INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

Deep Learning on Natural Language Text for Emotion Correlation Mining

Vudutha Raghav[1]

Chandu Delhi Police[2]

[1]MTech Student, Department of CSE, Priyadarshini Institute of Technology & Science, Chintalapudi-522306

[2]Assistant Professor MTech, Department of CSE, Priyadarshini Institute of Technology &Science, Chintalapudi-522 306

Abstract: The extraction of feelings from written communication is currently one of the most-discussed topics in contemporary natural language understanding. When it comes to deep learning methods, embedding and attention mechanisms are quite helpful in terms of emotion recognition. The study of people's feelings has recently garnered the interest of researchers. The majority of earlier research in the field of artificial intelligence have concentrated on emotion recognition rather than mining the reasons why emotions are either not identified or are perceived incorrectly. Both the CNN and the Collaborative algorithm are put to work in this situation. The data that are provided as input are first subjected to pre-processing, followed by classification, and finally, the final results are integrated. For example, the specified number of titles, emotions, hashtags, and commands will be compiled and presented in a distinct format as the output. The link that exists between feelings is one factor that contributes to the inability to recognise emotions. Three different types of characteristics and two different models of deep neural networks are provided in order to mine emotion correlation based on emotion identification through text. A process with an orthogonal basis is used to derive the emotion confusion law. As a significant and important traditional aspect of the field of knowledge mining, emotion analysis can be broken down into the following three levels: the word level, the sentence level, and the document level.

Keywords: Emotion analysis, artificial-intelligence, deep neural network, natural language processing, emotion recognition.

I. Introduction

The experience of emotion is multifaceted, unique, highly subjective, and highly sensitive to one's surroundings. Emotion serves as a driver for decision making, primes the body for action, and moulds behaviour in the long term [1]. The majority of philosophers come to the conclusion that emotions are a person's subjective reaction to the objective reality. This suggests that emotions are the result of the interaction between individuals and society. At a minimum, the following three facets contribute to the complexity of an individual's emotions. 1) Individual value can only be maintained over time by consistent experience. Even when faced with the identical circumstances, different people have different emotional responses. Regarding the Napoleonic War, for instance, there are two distinct points of view. The position held by those in favour of the war is that it is an attack on the French feudal force and that it hastens the progression of history, while the view held by those opposed to the war is that it is unjustified owing to its aggressive intent. 2) When people speak with one another, misunderstandings might arise. The context is understood differently by each person due to their unique histories and experiences. When a person has a deeper understanding of the topic at hand, their perspective is able to more accurately reflect the situation. When there is a discrepancy in the amount of prior knowledge held by the information sender and the information recipient, this might lead to a misunderstanding of first feelings. 3) There is turmoil in the individual's emotional life. The turbulence is affected by the instantaneous mood of the environment, whether it be happy or negative. Alterations in feeling go hand in hand with shifting circumstances pertaining to the same occurrence. The majority of people experience the phenomenon of their internal emotions being influenced by circumstances outside of themselves on a regular basis. For instance, a tweet that appears to be complimentary in tone but is actually critical of one's performance at work can be quite upsetting. On the one hand, the individual's emotional state is complicated as a result of their own unique long-term social experiences, interpersonal misunderstandings, and the instantaneous influence of the environment. On the other hand, the public's feelings regarding the social event are complicated for the reasons that will be discussed below. 1) The social event is complicated since it involves a number of different factors. The information regarding social events, which may cover a wide range of subjects, is published online. Individualized users of the internet have a tendency to focus on a variety of distinct aspects. It's possible that readers will react in a variety of ways, providing a variety of perspectives on the same social occurrence. If the reader is worried about more than one issue, then it is possible that his or her words will contain more than one type of feeling. 2) The feelings of the general public are complex and varied. In the field of social psychology research, the categorization of different feelings is a topic that is still fraught with debate. Love, joy, anger, sadness, fear, and surprise are the six emotions that make up the human emotional spectrum, which is one of the widely accepted classification systems for human feelings. In order to describe these fundamental feelings, Shaver et al. [2] utilised a tree-like framework. Ekman's idea, which was quite analogous to Shaver's theory, including the contentious feeling of surprise [3]. This article divides feelings into the aforementioned six classifications. The proposed strategy is appropriate in situations when different types of emotion categories are being considered.

II. Related work

Emotion analysis has been drawing researchers' attention in recent years. Text emotion distribution learning [5], [6], considered as one of the most important research areas, contributes to many applications. The research on emotion recognition [7], [8] is opening up numerous opportunities pertaining to social media in terms of understanding users' preferences, habits, and their contents. The application of the subjective and emotional data from social media includes but is not limited to sentiment analysis; sarcasm detection; event dissemination; user clustering; and user behaviour analysis. Some tasks of the applications are combined together, such as multitask assignment on sentiment classification and sarcasm detection, which employs the deep neural network in natural language processing (NLP) tasks [16]. Emotion analysis from text is one of the hot topics in modern natural language understanding. Embedding and attention mechanisms help a lot with emotion recognition in deep learning methods. Continuous word representations, including word2vec [17], weighted word embedding [18], and the derivatives [19] denoted words with dense embeddings and provided new ideas for automatic feature mining. Later, different kinds of attention mechanisms and pretrained models were proposed. Wang et al. [20] proposed an embedded recursive neural network for improving emotion recognition. Barros et al. introduced a personalized affective memory. In 2017, Vaswani et al. proposed a new network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. With the capability of modelling bidirectional contexts, denoising autoencoding-based pretraining, such as BERT, and an autoregressive pretraining method, such as XLNet [24], achieved better performance than many other pretrained models on GLUE tasks. Some researchers solved the emotion recognition task through graph modelling [25], such as a capsule network. Popular natural language understanding models, such as long short-term memory (LSTM) [26], convolutional neural network (CNN) [27], recursive autoencoders [28], adversarial learning [29], and attention mechanism [30], have been applied for emotion analysis and classification tasks. Electroencephalography signals and facial expression sequences were also used for emotion recognition with deep learning models. More complex and classification oriented deep learning models made it harder to understand the correlation among emotions even with remarkable recognition accuracy. Emotion analysis, as an important traditional branch of knowledge mining, is categorized into three levels, namely: word level, sentence level, and document level. In word level, emotion words were extracted mainly through three ways: 1) manual approach; 2) dictionary-based approach [34]; and 3) corpus-based approach [35]. Strapparava and Valitutti [36] developed WordNet affect through tagging a subset of synets with affective meanings in English WordNet (EWN). Staiano and Guerini [37] presented DepecheMood, an emotion lexicon produced automatically by harvesting social media data annotated with emotion scores. Then, Badaro et al. provided EmoWordNet by expanding DepecheMood with the synonymy semantic relation from EWN. Emotion lexicons for different languages were developed. In SemEval 2018 Task 1: Affect in Tweets, labelled data from English, Arabic, and Spanish tweets are created for each task. Badaro et al. achieved the best result in the SemEval 2018 emotion classification subtask for the Arabic language. Features that they used were word embeddings from AraVec, and emotion features extracted from ArSEL and NRC emotion lexicon. In the sentence-level analysis, intrasentential and intersentential emotion consistency were explored. Qiu et al. employed dependency grammar to describe relations for double propagation between features and opinions. Ganapathibhotla and Liu adopted dependency grammar for the emotion analysis of comparative sentences. The conditional random fields (CRFs) method was used as the sequence learning technique for extraction. A multitask multilabel (MTML) classification model was proposed to classify sentiment and topics concurrently. By doing this, the closely related tasks, that is, sentiment and topic classification have been improved. Machine-learning methods were widely used in both the

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

sentence and document levels. Naive Bayesian, maximum entropy classification, graphical model, and pattern recognition methods were employed frequently. Zhao et al. explored the correlations among different microblogs for social event detection. Hu and Flaxman provided multimodal sentiment analysis by combining visual analysis and NLP to predict various emotional states of the user in social media. Most of the previous works focused on recognizing emotions from text rather than why emotion was wrongly recognized.

III. Methodology Implemented

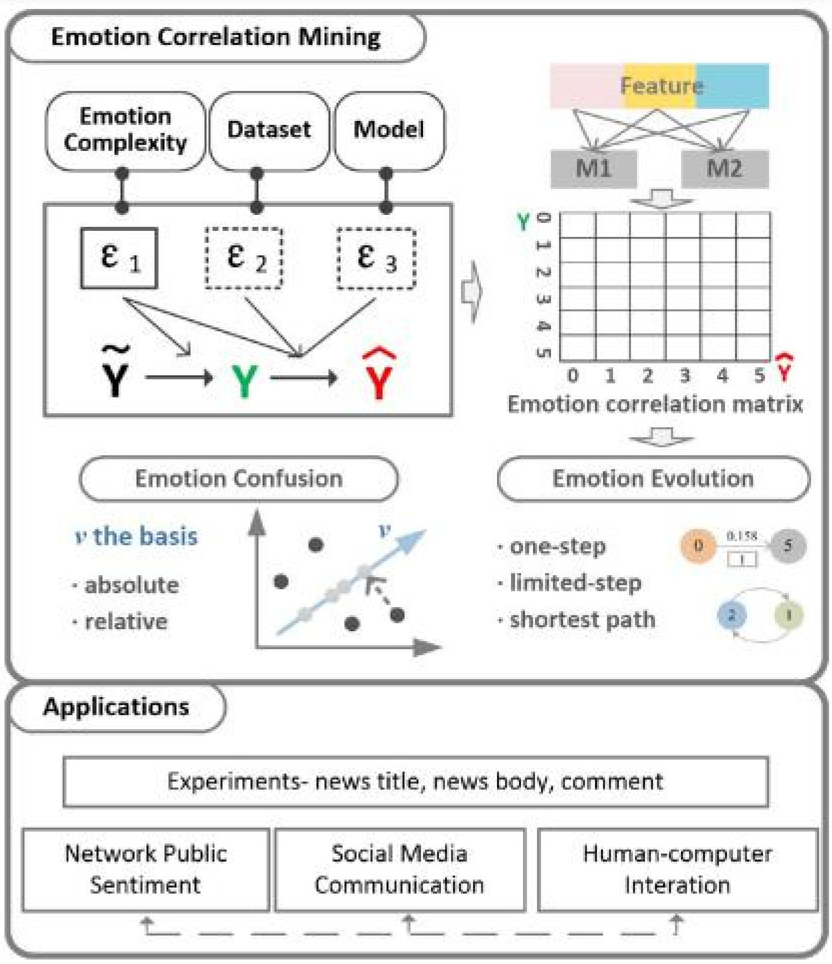

Figure 1: System Architecture

The data set must first be uploaded, and after it has been submitted, text pre-processing must be performed. The CNN algorithm and the COLLABORATIVE algorithm are both employed in this context. difficulty of discovering all instances of a query word that exist in an image of a scanned document without completely understanding the word being searched for. The removal of stop words, in addition to unnecessary data and duplicated data, also takes place. Tweets that are accessible to the public are retrieved and categorised, and the material that is expressed in tweets is extracted and standardised. This research classifies tweets based on a seed list that is specific to a domain. During the process of classifying tweets, semantic and syntactic analysis is performed on the tweets in order to limit the amount of information that is lost. There are three types for hashtags: a topic, which is directly related to a certain hashtag; sentiment hashtags, which express thoughts in a personal manner; and sentiment-topic hashtags, which identify a certain target phrase and the sentiment words. Clustering is the process of dividing a population or set of data points into a number of groups in such a way that the data points in each group are more similar to one another than they are to the data points in any other groups. To put it another way, the objective is to identify groups of people who share certain characteristics and organise them into clusters. Using the algorithm, the tweets in the data set are categorised, the hashtags are categorised, and the groupings are clustered, and then the result is eventually merged. A specific and detailed description of the results that a software module is expected to provide. Because it provides a definitive statement of the requirements that are imposed on the module, it can be utilised both by the implementer of the module and by the users of the module. Additionally, because it provides a precise statement of what the module provides, it can be utilised by both parties. A reliable module specification does not make any promises regarding the means by which the module's effects are realised. The only module that is included in this project is the administration module. To begin, you will need to register for an account on our system. When this step has been completed, you will then be able to log in to the system. Our system is restricted so that only authorised Admin can access it. This is a safety measure. After the steps of registering and logging in have been performed without a hitch. The compiled or connected set of data for which a certain output must be uploaded into the system in order to fulfil the requirements. After that, the processing of the data is finished. Data pre-processing is a technique for data mining that entails transforming raw data into a format that is understandable, where all of the processes, such as stop word removal, stemming, and tokenization, are also done to remove the unwanted data. This format can then be utilised for further analysis. The collaborative algorithm and the CNN algorithm are both utilised in the process of hashtag classification. There are three types for hashtags: a topic, which is directly related to a certain hashtag; sentiment hashtags, which express thoughts in a personal manner; and sentiment-topic hashtags, which identify a certain target phrase and the sentiment words. After that... On the basis of the provided input data and the user's name, feature extraction is performed with the goal of predicting the tweets keyword in group. The process of group recommendation helps to calculate the between value, which means it identifies users who belong to more than one group. This can be accomplished by determining which users belong to more than one group. This assists in identifying any false information or tweets that may be included in our input data. After then, the tweets are categorised into different groups, using a collaborative approach to get rid of any redundant data and the CNN algorithm to help with the clustering process. The process of clustering involves separating a population or data sets into a number of groups in such a way that the data points within each group are more similar to one another than they are to the data points within any other groups. To put it another way, the objective is to identify groups of people who share certain characteristics and organise them into clusters. The most essential phase is carrying out the performance analysis in order to determine the correctness of the algorithm through the application of the confusion matrix. At last, the findings are summarised in the form of a graph, with the size of each section determined by the total number of tweets.

A Convolutional Neural Network (CNN) is a type of Deep Learning algorithm that can take in an input image, give significance to various characteristics or objects within the image, and be able to discern one from the other. When compared to other classification methods, the amount of pre-processing work necessary for a ConvNet is significantly reduced. While filters in basic methods are hand-engineered, ConvNets have the ability, given sufficient training, to learn these filters and properties on their own given enough data. The process of collaborative filtering is a method that might eliminate content that a user might be interested in based on the responses of other users who are similar to that user. It does this by scanning across a big population of people and locating a subset of users who have preferences that are comparable to those of a specific user. It takes into account the things that the user like and compiles them into a ranked list of recommendations.

IV. Discussion and Analysis

ISSN: 2057-5688

The use of social networks as a tool for political engagement and public demonstrations is a relatively new and still developing phenomenon within the context of the current political system. The majority of earlier research in the field of artificial intelligence have concentrated on emotion recognition rather than mining the reasons why emotions are either not identified or are perceived incorrectly. The link that exists between feelings is one factor that contributes to the inability to recognise emotions. The newest is a research domain that is grabbing the attention of the academic community, and it refers to the misuse of hash tags for a different purpose that the one that was originally set, producing confusion to users who are interested in the topic that is related to the original hash tag. On the other hand, the system that is proposed to mine emotion correlation from emotion identification through text presents three different kinds of features in addition to two different deep neural network models. Through the use of an orthogonal basis, the emotion confusion law can be derived.

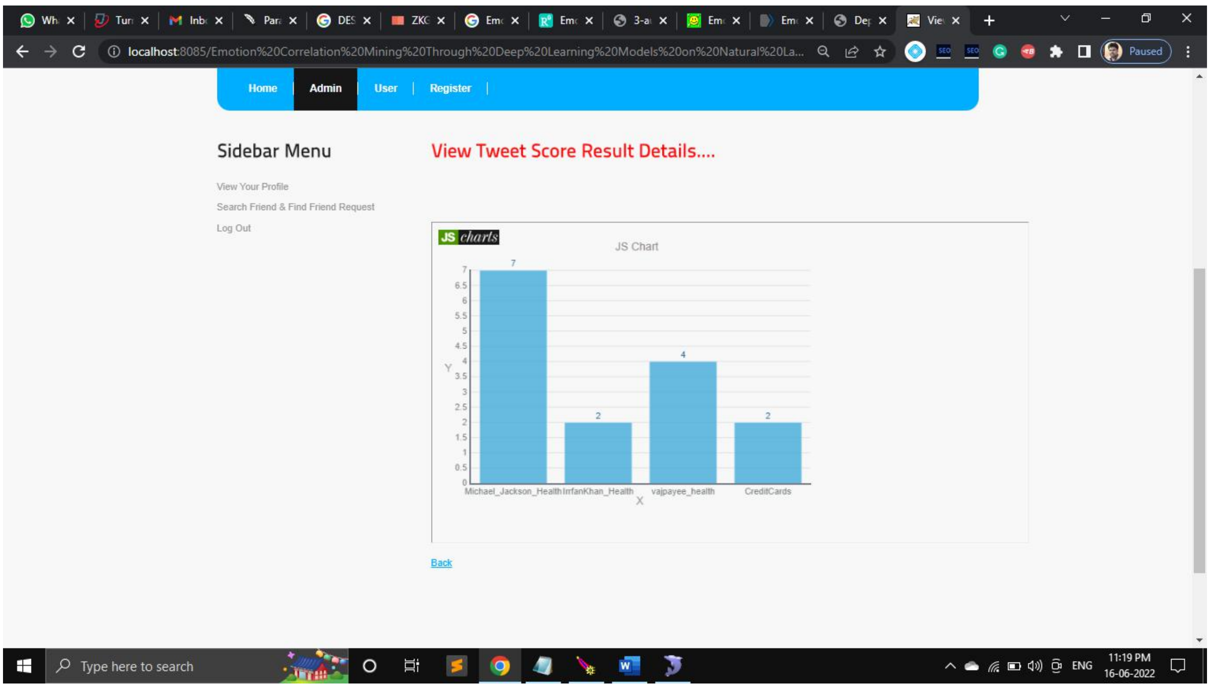

Figure 2: Emotion tweet chart results

The one-step shift, the limited-step shift, and the shortest path transfer are the three viewpoints that are utilised in the analysis of the emotion evolution law. The validity of the strategy is demonstrated through the utilisation of three different datasets, which encompass

ISSN: 2057-5688

both objective and subjective texts of diverse lengths (long and short). The findings of the experiment indicate that when it comes to subjective comments, emotions are easily confused with rage. The emotional cycle of love–anger and sadness–anger is one that is frequently triggered by comments. In this instance, the CNN and Collaborative algorithm are employed. The tweets that are pre-processed serve as the supplied input data. After that, a classification is performed, and the final results are integrated. For example, the given number of titles, emotions, hash tags, and instructions will be gathered and presented in a distinct format as the output.

V. Conclusion

The project Emotion Correlation Mining through Deep Learning Models on Natural Language Text mines the correlation of emotions based on the emotion recognition result of state-of-the- art deep learning models. The errors caused by the dataset and models are cut down by designing three kinds of features and two deep neural-network models. The emotion correlation is mined through an emotion confusion law, which is undirected, and an emotion evolution law, which is directed.

VI. References

[1] F. Strack and R. Deutsch, “The duality of everyday life: Dual-process and dual system models in social psychology,” in APA Handbook of Personality and Social Psychology, vol. 1. Washington, DC, USA: Amer. Psychol. Assoc., 2015, pp. 891–927.

[2] P. Shaver, J. Schwartz, D. Kirson, and C. O’Connor, “Emotion knowledge: Further exploration of a prototype approach,” Emotions Soc. Psychol. Essential Read., vol. 52, no. 6, pp. 26–56, 2001.

[3] P. Ekman, “An argument for basic emotions,” Cogn. Emotion, vol. 6, nos. 3–4, pp. 169–200, 1992.

[4] J. Zhao, L. Dong, J. Wu, and K. Xu, “MoodLens: An emoticon-based sentiment analysis system for Chinese tweets,” in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min., 2012, pp. 1528–1531.

[5] H.-M. Fu and P.-J. Cheng, “Learning unsupervised semantic document representation for fine-grained aspect-based sentiment analysis,” in Proc. ACM 42nd Int. SIGIR Conf. Res. Develop. Inf. Retrieval, 2019, pp. 1105–1108.

[6] Z. Zhao and X. Ma, “Text emotion distribution learning from small sample: A meta-learning approach,” in Proc. Conf. Empirical Methods Nat. Lang. Process. 9th Int. Joint Conf. Nat. Lang. Process. (EMNLPIJCNLP), 2019, pp. 3948–3958.

[7] E. Cambria, “Affective computing and sentiment analysis,” IEEE Intell. Syst., vol. 31, no. 2, pp. 102–107, Mar./Apr. 2016.

[8] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, “New avenues in opinion mining and sentiment analysis,” IEEE Intell. Syst., vol. 28, no. 2, pp. 15–21, Mar./Apr. 2013.

[9] X. Lin, M. Zhang, Y. Liu, and S. Ma, “A neural network model for social-aware recommendation,” in Proc. Asia Inf. Retrieval Symp., 2017, pp. 125–137.

[10] X. Wang, Z. Hui, S. Yuan, J. Wang, and Z. Yang, “Sentiment processing of social media information from both wireless and wired network,” EURASIP J. Wireless Commun. Netw., vol. 2016, no. 1, pp. 1–13, 2016. [Online]. Available: https://doi.org/10.1186/s13638-016-0661-x

[11] L. Peled and R. Reichart, “Sarcasm SIGN: Interpreting sarcasm with sentiment based monolingual machine translation,” 2017. [Online]. Available: arXiv:1704.06836.

[12] X. Gui et al., “Understanding the patterns of health information dissemination on social media during the ZIKA outbreak,” in Proc. AMIA Symp., 2017, pp. 820–829.

[13] S. Liang, Z. Ren, E. Yilmaz, and E. Kanoulas, “Collaborative user clustering for short text streams,” in Proc. 31st AAAI Conf. Artif. Intell., 2017, pp. 3504–3510.

[14] Y. Zhao, S. Liang, Z. Ren, J. Ma, E. Yilmaz, and M. de Rijke, “Explainable user clustering in short text streams,” in Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2016, pp. 155–164.

[15] J. Mao et al., “Understanding and predicting usefulness judgment in Web search,” in Proc. Int. ACM SIGIR Conf., 2017, pp. 1169–1172.

[16] N. Majumder, S. Poria, H. Peng, N. Chhaya, E. Cambria, and A. Gelbukh, "Sentiment and sarcasm classification with multitask learning," IEEE Intell. Syst., vol. 34, no. 3, pp. 38–43, Jan. 2019.

[17] T. Mikolov, I. Sutskever, C. Kai, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in Proc. Adv. Neural Inf. Process. Syst., vol. 26, 2013, pp. 3111–3119.

[18] S. Kuang and B. Davison, "Learning word embeddings with CHI-square weights for healthcare tweet classification," Appl. Sci., vol. 7, no. 8, p. 846, 2017.

[19] X. Wang, H. Zhang, and Y. Liu, "Sentence vector model based on implicit word vector expression," IEEE Access, vol. 6, pp. 17455–17463, 2018.

[20] C. Wang, B. Wang, W. Xiang, and M. Xu, "Encoding syntactic dependency and topical information for social emotion classification," in Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2019, pp. 881–884.

[21] P. V. A. Barros, G. Parisi, and S. Wermter, "A personalized affective memory model for improving emotion recognition," in Proc. Int. Conf. Mach. Learn., 2019, pp. 485–494.

[22] A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 5998–6008.

[23] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. Conf. North Amer. Assoc. Comput. Linguist. Human Lang. Technol., vol. 1, 2019, pp. 4171–4186.

[24] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNET: Generalized autoregressive pretraining for language understanding," 2019. [Online]. Available: arXiv:1906.08237.

[25] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," in Proc. AAAI Conf. Artif. Intell., vol. 33, 2019, pp. 7370–7377.

[26] Q. Qian, M. Huang, J. Lei, and X. Zhu, "Linguistically regularized LSTMs for sentiment classification," 2016. [Online]. Available: arXiv:1611.03949.

ISSN: 2057-5688

[27] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, “A convolutional neural network for modelling sentences,” 2014. [Online]. Available: arXiv:1404.2188.

[28] R. Socher et al., “Recursive deep models for semantic compositionality over a sentiment treebank,” in Proc. Conf. Empirical Methods Nat. Lang. Process., 2013, pp. 1631–1642.

[29] P. Liu, X. Qiu, and X. Huang, “Adversarial multi-task learning for text classification,” 2017. [Online]. Available: arXiv:1704.05742. [30] P. Chen, Z. Sun, L. Bing, and W. Yang, “Recurrent attention network on memory for aspect sentiment analysis,” in Proc. Conf. Empirical Meth. Natural Lang. Process., 2017, pp. 452–461.

[31] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, “EmotionMeter: A multimodal framework for recognizing human emotions,” IEEE Trans. Cybern., vol. 49, no. 3, pp. 1110–1122, Mar. 2019.

[32] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, “Spatial–temporal recurrent neural network for emotion recognition,” IEEE Trans. Cybern., vol. 49, no. 3, pp. 839–847, Mar. 2019.