IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 19, 2020, accepted July 8, 2020, date of publication July 13, 2020, date of current version July 23, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3008824

Deep Learning Model for Fine-Grained Aspect-Based Opinion Mining

AHMED R. ABAS, IBRAHIM EL-HENAWY, HOSSAM MOHAMED, AND AMR ABDELLATIF
Department of Computer Science, Faculty of Computer and Informatics, Zagazig University, Zagazig 44519, Egypt
Corresponding author: Hossam Mohamed (h.hawash.research@gmail.com)

ABSTRACT Despite the great manufactures' efforts to achieve customer satisfaction and improve their performance, social media opinion mining is still on the fly a big challenge. Current opinion mining requires sophisticated feature engineering and syntactic word embedding without considering semantic interaction between aspect term and opinionated features, which degrade the performance of most of opinion mining tasks, especially those that are designed for smart manufacturing. Research on intelligent aspect level opinion mining (AOM) follows the fast proliferation of user-generated data through social media for industrial manufacturing purposes. Google's pre-trained language model, Bidirectional Encoder Representations from Transformers (BERT) widely overcomes existing methods in eleven natural language processing (NLP) tasks, which makes it the standard way for semantic text representation. In this paper, we introduce a novel deep learning model for fine-grained aspect-based opinion mining, named as FGAOM. First, we train the BERT model on three specific domain corpora for domain adaption, then use adjusted BERT as embedding layer for concurrent extraction of local and global context features. Then, we propose Multi-head Self-Attention (MSHA) to effectively fuse internal semantic text representation and take advantage of convolutional layers to model aspect term interaction with surrounding sentiment features. Finally, the performance of the proposed model is evaluated via extensive experiments on three public datasets. Results show that performance of the proposed model outperforms performances of recent the-of-the-art models.

INDEX TERMS Deep learning, opinion mining, sentiment analysis, social media analytics.

I. INTRODUCTION

The continuous proliferation of digital social media and increased popularity of e-commerce technologies enlarged the amount of user-generated multimodality data daily. This data encapsulates a massive amount of emotional intelligence and critical information that is valuable for several beneficiaries, like industrial manufacturers, corporations, and individuals. Consequently, an evolving research area of opinion mining gains increased attention concerning using information retrieval, translation, text summarization, and sentiment analysis models for mining people's feelings and reactions from overwhelming social streams [1] aiming to improve the industrial manufacturing process. In this paper, we primarily perform opinion mining on social media textual data.

Opinion mining (Sentiment analysis) has been a vigorous NLP research area, where various models have been introduced. Generally speaking, opinion mining techniques can be primarily assorted into three levels of granularities depending on the granularity level adopted for data processing: sentence-level opinion mining [2], document level opinion mining [3], and feature level opinion mining [4]. Among all these opinion mining levels, detecting the opinion polarity corresponding to various features (aspects) of the target entity is a challenging task that has many beneficial applications, especially in the smart manufacturing environments. For instance, if we have a statement like "The coffee at this cafe is perfect; nevertheless, the waiter is insolent," The aspect term "coffee" has a positive polarity, while the contradiction of the aspect " waiter " is contradictory to the "coffee" polarity. The aspect-based opinion mining is considered a fine-grained level of opinion mining that seeks to detect the sentiment polarity corresponding to various target-aspects in the textual social media streams, which offers a further accurate and comprehensive opinion mining model. Dependently, our main focus in this paper is the aspect-based opinion mining (AOM). Two widely adopted models for AOM task are conventional machine learning techniques and deep learning techniques. Conventional machine learning techniques primarily use handcrafted features such as bag-of-words, opinion lexicons to learn an intelligent model (e.g., support vector machine, naive Bayes) for detecting sentiment polarity

The associate editor coordinating the review of this manuscript and approving it for publication was Kemal Polat.

for user-generated text in various social media [5]. However, the model performance of such models extremely heavily relies on the eminence of the underlying features. So, researchers extensively adopted deep neural networks as an alternative paradigm to effectively learn dormant features in high dimensional data without the need for sophisticated feature engineering.

Recently, numerous neural networks, like Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Generative models, and others, are employed for the AOM task. RNN-based models are widely adopted by researchers for different NLP tasks because of its powerfulness in sequence modeling by stacking multiple layers of Long-Short-Term Memory (LSTM) or Gated-Recurrent-Units (GRU) to capture sentiment polarity of text. However, these models treat every word equally without considering important clues about the targeted-aspect. From the lingual perspective, rules and position of context words play different roles in forecasting the opinion polarity of various dimensions as mentioned above, the vital clue to the aspect term "waiter" is "insolent," not "delicious." Consequently, usage of the neural network model alone fails to discriminate the importance of context words. Moreover, if a person asked to complete a task like this, he/she will concentrate on specific fragments of the context to attain information required to develop a comprehension of targeted-aspect [6]. In inspiration of this human behavior, researchers widely applied attention mechanism with neural network architectures for many AOM applications. For example, Han *et al.* [7] introduced an attention mechanism with bidirectional GRU (Bi-GRU) for aspect sentiment classification in drug review, also Jiang *et al.* [8] employed global and local self-attention (SA) mechanism to learn textual web comment embedded semantics.

Differently from computer vision where low-level feature extraction can effectively be accomplished with transfer learning, most of NLP tasks require training from scratch for every application with an exception for word embedding models [9] (e.g., word2vec, Glove, Fast Text, etc.) that is trained on huge wide-scaled unlabeled text corpora, however they only deliberate syntactic information and suffer from semantic independence since the same word have the same embedding despite the diversity in context. Lately, several pre-trained language models like XLnet [29], BERT [10], and GPT [11] have revealed superior performance as a strong semantic text representation in many NLP research challenges. Even with its growing popularity in various NLP applications, extensive adoption of these models has not been widely exploited for AOM.

In this paper, AOM for Industrial Manufacturing is implemented depending on semantic embedding. The BERT model [10], which has shown high performance on sentence-level polarity detection, is pre-trained using three specific domain corpora. Then, a continuous dynamic masking technique is adopted for discriminating aspects' local and global semantic features based on a predefined relative distance measure. Finally, a fine-tuning layer is presented, in which several Multi-head Attention (MHA) mechanisms and convolutional operation are employed to capture hidden semantic interaction. Results of experiments demonstrate that the proposed model reliably outperforms the recent cutting-edge models and show a stable performance.

The main contributions of this research work are epitomized as follows:

- For context embedding, the BERT model is fine-tuned and trained on three domain-related corpora for AOM.
- Several MHA mechanisms and convolutional operation are adopted for both local and global context fusion.
- A sub-model is developed that comprises an MHA mechanism followed by a convolution layer fusing local context features and global features concurrently to deduce opinion polarity of the targeted aspect.
- The proposed model is implemented using widely known baseline language models to demonstrate the effectiveness of domain-adaption on model performance.

The organization of the rest of this paper can be summarized as follows. In section II, the recent related studies are discussed. In section III, an exhaustive explanation of the proposed model is introduced. In section IV, experimental results are shown to evaluate and compare the proposed model with other recent deep learning studies. Finally, in section V, conclusions and future work are presented.

II. RELATED WORK

AOM is a branch of opinion mining for fine-grained classification of user opinions in generated text across various digital social media. This research direction attracts the intention of researchers in recent years to investigate deep learning techniques for AOM using two kinds of word embedding paradigms: syntactical embedding and semantic embedding models [12]–[18].

Concerning syntactical embedding, Kumar *et al.* [12], employed two CNN and gating mechanisms for non-separated target extraction and learned semantic feature interaction. Also, Shindu *et al.* [1] introduced two layers of the LSTM model for domain embedded input. The first layer employed for aspect detection and the later for polarity classification. Additionally, Liu *et al.* [13] developed a deep gated other network to learn the representation of sentiment clue based on the relative distance between and sematic words and corresponding aspects, then a gating mechanism employed for noise reduction. Although conventional deep network models yielded a good result, yet it cannot effectively detect the word significance in the context. This problem motivates researchers to adopt attention mechanisms for deep learning models to automatically recognize various relevant clues for specific aspects in source statements. In [14], Xie *et al.*, present SA Bidirectional LSTM (Bi-LSTM) model for classifying aspect polarity in short texts by effectively capturing significant sentence parts and contextual correlations between aspects and semantic words [30]–[33].

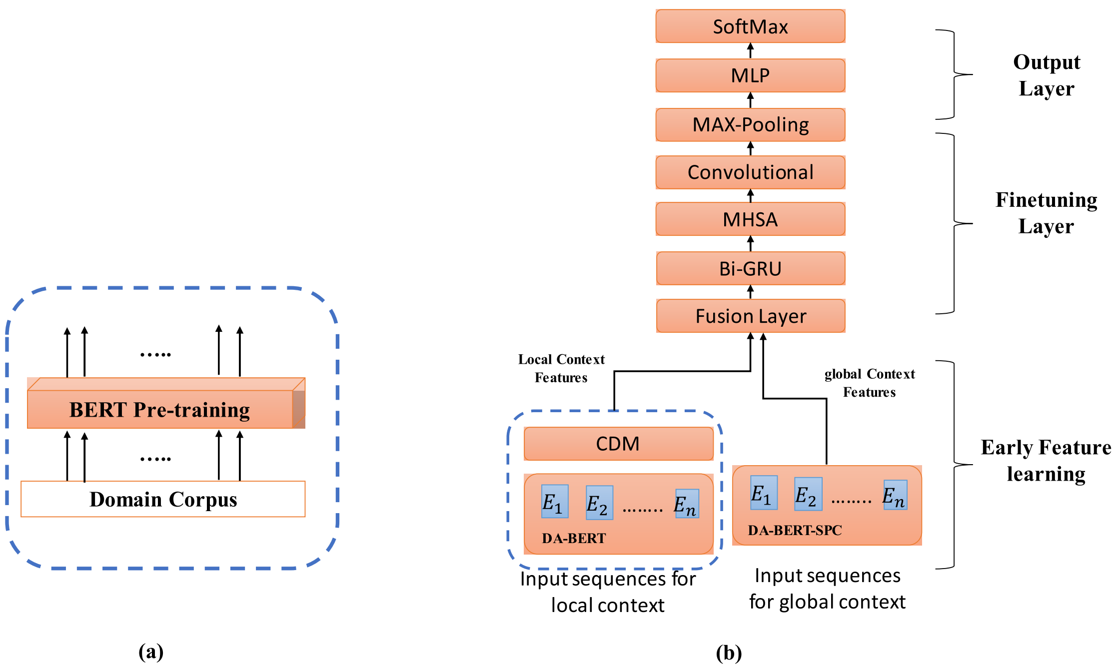


FIGURE 1. The FGAOM architecture (a) The BERT domain adaption pre-training (b) the architecture of the FGAOM model.

Meng *et al.* [9] adopted a CNN to extract high-level feature representation from improved word embedding layer. Bi-LSTM is used to capture local and global semantic information after that an attention layer is employed to highlight relevant aspect term features. Furthermore, Han *et al.* [15] applied attention mechanisms on top of Bi-GRU for multi-task learning AOM on pre-trained weight for online drug reviews. However, these studies rely on restricted window size for embedding, and cannot make use of semantic information locally or globally. This occurs due to converting words into a constant vector which incapable of representing the contextual meaning of words [34]–[39].

In [16], Yang *et al.* adopted BERT for generating embedding vector for each input sequence and then applied the CNN and Bi-GRU to learn and extract relative sentiment patterns in product reviews. These reviews were assigned an attention score with SA mechanism; however, the model trained cannot detect any polarity other than positive or negative and did not address aspect level classification. Zhang *et al.* [17] introduced a novel method for AOM by using a pre-trained BERT model for both aspect term embedding and word embedding and then adopted MHA and convolutional operation for learning hidden patterns. Even though the classification results obtained in [17] are good, two crucial matters are overlooked by authors: (a) the proposed model sometimes yields an incorrect analysis; (b) relevant semantic information about aspects and their neighbors are not exploited. In addition, Gao *et al.* [11] proposed three adapted versions for AOM concentrating on the target-related words other than the whole sentence; however, the proposed architectures yielded a high misclassification on the statement with neutral polarity class. In [18], Zhou applied the Gaussian kernel to generate influence vector for position-aware impact between BERT encoded contextual words and target aspects. Then, R-transformer was utilized to capture the relevant global and local information. The model depends on positional features.

In this paper, a novel model is proposed for AOM based on adapted BERT language model, which is pre-trained on relevant domain knowledge corpora, and a novel mechanism to discriminate local and global context features.

III. METHODOLOGY

In this section, the proposed model, namely deep learning model for AOM (FGAOM), is explained (see Figure 1 for a detailed visualization). The FGAOM primarily comprises five components: (1) BERT domain adaption, in which several corpora are used to train the BERT model. (2) Extraction of both local and global context features, in which the pre-trained BERT model (DA-BERT) is used as sequence embedding layer. (3) Learning long term dependencies representation, in which the GRU architecture is used as a feature extraction layer. (4) Capturing aspect relevant features, in which the generated sequential representation by the GRU is used as input to Multi-Head Self-Attention (MHSA) and convolution layer as a fine-tuning layer. (5) Polarity classification, in which fused output is fed into MLP and SoftMax function as an output layer. Assuming the input review/tweet is s $S = \{S_1, S_2, \cdots\cdots S_n\}$ the penalty area of our architecture is to forecast the targeted aspect term S_t polarity.

A. DOMAIN ADAPTION AND EMBEDDING LAYER

Because of the efficient performance of BERT [10] in feature representation compared to conventional syntax embeddings such as Glove [12] or word2vec [9] this research adopts and

TABLE 1. Description of pre-training corpora.

Corpus	Source	Sentences
Laptop	Yelp	100720930
Restaurant	Amazon	10000000
Twitter	Web crawler	9003100

fine-tunes the BERT model for the AOM task. As the first step of the FGAOM model, the language BERT model is fine-tuned by pre-training it using three domain-specific corpora (see, Table 1). The used corpora represent three domains that are laptops, restaurants, and twitter. These corpora are Yelp restaurants dataset [19], Amazon Laptop reviews [20], and tweets from our web crawler. SemEval 2014 data are omitted to prevent training bias for the test data. Also, short reviews/tweets with length less than two are removed. The BERT model is fine-tuned for ten epochs on each corpus. Then, the domain adapted BERT (DA-BERT) is adopted to form two concurrent embedding layers. One for extraction of local context sequence features S_l based on standard BERT; the other one is adopted for global context features S_j based on pair classification BERT. The embedding layers are represented in equations (1) and (2).

$$O^l_{BERT} = DA - BERT^l(S^l) \tag{1}$$
$$O^G_{BERT} = DA - BERT^g(S^g) \tag{2}$$

where O^l_{BERT} and O^g_{BERT} represent local and global input representation, respectively. Particularly, the input representation is produced by summation, segmentation and positional embeddings of tokens. For instance, given input sentence "[CLS] the soap is good [SEP] the waiter is insolent [CLS]" [CLS] is always denotes the first token of any sequence while language separator is represented with [SEP]. Thus, we adapt the given local and global context to match the data processing method of BERT.

Since previous studies [9] separately process aspect terms and context sequences and then model their interactions, the FGAOM adopts Semantic-Relative Distance (SRD) [21] to calculate if a semantic feature lies within local context interval of specific aspect term-based on the intuition that most significant context information belongs to aspect local context. Predominantly, the number of tokens between aspect term and semantic token considered as SRD of aspect-semantic pair (see, equation 3). For example, if SRD threshold is set to be seven, then every semantic token whose SRD in the direction of targeted aspect and less than or equal to seven will be considered as a local context; otherwise, it will be considered as a global context. In equation 3, P_w and P_a denote context and aspect word position, respectively, while L_a represents aspect term length.

$$SRD_i = |P_w - P_a| - (L_a/2) \tag{3}$$

B. EARLY FEATURE LEARNING LAYER

Early Feature Learning (EFL) layer is responsible for concurrently capturing both global and local context information. Global semantic information is extracted using DA-BERT-SPC based on a modified version of BERT model for sentence classification [22]. On the other hand, local context features obtained using continuous dynamic masking (CDM) layer [21] that mask non-local extracted semantic patterns learned by the DA-BERT layer. So, all positions of semantic features out of SRD of aspect words will be assigned zero values to balance the distribution of the features after the CDM operation as shown in equation 3. Hence, local context features can be calculated as shown in equation (4-6).

$$V_i = \begin{cases} ESRD_i \leq \alpha \\ OSRD_i > \alpha \end{cases} \tag{4}$$
$$M = [V^m_1, V^m_2, \ldots . V^m_n] \tag{5}$$
$$O^l_{CDM} = O^l_{BERT} \cdot M \tag{6}$$

where M denotes feature masking matrix, V^m_i represents token mask vectors within the sequence with length n, α denotes the SRD threshold value. Meanwhile, tokens that have SRD values from the corresponding aspect that are less than α are considered local contexts. The ones vectors denoted as $E \in \mathbb{R}^d$Ë while the zeros vectors denoted as $O \in \mathbb{R}^d$.

Both of global and local features are passed to the fusion layer. Despite of the existence of a variety of fusion strategies in the literature [40] the FGAOM utilizes the concatenation fusion strategy [40] because it doesn't change its input values and therefore it is suitable for the underlying nature of local and global features. The concatenation of local and global features can be formulated as shown in equation (7).

$$O = \left[O^l_{CDM}, O^G_{BERT}\right] \tag{7}$$

C. FINE-TUNING LAYER

Several recent studies for extracting hidden contextual states adopt RNN or relevant transformation techniques [9]. In this paper, hidden semantic representation is learned using a novel MHSA technique to learn and extract the contextual concealed patterns from concatenated local and global context features. The Finetuning layer primarily comprises three stacked layers. These layers are namely Bidirectional GRU (Bi-GRU), Attention mechanisms and the Convolution layer.

1) BI-GRU

The primary purpose of this layer is to capture the long-term interactions among fused input futures. GRU architecture is a different version of the RNN widely used for modeling long term sequential information. GRU is used in the FGAOM because of its time efficiency characteristics [16]. Also, it is used to merge the information of previous step in calculating the current output and to extract the context features in the sequence data. In linguistics, both of the previous and the succeeding words influence the current word [18]. For this reason, bidirectional GRU is adopted in the FGAOM to learn long term dependencies from accumulated features in the previous layer in both forward and backward directions.

Thus, for feature O_t at time t, the hidden states of forward and backward GRU are computed as shown in equations (8-10).

$$h_t^{'} = \overrightarrow{GRU}(O_t, h_{t-1}^{'}) \tag{8}$$

$$h_t^{''} = \overleftarrow{GRU}(O_t, h_{t-1}^{''}) \tag{9}$$

$$h_t = \left[h_t^{'}, h_t^{''}\right] \tag{10}$$

where $h_t^{'}$ and $h_t^{''}$ respectively represents the hidden state of the forward and backward GRU unit.

2) MULTI-HEAD SELF-ATTENTION PART

Inspired by the SA mechanism [14], we develop a MHSA that conducts several self-attentions to calculate attention scores for each semantic word, and to eschew the negative impact of long-distance correlation during feature learning. For faster and efficient calculation, scale-dot attention (SDA) is employed for score calculation [17]. Thus, given Bi-GRU output representation as $\{h_1, \cdots, h_t\}$ MHSA concurrently calculates SDA for each head hd_i, then concatenates the output and multiplies it with a weight matrix W^{MH} for feature transformation, as shown in equations (11-14).

$$\{\alpha_1, \cdots, \alpha_n\} = SofMax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{11}$$

$$hd_i\,(h_t) = \sum_{j=1}^{n} \alpha_j\, v_j \tag{12}$$

$$U = Concat\,(hd_1\,(h_t)\,, \ldots, hd_k\,(h_t)) * W^{MH} \tag{13}$$

$$U = \tanh(U) \tag{14}$$

where $Q, K,\ and\ V$ denote the query, key, and value matrix, correspondingly. The vector of $v_i, k_i\ and\ v_j$ in the mapping matrix can be formulated as shown in equation (14).

$$v_i, k_i, v_j = W^q h_t, W^k h_t, W^v h_t \tag{15}$$

where W^q, W^k, W^v represent the mapping matrices that vary for each attention head

3) CONVOLUTION LAYER

The output of MHSA mechanism is fed into two convolutional layers: the first layer has a *Relu* activation function, and the second layer has a *Linear* activation function to analyze the obtained contextual representation. Convolution layers are capable of capturing the most significant clues vector relevant to the targeted aspect, which in turn boost prediction of aspect polarity. The convolution operation is defined as shown in equation (16).

$$Conv\,(U) = Relu\left(U * W^1 + b^1\right) * W^2 + b^2 \tag{16}$$

While $*$ denotes convolution operation, W^1, W^2, b^1, b^2 represents the weight matrices and biases of the two convolutional layers, respectively.

4) MAX-POOLING LAYER

After the Convolutional layers, a max-pooling layer is adopted to detect the most important aspect-related opinion words within convolutional output representation. The extracted pattern by the max-pooling layer is considered as final and conclusive aspect-related feature representation. The most remarkable point about the max-pooling layer is its capability to transmute the inconstant-length vectors into the same-length vectors. The output of the max-pooling layer is computed as shown in equation (17).

$$h^p = MaxPooling(h^c|h, w) \tag{17}$$

where h^p represents max-pooling layer output, $h\ and\ w$ represent the sliding window height and width, and h^c represents the output of the convolutional layer.

D. OUTPUT LAYER

The last layer in the FGAOM model is a fully connected layer (FCL) that is used to predict the aspect sentiment class by converting the final context representation into a fine-grained sentiment h_f representation that is appropriate for forecasting the opinion polarity. Multi-layer perceptron (MLP) is adopted for the construction of fully-connected layer, and the output is computed as shown in equation (18).

$$X = Relu(W_m \cdot h_f + b_m) \tag{18}$$

where W_m and b_m are the weights and bias learning parameters that are updated through learning. $W_m \in \mathbb{R}^{c\times n}, b_m \in \mathbb{R}^{c\times 1}, h \in \mathbb{R}^{c\times 1} c$ is the number of polarity classes. Finally, a *SoftMax* activation function is employed to compute each sentiment class probabilities, where the class that gains the highest probability score is considered as the polarity of the targeted aspect. The computation of sentiment polarity class is shown in equations (19-20).

$$p = SoftMax(X) = \frac{exp(X)}{\sum_1^c exp(X)} \tag{19}$$

$$\hat{Y} = argmax(p) \tag{20}$$

where $p \in \mathbb{R}^{c\times 1}$ represents the probability of each class in c classes.

E. MODEL TRAINING

The proposed FGAOM model is trained using an end-to-end strategy by iteratively reducing the total cross-entropy loss to the lowest possible value with $L2$ regularization. In this paper, $\hat{y}_i \in \hat{Y}$ is used to represent the predicted polarity class and y_i is utilized to denote the actual polarity class of sentences. The training of the model aims to reduce the loss function calculated as shown in equation (21).

$$L_{Entopy} = \sum (y_i \log(\tilde{y}_i) + (1 - y_i)\log(1 - \tilde{y}_i)) + \omega L_2\,(\ominus) \tag{21}$$

where ω represents $L2$ regularization factor and $\ominus$ represents a set of model parameters.

TABLE 2. Proposed model hyperparameters.

Hyperparameters	Optimal values
DA-BERT epochs	10
Sequence length	256
learning rate	5e-5
batch size	96
training epochs	50
SRD $_{threshold}$	6 (Restaurant, Twitter)
	4(laptop)
optimizer	Adam
Number of GRU neurons	128
Number of head attention	16
Convolutional kernels	96
Convolutional size	3×3
MLP neurons	256,128

TABLE 3. Description of experimental datasets.

	Split	Restaurant	Laptop	Twitter
Positive	Train	2164	987	1411
	Test	728	340	173
Negative	Train	898	866	1411
	Test	196	128	173
Neutral	Train	637	460	2826
	Test	196	169	346
Total	------	4819	2950	6340

IV. EXPERIMENTS

This section shows the experimental settings used to evaluate the proposed FGAOM model; the data sets employed; the adopted evaluation metrics; and the results of the proposed model with different datasets. The proposed model is compared to the recent state-of-the-art models. The results are shown in Table 4. These results show the superior performance of the FGAOM in the AOM task.

A. EXPERIMENTAL SETTINGS

Pytorch library is utilized for model implementation, and for BERT domain adaption. Specifically, the uncased version of *BERTBASE with* 768 of embedding dimension is chosen. Bias values are initialized with zeros and Glorot uniform model [23] is adopted for weights initialization. Repeated experiments are performed to discover the optimal settings of hyper parameters. The optimal hyper parameters obtained from repeated experiments are shown in Table 2.

B. DATA SETS

The FGAOM model is trained and tested on three public datasets: Laptop, Restaurant, and Twitter datasets [24]. SemEval2014 [25] introduced both Laptop and Restaurant data sets. For data balance, the ‘‘conflict’’ polarities samples are removed from the first two data sets. These datasets contain three classes of polarity: positive, neutral, and negative. Every review/tweet comprises various aspect terms and their corresponding sentiment polarity. The training and testing samples for each sentiment category and the overall number of instances in each dataset are shown in Table 3.

C. PERFORMANCE METRICS

Accuracy and F1-measure are used in this paper as evaluation metrics as they are reliable measures of performance. The following parameters are used in computing the performance metrics:

- TP: the count of positive reviews/tweets that are labeled as positive.
- FP: the count of negative reviews/tweets that are labeled as positive.
- TN: the count of negative reviews/tweets that are labeled as a negative.
- FN: the count of positive reviews/tweets that are labeled as negative.
- Accuracy: the number of correctly classified reviews /tweets divided by the overall number of reviews/tweets.
- Precision: the percentage of correctly predicted positive reviews/tweets to the total positive reviews/tweets.
- Recall: the percentage of correctly predicted positive reviews/tweets to all reviews/tweets in the actual class.
- F1-measure: the weighted average of precision and recall.

D. COMPARED STUDIES

The performance of the proposed FGAOM model is compared with the performances of recent state-of-the-art deep learning models using three datasets mentioned in section B under the same experimental settings. These deep learning models are described as follows:

1) INTERACTIVE MULTI-HEAD ATTENTION NETWORKS (IMAN) [17]

This model uses the BERT model as an alternative to traditional embedding to simultaneously generate aspect embedding and context embedding. MHA mechanism and convolution operation are employed to further analyze the relevance of aspect and context words that passes through interactive pooling layer.

2) TRANSFORMER BASED MEMORY NETWORK (TF-MN) [8]

This model applies the transformer memory network to learn hidden contextual patterns in complex web comments via global attention and local attention mechanisms for constructing fine-grained semantic representation.

3) SELF-ATTENTION-BASED BiLSTM (SA-BiLSTM) [14]

This model adopts Bi-LSTM layers to model information from forward and backward directions for left and right context fusion. Then, a self-attention layer is used to assign higher weights for contextual words that highly correlate to aspect terms; the SoftMax layer is used for detecting polarity class of attention output.

4) INTERACTIVE GATED CONVOLUTIONAL NETWORK (IGCN) [12]

This model uses two CNN to concurrently extract and learn features from both aspect embedding and word embedding via convolutional kernels. Then, an interactive gating mechanism is employed to capture the impact of context information on Aspect and capture the most relevant feature for an aspect for efficient sentiment classification.

5) POSITION AND SELF-ATTENTION R-TRANSFORMER NETWORK (PSRTN) [18]

This model comprises three modules. First, the Gaussian kernel is used to calculate the position influence vector; second, the Bi-GRU network is used for modeling semantic representation of text; third, the MHA is used for scoring relevant aspect keywords.

6) FEATURE ENHANCED ATTENTION CNN-BiLSTM (FEA-NN)[9]

This model uses CNN for high-level feature extraction of input embedding. Then, the Bi-LSTM is used to capture both global and local context features that fed into SA for obtaining essential feature interaction with aspect.

7) TARGET-DEPENDENT BERT (TD-BERT) [11]

This model develops the BERT model for target sentiment classification by stacking a max-pooling layer and an FCL on top of the BERT model. Max pooling layer extracts targets from multiple target sequences, then the target position and surrounding semantic fed into FCL for classification.

8) RECURRENT MEMORY NEURAL NETWORK (ReMemNN) [27]

This model uses an adjustment module for enhancing embedding. The embedded input is fed into multi-element attention module to generate an accurate score for aspect relevant sentiment, and then attention outputs are passed to a memory module for hidden state generation.

9) MULTI-ATTENTION NETWORK (MAN) [6]

This model uses two attention modules, namely: Intra-attention and Inter-attention for AOM task. In the first module, transformer is used to capture long term semantic relations. In second module, global and local attention mechanisms are used for gathering a fine-tuned interaction between aspects and surrounding clues.

10) LSTM [4]

This model is a two-stage model. The first stage uses position-based attention to model the obvious position information between the aspect and its semantic words to deal with each aspect alone; in second stage, content attention technique is used to explore multi-aspects within one opinionated sentence.

11) HOLISTIC RECURRENT ATTENTION ON TARGET (HRT) ONE DIRECTION (HRT_ONE) [15]

This model uses two single layers of one direction LSTM to learn and memorize hidden semantic features. Then, it applies holistic recurrent content attention for scoring aspect relevant terms that are passed to SoftMax function for classification.

12) HRT BI-DIRECTIONAL (HRT_Bi) [15]

This model replaces LSTM layer in HRT_one with Bi-LSTM for both left and right feature learning. Also, it uses position attention to assign higher weights to the opinion word that is near to the aspect terms according to an intuition that opinion words close to the aspect terms are more likely to detect aspect polarities.

TABLE 4. Comparing performance of the proposed model to recent state-of-the-art models.

Model	Restaurant (%)		Laptop (%)		Twitter (%)	
	Acc	F1	Acc	F1	Acc	F1
IMAN [17]	83.59	75.63	80.53	76.91	75.72	74.5
TF-MN [8]	81.87	80.70	78.06	77.73	72.12	68.13
SA-BiLSTM [14]	75.14	67.32	66.55	65.17	69.79	63.27
IGCN [12]	81.34	75.32	75.24	63.28	72.8	67.73
PSRTN [18]	83.80	76.14	80.9	73.21	75.1	72.13
FEA-NN [9]	83.21	77.57	78.55	72.65	73.31	68.67
TD-BERT [11]	84.56	79.61	78.42	74.37	77.31	74.4
ReMemNN [27]	79.64	68.36	71.58	65.41	71.39	68.88
MAN [6]	84.38	71.31	78.13	78.13	76.56	72.19
LSTM [4]	80.10	72.30	73.1	69.56	73.78	70.31
HRT_one [15]	81.43	72.57	73.04	68.69	73.84	72.08
HRT_Bi [15]	81.96	74.09	74.45	70.83	73.27	71.98
GANN [13]	80.09	75.16	72.21	68.29	74.25	72.01
BERT [28]	86.16	78.83	79.18	74.86	73.25	70.12
FGAOM	**91.6**	**88.01**	**85.24**	**83.67**	**80.6**	**79.21**

13) GATED ALTERNATE NEURAL NETWORK (GANN) [13]

This model uses a new gated truncation RNN to learn the hidden representation of aspect related sentiment clues in input sequences. Also, it uses a noise filtering mechanism to alleviate unnecessary information. Then, the convolution and the max-pooling layers are used to learn important clue features and their positions with aspect.

14) BERT [28]

This model uses the BERT model [10] to generate aspect related word representation for opinion classification in input sequences based on a novel unified framework that amalgamates data from three corpora: word-level sentiment lexicons, aspect-level corpora, and sentence-level corpora.

E. RESULTS

Table 4 and Fig.2 show performance results of the FGAOM model compared with other recent state-of-the-art deep learning models. These results show that the performance of the proposed model outperforms other models compared in accuracy and F1-measure evaluation metrics using three benchmark data sets. The proposed model achieves new state- of-the-art results on restaurant dataset with **91.6% of** accuracy and **88.01%** of F1-measure, which are higher than all compared models. Table 6 and Fig.3 show that the proposed model yields an accuracy of **85.24%** and F1-measure

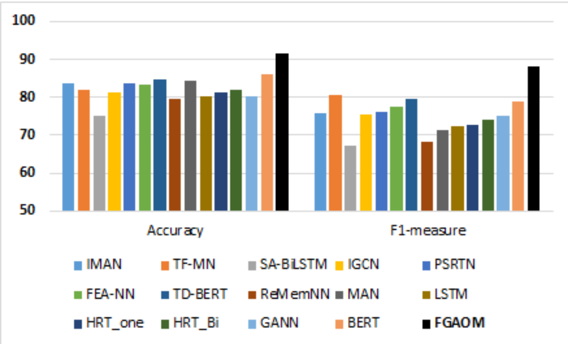


FIGURE 2. Results of comparing the proposed model with recent models using Restaurant data set.

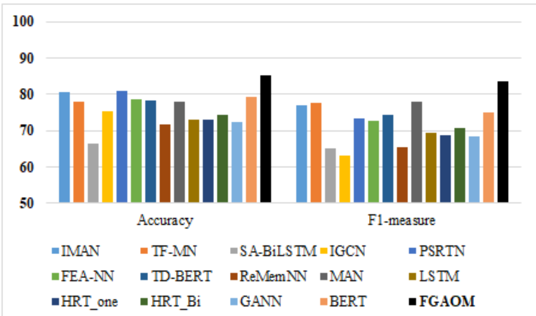


FIGURE 3. Results of comparing the proposed model with recent models using Laptop data set.

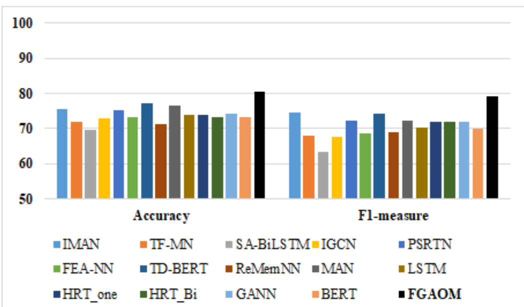


FIGURE 4. Results of comparing the proposed model with recent models using Twitter data set.

of **83.67% using the** laptops dataset, which outperforms the other models compared. Table 7 and Fig.4 show that the proposed model achieves **80.6 %** of accuracy and **79.21**% of F1-measure using the twitter dataset, which outperforms the models compared. Table 5 and Fig.5 show that the proposed model outperforms the models compared using the restaurant dataset with 5.44% of accuracy and 7.31% of F1-measure.

Reasons of performance superiority of the proposed model are as follows:

Domain adapted BERT: using pre-trained domain adapted BERT as an embedding layer rather than using baseline pre-trained language models enormously boosts the performance of the FGAOM compared with the performances of BERT $_{\text{BASE}}$ and XLnet $_{\text{BASE}}$ [29]. In order to demonstrate this point, the performance of the proposed

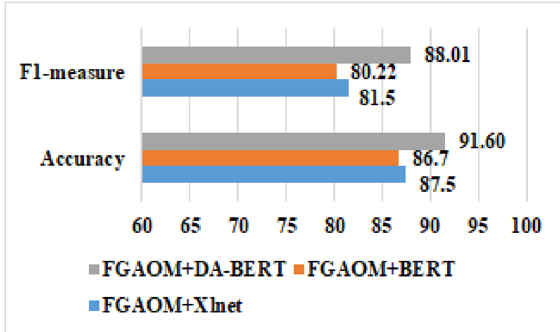


FIGURE 5. Comparing the performance of the proposed model using DA-BERT and using baseline pre-trained language models with Restaurant data set.

TABLE 5. The confusion matrix of FGAOM on Restaurant data set.

Model	Restaurant (%)		Laptop (%)		Twitter (%)	
	Acc	F1	Acc	F1	Acc	F1
FGAOM + BERT	86.7	80.22	80.53	76.91	75.72	74.5
FGAOM + Xlnet	87.5	81.5	78.06	77.73	72.12	68.13
FGAOM + DA-BERT	**91.6**	**88.01**	**85.24**	**83.67**	**80.6**	**79.21**

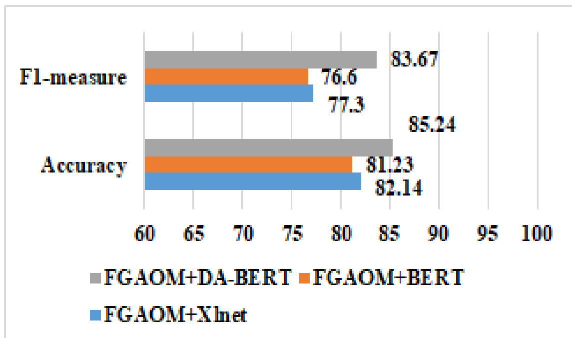


FIGURE 6. Comparing the performance of the proposed model using DA-BERT and using baseline pre-trained language models with Laptop.

model that uses domain-adapted BERT model is compared with its performance when the BERT-based architecture [10] and the XLnet-Based architecture [29] pre-trained models are used instead of the domain adapted BERT model using Restaurant, Laptop, and Twitter data sets and results are shown in Table 5, and Fig.5, Fig.6, and Fig.7. These results show that the performance of the proposed model outperforms the BERT-based architecture and the XLnet-Based architecture on accuracy and F1-measure respectively. Therefore, domain adjustment effectively enhances AOM and demands fewer parameters to be learned.

Multiple Attention Mechanisms: employing multiple attention mechanisms leads to grasping significant information of concatenated local and global context words and learning further interactive aspect and sentence representations. Consequently, the proposed model works better than the traditional neural network model and achieves better results.

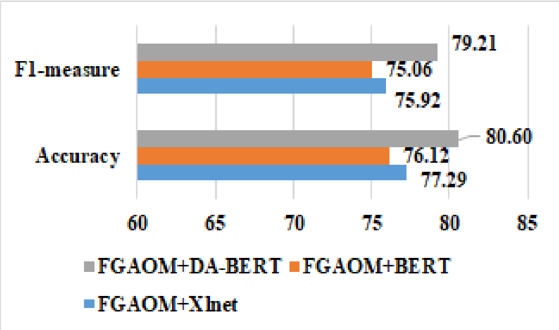


FIGURE 7. Comparing the performance of the proposed model using DA-BERT and using baseline pre-trained language models with Twitter data set.

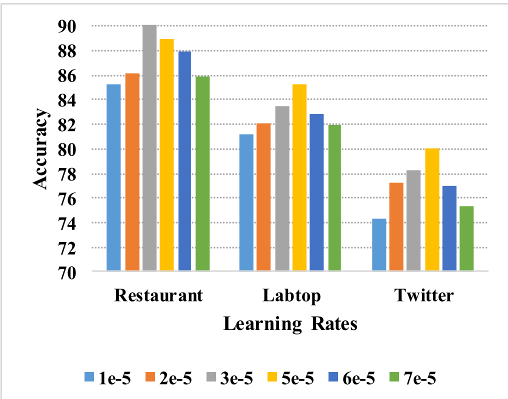


FIGURE 8. Impact of different learning rates on model accuracy.

On the other hand, DA-BERT's architecture helps the proposed model in implementing much attention to produce improved aspect-specific sequence representations.

The proposed model is tested using different learning rates and results are shown in Fig.8. It is observed that the highest accuracy is obtained with Restaurant dataset when the learning rate is equal to 3e-5. On the other hand, the highest accuracy is obtained with Laptop and Twitter datasets when the learning rate is equal to 5e-5. It is also noticed that changes in learning rate value significantly affect the performance of the proposed model and how different parameter configurations influence obtained results.

Furthermore, the effect of the dynamic masking technique on the proposed model is validated through analyzing the impact of different $SRD_{threshold}$ values on model accuracy. Results are shown in Fig 9. Results show that for restaurant data set the proposed model achieves the highest accuracy of 91.6% when $SRD_{threshold}$ is equal to 4. While for Twitter data set it achieves the maximum accuracy of 80.68% when the $SRD_{threshold}$ is equal to 4. Finally, for Laptop dataset it achieves the maximum accuracy of 85.24% when the $SRD_{threshold}$ is equal to 6. These results show the impact of $SRD_{threshold}$ on the proposed model performance.

Moreover, the confusion matrix corresponding to testing the proposed model using Restaurant, Laptop, and twitter datasets are presented in Tables 6, 7, and 8, respectively.

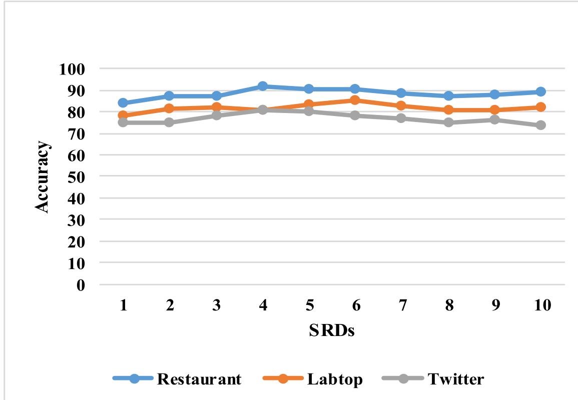


FIGURE 9. Impact of SRD on the proposed model accuracy.

TABLE 6. Comparing the performance of the proposed model using DA-BERT and using baseline pre-trained language models.

		Predicted values			
		Positive	**Negative**	**Neutral**	**RCL**
Actual values	**Positive**	691	9	28	94.91 %
	Negative	8	175	13	89.28 %
	Neutral	17	19	160	81.63 %
	PRC	96.50%	86.20%	79.60%	
	F1	95.70%	87.71%	80.60%	

TABLE 7. The confusion matrix of FGAOM on Laptop data set.

		Predicted values			
		Positiv e	**Negative**	**Neutra l**	**RCL**
Actual values	**Positive**	298	9	33	87.64 %
	Negative	4	112	12	87.50 %
	Neutral	19	17	133	78.69 %
	PRC	92.83%	81.15%	74.71%	
	F1	90.16%	84.21%	76.65%	

TABLE 8. The confusion matrix FGAOM on Twitter data set.

		Predicted values			
		Positive	**Negative**	**Neutral**	**RCL**
Actual values	**Positive**	122	9	42	70.52 %
	Negative	7	127	39	73.41 %
	Neutral	18	19	309	89.30 %
	PRC	82.99%	81.93%	79.23%	
	F1	76.25%	77.43%	83.96%	

For each class, corresponding precision (PRC), recall (RCL), and F1-measure (F1) are provided. Based on the differences between predicted labels and actual labels it is shown that the proposed model can discriminate the Positive and Negative sentiment statements. It is also shown that the proposed model achieves the lowest precision with the Neutral class. This indicates that it is quite difficult for the proposed model to discriminate the Neutral sentiment statements.

V. CONCLUSION AND FUTURE WORK

In this paper, it is proposed a novel cognitive analytics model for fine-grained aspect-based opinion mining. This model improves the industrial manufacturing process by exploiting interaction of semantic context words and aspect terms in the AOM process. First, the BERT model is pre-trained using three domain-related corpora to enhance model classification performance. Second, the domain-adapted BERT model is employed as an input embedding layer for independent local and global context fusion using the SRD technique. Third, MHSA is employed for extracting high-level hidden state representation. Fourth, convolution operation is used to perform fine-tuned learning of aspect interaction with surrounding sentiment features. Extensive experiments are performed using three public datasets. Results obtained show that the proposed model outperforms the compared recent state-of-the-art models. However, the performance of the proposed model in Neutral class prediction is low compared to Positive and Negative classes.

In future work, the proposed model might be investigated in cross-domain and multi-lingual cognitive analysis. Also, the proposed model might be developed such that an adaptive distance threshold parameter might be learned from context during training. Finally, the XLnet architecture might be developed with fine-grained cognitive mining and compared to various BERT architectures.

REFERENCES

[1] I. Sindhu, S. M. Daudpota, K. Badar, M. Bakhtyar, J. Baber, and M. Nurunnabi, "Aspect-based opinion mining on Student's feedback for faculty teaching performance evaluation," *IEEE Access*, vol. 7, pp. 108729–108741, 2019.

[2] M. Araújo, A. Pereira, and F. Benevenuto, "A comparative study of machine translation for multilingual sentence-level sentiment analysis," *Inf. Sci.*, vol. 512, pp. 1078–1102, Feb. 2020.

[3] L. Kong, C. Li, J. Ge, F. Zhang, Y. Feng, Z. Li, and B. Luo, "Leveraging Multiple Features for Document Sentiment Classification," *Inf. Sci.*, vol. 518, pp. 39–55, May 2020.

[4] A. Da'u, N. Salim, I. Rabiu, A. Osman, "Recommendation system exploiting aspect-based opinion mining with deep learning method," *Inf. Sci.*, vol. 512, pp. 1279–1292, Feb. 2020.

[5] X. Ma, J. Zeng, L. Peng, G. Fortino, and Y. Zhang, "Modeling multi-aspects within one opinionated sentence simultaneously for aspect-level sentiment analysis," *Future Gener. Comput. Syst.*, vol. 93, pp. 304–311, Apr. 2019.

[6] Q. Xu, L. Zhu, T. Dai, and C. Yan, "Aspect-based sentiment classification with multi-attention network," *Neurocomputing*, vol. 388, pp. 135–143, May 2020.

[7] Y. Han, M. Liu, and W. Jing, "Aspect-level drug reviews sentiment analysis based on double BiGRU and knowledge transfer," *IEEE Access*, vol. 8, pp. 21314–21325, 2020.

[8] M. Jiang, J. Wu, X. Shi, and M. Zhang, "Transformer based memory network for sentiment analysis of Web comments," *IEEE Access*, vol. 7, pp. 179942–179953, 2019.

[9] W. Meng, Y. Wei, P. Liu, Z. Zhu, and H. Yin, "Aspect based sentiment analysis with feature enhanced attention CNN-BiLSTM," *IEEE Access*, vol. 7, pp. 167240–167249, 2019.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*. [Online]. Available: http://arxiv.org/abs/1810.04805

[11] Z. Gao, A. Feng, X. Song, and X. Wu, "Target-dependent sentiment classification with BERT," *IEEE Access*, vol. 7, pp. 154290–154299, 2019.

[12] A. Kumar, V. T. Narapareddy, V. A. Srikanth, L. B. M. Neti, and A. Malapati, "Aspect-based sentiment classification using interactive gated convolutional network," *IEEE Access*, vol. 8, pp. 22445–22453, 2020.

[13] N. Liu and B. Shen, "Aspect-based sentiment analysis with gated alternate neural network," *Knowl.-Based Syst.*, vol. 188, Jan. 2020, Art. no. 105010.

[14] J. Xie, B. Chen, X. Gu, F. Liang, and X. Xu, "Self-attention-based BiLSTM model for short text fine-grained sentiment classification," *IEEE Access*, vol. 7, pp. 180558–180570, 2019.

[15] H.-J. Park, M. Song, and K.-S. Shin, "Deep learning models and datasets for aspect term sentiment classification: Implementing holistic recurrent attention on target-dependent memories," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104825.

[16] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for E-commerce product reviews in chinese based on sentiment lexicon and deep learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020.

[17] Q. Zhang, R. Lu, Q. Wang, Z. Zhu, and P. Liu, "Interactive multi-head attention networks for aspect-level sentiment classification," *IEEE Access*, vol. 7, pp. 160017–160028, 2019.

[18] Z. Zhou, F. Liu, and Q. Wang, "R-transformer network based on position and self-attention mechanism for aspect-level sentiment classification," *IEEE Access*, vol. 7, pp. 127754–127764, 2019.

[19] J.-D. Zhang and C.-Y. Chow, "MOCA: Multi-objective, collaborative, and attentive sentiment analysis," *IEEE Access*, vol. 7, pp. 10927–10936, 2019.

[20] Q. Ma, L. Yu, S. Tian, E. Chen, and W. W. Y. Ng, "Global-local mutual attention model for text classification," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 12, pp. 2127–2139, Dec. 2019.

[21] B. Zeng, H. Yang, R. Xu, W. Zhou, and X. Han, "LCF: A local context focus mechanism for aspect-based sentiment classification," *Appl. Sci.*, vol. 9, no. 16, p. 3389, Aug. 2019.

[22] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "Attentional encoder network for targeted sentiment classification," 2019, *arXiv:1902.09314*. [Online]. Available: http://arxiv.org/abs/1902.09314

[23] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[24] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive recursive neural network for target-dependent Twitter sentiment classification," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics (Short Papers)*, vol. 2, 2014, pp. 49–54.

[25] S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, "NRC-Canada-2014: Detecting aspects and sentiment in customer reviews," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, 2014, pp. 437–442.

[26] X. Tan, Y. Cai, J. Xu, H.-F. Leung, W. Chen, and Q. Li, "Improving aspect-based sentiment analysis via aligning aspect embedding," *Neurocomputing*, vol. 383, pp. 336–347, Mar. 2020.

[27] N. Liu and B. Shen, "ReMemNN: A novel memory neural network for powerful interaction in aspect-based sentiment analysis," *Neurocomputing*, vol. 395, pp. 66–77, Jun. 2020.

[28] F. Chen, Z. Yuan, and Y. Huang, "Multi-source data fusion for aspect-level sentiment classification," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104831.

[29] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 5754–5764.

[30] L. Caviglione, M. Podolski, W. Mazurczyk, and M. Ianigro, "Covert channels in personal cloud storage services: The case of dropbox," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1921–1931, Aug. 2017.

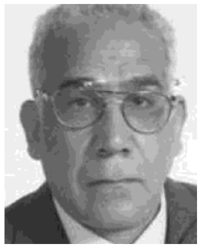
[31] M. Skowron, A. Janicki, and W. Mazurczyk, "Traffic fingerprinting attacks on Internet of Things using machine learning," *IEEE Access*, vol. 8, pp. 20386–20400, 2020.

[32] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, and T. Baker, "Analysis of dimensionality reduction techniques on big data," *IEEE Access*, vol. 8, pp. 54776–54788, 2020.

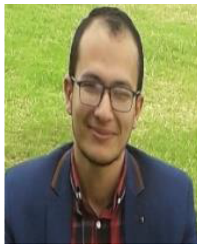
[33] M. Tariq, M. Adnan, G. Srivastava, and H. V. Poor, "Instability detection and prevention in smart grids under asymmetric faults," *IEEE Trans. Ind. Appl.*, early access, Jan. 7, 2020, doi: 10.1109/TIA.2020.2964594.

[34] J. Sakhnini, H. Karimipour, A. Dehghantanha, R. M. Parizi, and G. Srivastava, "Security aspects of Internet of Things aided smart grids: A bibliometric survey," *Internet Things*, Sep. 2019, Art. no. 100111, doi: 10.1016/j.iot.2019.100111.
[35] M. Aloqaily, I. A. Ridhawi, H. B. Salameh, and Y. Jararweh, "Data and service management in densely crowded environments: Challenges, opportunities, and recent developments," *IEEE Commun. Mag.*, vol. 57, no. 4, pp. 81–87, Apr. 2019.
[36] I. Al Ridhawi, Y. Kotb, M. Aloqaily, Y. Jararweh, and T. Baker, "A profitable and energy-efficient cooperative fog solution for IoT services," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3578–3586, May 2020.
[37] Z. Ullah, F. Al-Turjman, and L. Mostarda, "Cognition in UAV-aided 5G and beyond communications: A survey," *IEEE Trans. Cognit. Commun. Netw.*, early access, Jan. 21, 2020, doi: 10.1109/TCCN.2020.2968311.
[38] F. Al-Turjman, B. D. Deebak, and L. Mostarda, "Energy aware resource allocation in multi-hop multimedia routing via the smart edge device," *IEEE Access*, vol. 7, pp. 151203–151214, 2019.
[39] F. Al-Turjman, H. Zahmatkesh, and L. Mostarda, "Quantifying uncertainty in Internet of medical things and big-data services using intelligence and deep learning," *IEEE Access*, vol. 7, pp. 115749–115759, 2019.
[40] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1933–1941.

AHMED R. ABAS received the bachelor's degree in electronics and communications engineering from Zagazig University, Egypt, in 1993, the M.Sc. degree in computer engineering from Cairo University, Egypt, in 1999, and the Philosophy Doctorate in computer science from Exeter University, U.K., in 2005. He is currently an Assistant Professor and Master's Supervisor with the Computer Science Department, Faculty of Computer and Informatics, Zagazig University, Egypt. He is a middle-aged and young expert. His research interests include deep learning, neural networks, artificial intelligence, natural language processing, data mining, parallel processing, and pattern recognition.



IBRAHIM EL-HENAWY is currently a Professor of computer science and the Head of the Department of Computer Science, Faculty of Computer and Informatics, Zagazig University. His research interests include computer science, bioinformatics, deep learning, soft computing, and optimization.



HOSSAM MOHAMED was born in 1995. He received the bachelor's degree from Zagzi University, in 2016. He is currently pursuing the master's degree with the Faculty of Computer and Informatics, Zagazig University, Egypt. His research interests include deep learning, natural language processing, data mining, and cognitive analysis.

AMR ABDELLATIF was born in 1982. He received the Ph.D. degree from the Faculty of Computer Science, Cairo University, in 2016. He is currently a Lecturer with the Faculty of Computer Science, Zagazig University. His research interests include compiler construction, parsers, automata, parallel processing natural language processing, and artificial intelligence.