Journal of Electrical and Computer Engineering Innovations (JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Predicting the Sentiment of Tweet Replies Using Attentive Graph Convolutional Neural Networks

S. Nemati [*]

Department of Computer Engineering, Shahrekord University, Shahrekord, Iran.

Article Info

Article History:
Received 18 April 2023
Reviewed 20 May 2023
Revised 09 July 2023
Accepted 15 August 2023

Keywords:
Sentiment analysis
Deep leaning
Social media
Twitter
Graph convolutional neural networks

[*]Corresponding Author's Email Address: *s.nemati@sku.ac.ir*

Abstract

Background and Objectives: Twitter is a microblogging platform for expressing assessments, opinions, and sentiments on different topics and events. While there have been several studies around sentiment analysis of tweets and their popularity in the form of the number of retweets, predicting the sentiment of first-order replies remained a neglected challenge. Predicting the sentiment of tweet replies is helpful for both users and enterprises. In this study, we define a novel problem; given just a tweet's text, the goal is to predict the overall sentiment polarity of its upcoming replies.
Methods: To address this problem, we proposed a graph convolutional neural network model that exploits the text's dependencies. The proposed model contains two parallel branches. The first branch extracts the contextual representation of the input tweets. The second branch extracts the structural and semantic information from tweets. Specifically, a Bi-LSTM network and a self-attention layer are used in the first layer for extracting syntactical relations, and an affective knowledge-enhanced dependency tree is used in the second branch for extracting semantic relations. Moreover, a graph convolutional network is used on the top of these branches to learn the joint feature representation. Finally, a retrieval-based attention mechanism is used on the output of the graph convolutional network for learning essential features from the final affective picture of tweets.
Results: In the experiments, we only used the original tweets of the RETWEET dataset for training the models and ignored the replies of the tweets in the training process. The results on three versions of the RETWEET dataset showed that the proposed model outperforms the LSTM-based models and similar state-of-the-art graph convolutional network models.
Conclusion: The proposed model showed promising results in confirming that by using only the content of a tweet, we can predict the overall sentiment of its replies. Moreover, the results showed that the proposed model achieves similar or comparable results with simpler deep models when trained on a public tweet dataset such as ACL 2014 dataset while outperforming both simple deep models and state-of-the-art graph convolutional deep models when trained on the RETWEET dataset. This shows the proposed model's effectiveness in extracting structural and semantic relations in the tweets.

Introduction

Twitter is a social networking and microblogging platform with nearly 400 million users worldwide. Its interesting characteristics have made Twitter a prominent communication tool for not only ordinary people but also particular users, including students [1], [2], politicians [3], [4] medical specialists [5], [6], athletes [7] and traders [8],

[9]. Users post their tweets to share daily updates, talk about their opinion and emotions, and keep in contact with friends and family [10], [11]. Moreover, people use Twitter to connect with others to discuss common interests and concerns with other users worldwide. This may be achieved by tweeting, retweeting, mentioning, or replying to other users' tweets [11]. Therefore, in recent years, mining Twitter for information about people and their opinion, sentiment, preferences, and reactions to events has attracted the increasing attention of researchers, companies, and media organizations [10].

Many research studies have addressed Twitter sentiment analysis since 2011 [10]. The main goal of Twitter sentiment analysis is to detect the sentiment polarity of a given tweet in terms of positive, negative, or neutral [11], [12]. In addition to standard tweet polarity detection, more fine-grained tasks, including emotion detection [13], personality detection [14], event detection [15], stock prediction [16], and election prediction [17], have also been investigated in recent years. Sentiment analysis of Twitter data has more challenges than sentiment analysis of reviews or similar texts [12], [18]. This has several reasons, including the limited length of tweets, the use of informal language and unique abbreviations, and complex relations formed using the mention, retweet, and reply mechanisms of Twitter. Therefore, several studies addressed sentiment analysis of Twitter data for English and other languages in the last decade [19]-[22].

With the growing number of Twitter users and the increase of tweets' impact, users' desire to capture other users' attention via their tweets has increased [23]. High-quality tweets (i.e., those that capture others' attention) can increase users' reputations [24]. Therefore, predicting other users' reactions to a tweet is essential for users, especially before they post their tweets [24], [25]. The number of times others like a tweet or, similarly, the number of retweets may be signs of a good impression and hence may be used as popularity metrics for a tweet [24]. In addition to these metrics, tweet replies may also be analyzed to detect the sentiment of repliers expressed in their replies as a measure of popularity for the source tweet. To predict other users' reactions to a tweet before posting, it is necessary to analyze the textual content of the tweet. This is a challenging problem among natural language processing tasks because tweets have limited length, forcing users to abbreviate words, invent acronyms on the fly, or even omit words [10], [26].

Every tweet may produce positive, negative, or neutral sentiments and reactions in its readers [10]. Such responses may be shown in terms of likes or dislikes, retweets, or posting textual replies. For likes and retweets, the number of users who like a tweet or retweet it may be considered a factor for measuring the positive reaction of other users [24]. However, for replies, the number of replies does not necessarily show the popularity and the positive responses of others. In this case, the tweet replies' content must be considered to determine how positive/negative the reactions are [27]. Several studies have addressed the problem of predicting the number of likes and retweets in recent years [23], [28]-[30]. These studies usually model the issue as a regression problem in which the model's output is the predicted number of likes or retweets over time [29]. However, indicating other users' sentiments shown in their replies has been neglected in previous studies [27].

Recently, Arasteh et al. [27] addressed the problem of predicting the overall sentiment of tweet replies and proposed a deep learning-based method for this problem. Specifically, they created a relatively large dataset of tweets and their first-order replies, RETWEET, and trained a bi-directional long short-term memory (Bi-LSTM) deep model on manually labeled tweets from the SemEval datasets [31]. Then, using this trained model, they predict the sentiment polarity of all tweet replies without considering the source tweets. Finally, using a heuristic averaging algorithm, they assigned a label to each source tweet according to its replies' polarity labels [27]. Although this study presented the problem of predicting tweet replies' sentiment for the first time, the main shortcoming is the need for having all replies for labeling a tweet. Ignoring the content of the source tweet and assigning a sentiment polarity label using its replies necessitate waiting for others' reactions in terms of their reply to predict the overall sentiment of replies. This seems to be the main weakness of their proposed solution to the problem [27].

In this study, we define a new problem as follows. Given only the textual content of a source tweet, the task is to predict the overall sentiment polarity of upcoming replies. To address this problem, we propose a new deep learning-based model for processing tweets' textual content and predicting their replies' overall sentiment. To this end, we used the RETWEET dataset [32], which contains several tweets and their corresponding first-order replies. In our proposed model, unlike [32], we do not use replies' textual content and only exploit the source tweets' content. Specifically, we trained a graph convolutional network (GCN) on the textual content of source tweets to learn the structural and semantic relations in the text. Then, we evaluate the trained network on unseen tweets in the dataset. In summary, the main contributions of the current study are as follows:

- Defining the problem of predicting the overall sentiment polarity of tweet replies only based on the textual content of the source tweet.
- Proposing a graph convolutional network model for exploiting structural and semantic relations in the tweets.

• Comparing the baseline and state-of-the-art graph convolutional network models with the proposed model on three versions of the RETWEET dataset.

The remainder of the paper continues as follows. In the next section, a brief overview of related studies will be presented. The proposed model will be described in section III. Experimental results are shown in section IV. Conclusions and directions for future work will be discussed in the last section.

Literature Review

In this section, a brief overview of related studies is presented in two subsections as follows. Some Twitter data analysis studies are shown in the first subsection, and deep learning-based models for sentiment analysis are presented in the following subsection.

A. Twitter Data Analysis

Kouloumpis et al. [10] investigated using linguistic features for message-level tweet sentiment analysis. They used a machine learning method and utilized lexical resources and hashtag information in training. They showed that part-of-speech (POS) features were not helpful, while sentiment linguistic features and emoticons are helpful for classification [10]. Agarwal et al. [33] proposed a machine learning approach for sentiment analysis of Twitter data and modeled the problem as binary and 3-way classification problems. They evaluated unigram, feature-based, and tree-based models and showed that the combination of these models outperformed the baseline and each model in isolation [33]. Mohammad et al. [34] designed two sentiment lexicons and proposed a machine learning-based classifier for message-level and term-level sentiment classification of Twitter data. They showed that their lexicon-based approach outperformed the machine learning-based method.

Some recent studies investigated problems that use sentiment analysis to address other issues. For example, Abdar et al. [26] proposed a model for detecting people's attitudes toward energy in Alaska. They used Twitter as a data source in which people express their sentiments and emotion towards different subjects, including Energy. Gagne et al. [1] analyzed nursing student tweets in three countries during COVID-19. They investigated the opinion of students in their tweets to help nurse educators better understand the students. Basiri et al. [12] proposed a deep learning-based model for sentiment analysis of tweets in eight countries during the COVID-19 pandemic. They offered a fusion model and showed that the sentiment intensity expressed by people at different times and governments was not identical. Ali et al. [35] proposed a deep learning model for sentiment analysis of tweets in Pakistan. They aimed to predict the results of the Pakistan general election in 2018 using Twitter data.

Hong et al. [36] investigated the problem of predicting the popularity of tweets and used the number of retweets as the measure of popularity. They employed tweet contents and metadata of tweets, including temporal data and user data. In a similar study, Petrovic et al. [29] investigated the problem of predicting the number of retweets and proposed a machine, learning-based model. They showed that although social features performed very well, tweet features could also be used in the model to reach human-level accuracy. Daga et al. [24] evaluated some machine learning methods learned on a bag-of-words model and word embedding features to predict the number of likes and retweets for a source tweet. They showed that bag-of-words features were more helpful than embedding features for this task [24].

Lou et al. [37] introduced the problem of predicting the users who retweet a source tweet. They proposed a machine learning-based method and used features such as retweet history and followers-related features. They showed that common interests and the history of retweeting were factors that could be used for predicting future retweets [37]. Wang et al. [38] proposed a deep learning-based model to analyze users' retweeting behavior. They integrated user-based and message-based features to model the group retweeting behavior and tweets' content. In a similar study, Firdaus et al. [39] explored the problem of the retweeting behavior of users and focused on the topic's impact. Specifically, they investigated the effect of a user's topic-related sentiment on their retweet decision. They concluded that the topic and users' sentiment toward the topic were important for modeling their retweet behavior [39].

Some recent studies addressed the problem of tweet popularity prediction using novel approaches. For example, Lymperopoulos [40] proposed a model based on electronic circuits for predicting the popularity of tweets in terms of their number of retweets. As another example, Garvey et al. [41] proposed an artificial intelligence probabilistic model for generating popular tweets. Specifically, they used econometrics, machine learning, and Bayesian theory to create the structure of high-impact tweets. Gao et al. [28] proposed a heterogeneous bass model for the prediction of the popularity of tweets. They considered tweets with a similar topic, using a clustering approach and linear regression to improve the system's performance. Rivadeneira et al. [23] proposed an evidential reasoning model for predicting tweets' impact. Specifically, they used five features of tweets to indicate the number of electoral-related retweets.

B. Deep Learning

Several studies applied deep learning techniques to sentiment analysis problems in different domains in recent years. For example, as one of the first applications of the deep model in the sentiment analysis domain, Poria et al. [42] proposed a feature extraction method based on convolutional networks. They used the extracted features for multimodal sentiment analysis of short video clips. They reported a 14% improvement over existing methods for the same task. Edara et al. [43] applied LSTM to the problem of sentiment analysis of cancer-affected patients' tweets. They showed that their deep model outperforms traditional machine learning models. Basiri et al. [44] proposed a 3-way fusion model of deep and conventional learning techniques for sentiment analysis of drug reviews. They showed that their model outperformed traditional and deep models and considered classifier confidence in its decisions. Muhammad Shah et al. [45] proposed a deep model for multimodal patient review sentiment analysis. They processed both textual and image content of patients' reviews published on the Yelp.com platform.

Parimala et al. [46] proposed an LSTM deep model for sentiment analysis of tweets collected before, after, and during disasters. They compared their model with traditional learning models and reported a slightly better performance for the binary classification scenario. Basiri et al. [12] proposed a deep fusion model consisting of four deep and one traditional learning method for analyzing COVID-19 tweets in different countries. Serrano-Guerrero et al. [47] addressed the problem of sentiment analysis and emotion recognition of patients' reviews. They proposed a hybrid of bidirectional gated recurrent unit (Bi-GRU) and convolutional network to classify reviews. They also evaluated different word embeddings for their models and showed that their clinical-domain word embedding model outperformed other deep and traditional learning models. Basiri et al. [48] proposed a Bi-LSTM model for sentiment analysis of online doctor reviews. They introduced the PODOR dataset containing Persian online doctor reviews and showed that their proposed deep model outperforms traditional learning models for the polarity detection of online doctor reviews.

Some recent studies applied deep learning models to the problem of sentiment analysis in other languages. For example, Shehu et al. [49] evaluated different data augmentation and deep learning models on Turkish tweets. They compared their models with traditional machine learning models and concluded that conventional models outperformed deep models in speed, but deep models performed better. Several studies applied deep learning models to Arabic sentiment analysis [50]. For example, Elfaik et al. [51] used Bi-LSTM, Saleh et al. [52] used a hybrid of CNN and LSTM models, and Al-Dabet et al. proposed a CNN-based model for aspect-based sentiment analysis of Arabic texts. Dashtipour et al. [53] used LSTM and CNN for Persian sentiment analysis of movie reviews. Bokaee Nezhad et al. [54] applied a combination of CNN and LSTM models to COVID-19 tweets in the Persian language. Gonzalez et al. used pre-trained Bert models for Spanish tweet sentiment analysis. Gan et al. used an attention mechanism on a CNN-BiLSTM model for Chinese sentiment analysis.

Smetanin et al. [55] applied a transformer-based deep model to Russian sentiment analysis. In recent years, other languages have also been the target of deep learning methods for sentiment analysis problems. In summary, compared to the previous studies, the novelty of the current research is two-fold. First, we introduce a new problem in the domain of sentiment analysis of Twitter data. Second, we propose a new attentive graph convolutional deep model for solving the problem. The proposed model will be described in the next section in more detail.

Proposed Model

We exploit graph-based convolutional neural networks in the proposed model to consider tweets' structural and semantic information. The overall structure of the proposed model is shown in Fig. 1.

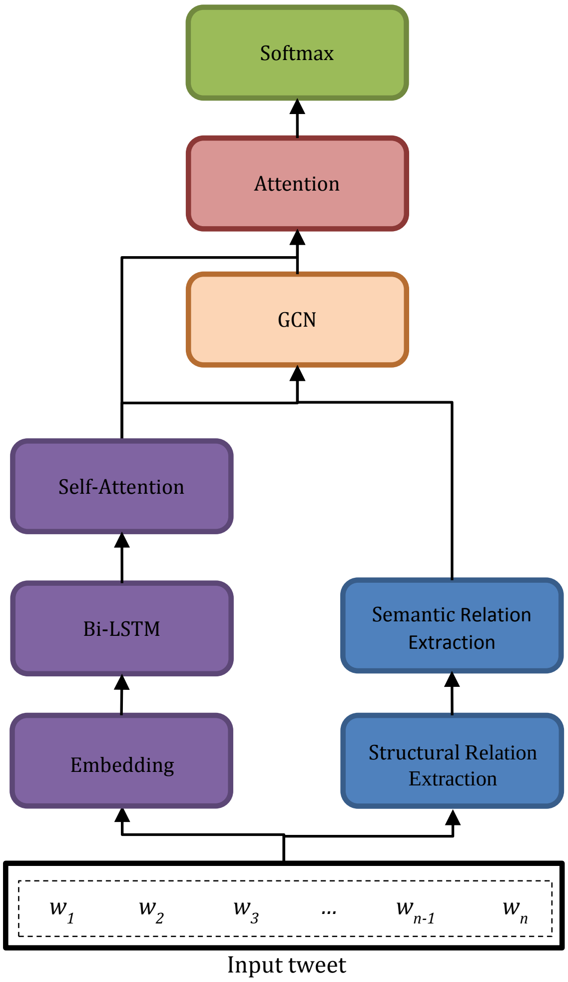


Fig. 1: The overall structure of the proposed model.

There are two parallel branches in the proposed model as follows. The first branch, which starts with the embedding module, extracts the contextual representation of the input tweets. The second branch extracts the structural and semantic information from tweets. Specifically, the input tweets $t = \{w_1, w_2, \dots, w_n\}$, containing *n* words, is sent to both embedding and structural relation extraction modules for converting to a numerical vector and extracting the structural graph, respectively.

A. Contextual Representation Branch

In the embedding module, each tweet is converted to a numerical matrix using a lookup table which is usually derived from transformer-based pre-trained word embeddings such as BERT [56], Elmo [57], or Glove [58]. The numerical representation of each tweet, *t*, contains *n* vectors of length *m*, where *m* is the dimension of the word vectors in the lookup table. In the current study, we used 300-dimensional vectors of Glove trained on 42 billion words from Wikipedia pages and newswires as the lookup table [58].

The Bi-LSTM module takes the embedding matrix of each tweet as input and derives its hidden contextual representations as follows.

$$H^c = \{h_1^c, h_2^c, \dots, h_n^c\} = Bi - LSTM(x) \tag{1}$$

where,

$$x = [x_1, x_2, \dots, x_n] \text{ and } x_i \in \mathbb{R}^m \tag{2}$$

The self-attention module is used on top of the Bi-LSTM module to learn syntactical dependencies [59], [60]. Using this module, each word in the tweet pays attention to other words regardless of their position. To achieve this, three parameters, namely *Q* (queries), *K* (keys), and *V* (values), are combined as follows [59]:

$$Att(Q, K, V) = softmax(QK^T)V \tag{3}$$

To obtain the values of the above three parameters, three randomly initialized weight matrices, W_Q, W_K, W_V and the input of the self-attention module, which is here the output of the Bi-LSTM module, are used as follows:

$$\begin{aligned} SelfAtt(H^c) &= Att\left(H^c \mathrm{W}_Q, H^c \mathrm{W}_K, H^c \mathrm{W}_V\right)V \\ &= softmax(H^c \mathrm{W}_{QK} H^{c^T}) H^c W_V \end{aligned} \tag{4}$$

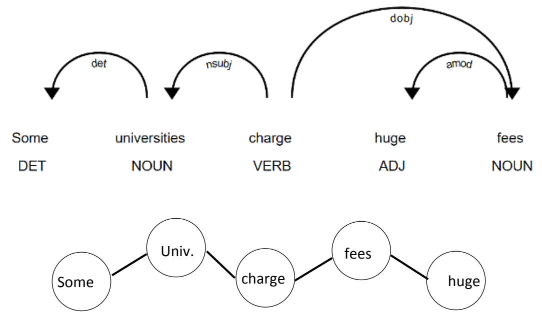
where $\mathrm{W}_{QK} = \mathrm{W}_Q W_K^T$.

B. Relation Extraction Branch

The structural relation extraction module is used in the proposed method to construct the dependency graph of tweets. To this aim, we first build the dependency tree of an input tweet using the SpaCy module [61]. Then, we make the adjacency matrix $D \in \mathbb{R}^{n \times n}$ of the tweet using the dependency tree by setting $D_{i,j}$ to one if there is a dependency between the i^{th} and j^{th} words and assigning it to zero otherwise. This strategy is proposed in [62] to create an undirected dependency graph. An illustrative example of converting a sample tweet "*Some universities charge huge fees*" to its corresponding adjacency matrix, is shown in Fig. 2. As shown in the figure, the undirected dependency graph is created based on the dependency tree relations.

The semantic relation extraction module adds external knowledge to the dependency graph. This knowledge may be in the form of sentiment scores stored in a lexicon or an affective resource such as SenticNet [63]. In the current study, we used sentiment scores from SenticNet as follows. For each pair of dependent words w_i and w_j in the adjacency matrix, we compute $S_{i.j}$ as:

$$S_{i,j} = Sentic(w_i) + Sentic(w_j) \tag{5}$$



$$D = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Fig. 2: A sample tweet and its corresponding dependency diagram, undirected dependency tree, and dependency graph.

where *Sentic*(w_i) is a real number in the range [-1,1] showing the sentiment intensity of w_i according to the following rule:

$$Sentic(w_i) \in \begin{cases} [-1,0), & w_i is\ a\ negative\ word \\ (0,+1], & w_i is\ a\ positive\ word \end{cases} \tag{6}$$

and if *Sentic*(w_i) = 0, w_i is a neutral word, or it is absent in SenticNet. Having computed $S_{i,j}$ for all dependent words, we enhance the adjacency matrix by:

$$A_{i,j} = D_{i,j} \times (S_{i,j} + 1) \tag{7}$$

C. GCN and Attention

The next module, GCN, takes the enhanced adjacency matrix and H^c as inputs and computes $\widetilde{H}$, which is the learned representation of the tweet as follows:

$$\widetilde{H}_i = relu\left(\tilde{A}_i g_i W + b\right) \tag{8}$$

$$g_i = \mathcal{F}(h_i) \tag{9}$$

$$\tilde{A}_i = \frac{A_i}{1 + \sum_{j=1}^{n} A_{i.j}} \tag{10}$$

where g is the hidden representation from the previous layer of GCN and $\mathcal{F}(.)$ is a transformation function, as suggested in [62].

The attention module takes H^c and $\widetilde{H}$ as inputs and computes the final representation of the tweets as follows:

$$r = \sum_{i=1}^{n} \alpha_i h_i^c \tag{11}$$

$$y = softmax(W_o r + b_o) \tag{12}$$

where α_i is the attention weight calculated as follows [62]:

$$\alpha_i = \frac{e^{\beta_j}}{\sum_{j=1}^{n} e^{\beta_j}} \tag{13}$$

$$\beta_i = \sum_{j=1}^{n} h_i^{c^T} \tilde{h}_i \tag{14}$$

Here, the attention mechanism is a retrieval-based method proposed by [62] and adopted in [64] for learning the affective and semantic information from a sentence.

Experimental Results

A. Datasets and Settings

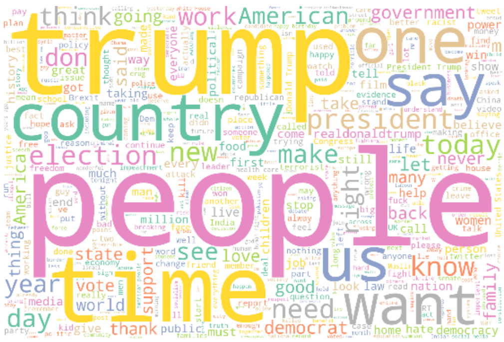
We used the RETWEET dataset [32] for experiments in the current study. The tweets in this dataset were downloaded using a pre-defined list of keywords. Word clouds of the train and test parts of RETWEET are shown in Fig. 3.

Because the public version of the RETWEET dataset only contained tweet IDs, we downloaded the tweets using the provided IDs. However, from 35020 training tweets in RETWEET, only 17613 tweets were and from 1519 test tweets, only 1037 tweets were downloaded. As discussed in the introduction section, unlike [32], we do not use the replies' textual content and only exploit the source tweets' content.

Therefore, we only need the test part of the RETWEET dataset. This dataset contains 1037 tweets, and we named it "Original". Because the Original dataset is unbalanced, we created a "Balanced" version by selecting positive, neutral, and negative tweets according to the number of tweets in the minority class (i.e., neutral class) in the Original dataset. Moreover, we created a "Resampled" version of the Original dataset by resampling the classes according to the distribution of the classes in the train part of the RETWEET dataset introduced in [32].

The detailed specifications of the datasets are shown in Table 1 and the histograms of the distribution of tweets based on their word count in the datasets are shown in Fig. 4.

In the experiments, we used the 300-dimensional vectors of Glove trained on 42 billion words from Wikipedia pages and newswires as the lookup table [58] for the proposed model. Also, in the GCN module, we used two layers, and the dimensionality of all hidden states was set to 300. The learning rate was 0.00002, the batch size was four, and the Adam optimizer with a learning rate of 0.001 was used for optimization.



(a)

(b)

Fig. 3: Word clouds of (a) the train and (b) the test parts of the RETWEET dataset.

Table 1: Specification of datasets used in the current study

	Original		Balanced		Resampled	
	train	test	train	test	train	test
Negative	318	106	226	75	167	56
Neutral	226	75	226	75	226	75
Positive	234	78	226	75	120	40
Total	778	259	678	225	513	171

B. Comparison Models

To evaluate the proposed model, the following methods were used for comparison:

- **2-BiLSTM** [32] uses two BiLSTM layers on the top of an embedding layer equipped with dropout layers.
- **2-LSTM** [65] uses two serial LSTM layers on the top of an embedding layer.

- **SenticGCN** [64] uses a GCN with depth two on the top of an LSTM layer. This model uses structural and semantic information from tweets.
- **AffectiveGCN** [64] is similar to SenticGCN but only employs semantic information to construct the dependency graph.
- **DSenticGCN** [64] similar to SenticGCN but uses directed structural graphs and GCN with depth four.

C. Evaluation Criteria

To assess the performance of models, accuracy and F1 evaluation criteria are used in the experiments.

$$F1 = \frac{2\times\pi\times\rho}{(\pi+\rho)} \tag{15}$$

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{16}$$

$$\pi = \frac{TP}{TP+FP} \tag{17}$$

$$\rho = \frac{TP}{TP+FN} \tag{18}$$

where π and ρ are precision and recall, respectively. TP, TN, FP, and FN are true positive, true negative, false positive, and false negative, respectively.

Results

A. Preliminary Results

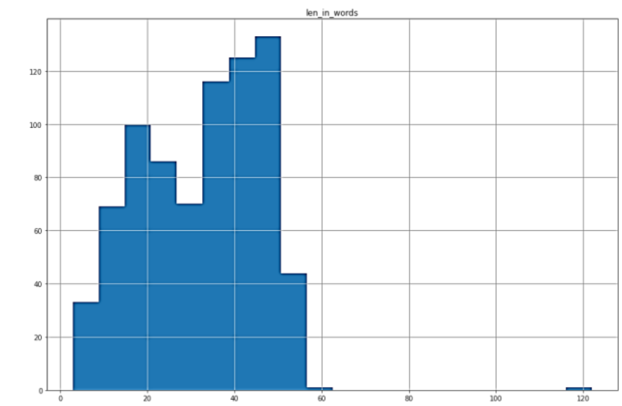
This section reports the results obtained by training the models on the ACL 2014 Twitter dataset.

As we pointed out earlier, in [32], the RETWEET model was trained on the replies posted to the original tweets (i.e., the test set of the RETWEET dataset) and evaluated on the original tweets.

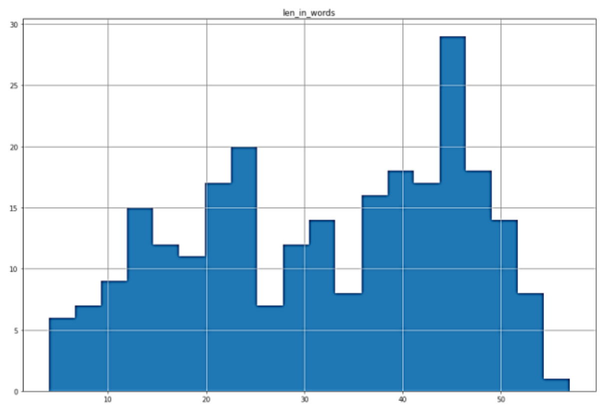
There are better methods for assessing the models than this because the final goal is to predict the overall sentiment of the first-order replies. Using these replies in the training process is unfair. Therefore, we selected the ACL 2014 Twitter dataset for the first round of experiments. This dataset contains 6248 tweets labeled as positive, neutral, or negative. The main reason for choosing this dataset for training the models is its conceptual and syntactic similarity with the RETWEET dataset.

Fig. 5 shows the results obtained using the ACL 2014 dataset for the train and test sets of three versions of the RETWEET dataset for the test.

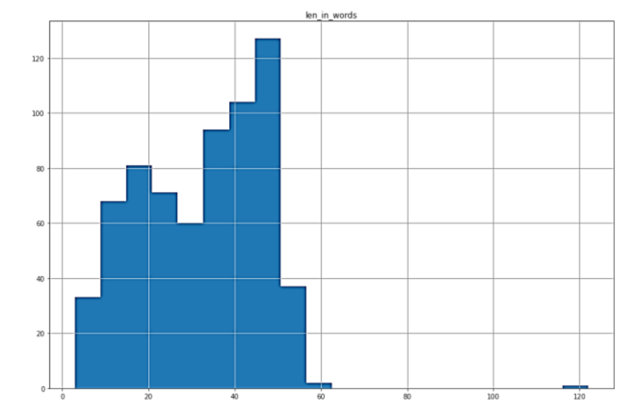
As shown in Fig. 5, when the models were trained on the ACL 2014 dataset, the proposed and other GCN-based models did not show a significant advantage. For example, on the balanced RETWEET dataset, the 2-LSTM and 2-BiLSTM models outperform other models. Also, on all test sets of the RETWEET dataset, the performance of the proposed model could have been better compared to other models when trained on the ACL 2014 dataset.



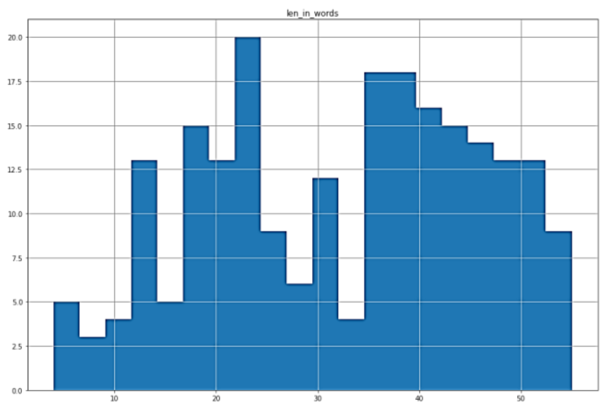
(a)



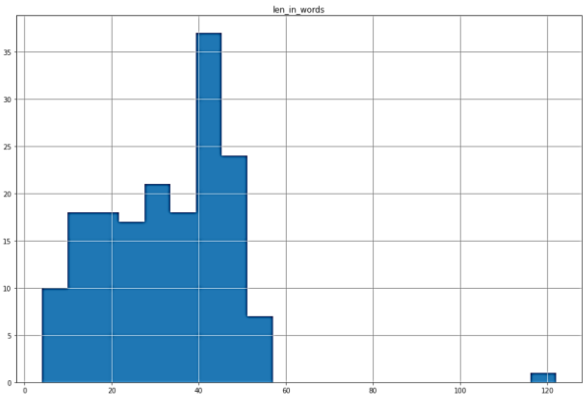
(b)



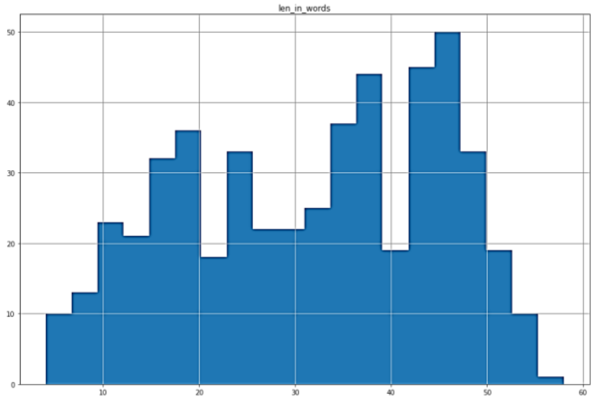
(c)



(d)



(e)



(f)

Fig. 4: Histograms of the distribution of tweets based on their word count in (a) original train, (b) original test, (c) balanced train, (d) balanced test, (e) resampled train, and (f) resampled test datasets.

This lower performance of the proposed model may be justified because the proposed model's structural and semantic relation extraction modules (See Fig. 1) cannot extract meaningful relations when trained on a different dataset (i.e., the ACL 2014). To show the utility of the proposed model, in the next section, we report the results obtained using the train and test parts of the REWTWEET dataset.

B. Main Results

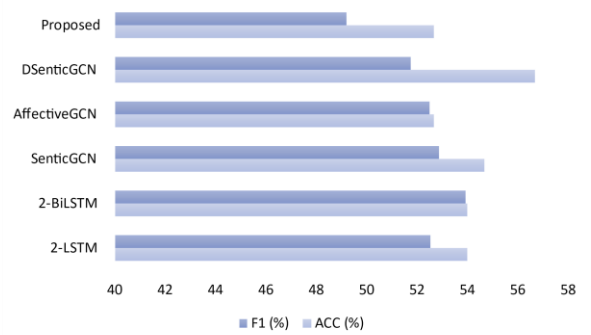
As shown in Fig. 6, the proposed model outperforms 2-LSTM and 2-BiLSTM models on all versions of the RETWEET dataset.

This shows the effectiveness of utilizing structural and semantic information in the proposed method. Moreover, the results show that the proposed method has similar results with other GCN-based methods on the Original RETWEET dataset while outperforming all other models on the Balanced and Resampled RETWEET datasets.

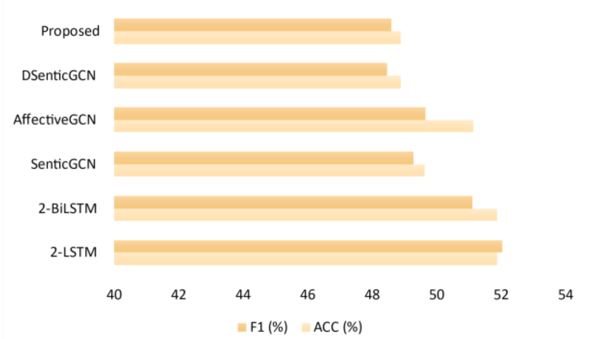
This shows the power of the proposed model for better retrieving syntactical information via the self-attention module and semantic information via external knowledge for constructing the dependency graph of the tweets. For the Balanced and Resampled datasets, the second-best method is AffectiveGCN which employs affective information in making the adjacency matrix of the tweets. This also verifies the effect of using external knowledge to enhance the system.

C. Discussion

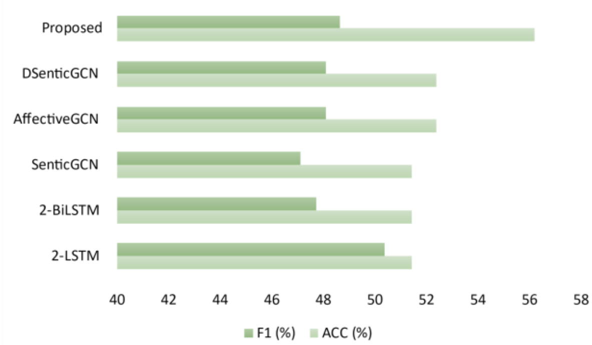
As we pointed out in the previous section, we created three versions of the RETWEET dataset and evaluated our proposed and other deep models on these three versions. To discuss the results obtained on these three versions, we should briefly describe the process of creating the original RETWEET dataset published in [32].



(a)

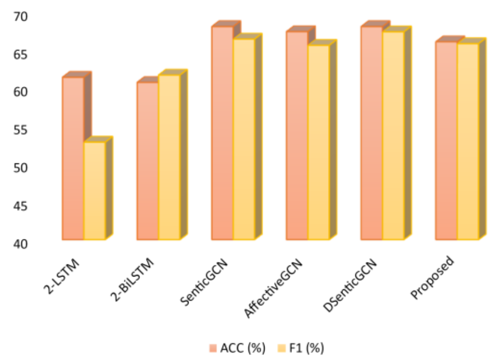


(b)

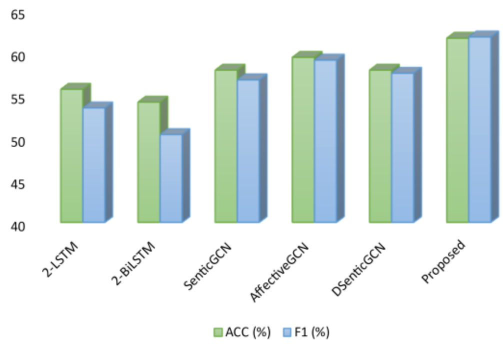


(c)

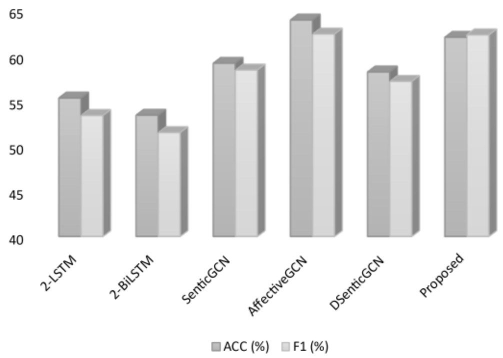
Fig. 5: Results of training the models on the ACL 2014 dataset and testing on the (a) Original, (b) Balanced, and (c) Resampled versions of the RETWEET dataset.



(a)



(b)



(c)

Fig. 6: Comparison of the results obtained by training and testing on the (a) Original, (b) Balanced, and (c) Resampled versions of the RETWEET dataset.

The first version of the RETWEET contains train and test parts created as follows [32] The training part contains first-order replies to the tweets of the test part. In the manual labeling process of the test set, three human annotators were asked to assign a three-class label to an unseen tweet based on its first-order replies in the training set. In their proposed deep model in [32], the training tweets were automatically labeled using a deep learning method (2-Bi-LSTM method in our comparisons). These labels were sent to a heuristic algorithm for assigning an overall sentiment label to the source tweet (i.e., the corresponding tweet in the test set) based on its first-order tweets. However, the original train/test separation by [32] is not helpful in the current study because we introduced a different problem in the present study. Specifically, in contrast to [32], we defined the problem as predicting the overall sentiment of replies to a tweet based on its content. Therefore, we could not use the content of the replies in the training process of the models. Hence, the original training set of [32] was useless in our study.

According to the points mentioned above, we used the original test set of RETWEET as the dataset for our experiments. We created three versions of this dataset, as described in section IV. According to this separation, the results reported in the previous section have several points which should be clarified. First, when training on another dataset (i.e., the ACL 2014 dataset), the performance of simple deep models such as 2-LSTM and 2-BiLSTM models are better (on the Balanced version) or at least comparable with more sophisticated deep models (on the Original and Resampled versions). This shows that the main reason for the effectiveness of the proposed model and similar knowledge-enhanced models is their ability to utilize the structural and semantic information in modeling the tweets in the form of an affective knowledge graph. When these models are trained on a different dataset (i.e., the ACL 2014 dataset), these models are unable to form suitable graphs and hence have weak performance.

On the other hand, when the models are trained on the RETWEET dataset, the results are more comparable and justifiable. Therefore, as the second point, it should be noted that the difference in ranking of the models in the three versions of the RETWEET dataset is due to the differences in the number of neutral, positive, and negative tweets in the datasets (i.e., see Table 1). Third, all the GCN-based models (including the proposed model) outperform the LSTM-based models, which shows the effectiveness of the graph-based convolutional models in utilizing both the tweets' textual content and their dependencies. Forth, the most confident results are for the balanced version of the dataset where the proposed model significantly outperforms all the other methods. This may be due to the simultaneous use of external knowledge and the self-attention mechanism in the proposed model.

Conclusion

Predicting the sentiment of tweet replies is an interesting problem for people and companies who want to capture other users' attention via influencing tweets. The previous studies used the replies to train deep models that can predict the tweets' overall sentiment. The content of the tweets was ignored in the process of training the deep predictive models. In the current study, we defined a new problem as follows. Given only the textual content of a source tweet, the task is to predict the overall sentiment polarity of upcoming replies. To address this problem, we proposed a new deep model including two branches for extracting syntactical (using Bi-LSTM layers) and semantic relations (using dependency trees enhanced with an external affective source of knowledge) from the text body of the tweets and a graph convolutional network for learning the joint feature representation. Moreover, we utilized two attention mechanisms in the proposed model; first, a self-attention mechanism on the top of the Bi-LSTM module of the first branch to extract the importance of different parts of the learned representation of tweets. Second, a retrieval-based attention mechanism on the output of the graph convolutional network for learning essential features from the final affective picture of tweets.

To show the performance of the proposed model, we used the recently published RETWEET dataset, which contains manually labeled tweets in a three-class form (i.e., negative, neutral, positive) based on their content. We divided the experiments into two parts; In the first part of the experiments, we trained the models on a general tweet dataset, ACL 2014. In the second part of the experiments, we trained the models on the RETWEET dataset. The point is that when the models are trained on the RETWEET dataset, the results are more comparable and justifiable. The general ACL 2014 datasets cover several topics and users, while the RETWEET dataset contains a limited number of users and replies to their tweets. Moreover, the context of the training set is an essential factor in tuning the proposed model and similar knowledge-enhanced models because the main reason for the effectiveness of such knowledge-enhanced models is their ability to utilize the structural and semantic information in modeling the tweets in the form of affective knowledge-graph.

The results showed that the proposed model achieves similar or comparable results with simpler deep models when trained on a general tweet dataset such as ACL 2014 dataset while outperforming both simple deep models and state-of-the-art graph convolutional deep models when trained on the RETWEET dataset. This shows the proposed model's effectiveness in extracting structural and semantic relations in the tweets. For the future study, we plan to create a large dataset of tweets to address the new problem we defined in the current study. Also, designing deep ensemble models for this task may be a promising line of research for future studies.

Author Contributions

S. Nemati designed the experiments, analyzed the data, interpreted the results, and wrote the manuscript.

Acknowledgment

This work was financially supported by the research deputy of Shahrekord University (grant number 0GRD34M44264).

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work.

Abbreviations

Bi-GRU	Bidirectional Gated Recurrent Unit
Bi-LSTM	Bi-directional Long Short-Term Memory
GCN	Graph Convolutional Network
POS	Part-Of-Speech

References

[1] J. C. de Gagne, E. Cho, H. K. Park, J. D. Nam, D. Jung, "A qualitative analysis of nursing students' tweets during the COVID-19 pandemic," Nurs. Health Sci., 23(1): 273–278, 2021.

[2] M. T. Rajeh, S. N. Sembawa, A. A. Nassar, S. A. al Hebshi, K. T. Aboalshamat, M. K. Badri, "Social media as a learning tool: Dental students' perspectives," J. Dent. Educ., 85(4): 513–520, 2021.

[3] S. Rill, D. Reinel, J. Scheidt, R. V. Zicari, "Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis," Knowl. Based Syst., 69: 24–33, 2014.

[4] B. Castanho Silva, S. O. Proksch, "Politicians unleashed? Political communication on Twitter and in parliament in Western Europe," Political Sci. Res. Methods, 10(4): 776–792, 2022.

[5] N. Corsi, D. Dan Nguyen, M. Butaney, S. E. Majdalany, M. P. Corsi, T. Malchow, A. J. Piontkowski, Q. D. Trinh, S. Loeb, F. Abdollah, "Top 100 Urology Influencers on Twitter: Is Social Media Influence Associated with Academic Impact?," Eur. Urol. Focus, 9(2): 396-402, 2022.

[6] B. Mishra et al., "Use of twitter in Neurology: boon or bane?," J. Med. Internet Res., 23(5): e25229, 2021.

[7] Y. Wang, "Building relationships with fans: how sports organizations used twitter as a communication tool," Sport Soc., 24(7): 1055–1069, 2021.

[8] L. Ante, "How Elon Musk's twitter activity moves cryptocurrency markets," Technol. Forecast Soc. Change, 186: 122112, 2023.

[9] S. Duz Tan, O. Tas, "Social media sentiment in international stock returns and trading activity," J. Behav. Finance, 22(2): 221–234, 2021.

[10] E. Kouloumpis, T. Wilson, J. Moore, "Twitter sentiment analysis: The good the bad and the omg!," in Proc. the international AAAI conference on web and social media, 5(1): 538–541, 2011.

[11] A. Giachanou, F. Crestani, "Like it or not: A survey of twitter sentiment analysis methods," ACM Comput. Surv. (CSUR), 49(2): 1–41, 2016.

[12] M. E. Basiri, S. Nemati, M. Abdar, S. Asadi, U. R. Acharrya, "A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets," Knowl. Based Syst., 228: 107242, 2021.

[13] K. Sailunaz, R. Alhajj, "Emotion and sentiment analysis from Twitter text," J. Comput. Sci., 36: 101003, 2019.

[14] J. T. Yun, U. Pamuksuz, B. R. L. Duff, "Are we who we follow? Computationally analyzing human personality and brand following on Twitter," Int. J. Advert, 38(5): 776–795, 2019.

[15] M. Hasan, M. A. Orgun, R. Schwitter, "A survey on real-time event detection from the Twitter data stream," J. Inf. Sci., 44(4): 443–463, 2018.

[16] N. Oliveira, P. Cortez, N. Areal, "The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices," Expert Syst. Appl., 73: 125–144, 2017.

[17] M. H. Tsai, Y. Wang, M. Kwak, N. Rigole, "A machine learning based strategy for election result prediction," in Proc. International Conference on Computational Science and Computational Intelligence (CSCI): 1408-1410, 2019.

[18] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, U. R. Acharya, “ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis,” Future Gener. Comput. Syst., 115: 279–294, 2021.

[19] M. Cieliebak, J. Deriu, D. Egger, F. Uzdilli, “A Twitter corpus and benchmark resources for German sentiment analysis,” in Proc. 4th International Workshop on Natural Language Processing for social media: 45–55, 2017.

[20] S. Özsoy, “Use of new media by Turkish fans in sport communication: Facebook and Twitter,” J. Hum. Kinet., 28: 165, 2011.

[21] S. O. Alhumoud, M. I. Altuwaijri, T. M. Albuhairi, W. M. Alohaideb, “Survey on arabic sentiment analysis in twitter,” Int. J. Comput. Inf. Eng., 9(1): 364–368, 2015.

[22] E. S. Tellez, S. Miranda-Jiménez, M. Graff, D. Moctezuma, O. S. Siordia, E. A. Villaseñor, “A case study of Spanish text transformations for twitter sentiment analysis,” Expert Syst. Appl., 81: 457–471, 2017.

[23] L. Rivadeneira, J. B. Yang, M. López-Ibáñez, “Predicting tweet impact using a novel evidential reasoning prediction method,” Expert Syst. Appl., 169, 2021.

[24] I. Daga, A. Gupta, R. Vardhan, P. Mukherjee, “Prediction of likes and retweets using text information retrieval,” Procedia Comput. Sci., 168: 123–128, 2020.

[25] S. Butt, S. Sharma, R. Sharma, G. Sidorov, A. Gelbukh, “What goes on inside rumour and non-rumour tweets and their reactions: A psycholinguistic analyses,” Comput. Human Behav., p. 107345, 2022.

[26] M. Abdar et al., “Energy choices in Alaska: Mining people’s perception and attitudes from geotagged tweets,” Renewable Sustainable Energy Rev., 124: 109781, 2020.

[27] S. T. Arasteh et al., “How will your Tweet be received? Predicting the sentiment polarity of Tweet replies,” in Proc. 2021 IEEE 15th International Conference on Semantic Computing, ICSC 2021: 370–373, 2021.

[28] X. Gao, Z. Zheng, Q. Chu, S. Tang, G. Chen, Q. Deng, “Popularity prediction for single Tweet based on heterogeneous bass model,” IEEE Trans. Knowl. Data Eng., 33(5): 2165–2178, 2021.

[29] S. Petrovic, M. Osborne, V. Lavrenko, “Rt to win! predicting message propagation in twitter,” in Proc. the international AAAI conference on web and social media, 5(1): 586–589, 2021.

[30] M. Mahdikhani, “Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of Covid-19 pandemic,” Int. J. Inf. Manage. Data Insights, 2(1): 100053, 2022.

[31] M. Pontiki et al., “SemEval-2016 task 5: Aspect based sentiment analysis,” in ProWorkshop on Semantic Evaluation (SemEval-2016): 19–30, 2016.

[32] S. T. Arasteh, M. Monajem, V. Christlein, P. Heinrich, A. Nicolaou, H. Naderi Boldaji, M. Lotfinia, S. Evert, “How will your Tweet be received? Predicting the sentiment polarity of Tweet replies,” in Proc. 2021 IEEE 15th International Conference on Semantic Computing, ICSC 2021: 370–373, 2021.

[33] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. J. Passonneau, “Sentiment analysis of twitter data,” in Proc. the workshop on language in social media (LSM 2011): 30–38, 2011.

[34] S. M. Mohammad, S. Kiritchenko, X. D. Zhu, “NRC-Canada: Building the State-of-the-Art in sentiment analysis of Tweets,” in Proc. the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013), 2013.

[35] H. Ali, H. Farman, H. Yar, Z. Khan, S. Habib, A. Ammar, “Deep learning-based election results prediction using Twitter activity,” Soft. Comput., 26(16): 7535–7543, 2022.

[36] L. Hong, O. Dan, B. D. Davison, “Predicting popular messages in Twitter,” in Proc. the 20th international conference companion on World wide web: 57–58, 2011.

[37] Z. Luo, M. Osborne, J. Tang, T. Wang, “Who will retweet me? Finding retweeters in Twitter,” in Proc. the 36th international ACM SIGIR conference on Research and development in information retrieval: 869–872, 2013.

[38] L. Wang, Y. Zhang, J. Yuan, K. Hu, S. Cao, “FEBDNN: fusion embedding-based deep neural network for user retweeting behavior prediction on social networks,” Neural Comput. Appl., 34(16): 13219–13235, 2022.

[39] S. N. Firdaus, C. Ding, A. Sadeghian, “Topic specific emotion detection for retweet prediction,” Int. J. Mach. Learn. Cybern., 10(8): 2071–2083, 2019.

[40] I. N. Lymperopoulos, “RC-Tweet: Modeling and predicting the popularity of tweets through the dynamics of a capacitor,” Expert Syst. Appl., 163: 113785, 2021.

[41] M. D. Garvey, J. Samuel, A. Pelaez, “Would you please like my tweet?! An artificially intelligent, generative probabilistic, and econometric based system design for popularity-driven tweet content generation,” Decis. Support Syst., 144: 113497, 2021.

[42] S. Poria, E. Cambria, A. Gelbukh, “Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis,” Association for Computational Linguistics, 2015.

[43] D. C. Edara, L. P. Vanukuri, V. Sistla, V. K. K. Kolli, “Sentiment analysis and text categorization of cancer medical records with LSTM,” J. Ambient Intell. Hum. Comput., 14: 5309-5325, 2023.

[44] M. E. Basiri, M. Abdar, M. A. Cifci, S. Nemati, U. R. Acharya, “A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques,” Knowl. Based Syst., 198: 105949, 2020.

[45] A. M. Shah, X. Yan, S. A. A. Shah, G. Mamirkulova, “Mining patient opinion to evaluate the service quality in healthcare: a deep-learning approach,” J. Ambient Intell. Hum. Comput., 11: 2925-2942, 2020.

[46] M. Parimala, R. M. Swarna Priya, M. Praveen Kumar Reddy, C. Lal Chowdhary, R. Kumar Poluru, S. Khan, “Spatiotemporal-based sentiment analysis on tweets for risk assessment of event using deep learning approach,” Softw.: Pract. Exper., 51(3): 550-570, 2020.

[47] J. Serrano-Guerrero, M. Bani-Doumi, F. P. Romero, J. A. Olivas, “Understanding what patients think about hospitals: A deep learning approach for detecting emotions in patient opinions,” Artif. Intell. Med., 128: 102298, 2022.

[48] M. E. Basiri, R. S. Chegeni, A. N. Karimvand, S. Nemati, “Bidirectional LSTM deep model for online doctor reviews polarity detection,” in Proc. 2020 6th International Conference on Web Research (ICWR), 2020.

[49] H. A. Shehu et al., “Deep sentiment analysis: a case study on stemmed Turkish twitter data,” IEEE Access, vol. 9, pp. 56836–56854, 2021.

[50] A. B. Nassif, A. Elnagar, I. Shahin, S. Henno, “Deep learning for arabic subjective sentiment analysis: Challenges and research opportunities,” Appl. Soft. Comput., 98: 106836, 2021.

[51] H. Elfaik, E. H. Nfaoui, “Deep bidirectional lstm network learning-based sentiment analysis for arabic text,” J. Intell. Syst., 30(1): 395–412, 2021.

[52] H. Saleh, S. Mostafa, L. A. Gabralla, A. O. Aseeri, S. El-Sappagh, “Enhanced arabic sentiment analysis using a novel stacking

ensemble of hybrid and deep learning models," Appl. Sci., 12(18): 8967, 2022.

[53] K. Dashtipour, M. Gogate, A. Adeel, H. Larijani, A. Hussain, "Sentiment analysis of persian movie reviews using deep learning," Entropy, 23(5): 596, 2021.

[54] Z. B. Nezhad, M. A. Deihimi, "Twitter sentiment analysis from Iran about COVID 19 vaccine," Diabetes Metab. Syndr., 16(1): 102367, 2021.

[55] S. Smetanin, M. Komarov, "Deep transfer learning baselines for sentiment analysis in Russian," Inf. Process Manag., 58(3): 102484, 2021.

[56] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.

[57] N. Reimers, I. Gurevych, "Alternative weighting schemes for elmo embeddings," arXiv preprint arXiv:1904.02954, 2019.

[58] J. Pennington, R. Socher, C. D. Manning, "Glove: Global vectors for word representation," in Proc. the 2014 conference on empirical methods in natural language processing (EMNLP): 1532–1543, 2014.

[59] G. Letarte, F. Paradis, P. Giguère, F. Laviolette, "Importance of self-attention for sentiment analysis," in Proc. the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP: 267–275, 2018.

[60] M. H. Phan P. O. Ogunbona, "Modelling context and syntactical features for aspect-based sentiment analysis," in Proc. the 58th Annual Meeting of the Association for Computational Linguistics: 3211–3220, 2020.

[61] Y. Vasiliev, Natural Language Processing with Python and SpaCy: A Practical Introduction. No Starch Press, 2020.

[62] C. Zhang, Q. Li, D. Song, "Aspect-based sentiment classification with aspect-specific graph convolutional networks," arXiv preprint arXiv:1909.03477, 2019.

[63] E. Cambria, R. Speer, C. Havasi, A. Hussain, "Senticnet: A publicly available semantic resource for opinion mining.," in Proc. AAAI fall symposium: commonsense knowledge, 10(0), 2010.

[64] B. Liang, H. Su, L. Gui, E. Cambria, R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," Knowl. Based Syst., 235: 107643, 2022.

[65] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, P. Agrawal, "Understanding emotions in text using deep learning and big data," Comput. Human Behav., 93: 309–317, 2019.

Biographies

Shahla Nemati was born in Shiraz, Iran, in 1982. She received the B.S. degree in hardware engineering from Shiraz University, Shiraz, in 2005, the M.S. degree from the Isfahan University of Technology, Isfahan, Iran, in 2008, and the Ph.D. degree in computer engineering from Isfahan University, Isfahan, in 2016. Since 2017, she has been an Assistant Professor with the Computer Engineering Department, Shahrekord University, Shahrekord, Iran. She has written several articles in the fields of data fusion, emotion recognition, affective computing, and audio processing. Her research interests include data fusion, affective computing, and data mining.

- Email: s.nemati@sku.ac.ir
- ORCID: 0000-0003-2906-5871
- Web of Science Researcher ID: AAA-3341-2019
- Scopus Author ID: 24512475100
- Homepage: https://www.sku.ac.ir/~snemati#

How to cite this paper:
S. Nemati, "Predicting the sentiment of tweet replies using attentive graph convolutional neural networks," J. Electr. Comput. Eng. Innovations, 12(1): 57-68, 2024.

DOI: 10.22061/jecei.2023.9611.644

URL: https://jecei.sru.ac.ir/article_1926.html