International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

Review Paper on Emotion Detection from social media using Machine Learning

Uzma[1], Anshul Khurana[2]

[1] M.Tech Scholar(CSE), [2]HoD MTech(CSE)

Department of Computer Science & Engineering, Shri Ram Institute of Technology, Jabalpur,(M.P.)

Abstract

Social networking platforms have revolutionized the way people communicate their feelings and opinions to the world. These platforms provide a convenient and accessible way for individuals to express themselves through various forms of such as text, images, audio, and video, thoughts, emotions, and experiences with others, regardless of geographic or cultural boundaries. The vast amount of data generated on social media platforms every second is largely unstructured, which makes it difficult to analyze and interpret using traditional methods. Sentiment analysis is a useful tool that can help process this data and gain insights into human psychology and behavior. sentiment analysis may not always be sufficient for understanding an individual's emotional/mental state. Emotion detection is a more advanced technique that can help determine an individual's emotional state more precisely by detecting specific emotions, such as joy, anger, sadness, fear, and disgust. Emotion detection goes beyond just identifying the polarity of a text and instead focuses on identifying specific emotions expressed in the text. There are various models and techniques used in emotion detection. However, both sentiment analysis and emotion detection face various challenges, such as dealing with sarcasm, irony, and other forms of figurative language, handling multilingual data, and addressing cultural differences in emotional expression. Overall, this type of review paper can be valuable in helping researchers and practitioners gain a better understanding of sentiment analysis and emotion detection, as well as the challenges and limitations of these techniques. By synthesizing and summarizing existing research and knowledge, review papers can provide a useful starting point for further exploration and investigation in these areas.

Introduction

Emotion detection, affective computing, emotion analysis, and emotion identification are all terms that are used interchangeably to describe the process of identifying and analyzing human emotions through various methods, including text analysis, facial recognition, and physiological sensors. With the increasing use of online social media (OSM) platforms, opinion mining has become an important area of research. Opinion mining, also known as sentiment analysis, refers to the process of analyzing user-generated content to determine the sentiment. Currently, 4.76 billion people utilize social media globally, or little under 60% of the entire world's population. However, this year's net addition of 137 million new users equates to an annual growth of about 3%, slowing down over recent months. Recently, with the surge in social media usage, there has been a growing interest in using AI and machine learning techniques to infer users' emotional states from their social media content. By analyzing social media posts, comments, and other online activities, machine learning algorithms can identify patterns and provide insights into individuals' emotional states. In text-based emotion detection, machine learning algorithms are used to infer the users' emotions from text data by utilizing natural language processing (NLP) approaches to extract significant patterns. Automatic emotion detection from user-generated data on online social media (OSM) platforms can be challenging due to several factors. One of the primary challenges is the scarcity of publicly available datasets labeled with emotion categories, which can limit the accuracy of machine learning algorithms trained on these datasets. Overall, the paper proposes a novel approach to emotion detection from micro-blogs that addresses the challenges associated with this task. The approach combines word and character-level embeddings to capture the rich information present in micro-blog text, achieving high accuracy and F1-score. The proposed approach has important implications for various applications, including marketing, healthcare, and social media analysis.

LITERATURE REVIEW

Kavitha et al. proposed an ensemble classifier (EC) based emotion detection model. They extracted features from micro-blog text using two traditional techniques, Term Frequency-Inverse Document Frequency (TF-IDF) and Bag-of-Words (BoW). They used EC-based model outperformed traditional machine learning models in terms of accuracy, F1-score, and other evaluation metrics.

Sundaram et al.'s study highlights the effectiveness of classical machine learning techniques for emotion detection from tweets and the importance of using appropriate feature extraction techniques such as TF-IDF(traditional technique of Term Frequency-Inverse Document Frequency).

Yousaf et al. proposed a voting classifier that combined the predictions of Logistic Regression (LR), and Stochastic Gradient Descent(SGD) classifiers. The authors evaluated their proposed approach on a dataset of tweets labeled with discrete emotion categories and reported that their voting classifier outperformed the individual classifiers in terms of accuracy, precision, recall, and F1-score.

Al Ajrawi et al. (2021) conducted a study on the application of sentiment analysis in social media marketing. The study aimed to provide insights into the effectiveness of sentiment analysis in identifying consumer opinions and preferences on social media platforms, which can help marketers develop more targeted and effective marketing campaigns.

M. Suhasini et al. aimed to develop a framework for emotion detection from Twitter data using supervised classifiers, specifically Naïve Bayes and k-nearest neighbor (KNN) algorithms. They collected a dataset of tweets related to various events and classified them into four emotion categories: happy-active, happy-inactive, unhappy-active, and unhappy-inactive. The study evaluated the performance of the classifiers using various metrics and found that Naïve Bayes outperformed KNN in classifying the tweets into the four emotion categories.

S. Patil et al. provides an overview of various approaches used for emotion detection, including rule-based, lexicon-based, and machine learning-based techniques. Their methodology involves several steps, such as pre-processing the textual data, creating an embedding layer, using a recurrent neural network (RNN) model for feature extraction, and a multi-layer perceptron (MLP) model for classification. The RNN model helps to capture the context and sequence of words in the text, while the MLP model is used to classify the text into different emotion categories.

W. Deng et al. review various aspects of deep facial expression recognition, including the use of convolutional neural networks (CNNs) for feature extraction, the importance of data augmentation and pre-processing, and the challenges associated with dataset bias and class imbalance.

Anchieta et al. extracted a total of 93 features related to users' writing styles, including the length of tweets in terms of characters and words, frequency of hashtags, frequency of shared hyperlinks, emoticons, ellipses, question marks, and exclamation marks. They compared these features with TF-IDF and Delta TF-IDF to build a sentiment classification system. Three classical machine learning models, namely SVM, Naïve Bayes, and Jun Li et al. proposed a multi-label maximum entropy (MME) model for user emotion classification on short text. The model utilized multiple emotion labels and valence scores assigned by numerous users to generate rich features. The scheme was able to identify entities and provide relevant social emotions for the generated lexicons. The experiments demonstrated the effectiveness of the method on social emotion classification using sparse features. However, the scheme had overfitting issues that needed to be addressed.

D. Kollias et al. proposed a dimensional emotion recognition approach on the OMG in-the-wild dataset. They utilized a combination of

J48, were trained using the extracted features for a binary (positive and negative) sentiment classification task.

Dongliang Xu et al. developed a model was specifically designed for microblog emotion classification and utilized a CNN for feature extraction and classification using of Word2Vec. This method was successful in extracting important features and achieved a high classification accuracy compared to other models, including SVM, RNN, and LSTM.

Srishti Vashishtha et al. developed a sentiment analysis system for social media posts using fuzzy rules and multiple lexicons and datasets. The system integrated natural language processing techniques and word sense disambiguation to create a novel unsupervised nine fuzzy rule-based system for classification. The system provided precise sentiment values and addressed linguistic problems, resulting in better performance than other state-of-the-art methods. However, the system had a high error rate, leading to inaccurate classification.

convolutional neural network (CNN) and recurrent neural network (RNN) models to extract both low-level and high-level features from facial expressions. The method involved using multiple CNNs to extract features from different regions of the face, which were then combined and fed into an RNN model for dimensional emotion recognition. The approach outperformed other state-of-the-art methods on the dataset, achieving high accuracy and correlation with human ratings of emotional valence and arousal.

A. Amin et al. proposed a modified version of the VADER (Valence Aware Dictionary and sEntiment Reasoner) algorithm for sentiment analysis of

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

Bengali language text. They used a Bengali sentiment lexicon, which was created by manually translating the English words from the VADER lexicon into Bengali and then adding new words specific to Bengali language. They also modified the scoring function of VADER to take into account the negation and intensification of sentiment in Bengali language. The performance of the proposed Bengali VADER approach was evaluated on a dataset of Bengali movie reviews and compared with other existing sentiment analysis approaches. The results showed that Bengali VADER outperformed the other approaches in terms of accuracy and F1-score.

Fazeel Abid et al. developed a scheme that used a combination of Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN) models with weighted attentive pooling (WAP) to address issues related to syntactic and semantic regularities and out-of-vocabulary (OOV) words. The scheme also incorporated distributed word representations (DWRs) through a weighted mechanism.

Muhammad Asif et al. developed a sentiment analysis scheme for multilingual textual data from social media, specifically to detect the intensity of extremism sentiments. The proposed method achieved an overall accuracy rate of 82% in identifying extreme sentiment from the multilingual data.

The proposed approach for emotion detection combines multiple types of features to better identify patterns in Twitter data and outperforms traditional machine learning-based emotion detection systems. The approach has potential to improve our understanding of emotions expressed in short, informal tweets.

PROPOSED METHODOLOGY

The SSEL feature representation is used as input for an ensemble emotion detection (ED) model, which combines multiple classifiers to improve the accuracy of the emotion predictions. The ED model is trained using a genetic algorithm (GA)-based feature selection method, which selects the most informative features for the emotion detection task. The GA-based feature selection reduces the dimensionality of the feature space and eliminates features that are not useful for the task at hand.

In summary, the proposed approach involves a combination of feature engineering, machine learning, and ensemble modeling techniques to detect emotions from tweets. The SSEL feature representation and GA-based feature selection are novel contributions of the paper.

PROPOSED EMOTION DETECTION MODEL

Once the SSEL feature representation is created, the paper trains different machine learning models to classify emotions in tweets. To efficiently discover the top-performing classification model, the paper uses a genetic programming (GP)-based approach. The GP technique identifies the top-performing models as random forest (RF), extreme gradient boosting (XGBoost), and linear support vector machine (SVM) classifiers.

To further improve the performance of the emotion classification, the paper develops an ensemble classifier composed of RF, XGBoost, and linear SVM classifiers. The ensemble classifier combines the predictions of multiple classifiers to improve the overall accuracy of the emotion classification task. By leveraging the strengths of multiple models, the ensemble classifier is more robust and less prone to overfitting compared to a single model. Overall, the proposed approach combines feature engineering, machine learning, and ensemble modeling techniques to accurately classify emotions in tweets.

REFERENCES

[1] P. A. Rauschnabel, P. Sheldon, and E. Herzfeldt, ‘‘What motivates users to hashtag on social media?’’ Psychol. Marketing, vol. 36, no. 5, pp. 473–488, May 2019.

[2] S. Kemp. (2022). Digital 2022: Digital Adoption Doubled Over the Past Decade. [Online]. Available: https://datareportal.com/reports/digital2022-digital-adoption-doubled-over-the-past-decade

[3] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, and V. Varadarajan, ‘‘AI based emotion detection for textual big data: Techniques and contribution,’’ Big Data Cognit. Comput., vol. 5, no. 3, p. 43, Sep. 2021.

[4] M. L. Gavrilova, F. Anzum, A. Hossain Bari, Y. Bhatia, F. Iffath, Q. Ohi, M. Shopon, and Z. Wahid, ‘‘A multifaceted role of biometrics in online security, privacy, and trustworthy decision making,’’ in Breakthroughs in Digital Biometrics and Forensics. Cham, Switzerland: Springer, 2022, pp. 303–324.

[5] F. Anzum, A. Z. Asha, and M. L. Gavrilova, ‘‘Biases, fairness, and implications of using AI in social media data mining,’’ in Proc. Int. Conf. Cyberworlds (CW), Sep. 2022, pp. 251–254.

[6] N. Andalibi and J. Buss, ‘‘The human in emotion recognition on social media: Attitudes, outcomes, risks,’’ in Proc. CHI Conf. Hum. Factors Comput. Syst., Apr. 2020, pp. 1–16.

[7] S. Li and W. Deng, ‘‘Deep facial expression recognition: A survey,’’ IEEE Trans. Affect. Comput., vol. 13, no. 3, pp. 1195–1215, Jul. 2022.

[8] M. L. Gavrilova, and P. S. P. Wang, ‘‘Facial metamorphosis using geometrical methods for biometric applications,’’ Int. J. Pattern Recognit. Artif. Intell., vol. 22, no. 3, pp. 555–584, May 2008.

[9] Y. Bhatia, A. H. Bari, and M. Gavrilova, ‘‘A LSTM-based approach for gait emotion recognition,’’ in Proc. IEEE 20th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI*CC), Oct. 2021, pp. 214–221.

[10] P. Nandwani and R. Verma, ‘‘A review on sentiment analysis and emotion detection from text,’’ Social Netw. Anal. Mining, vol. 11, no. 1, pp. 1–19, Dec. 2021.

[11] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, ‘‘The influence of preprocessing on text classification using a bag-of-words representation,’’ PLoS ONE, vol. 15, no. 5, May 2020, Art. no. e0232525.

[12] J. Fürnkranz, ‘‘A study using n-gram features for text categorization,’’ Austrian Res. Inst. Artif. Intell., vol. 3, no. 1998, pp. 1–10,1998.

[13] S. Qaiser and R. Ali, ‘‘Text mining: Use of TF-IDF to examine the relevance of words to documents,’’ Int. J. Comput. Appl., vol. 181, no. 1, pp. 25–29, Jul. 2018.

[14] K. W. Church, ‘‘Word2Vec,’’ Nat. Lang. Eng., vol. 23, no. 1, pp. 155–162, 2017.

[15] J. Pennington, R. Socher, and C. Manning, ‘‘GloVe: Global vectors for word representation,’’ in Proc. Conf. Empirical Methods Natural Lang.

Process. (EMNLP), 2014, pp. 1532–1543.

[16] V. Anusha and B. Sandhya, ''A learning based emotion classifier with semantic text processing,'' in Advances in Intelligent Informatics. Springer, 2015, pp. 371–382.

[17] K. P. Kumar and M. L. Gavrilova, ''Latent personality traits assessment from social network activity using contextual language embedding,'' IEEE Trans. Computat. Social Syst., vol. 9, no. 2, pp. 638–649,Apr. 2022.

[18] P. S. Sreeja and G. Mahalakshmi, ''Emotion models: A review,'' Int. J. Control Theory Appl., vol. 10, no. 8, pp. 651–657, 2017.

[19] P. Ekman, ''Basic emotions,'' in Handbook of Cognition and Emotion. Hoboken, NJ, USA: Wiley, 1999, pp. 45–60.

[20] R. Plutchik, ''A general psychoevolutionary theory of emotion,''in Theories Emotion. Amsterdam, The Netherlands: Elsevier, 1980, pp. 3–33.

[21] D. Kollias and S. Zafeiriou, ''Exploiting multi-CNN features in CNNRNN based dimensional emotion recognition on the OMG in-the-wild dataset,'' IEEE Trans. Affect. Comput., vol. 12, no. 3, pp. 595–606, Jul. 2021.

[22] J. A. Russell, ''A circumplex model of affect,'' J. Personality Social Psychol., vol. 39, no. 6, p. 1161, Dec. 1980.

[23] J. A. Russell and A. Mehrabian, ''Evidence for a three-factor theory of emotions,'' J. Res. Pers., vol. 11, no. 3, pp. 273–294, 1977.

[24] D. Kavitha, P. P. Reddy, and K. V. Rao, ''Emotion recognition in tweets using optimized ensemble classifiers,'' in Proc. 7th Int. Conf. Commun. Electron. Syst. (ICCES), Jun. 2022, pp. 1728–1731.

[25] V. Sundaram, S. Ahmed, S. A. Muqtadeer, and R. Ravinder Reddy, ''Emotion analysis in text using TF-IDF,'' in Proc. 11th Int. Conf. Cloud Comput., Data Sci. Eng. (Confluence), Jan. 2021, pp. 292–297.

[26] A. Yousaf, M. Umer, S. Sadiq, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, ''Emotion recognition by textual tweets classification using voting classifier (LR-SGD),'' IEEE Access, vol. 9, pp. 6286–6295, 2021.

[27] M. Suhasini and B. Srinivasu, ''Emotion detection framework for Twitter data using supervised classifiers,'' in Proc. 3rd Data Eng. Commun. Technol. (ICDECT-2K19). Singapore: Springer, 2020, pp. 565–576.

[28] A. Dvoynikova, O. Verkholyak, and A. Karpov, ''Emotion recognition and sentiment analysis of extemporaneous speech transcriptions in Russian,'' in Proc. Int. Conf. Speech Comput. Cham, Switzerland: Springer, 2020, pp. 136–144.

[29] O. Perepelkina, E. Kazimirova, and M. Konstantinova, ''RAMAS: Russian multimodal corpus of dyadic interaction for affective computing,'' in Proc. Int. Conf. Speech Comput. Cham, Switzerland: Springer, 2018, pp. 501–510.

[30] R. T. Anchiêta, F. A. R. Neto, R. F. D. Sousa, and R. S. Moura, ''Using stylometric features for sentiment classification,'' in Proc. Int. Conf. Intell. Text Process. Comput. Linguistics. Cham, Switzerland: Springer, 2015, pp. 189–200.