Al-Negheimish, H. (2019). Surface and deep features ensemble for sentiment analysis of Arabic tweets. IEEE Access, 7: 84122-84131. https://doi.org/10.1109/ACCESS.2019.2924314

[4] Zhao, J.Q., Gui, X.L., Zhang, X.J. (2018). Deep convolution neural networks for twitter sentiment analysis. IEEE Access, 6: 23253-23260. https://doi.org/10.1109/ACCESS.2017.2776930

[5] Ye, Z., Li, F., Baldwin, T. (2018). Encoding sentiment information into word vectors for sentiment analysis. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, USA, pp. 997-1007. https://aclanthology.org/C18-1085

[6] Araque, O., Corcuera-Platas, I., Sánchez-Rada, J.F., Iglesias, C.A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications, 77: 236-246. https://doi.org/10.1016/j.eswa.2017.02.002

[7] Giatsoglou, M., Vozalis, M.G., Diamantaras, K., Vakali, A., Sarigiannidis, G., Chatzisavvas, K.C. (2017). Sentiment analysis leveraging emotions and word embeddings. Expert Systems with Applications, 69: 214-224. https://doi.org/10.1016/j.eswa.2016.10.043

[8] Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., Qin, B. (2014). Learning sentiment-specific word embedding for twitter sentiment classification. In ACL, pp. 1555-1565.

[9] Hassan, A., Mahmood, A. (2018). Convolutional recurrent deep learning model for sentence classification. IEEE Access, 6: 13949-13957. https://doi.org/10.1109/ACCESS.2018.2814818

[10] Rehman, A.U., Malik, A.K., Raza, B., Ali, W. (2019). A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. Multimedia Tools and Applications, 78(18): 26597-26613. https://doi.org/10.1007/s11042-019-07788-7

[11] Rezaeinia, S.M., Rahmani, R., Ghodsi, A., Veisi, H. (2019). Sentiment analysis based on improved pre-trained word embeddings. Expert Systems with Applications, 117: 139-147. https://doi.org/10.1016/j.eswa.2018.08.044

[12] Narayanan, R., Liu, B., Choudhary, A. (2009). Sentiment analysis of conditional sentences. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pp. 180-189.

[13] Farías, D.I.H., Sulis, E., Patti, V., Ruffo, G., Bosco, C. (2015). Valento: Sentiment analysis of figurative language tweets with irony and sarcasm. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pp. 694-698.

[14] Ganapathibhotla, M., Liu, B. (2008). Mining opinions in comparative sentences. In Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), pp. 241-248.

[15] Phan, H.T., Nguyen, N.T., Tran, V.C., Hwang, D. (2019). A method for detecting and analyzing the sentiment of tweets containing conditional sentences. In Asian Conference on Intelligent Information and Database Systems, pp. 177-188. https://doi.org/10.1007/978-3-030-14799-0_15

[16] Abdel-Hamid, O., Mohamed, A.R., Jiang, H., Deng, L., Penn, G., Yu, D. (2014). Convolutional neural networks for speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(10): 1533-1545. https://doi.org/10.1109/TASLP.2014.2339736

[17] Go, A., Bhayani, R., Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N project report, Stanford, 1(12): 2009. https://www.kaggle.com/adityajn105/flickr8k/.

[18] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Zitnick, C.L. (2014). Microsoft coco: Common objects in context. Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48

[19] Livingstone, S.R., Russo, F.A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PloS One, 13(5): e0196391. https://doi.org/10.1371/journal.pone.0196391

[20] Kim, H., Jeong, Y.S. (2019). Sentiment classification using convolutional neural networks. Applied Sciences, 9(11): 2347. https://doi.org/10.3390/app9112347

[21] Parmar, H., Bhanderi, S., Shah, G. (2014). Sentiment mining of movie reviews using random forest with tuned hyperparameters. In International Conference on Information Science, pp. 1-6.

[22] Abdi, A., Shamsuddin, S. M., Hasan, S., Piran, J. (2019). Deep learning-based sentiment classification of evaluative text based on multi-feature fusion. Information Processing & Management, 56(4): 1245-1259. https://doi.org/10.1016/j.ipm.2019.02.018

[23] Kotelnikov, E.V., Pletneva, M.V. (2016). Text sentiment classification based on a genetic algorithm and word and document co-clustering. Journal of Computer and Systems Sciences International, 55(1): 106-114. https://doi.org/10.1134/S1064230715060106

[24] Araque, O., Corcuera-Platas, I., Sánchez-Rada, J.F., Iglesias, C.A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications, 77: 236-246. https://doi.org/10.1016Zj.eswa.2017.02.002

[25] Raczko, E., Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. European Journal of Remote Sensing, 50(1): 144-154. https://doi.org/10.1080/22797254.2017.1299557

[26] Behera, R.K., Jena, M., Rath, S.K., Misra, S. (2021). Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. Information Processing & Management, 58(1): 102435. https://doi.org/10.1016/jipm.2020.102435

[27] Graesser, L., Gupta, A., Sharma, L., Bakhturina, E. (2017). Sentiment Classification using Images and Label Embeddings. arXiv preprint arXiv:1712.00725.