International Journal of Informatics, Information System and Computer Engineering

NEW MODERN APPROACH TO PREDICT USERS' SENTIMENT USING CNN AND BLSTM

*R. Sathish Kumar**

* Department of Computer Science and Engineering
Manakula Vinayagar Institute of Technology, Kalitheerthalkuppam, Puducherry, India.
*Corresponding Email: sathishmail8@gmail.com

A B S T R A C T S

In Today's world social network play a vital role and provides relevant information on user opinion. This paper presents emotional health monitoring system to detect stress and the user mood. Depending on results the system will send happy, calm, relaxing or motivational messages to users with phycological disturbance. It also sends warning messages to authorized persons in case a depression disturbance is detected by monitoring system. This detection of sentence is performed through convolution neural network (CNN) and bi-directional long-term memory (BLSTM). This method reaches accuracy of 0.80 to detect depressed and stress users and also system consumes low memory, process and energy. We can do the future work of this project by also including the sarcastic sentences in the dataset. We can also predict the sarcastic data with the proposed algorithm.

A R T I C L E I N F O

Article History:
Received 18 Dec 2022
Revised 20 Dec 2022
Accepted 25 Dec 2022
Available online 26 Dec 2022

Keywords:
Sentimental Analysis, Recommendation System, Deep Learning, CNN, BLSTM, Social Networks.

1. INTRODUCTION

Nowadays, the number of active social network users has grown drastically. This high number of users on social networks is mainly due to the increase of the number of mobile devices, such as smart phones and tablets. Currently OSN have become a universal means of opinion, expression, feelings, and they reflect the bad habits or wellness practices of each user. In recent years, the analysis of the messages posted on social networks have been used by many applications in the industry of healthcare informatics. At first, social media existed to help end users connect digitally with friends, colleagues, family members, and like-minded individuals they might never have met in person. Desktop access to bulletin board services such as CompuServe and Prodigy made it easier to grow free online communities without ever

leaving the house. As social media companies grew their user bases into the hundreds of millions, the business applications of Facebook, Twitter, and other social platforms began to take shape. Social media companies had access to some of the richest trackable user data ever conceived. Another downside of many of the Internet's segmented communities is that users tend to be exposed only to information they are interested in and opinions they agree with. This lack of exposure to novel ideas and contrary opinions can create or reinforce a lack of understanding among people with different beliefs, and make political and social compromise more difficult to come by.

2. LITERATURE SURVEY

2.1. The Use of social media for Communication

Social media takes on many different forms including magazines, Internet forums, weblogs, social blogs, micro blogging, wikis, podcasts, photographs or pictures, video, rating and social bookmarking (Glavan et al., 2016; WHO, 2016; Zhang et al., 2018). With the world in the midst of a social media revolution, it is more than obvious that social media like face book, twitter, orkut, myspace, skype etc., are used extensively for the purpose of communication.

2.2. Music Recommendation System Based on User's Sentiments Extracted from Social Networks.

This paper presents a music recommendation system based on a sentiment intensity metric, named enhanced Sentiment Metric (eSM) that is the association of a lexicon-based sentiment metric with a correction factor based on the user's profile (Rosa et al., 2018; Al-Qurishi et al., 2018; Sathish Kumar & Pariselvam, 2012). This correction factor is discovered by means of subjective tests, conducted in a laboratory environment. Based on the experimental results.

2.3. Hunting Suicide Notes in Web2.0 - Preliminary Findings:

Y. P. Huang, T. Goh, and C. L. Liew. This paper will explore the techniques used by other researchers in the process of identifying emotional content in unstructured data, and will make use of existing technologies to attempt to identify at-risk bloggers (Huang et al., 2007; Sathish et al., 2016; Sathish et al., 2016; Sathish et al., 2017). Using a selection of real blog entries harvested from MySpace.com, supplemented with artificial entries from our research, we test the accuracy of a simple algorithm for scoring the presence of certain key words and phrases in blog entries. Despite the simplistic approach taken, the preliminary results of this study were very promising.

2.4. Detecting Stress Based on Social Interactions in Social Networks:

In this paper, we find that users stress state is closely related to that of his/her friends in social media, and we employ a large-scale dataset from real-world social platforms to systematically study the correlation of users' stress states and social interactions (Lin et al., 2017; Sathish et al., 2019; Thapliyal et al., 2017; Berbano et al., 2017; Sathish et al., 2019). We first define a set of stress-related textual, visual, and social attributes from various aspects, and then propose a novel hybrid model - a factor graph model combined with Convolution Neural Network to leverage tweet content and social interaction information for stress detection.

2.5. Deep Learning Based Document Modeling for Personality Detection from Text:

The authors train a separate binary classifier, with identical architecture, based on a novel document modeling technique. Namely, the classifier is implemented as a specially designed deep convolution neural network, with injection of the document-level Marissa features, extracted directly from the text, into an inner layer (Majumeder et al., 2017; Xue et al, 2014; Tsugawa et al., 2015; Sathish et al., 2020; Rodrigues et al., 2016). The first layers of the network treat each sentence of the text separately; then the sentences are aggregated into the document vector. Filtering out emotionally neutral input sentences improved the performance. This method outperformed the state of the art for all five traits, and the implementation is freely available for research purposes.

3. EXISTING SYSTEM

Machine Learning is that field of study that provides computers the aptitude to find out while not being expressly programmed. Millilitre is one of the foremost exciting technologies that one would have ever come upon. Because it is clear from the name, it provides the pc that creates it additional the same as humans the power to find out. Machine learning is actively being employed these days, maybe in more places than one would expect (Ma & Hovy, 2016; Lample et al., 2016; Khodayar et al., 2017; Guimaraes et al., 2017; Araque et al., 2017).

3.1. Types of Machine Learning:

Machine learning implementations are classified into 3 major classes, betting on the character of the training “signal” or “response” on the market to a learning system.

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning
4. Semi-Supervised learning

1. Supervised Learning

When associate degree algorithmic rule learns from example knowledge and associated target responses which will carry with it numeric values or string labels, like categories or tags, so as to later predict the proper response once displayed with new examples comes beneath the class of supervised learning.

2. Unsupervised Learning

Whereas once associate degree algorithmic rule learns from plain examples with none associated response, going away to the algorithmic rule to work out the info patterns on its own. This sort of algorithmic rule tends to reconstitute the info into one thing else, like new options that will represent a category or a brand-new series of un-correlated values.

3. Reinforcement Learning.

When you give the rule with examples that lack labels, as in unsupervised learning. However, you may accompany associate example with positive or feedback per the solution the rule proposes comes beneath the category of Reinforcement learning, that's connected to applications that the rule ought to produce picks (so the merchandise is prescriptive.

4. SEMI-SUPERVISED LEARNING.

Where associate degree incomplete coaching signal is given: a coaching set with some (often many) of the target outputs missing. There's a special case of this principle called Transduction wherever the complete set of downside instances is understood as learning time, except that a part of the target area unit is missing (Fig. 1).

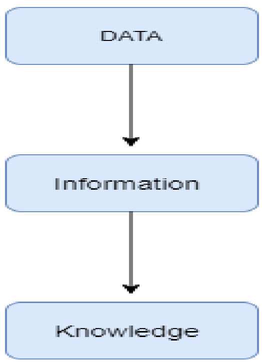

Fig. 1. Flow Diagram – ML Working

A common example of an application of semi-supervised learning is a text document classifier. This is the type of situation where semi-supervised learning is ideal because it would be nearly impossible to find a large amount of labeled text documents. This is simply because it is not time efficient to have a person read through entire text documents just to assign it a simple classification. So, semi-supervised learning allows for the algorithm to learn from a small amount of labeled text documents while still classifying a large amount of unlabeled text documents in the training data (Fig. 2).

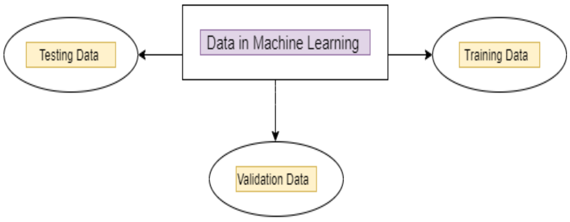

Fig. 2. Flow Diagram – Data Splitting

3.2. Limitations of Existing Work:

- The existing system shows accuracy of only 60%
- It uses only Random Forest Algorithm to process the data.
- Its efficiency of processing the results is slow and so the results appear
- Its user interface is not friendly.
- It shows error in some results.
- It does not process all the data; it leaves some of the data.

5. PROPOSED WORK

5.1. Algorithms Used

1) Random Forest

Random Forest may well be a machine learning formula that belongs to the supervised learning technique. It is typically used for every Classification and Regression problem in cubic centimeter. It's supported the conception of ensemble learning. Random Forest is one of the best high-performance strategies widely applied in numerous industries due to its effectiveness. It can handle data very effectively, whether it is binary, continuous, or categorical. Random forest is difficult to beat in terms of performance. Of course, you can always discover a model that performs better, for example, neural networks. Still, they take longer to construct and can handle a wide range of data types, including binary, category, and numerical. One of the finest aspects of the Random Forest is that it can accommodate missing values, making it an excellent solution for anyone who wants to create a model quickly and efficiently (Fig. 3).

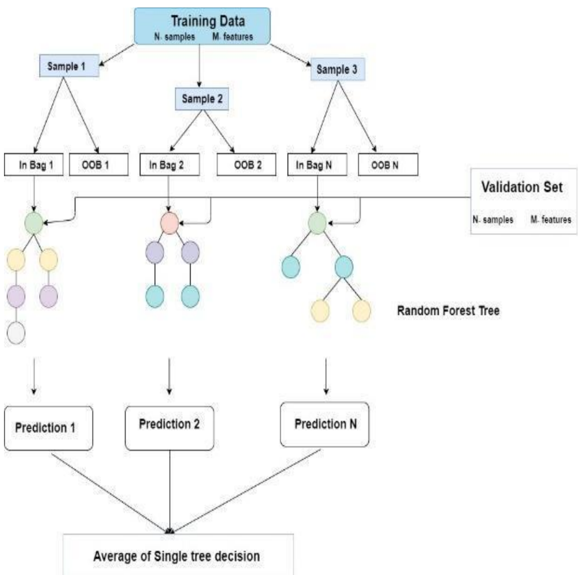

Fig. 3. Architecture Diagram – Diagram Random Forest

2) BLSTM Algorithm:

BLSTM is an associate degree extension of ancient LSTM. It will improve model performance on sequence classification issues. In issues wherever all time steps of the input sequence area unit offered, BLSTM train 2 rather than one LSTM on the input sequence. LSTM in its core preserves information from inputs that has already passed through it using the hidden state. Unidirectional LSTM only preserves information of the past because the only inputs it has seen are from the past. Using bidirectional will run your inputs in two ways, one from past to future and one from future to past and what differs this approach from unidirectional is that in the LSTM that runs backwards you preserve information from the future and using the two hidden states combined you are able in any point in time to preserve information from both past and future.

3) Convolutional Neural Network (CNN):

A convolution is the straightforward application of a filter to associate degree input that leads to activation. Indicating the locations associate degree strength of a detected feature in an input, like a picture. The innovation of CNN is the ability to mechanically learn an outsized range of filters in parallel specific to a coaching dataset below the constraints of a particular prognostication modelling drawback, like image classification. The experimental analysis shows that the model gives the accuracy of 84%.

6. RESULTS AND DISCUSSION

6.1. Upload OSN Dataset

Using this module, we will upload dataset to application. User profile and user data: database built from the data captured from OSNs. Messages: there is a database with 360 messages, 90 messages for each kind (relaxing, motivational, happy, or calm messages) to be suggested to the user by the recommendation engine. The users can previously choose one or two kinds of messages when they undergo a period of stress or depression. The messages were written by 3 Specialists in psychology and validated by 3 other Specialists.

6.2. Generate Train & Test Model from OSN Dataset:

Using this module, we will read all messages from dataset and build a train and test model by extracting features from dataset. Depression or stress detection by machine learning: the sentences are extracted from OSN and they are filtered by machine learning to detect depression or stress conditions. It is implemented in the emotional health monitoring system.

6.3. Build CNN BLSTM-RNN Model Using SoftMax:

Using this module, we will build deep learning BLSTM model on dataset and then using test data we will calculate BLSTM prediction accuracy. The CNNs have several different filters/kernels consisting of trainable parameters which can convolve on a given image spatially to detect features like edges and shapes. Hence, they can successfully boil down a given image into a highly abstracted representation which is easy for predicting.

6.4. Upload Test Message & Predict Sentiment & Stress:

Using this module, we will upload test messages and then application will detect stress by applying BLSTM model on test data. In the proposed system, users' personal information and context information is used. However, users do not always post this related information. In case users do not post personal information, standard information is used, such as sleep routine of 8 hours, no unhealthy habits, no preferences about work or study. It is important to note that in our tests only 5% of the users do not post this information. A traditional RS is also implemented, in which only the words searched by a person on OSN are used to feed the system, forming a content-based RS. For the sake of simplicity, the traditional content-based RS will not be explained in this section (Figs. 4-10).

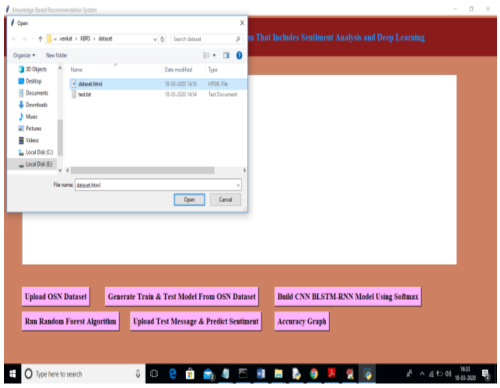

Fig. 4. In Fig. 4 I Am Uploading the Dataset File Which Contains Messages.

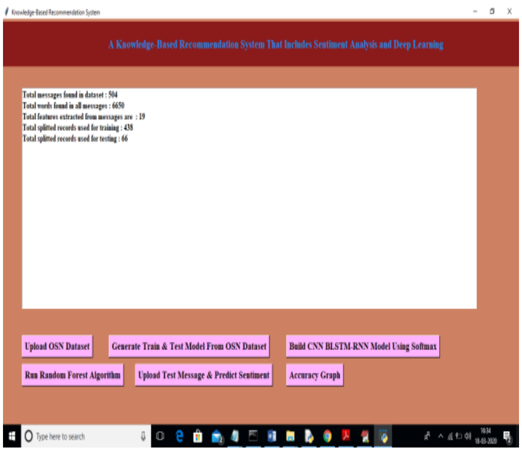

Fig. 5. In The Above Screen We Can See the Records for to Test the Prediction Performance.

Fig. 6. BLSTM Model Generated and The Accuracy Is Shown As 83.94%.

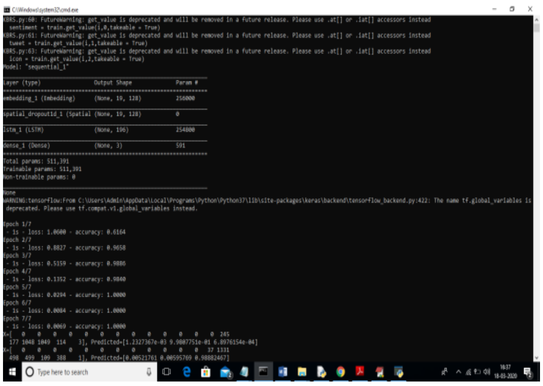

Fig. 7. In The Above Screen We Can See the Iterations to Generate Prediction Layers.

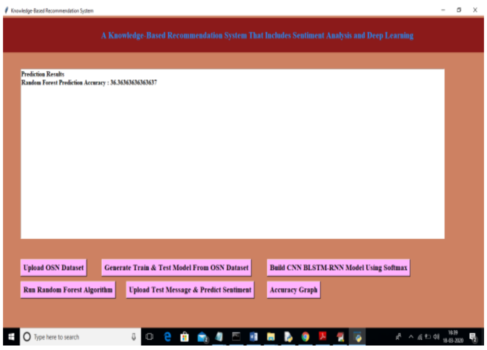

Fig. 8. The Random Forest Prediction Accuracy Is 36% Which Is Lower Than Proposed BLSTM Accuracy.

Fig. 9. In The Above Screen We Can See Each Message Application Detected and Mark with Stress or Non-Stress Status.

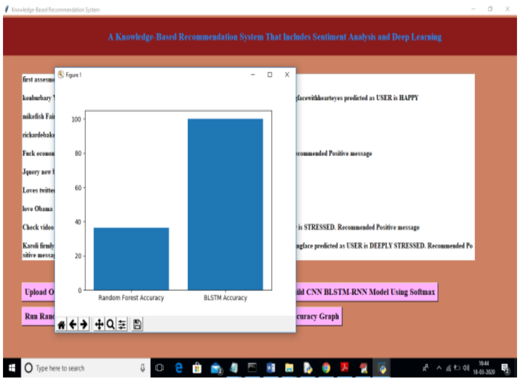

Fig. 10. Accuracy Of Both the Algorithms.

7. CONCLUSION

Various deep-learning techniques can be used for the prediction of Sentimental analysis and recommendation. The challenge is to develop accurate and computationally efficient medical data classifiers. In this paper the model contains the emotional health monitoring system, which uses the deep learning model and the sentiment metric named eSM2. The sentences are extracted from an OSN and then emotional health monitoring system identifies which sentences present a stress or depression content using machine learning algorithms and the emotion of the sentence content.

REFERENCES

Al-Qurishi, M., Hossain, M. S., Alrubaian, M., Rahman, S. M. M., & Alamri, A. (2017). Leveraging analysis of user behavior to identify malicious activities in large-scale social networks. *IEEE Transactions on Industrial Informatics*, *14*(2), 799-813.

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.

Berbano, A. E. U., Pengson, H. N. V., Razon, C. G. V., Tungcul, K. C. G., & Prado, S. V. (2017, September). Classification of stress into emotional, mental, physical and no stress using electroencephalogram signal analysis. In *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (pp. 11-14). IEEE.

Glavan, I. R., Mirica, A., & Firtescu, B. N. (2016). The Use of Social Media for Communication In Official Statistics at European Level. *Romanian Statistical Review*, *64*(4), 37-48.

Guimaraes, R. G., Rosa, R. L., De Gaetano, D., Rodriguez, D. Z., & Bressan, G. (2017). Age groups classification in social network using deep learning. *IEEE Access*, *5*, 10805-10816.

Huang, Y. P., Goh, T., & Liew, C. L. (2007, December). Hunting suicide notes in web 2.0-preliminary findings. In *Ninth IEEE International Symposium on Multimedia Workshops (ISMW 2007)* (pp. 517-521). IEEE.

Khodayar, M., Kaynak, O., & Khodayar, M. E. (2017). Rough deep neural architecture for short-term wind speed forecasting. *IEEE Transactions on Industrial Informatics*, *13*(6), 2770-2779.

Kumar, R. S., & Pariselvam, S. (2012). Formative impact of Gauss Markov mobility model on data availability in MANET‖. *Asian Journal of Information Technology*, *11*(3), 108-116.

Kumar, R. S., Dhinesh, T., & Kathirresh, V. Consensus Based Algorithm to Detecting Malicious Nodes in Mobile Adhoc Network. *International Journal of Engineering Research & Technology (IJERT)* Vol, 6.

Kumar, R. S., Koperundevi, S., & Suganthi, S. (2016). Enhanced Trust Based Architecture in MANET using AODV Protocol to Eliminate Packet Dropping Attacks. *International Journal of Engineering Trends and Technology*, *34*, 21-27.

Kumar, R., Logeswari, R., Devi, N., & Bharathy, S. (2017). Efficient clustering using ECATCH algorithm to extend network lifetime in wireless sensor networks. *Int. J. Eng. Trends Technol*, *45*, 476-481.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*.

Lin, H., Jia, J., Qiu, J., Zhang, Y., Shen, G., Xie, L., ... & Chua, T. S. (2017). Detecting stress based on social interactions in social networks. *IEEE Transactions on Knowledge and Data Engineering*, *29*(9), 1820-1833.

Ma, X., & Hovy, E. (2016). End-to-end sequence labeling via bi-directional lstm-cnns-crf. *arXiv preprint arXiv:1603.01354*.

Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, *32*(2), 74-79.

Rodrigues, R. G., das Dores, R. M., Camilo-Junior, C. G., & Rosa, T. C. (2016). SentiHealth-Cancer: a sentiment analysis tool to help detecting mood of patients in online social networks. *International journal of medical informatics*, *85*(1), 80-95.

Rosa, R. L., Rodriguez, D. Z., & Bressan, G. (2015). Music recommendation system based on user's sentiments extracted from social networks. *IEEE Transactions on Consumer Electronics*, *61*(3), 359-367.

Sathish Kumar R, Abdulla M.G. (2019). "Head Gesture and Voice Control Wheel Chair System using Signal Processing", *Asian Journal of Information Technology,* 18(18), Issue 8, 1682-3915.

Sathish Kumar R, Girivarman R, Parameshwaran S, Sriram V. (2020) "STOCK PRICE PREDICTION USING DEEP LEARNING AND SENTIMENTAL ANALYSIS", *International Journal of Emerging Technologies and Innovative Research* 7(5), 346-354.

Sathishkumar, R., Kalaiarasan, K., Prabhakaran, A., & Aravind, M. (2019, March). Detection of lung cancer using SVM classifier and KNN algorithm. In 2019 *IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)* (pp. 1-7). IEEE.

Thapliyal, H., Khalus, V., & Labrado, C. (2017). Stress detection and management: A survey of wearable smart health devices. *IEEE Consumer Electronics Magazine*, *6*(4), 64-69.

Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., & Ohsaki, H. (2015, April). Recognizing depression from twitter activity. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 3187-3196).

World Health Organization. (2016). *World health statistics 2016: monitoring health for the SDGs sustainable development goals*. World Health Organization.

Xue, Y., Li, Q., Jin, L., Feng, L., Clifton, D. A., & Clifford, G. D. (2014, April). Detecting adolescent psychological pressures from micro-blog. In *International Conference on Health Information Science* (pp. 83-94). Springer, Cham.

Zhang, Y., Xu, C., Li, H., Yang, K., Zhou, J., & Lin, X. (2018). HealthDep: An efficient and secure deduplication scheme for cloud-assisted eHealth systems. *IEEE Transactions on Industrial Informatics*, *14*(9), 4101-4112.