Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776758

DEEP SOCIAL ATTENTION NETWORK ARCHITECTURE TO DETECT AND PREVENT THE UNWANTED INFORMATION PROPAGATION IN SOCIAL MEDIA NETWORK

Ms. D. Kalaivani
Assistant Professor, Department of Computer Science with Data Analytics,
Vellalar College for Women (Autonomous), Erode.
knhvhs2020@gmail.com

Dr. A. Revathi
Assistant Professor, Department of Computer Science with Data Analytics,
Vellalar College for Women (Autonomous), Erode.
revathi16@gmail.com

Abstract

Online Social Media and Event-based Social Networks are exploring a vast volume of unwanted information on the proliferation of user-generated content. Social Media platform is envisioned to provide a member to generate and exchange range with other members in social media for socializing and knowledge sharing. However, it exhibits numerous complications in the network, especially to members' walls, with the propagation of unwanted information from cyber bullies and haters to their information posted as it spreads among various groups and communities. However, many traditional techniques are employed to manage these challenges, but it produces the wrong misinterpretation. To mitigate the challenges mentioned above, a new deep learning architecture named a deep social attention network is designed to detect and prevent the propagation of unwanted content in social media. In this architecture, the dataset is collected from Facebook. It contains various posts for various statuses and is available in a CSV file. CSV file is preprocessed using data normalization, stop removal, and stemming process for removing stop words, numbers, hashtags, and emojis. Preprocessed data is employed in the vector space model to process the data as the feature vector.

Further, the feature vector is employed in the BERT model to extract the sentiment of the feature vector as a positive instance or negative instance. The positive polarity of the example of the vector is classified as a regular comment, and the negative polarity of the feature vector model is classified as an unwanted comment using a convolution Neural Network learning classifier considered as deep learning architecture. Negatively Classified content is eliminated and prevented from posting in the user wall. Instead, it generates a warning message to the unwanted content posted to members, and unwanted information is placed in the LTSM model as hidden data against exposing the member and other audiences. Experimental analysis of the proposed approach is compared with conventional methods to evaluate the performance. Further performance of the system is computed using performance metrics such as precision, recall, and fmeasure. The proposed model produces 99% accuracy in identifying unwanted information compared to conventional approaches.

Keywords: Unwanted Content Classification, Convolution Neural Network, Deep Learning, Sentiment Analysis

1. Introduction

Online Social Media is exploring a massive volume of unwanted information on the proliferation of user-generated content. Social Media platform is functionalized to generate social data and exchange the generated content among other community members as knowledge sharing task. However, it leads to numerous challenges to social network members on the propagation of unwanted information from cyber bullies and haters to their shared status as it spreads to various community members. However, many conventional approaches have implemented several strategies and constraints to handle these complications, but it leads to false misinterpretation.

To mitigate those complications, a new deep learning architecture entitled a deep social attention network is designed to detect and prevent abusive content from propagating in the social media network. In this architecture, the dataset is collected from Facebook. It contains comments for various statuses and is organized in a CSV file. The dataset file is preprocessed using data normalization, stop removal, and stemming process for removing stop words, hashtags, ings words, numbers, and emojis. Preprocessed data is employed in the vector space model, which processes the data as a vector.

The feature vector is applied to the BERT model to determine the sentiment of the instance. The positive polarity of the feature vector is classified as a regular comment. The negative polarity of the feature vector is classified as an abusive comment using a deep learning classifier considered a convolution neural network on processing the vector in the softmax layer with an activation function. Classified content with a negative comment is eliminated from the user wall and stored in the LTSM model as confidential information, generating a warning message to the user against posting a particular type of Comment.

The rest of the paper is sectioned as follows, section 2 illustrates the related work, and Section 3 represents the design of the profound social attention network using a Convolution Neural Network for classification and prediction and the LSTM model for storing the abusive content as confidential information It processes the data collected from the FACEBOOK on the incorporation of the sentiment framework named as BERT model. Section 4 evaluates the current model's performance against the traditional model on accuracy measures in the mentioned experimental setup.

2. Related work

In this section, various related work using machine learning models to filter and handle abusive content is analyzed and detailed on the functional specification of the dataset of the social media network as follows.

2.1. Unwanted Message Filtering using Support Vector Machine

In this architecture, unwanted content in social media is extracted and filtered using machine learning algorithms. It employs the support vector machine algorithm to detect an unwanted messages from the content posted on the user walls of Twitter and Facebook. Specific models process the dataset with preprocessing, feature extraction, and classification processes. It yields an accuracy of 88% in predicting the unwanted content from the regular content on the user's walls

3. Proposed model

In this part, a detailed specification of the current deep social attention network architecture for classification along sentiment analysis architecture and content management architecture such as LTSM to identify the polarity of the extracted word and primary data processing step of the Facebook dataset is presented as follows:

3.1. Data Transformation

Initially, the dataset extracted from Facebook is transformed into CSV file format as it is a highly reliable information format for matrix operation to the task of vital feature extraction and feature classification. In addition, it is a supported format for prediction and Sentiment Analysis.

3.2. Data Preprocessing

The transformed dataset is preprocessed with normalization, stop word removal, emojis removal, and stemming process for processing the collected data for prediction and prevention of the unwanted information through the following processes

- Stop Word Removal: It is to eliminate the stop words like and, is, was, are, were, etc., as it is represented as non-actionable words
- Stemming process: It is to eliminate "ing" and "ed" words in the comments as it is also described as non-actionable words
- Emoji removal: it is to eradicate emoji in the Comment as it is represented as non-actionable information for the prediction task
- Number Removal: It is to eliminate the numbers in the comments

3.3. Tokenization

It is to separate the words into tokens. The classifier quickly processed tokens. Tokens are represented as words. Tokenization also eliminates the punctuation, whitespaces, and delimiters from the comments, as it is described as meaningless.

3.4. Feature Extraction

Feature Extraction uses the vector space model and n-gram operation to process the data containing the information as an instance. Those instances are represented in vector form by employing a vector space model. The vector comprises the high-frequency words computed using the term frequency. It is the representation of frequently occurring words in a particular comment.

Vector Space model VS of Comment = {Term 1, Term 2, Term 3.....Term n}

3.5. Sentiment Analysis -BERT model

The sentiment is computed to the word vector. In this work, the BERT model is employed to analyze the sentiment of the word vector. It yields the polarity to the words as positive and negative on contextual analysis of the terms in the vector. BERT processes the word on utilization of the sentiment base and modifier base along the rule base to identify the polarity of the word in the vector. It is highly employed in the NLP model.

3.6. Classification - Convolution Neutral Network

Convolution Neural Network is employed to classify the feature vector containing polarity into classes as regular comments and unwanted comments. Convolution Neural Network operates with a convolution layer and a fully connected layer for feature processing Convolution layer process the feature to generate the feature map by utilizing the activation

function. Feature Map is classified using the softmax layer in the fully connected layer. It is activated by ReLu based activation function.

The softmax layer classifies the feature map into a class using the KNN classifier. KNN classifier is a classifier employed to classify the polarity of the word. It uses Euclidean space and distance to process the word vector with polarity to type the Comment. The word vector with positive polarity is considered a regular comment, and the word vector with negative polarity is regarded as an abusive comment.

3.6.1. Convolution layer

The convolution layer processes the word vector with polarity into the feature map on processing using the activation function along the kernel and stride functions. Feature mapping to the instance of the vector with polarity makes the classification task more manageable. Figure 1 represents the architecture of the proposed model using the detection and prevention units.

Word vector = {v1, v2, v3...vn}

Polarity of the vector = {Positive | Negative |

Feature Map = {Positive Feature Vector as Node 1 and Negative Feature Vector as Node 2}

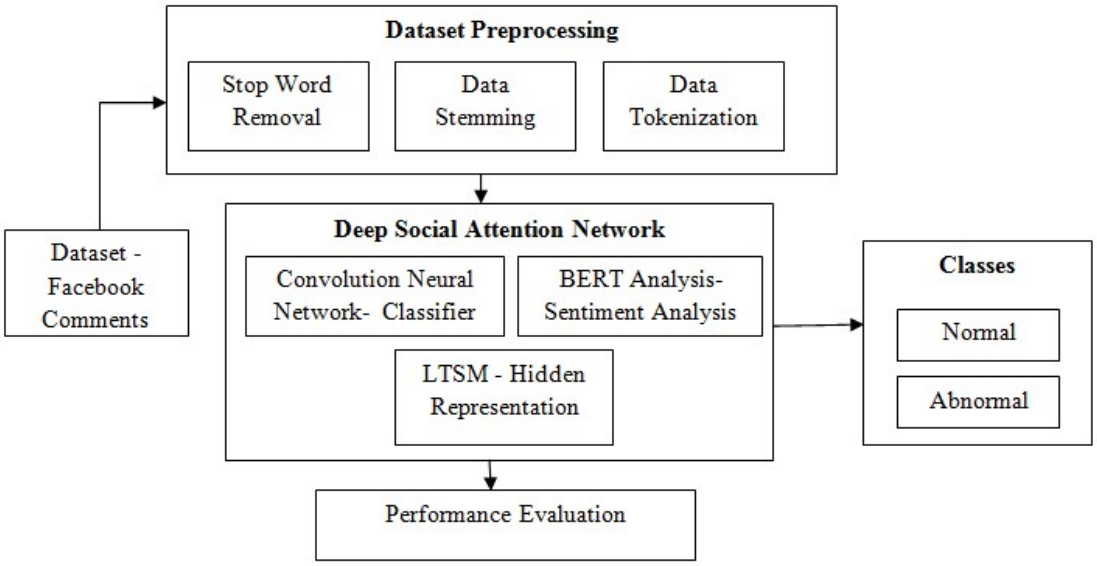

Figure 1: Architecture Diagram of the Proposed model

3.6.2. Fully Connected Layer

In this layer, the feature map generated in the convolution layer is aggregated and processed using the softmax layer. The softmax layer is activated using the ReLu activation function Softmax layer process the polarity-based sentiment word to classify it into a regular and unwanted comment. Polarity-based words from the feature map are classified using a support vector machine. It uses the hyper-plane to classify the feature vector.

3.7. Long-Term Short Memory

Long-Term Short Memory stores the classified unwanted Comment in the hidden states of the model. It keeps the unwanted comment and further trains the model to identify the unwanted

Comment without using sentiment analysis. On detecting the unwanted word by the classifier, the Hidden layer of LTSM transforms the text to encoded form and serves in the user wall.

Algorithm 1: Unwanted Comment Detection and Prevention

Input: User Comment
Output: Unwanted Information Classification
Process:
Normalize Data N = CSV(User Comment)
Preprocess()
S_w = Stop Word Removal (Transformed data)
S_t = Stemming (S_w)
R_t = Remove (Emoji on S_t)
H_t = Removal of Hash tag(R_t)
T = Tokenize(R_t)
Resultant Data R= Tokenized Data T
Compute Term Frequency TF = Resultant Data
TF= frequency of term in the Comment
Vector Space Model
Feature Vector FV = Transform(TF).
BERT ()
Sentiment polarity P = Sentiment base (FV) * Modifier Base (FV)
Classify CNN
Convolution (kernel, Stride)
Feature Map FM = feature Vector
Fully Connected Layer (Feature Map)
Activation Function (FN)
Softmax(SVM)
Class = {Normal, Abusive}
Prevention ()
if (Comment = Abusive)
Store the Comment in LTSM
Encode text = Hidden Layer (Comment)
Display Hidden representation of the LTSM
Else
Display Normal Comment

4. **Experimental results**

The experiment of the proposed approach is carried out on the Facebook dataset [10], and the performance of the proposed method is evaluated using various performance measures such as precision, recall, and F-measure to evaluate the efficiency and accuracy against the conventional approaches.

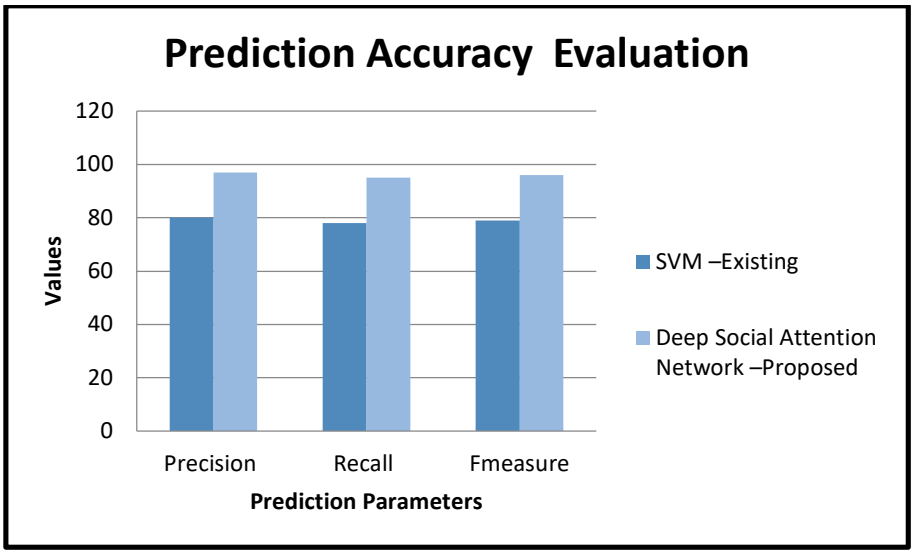

Figure 2: Performance evaluations of the Abusive Detection in Comment

The proposed approach generates high performance by incorporating Sentiment Analysis on the classifier with an encryption technique. Unwanted Comment detection accuracy is computed and depicted in Figure 2 and Table 2.

Table 2: Performance Evaluation

S.No	Technique	Precision	Recall	Fmeasure
1	SVM –Existing	80	78	79
2	Deep Social Attention Network – Proposed	97	95	96

As the proposed approach possessed higher accuracy when compared with the existing system, we opted for the proposed approach to assigning labels for the Comment as unwanted and usual. On detection of unwanted content, it is hidden and displayed on the hidden state of the LTSM model. Figure 3 represents the output of the model.

(a) (b)

Figure 3: Output of the model (a) Comment Input (b) Comment Encrypted Form

Conclusion

We designed and implemented a Deep Social Attention Network-based learning model using Convolution Neural Network, BERT-based Sentiment Analysis, and a long-term short memory for data representation. The proposed model detects unwanted comments on the Facebook page posted by the user. A particular model incorporates the BERT model for Sentiment Analysis of content as it reduces the complexity of the data processing and enhances the detection accuracy of the Comment with the classification of the unwanted content. Further, it hides the unwanted Comment in the hidden state of the LTSM model instead of posting back on the user wall of Facebook. Experimental analysis is considered highly accurate compared to the state of art approaches. It is required to incorporate the constraints to reduce the overfitting issues is regarded as the future direction of the particular work.

References

1. E. F. Cardoso, R. M. Silva, and T. A. Almeida, “Towards automatic filtering of fake reviews, Neuro Computing, Vol. 309, pp. 106116, Oct. 2018.
2. L. Da Xu, W. He, and S. Li, ``Internet of Things in Industries: A Survey," IEEE Trans. Ind. Informat., Vol. 10, No. 4, pp. 22332243, Nov. 2014.
3. Y. Ren and Y. Zhang, ``Deceptive Opinion Spam Detection using Neural Network," in Proc. 26th Int. Conf. Comput. Linguistics: Tech. Papers (COLING), 2016, pp. 140150.
4. N. Jindal and B. Liu, ``Opinion Spam and Analysis," in Proc. Int. Conf. Web Search Web Data Mining (WSDM), 2008, pp. 219230.
5. A. Heydari, M. Tavakoli, and N. Salim, “Detection of Fake Opinions using Time Series," Expert Syst. Appl., Vol. 58, pp. 8392, Oct. 2016.
6. L. Li, W. Ren, B. Qin, and T. Liu, “Learning Document Representation for Deceptive Opinion Spam Detection," in Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Nanjing, China: Springer, 2015, pp. 393404.
7. H. Aghakhani, A. Machiry, S. Nilizadeh, C. Kruegel, and G. Vigna, “Detecting Deceptive Reviews using Generative Adversarial Networks,” in Proc. IEEE Secur. Privacy Workshops (SPW), May 2018, pp. 8995.

8. Yookesh, T. L., et al. “Efficiency of Iterative Filtering Method for Solving Volterra Fuzzy Integral Equations with a Delay and Material Investigation ."Materials Today: Proceedings 47 (2021): 6101-6104.
9. A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, “Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews," Univ. Illinois Chicago, Chicago, IL, USA, Tech. Rep. UIC-CS-03-2013, 2013.
10. E.B.Kumar and V.Thiagarasu, “Color Channel Extraction in RGB images for Segmentation," *2017 2nd International Conference on Communication and Electronics Systems (ICCES),* Coimbatore, India, 2017, pp. 234-239, doi: 10.1109/CESYS.2017.8321272.
11. Reddy, Ch Subba, T. L. Yookesh, and E. Boopathi Kumar. “A Study on Convergence Analysis of Runge-Kutta Fehlberg Method to Solve Fuzzy Delay Differential Equations ."JOURNAL OF ALGEBRAIC STATISTICS 13.2 (2022): 2832-2838.
12. R. Yafeng, J. Donghong, Z. Hongbin, and Y. Lan, “Deceptive Reviews Detection based on Positive and Unlabeled Learning," J. Comput. Res. Develop., Vol. 52, No. 3, pp. 639, 2015.
13. Kumar, E. Boopathi, and V. Thiagarasu. "Comparison and Evaluation of Edge Detection using Fuzzy Membership Functions." International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRCSCE), ISSN (2017): 2454-4248.
14. R. Y. K. Lau, S. Y. Liao, R. C.-W. Kwok, K. Xu, Y. Xia, and Y. Li, ``Text Mining and Probabilistic Language Modeling for Online Review Spam Detection," ACM Trans. Manage. Inf. Syst., Vol. 2, No. 4, pp. 130, Dec. 2011