Attention-based aspect sentiment classification using enhanced learning through CNN-BiLSTM networks

Eniafe Festus Ayetiran

Department of Mathematical Sciences, Achievers University, Owo, Nigeria

ARTICLE INFO

Keywords:
Transfer learning
Attention
CNN
BiLSTM
Joint learning

Abstract

Deep neural networks (DNN) techniques for aspect-based sentiment classification have been widely studied. The success of these methods depends largely on training data which are often inadequate because of the rigor involved in manually tagging large collection of opinionated texts. Attempts have been made to transfer knowledge from document-level to aspect-level sentiment task. However, the success of this approach is also dependent on the model because aspect sentiment data like other type of texts contain complex semantic features. In this paper, we present an attention-based deep learning technique which jointly learns on document and aspect-level sentiment data and which also transfers learning from the document-level data to aspect-level sentiment classification. It basically consists of a convolutional layer and a bidirectional long short-term memory (BiLSTM) layer. The first variant of our technique uses convolutional neural network (CNN) to extract high-level semantic features. The output of the feature extraction is then fed into the BiLSTM layer which captures the contextual feature representation of the texts. The second variant applies the BiLSTM layer directly on the input data. In both variants, the output hidden representation is passed to an output layer using softmax activation function for sentiment polarity classification. We evaluate our model on four standard benchmark datasets which shows the effectiveness of our approach with improvements over baselines. We also conduct ablation studies to show the effect of the different document-level weights on the learning techniques.

1. Introduction

Aspect-based sentiment analysis (ABSA) involves mining opinions from text about specific entities (targets) and their aspects [43] which provides valuable insights to both consumers and businesses for decision making [9]. Current tasks in aspect-based sentiment analysis (ABSA) includes aspect category detection, opinion target extraction and aspect sentiment polarity classification [30]. However, the main task of the three is the sentiment polarity classification, which actually assists organizations in decision making. The possibility of multiple aspects in an opinionated text implies equivalent sentiment polarities, one for each aspects in the text. Sentiment polarity classes have evolved over the years from the traditional two-way *positive* and *negative* classes with the addition of the *neutral* class. Furthermore, aspect sentiment can be expressed at sentence and document levels.

Schouten and Frasincar [33] identifies three major approaches to aspect-based sentiment analysis namely: dictionary-based, supervised machine learning and unsupervised machine learning. In the dictionary-based approach, a sentiment dictionary is created from lexical resources such as WordNet [13] by tagging sentiment words (mainly adjectives) into polarity classes leveraging the synonymy/antonymy relationships. The synonymy/antonymy relations assist in accurately classifying the opinion words into *positive* and *negative* classes. The opinion texts are classified into sentiment classes using an algorithm centred around the polarity classes of the sentiment words in each of the texts. In contrast to dictionary-based approaches, supervised machine learning approaches use labeled examples to predict and classify unlabeled texts into sentiment classes. It can use any of the widely known supervised algorithms such as support vector machines (SVM), neural networks etc. A recent trend in supervised machine learning approach for sentiment analysis is the use of emotional recurrent units [21] and affective knowledge with neural networks [23]. Unsupervised machine learning approaches on the other hand, include algorithms which are self-learning on unlabeled data to predict sentiment classes for opinionated texts.

In supervised machine learning approach, attention mechanisms have been shown to achieve good results on aspect sentiment analysis [39, 24, 34, 25]. This is due to its ability to focus on different parts of the text and assign different weights. A major problem with supervised systems is the lack of adequate labeled training data. As a result of this reason, in order to enhance learning, knowledge transfer techniques from documents to attention mechanism have also been studied [17, 16]. On the other hand, the traditional long short-term memory LSTM networks have been applied to several natural language processing (NLP) techniques, including aspect-based sentiment analysis due to their ability to learn long-term dependencies. In this work, we apply bidirectional LSTM to joint/transfer learning in order to capture past (using backward layer) and future (using forward layer) contextual information inherent in opinionated texts. Furthermore, we also experiment with a variant which first extracts high-level features using a convolutional neural network (CNN) layer before capturing the contextual information.

The significant contributions of this work are as follow:

- We develop an attention-based CNN-embedded BiLSTM model for aspect sentiment classification which learns

eniafe.ayetiran@achievers.edu.ng (E.F. Ayetiran)
ORCID(s): 0000-0002-6816-2781 (E.F. Ayetiran)

in both direction of an opinionated text and which allows capturing of high-level semantic and contextual features

- In a departure from previous works in which opinion target representation is usually an average of word embeddings of the target, we compute the target representation as an average of the BiLSTM hidden vectors of the target. This can make a remarkable difference with a considerable number of multi-word targets.

- Experimental evaluation of our hypothetical model achieves predominant superior performance over the baselines

The rest of the paper is structured as follows: Section 2 reviews related work. The knowledge-enhanced model using attention-based CNN-BiLSTM algorithm is discussed in Section 3. Section 4 describes the experiments. In section 5, we present and discuss the results. Section 6 presents and discusses the results of ablation experiments. Section 7 concludes the paper.

2. Related work

2.1. Attention-based network approaches to aspect-based sentiment analysis

One of the earliest work on attention-based LSTM for aspect sentiment classification is the work of [39]. They specifically proposed two ways of taking aspect information into account. The first method is to concatenate aspect vectors into sentence hidden vectors for computing attention weights. The second method is to affix the aspect vector into the input word vectors. Evaluation of their method on SemEval-2014 Task 4 dataset [32] showed improvement over previous techniques. In another work, Liu and Zhang [24] used vanilla LSTM-based model to induce the attention value of different parts of a whole sentence with respect to the specific target mentioned in the sentence. They further modify their model to differentiate the left and right contexts of the target (aspect) words. They showed improvement on two benchmark datasets. Ma et al. [25] chose to model opinion targets and their context separately through what they call *interactive attention networks* (IAN). Evaluation of their method was on SemEval-2014 Task 4 dataset improved baselines excluding [39] which used the same dataset. Chen et al. [7] proposed a multiple attention framework to detect the sentiment of opinion targets. The framework consists of five modules: input module, memory module, position-weighted memory module, recurrent attention module, and output module. Evaluation on four benchmark datasets showed improvement in performance. Fan et al. [12] developed a multiple-attention neural network model for identification of sentiment polarity in comments/reviews. Experiments were carried out on four benchmark datasets. Huang et al. [19] introduced a method based on neural networks which learns using an attention-over-attention mechanism. This method allows the model to jointly learn representations for both aspects and sentences and was evaluated on two benchmark datasets covering restaurant and laptop domains. Advancing their previous work, He et al. [16] proposed two techniques for improving the effectiveness of attention of contexts towards opinion targets. The first proposal introduced a target representation that better captures the semantics of the opinion target. The second introduced syntactic information into the model. These techniques further improved their previous method when evaluated on the same datasets. Yang et al. [42] proposes a co-attention mechanism which models both target and contexts attention alternately so as to focus on those key words of targets in order to learn more effective context representation. They implemented this technique using LSTM and another variation using a co-attention-MemNet network which adopts a multiple-hops co-attention mechanism The result of their experiment showed effectiveness of the proposed methods.

Wu et al. [40] have shown that knowledge from multiple sources can greatly improve aspect sentiment classification. Their model specifically incorporates sentiment knowledge from document-level sentiment labels and conjunction relations at clause level. Du et al. [11] proposed an interactive attention capsule network to tackle the issue of overlapping feature representations which usually result from multi-aspect sentences. Experiments on three datasets from SemEval-2014 task 4 and Twitter collection achieved superior performance over baselines at the time of reporting. To take the position of each aspect into consideration, Zhou et al. [44] proposes position-aware hierarchical transfer (PAHT) model which consists a combination of components: aspect-based word position modeling, word encoder, aspect-based word positional attention, aspect-based segment position modeling, segment encoder, aspect-based segment positional attention and sentiment classification. They augmented training data through hierarchical transfer from sentence-level sentiment with improved performance. Due to the fact that words may possibly carry different sentiments for different aspects and that an aspect's sentiment might be highly influenced by the domain-specific knowledge, Meskele and Frasincar [27] proposed a hybrid solution using a lexicalized domain ontology and a regularized neural attention model. A bidirectional context attention mechanism is introduced to measure the weight of each aspect word in an opinionated text to determine its polarity. They reported superior results on some standard benchmark datasets. Yadav et al. [41] proposed a positionless attention-based sentiment classification using bidirectional recurrent gated units. The crux of their work is the elimination of aspect words' positions and positional embeddings.

2.2. Knowledge-enhanced approaches to aspect-based sentiment analysis

In recent times, due to the problem of inadequate training data, incorporation of lexical and/or external knowledge into several NLP areas is attracting increasing attention. These areas include but not limited to text classification [2], word sense disambiguation [4, 3] and sentiment analysis. In order to enhance learning, He et al. [17] introduced an enhanced attention-based LSTM model for aspect sentiment classification using pretrained and multitask learning on document and aspect-level data. Experimentation on a combination of the two approaches is done by obtaining the average of results from five runs using random weight for initialization. Evaluation of this approach including a series of ablation studies showed improvement over baselines. Chen and Qian [8] also developed a transfer learning method called *Transfer Capsule Network* (TransCap) which transfers document-level knowledge to aspect-level sentiment classifier. Their method involves what they referred to as *aspect routing* approach which encapsulates the sentence-level semantic representations into semantic capsules from both aspect-level and document-level data. They evaluated their model on SemEval-2014 Task 4 [32]. Kazmaier and van Vuuren [20] leverage unstructured, opinionated data combined with structured sources for sentiment classification. In a shift from general sentiment analysis, Bao et al. [6] leveraged lexical information incorporated into a baseline attention-based LSTM model for aspect-based sentiment analysis. This approach was introduced to enhance the flexibility and robustness of the model. In the first stage of their algorithm, a hybrid lexicon is built from the merger of four different lexicons. In the second stage, the lexical information obtained from the merger was first linearly transformed for dimensional compatibility and then the attention vector learned from the baseline LSTM model was applied to the transformed lexical information. In the final stage, a regularizer was applied to the attention vector to prevent potential overfitting. Experiments conducted on SemEval-2014 Task 4 [32] resulted in improvement over baselines. Li et al. [22] proposed a novel padding method for sentiment classification of user reviews. This is in contrast to the traditional zero padding technique. The underlying idea is to make input data consistent in size and to improve the composition of sentiment information in the reviews. Their method integrates lexicon into a combined model comprising convolutional neural networks (CNN) and conventional LSTM or Bidirectional LSTM (BiLSTM) in a two-channel mode. Experimental evaluation on Stanford Sentiment Treebank (SST) dataset [35] and Chinese tourism review dataset shows improvement over baselines of LSTM, BiLSTM, Dependency Tree-structured LSTM among others.

In order to address the issue of sparsity of training data and complex semantic features in opinionated texts, we propose a knowledge-enhanced deep neural architecture. Our algorithm includes a transfer learning module which enables knowledge transfer from a model pretrained on document-level data to the aspect-level model. The algorithm also includes joint learning module which enables the model to learn on both aspect-level and document-level data simultaneously. Other details are discussed in Section 3.

3. Model description

We describe the problem, the attention-based CNN-BiLSTM model, the transfer and joint learning techniques. The base CNN-BiLSTM architecture is presented in Figure 1. Detailed description of the components and their interrelationships are presented in Section 3.2 to 3.4.

r

3.1. Problem formulation

Given an opinionated sentence s consisting of n words $\left(s = \{w_1, w_2, ..., w_n\}\right)$ and an opinion target a consisting of a word or sequence of words of size m and occurring in sentence s, $\left(a = \{a_1, a_2, ..., a_m\}\right)$, aspect-level sentiment classification aims to determine the sentiment polarity $\boldsymbol{P}$ of the sentence s towards the opinion target a.

3.2. Convolutional neural networks

Convolutional neural network CNN [14] has become one of the most popular choice in the field of deep learning. Computer vision based on convolutional neural networks has enabled accomplishments that had been considered impossible in the past decades. These areas include face recognition, autonomous vehicles and intelligent medical treatment. They have also recently found applications in sequence modeling problems such as text classification, sentiment analysis, prediction tasks among others. Convolutional neural network is a kind of feedforward neural network that is able to extract features from data with convolution structures. Different from the traditional feature extraction methods, CNN does not need to extract features manually. The CNN component of our architecture uses a single-dimensional convolutional layer with m filters and k kernels. For each word w in sentence s, the word embeddings vector e_w from an embedding matrix $\boldsymbol{E} \in \mathbb{R}^{V \times d}$, (where V is the vocabulary size of the embedding matrix and d, the dimension) is fed into the convolutional layer and activated with Rectified Linear Unit (ReLU) activation. The output is then fed into the BiLSTM layer.

3.3. Bidirectional LSTM

The original long short-term memory LSTM [18] was developed to address the exploding and vanishing gradient problems inherent in traditional feed-forward neural networks. The LSTM architecture has three gates; an input gate i_t , a forget gate f_t, an output gate o_t. It also has a memory cell c_t which enables it to learn long-distance dependencies in a sequence and a hidden state h_t. The transition equations of the LSTM are as follows:

$$i_t = \sigma\Big(w_i \left[h_{t-1}, x_t\right] + b_i\Big) \tag{1}$$

$$f_t = \sigma\Big(w_f \left[h_{t-1}, x_t\right] + b_f\Big) \tag{2}$$

$$o_t = \sigma\Big(w_o \left[h_{t-1}, x_t\right] + b_o\Big) \tag{3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh\Big(w_c \left[h_{t-1}, x_t\right] + b_c\Big) \tag{4}$$

$$h_t = o_t \odot \tanh\left(c_t\right) \tag{5}$$

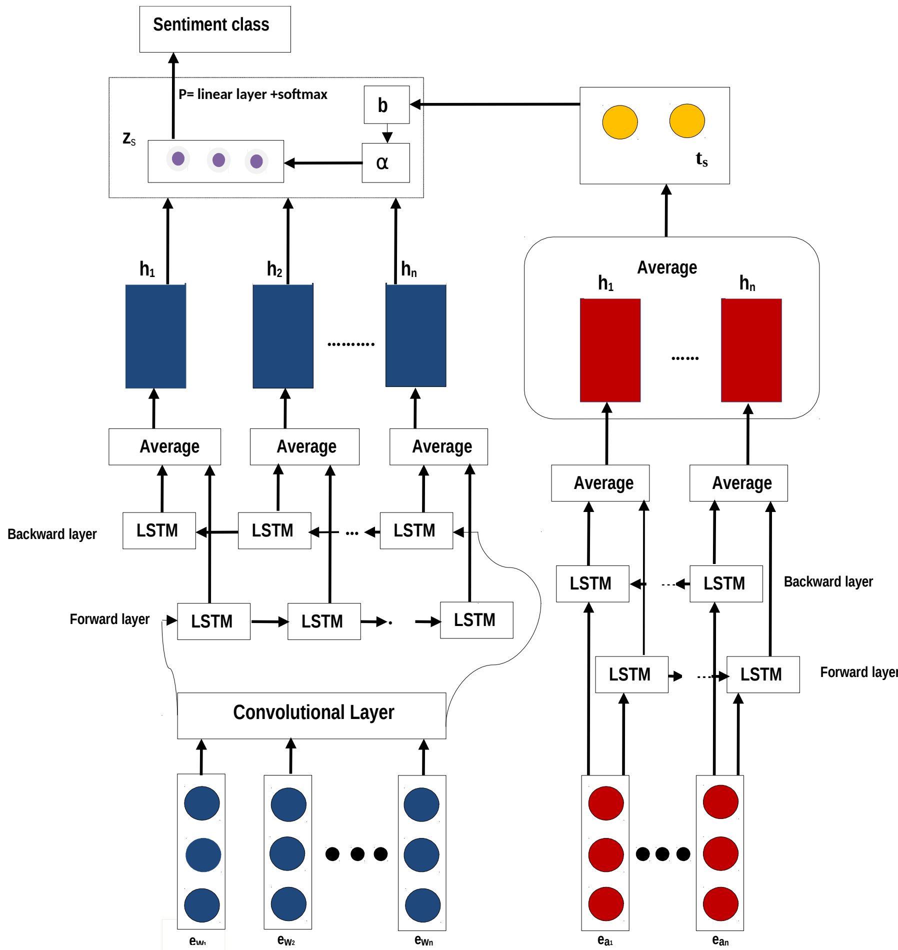

Figure 1: CNN-BiLSTM framework for aspect sentiment classification

where w_i, w_f and w_o are the weights of the neurons. b_i, b_f and b_o are the biases to be learned during training. σ denotes a logistic sigmoid function and $\odot$ denotes element-wise multiplication. $tanh()$ is a hyperbolic tangent function. Bidirectional LSTM [15] is a variant of the conventional LSTM which consists of two LSTMs that are run forward and backward simultaneously on the input sequence. The backward LSTM is used to capture the past contextual information while the forward LSTM is used to capture future contextual information. In our algorithm, the input sequences are either from the extracted feature sequence from the convolutional layer or directly from the word embeddings of the opinionated sentences. Therefore, the final hidden state is obtained using equation 6:

$$h_t = \mu\left(\overrightarrow{h_t}, \overleftarrow{h_t}\right) \tag{6}$$

where μ is the average of the two hidden states.

3.4. Attention-based BiLSTM

Following the work of [17], each opinionated text which comprises words w_i are represented by their word embeddings e_{w_i} still from $E \in \mathbb{R}^{V \times d}$, where V is the vocabulary size of the embedding matrix and d is the dimension.The bi-directional LSTM (BiLSTM) with trainable parameters θ_{bilstm} is used to capture the sequential and contextual information. The output are hidden representations of the input as presented in equation 7:

$$[h_1, ..., h_n] = Bi\text{LSTM}([e_{w_1}, ..., e_{w_n}], \theta_{bilstm}) \tag{7}$$

The final target-specific representation for a sentence s of length n is given by z as in equation 8:

$$z_s = \sum_{i=1}^{n} \alpha_i h_i \tag{8}$$

where α is an attention layer which assigns weight to each word in the sentence and h is the hidden representation of the sentence. α is computed as in equation 9:

$$\alpha = \frac{exp(b_i)}{\sum_{i=j}^{n} exp(b_j)} \tag{9}$$

$$b_i = f_{score}\left(h_i, t_s\right) \tag{10}$$

$$t_s = \frac{1}{m}\sum_{i=1}^{m} h_{a_i} \tag{11}$$

f_{score} is a function which computes the weight of each word w_i in sentence s based on the semantic and contextual relationships between h_i and t_s. t_s is the average of the hidden representation of the target.

The sentence representation z is fed into output layer using *softmax* activation function to predict the probability distribution P over sentiment labels as presented in equation 12.

$$P = softmax\left(w_o z + b_o\right) \tag{12}$$

where w_o and b_o are the output weight and bias respectively. The attention-based model is trained through cross-entropy minimization as given by equation 13:

$$J = -\sum_{i \in D} \log P_i\left(C_i\right) \tag{13}$$

where D is the training dataset, C_i is the true label for sample i and $P_i\left(C_i\right)$ is the probability of true label given sample i.

3.5. Learning techniques

The attention-based BiLSTM is the base model with parameters given in equation 14:

$$\theta_{aspect} = \left\{E, \theta_{bilstm}, w_a, w_o, w_b\right\} \tag{14}$$

where E, $bilstm$, w_a, w_o and w_b are embeddings, bidirectional LSTM, aspect attention, output and bias weights respectively.

3.5.1. Joint learning

In this technique, a model is jointly trained on the aspect-level data and document-level data with the embeddings and BiLSTM layers shared between the two tasks. A document is represented as an average vector over BiLSTM outputs. The loss function is a linear combination of losses from the individual tasks given by equation 15:

$$L = L_a + \lambda L_d \tag{15}$$

where $\lambda \in [0, 1]$ is a hyperparameter which controls the weight of L_d

3.5.2. Transfer learning

A model is first trained on document-level samples with the same parameters as that of the base model without the attention and aspect layers. The training objective is also cross entropy minimization. The pretrained weights of the word embeddings, BiLSTM and the output layers are initialized and used to fine-tune training on the aspect-level data. For the BiLSTM layer, the last hidden vector output is used as the document representation.

3.5.3. Joint and transfer learning

In this module, the joint and transfer learning techniques are combined with the pretrained weights from document-level data used to fine-tune training on both aspect-level and document-level data.

4. Experiments

4.1. Datasets description

Experiments were run on four benchmark datasets covering restaurant and laptop domains from SemEval-2014 Task 4 [32], SemEval-2015 task 12 [31] and SemEval-2016 task 5 [30]. The statistics of the datasets are shown in Table 1 as RES14 (restaurant domain of SemEval-2014 Task 4), LAP14 (laptop domain of SemEval-2014 Task 4), RES15 (restaurant domain of SemEval-2015 task 12) and RES16 (restaurant domain of SemEval-2016 task 5)

For the document-level dataset, we use the dataset derived by [17] from YELP2014 dataset [37] covering restaurant domain and Amazon electronics dataset [26] covering electronics domain. Each of the datasets contain 30,000 entries with equal number of polarity labels. The polarity labels are presented as ratings ranging from 1 to 5. In order to have a uniform labels as the aspect-level data, these ratings were converted to a 3-class classification with ratings < 3, $= 3$ and > 3 taken as negative, neutral and positive polarities respectively.

4.2. Experimental settings

The model was developed using Keras [10] with Tensorflow [1] as the backend. RMSprop [38] was chosen as the optimization algorithm with learning rate of $1\times e^{-3}$ and decay rate of 0.9. The model was regularized with a Dropout [36] probability of 0.2. The batch size for the document-pretrained and the aspect-level models are 25 and 16 respectively. Pretrained language models including ones based on word senses [5] and subword information [28] are usually employed in initializing deep learning models. In this case, in both the pretrained document-level and aspect-level models, a 300-dimension GloVe embeddings [29] (trained on 840 billions word tokens) was used to initialize words. The models were trained for 10 epochs to select the best results. λ is set to 1 for the adjustment of the loss function at the document-level. The aspect-level training data was split by random sampling into 20% validation set and 80% training set.

5. Results and discussions

The results obtained from CNN-BiLSTM and BiLSTM variants of our algorithm are presented in Table 2. It comprises results for the different learning techniques; CNN-BiLSTM and BiLSTM on only aspect data without joint and/or transfer learning (denoted CNN-BiLSTM(S) and BiLSTM(S) respectively), CNN-BiLSTM and BiLSTM with joint learning (denoted CNN-BiLSTM(J) and BiLSTM(J) respectively), CNN-BiLSTM and BiLSTM with transfer learning (denoted CNN-BiLSTM(T) and BiLSTM(T) respectively) and CNN-BiLSTM and BiLSTM with both joint and transfer learning (denoted

Table 1
Statistics of the experimental datasets

	Positive		Neutral		Negative	
Dataset	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
REST14	2164	728	637	196	807	196
LAP14	994	341	464	169	870	128
REST15	1178	439	50	35	382	328
REST16	1620	597	88	38	709	190

Table 2
Accuracy (Acc) (%) and macro-F1 (%) of the CNN-BiLSTM algorithm on the learning techniques with the best performance in bold

	RES14		LAP14		RES15		RES16	
Method	**Acc**	**Macro-F1**	**Acc**	**Macro-F1**	**Acc**	**Macro-F1**	**Acc**	**Macro-F1**
BiLSTM(S)	77.77	68.50	67.71	65.10	76.68	65.38	83.88	68.65
CNN-BiLSTM(S)	77.05	66.78	72.26	67.98	77.58	60.34	83.39	63.83
BiLSTM(J)	78.75	67.66	71.94	68.22	83.04	**71.13**	86.55	73.29
CNN-BiLSTM(J)	77.77	68.18	71.47	68.29	80.55	68.34	85.33	71.79
BiLSTM(T)	78.84	68.37	71.32	67.06	**83.29**	68.74	86.67	**74.44**
CNN-BiLSTM(T)	77.59	67.51	**72.41**	**68.95**	79.05	69.06	84.61	70.80
BiLSTM(J/T)	80.18	71.40	71.63	67.88	81.55	66.34	**87.27**	72.92
CNN-BiLSTM(J/T)	**81.96**	**74.16**	69.75	65.18	81.17	69.43	86.42	72.35

CNN-BiLSTM(J/T) and BiLSTM(J/T) respectively). The results reveals that CNN-BiLSTM has the best performance on RES14 using joint/transfer learning and on LAP14 using only aspect data though with better Macro-F1 using joint learning. The BiLSTM performs best on RES15 in terms of accuracy using transfer learning. However, in terms of Macro-F1, the joint learning outperforms others. On RES16, BiLSTM with joint/transfer learning has the best accuracy but has the best Macro-F1 with transfer learning. The model generally benefits from the joint and/or transfer learning.

Table 3 compares the performance of our model with other methods based on attention mechanisms and transfer/augmented learning for aspect-sentiment classification. A brief highlights of the compared related methods are as follow:

- **EDK**: EDK [17] enhanced aspect sentiment classification using an LSTM-based model through pretrained and multitask learning. It presented results of several variants obtained by calculating the average of five runs with random weight initializations. The variants with the best results are used for comparison.
- **PABS-BiGRU**: PABS-BIGRU [41] proposed a positionless attention-based sentiment classification using bidirectional recurrent gated units.
- **ATAE-LSTM**: ATAE-LSTM [39] computed attention weights by appending the aspect vector into the input word vectors.
- **IAN**: IAN [25] modeled opinion targets and their context separately through interactive attention networks.
- **RAM**: RAM [7] proposed a multiple attention framework to detect the sentiment of opinion targets.
- **MGAN**: MGAN [12] developed a multiple-attention neural network model for identification of sentiment polarity in comments/reviews
- **AOA**: AOA [19] introduced a method based on neural networks which uses an attention-over-attention in learning.
- **TransCap**: TRANSCAP [8] transferred document-level knowledge to aspect-level sentiment classifier using *Transfer Capsule Network*.
- **IACapsNet**: IACAPSNET [11] proposed an interactive attention capsule network to tackle the issue of overlapping feature representation which usually results from multi-aspect sentences.

A juxtaposition of the performance of the compared methods reveals for RES14 dataset, that our method achieves state-of-the-art accuracy and Macro-F1 with 81.96% and

Table 3
Results comparison with baseline models. Results with † means the results were obtained directly as published while those without were obtained from open source codes. Best results in bold.

	RES14		LAP14		RES15		RES16	
Method	**Acc**	**Macro-F1**	**Acc**	**Macro-F1**	**Acc**	**Macro-F1**	**Acc**	**Macro-F1**
EDK[17]	79.11†	69.73†	71.32†	68.53†	81.30†	68.74†	85.58†	70.73†
PABS-BiGRU[41]	81.37†	72.06†	75.39†	70.50†	80.88†	62.48†	**89.30†**	66.93†
ATAE-LSTM[39]	77.20†	67.02	68.70†	63.93	78.48	60.53	83.77	61.71
IAN[25]	79.26	70.09	72.05	67.38	78.54	52.65	84.74	55.21
RAM[7]	80.23†	70.80†	74.49†	71.35†	79.98†	60.57†	83.88†	62.14†
MGAN[12]	81.25†	71.94†	75.39†	72.47†	79.36	57.26	87.06	62.29
AOA[19]	79.97	70.42	72.62	67.52	78.17	57.02	87.50	66.21
TransCap[8]	79.27†	70.85†	73.87†	70.10†	-	-	-	-
IACapsNet[11]	81.79†	73.40†	**76.80†**	**73.29†**	-	-	-	-
Ours	**81.96**	**74.16**	72.26	68.95	**83.29**	**71.13**	87.27	**74.44**

Table 4
Results of ablation studies with various document-level weights (by varying λ) on the datasets. The best indicated in bold

	RES14		LAP14		RES15		RES16	
λ	**Acc**	**Macro-F1**	**Acc**	**Macro-F1**	**Acc**	**Macro-F1**	**Acc**	**Macro-F1**
0.1	79.20	69.01	68.54	65.22	81.55	64.83	84.68	68.23
0.2	79.10	69.17	68.32	66.79	81.30	65.28	83.22	68.65
0.3	78.75	68.14	67.86	66.55	81.22	65.44	85.22	67.93
0.4	78.12	67.86	68.04	67.11	81.49	66.12	83.66	69.45
0.5	79.22	68.11	69.01	66.68	82.92	67.41	84.91	70.26
0.6	79.79	69.02	68.35	67.92	82.01	66.72	85.25	69.83
0.7	79.45	67.87	69.71	68.05	81.28	67.53	83.66	71.34
0.8	80.04	68.35	70.69	67.58	81.67	68.02	85.72	70.94
0.9	79.36	70.14	69.57	68.17	82.79	66.96	85.49	71.33
1.0	**81.96**	**74.16**	**72.26**	**68.95**	**83.29**	**71.13**	**87.27**	**74.44**

74.16% respectively. IACapsNet is the best performing on LAP14 dataset with accuracy of 76.80% and Macro-F1 of 73.29%. On RES15 dataset, our method outperforms other methods with accuracy and Macro-F1 of 83.29% and 71.13% respectively. PABS-BIGRU performs best on RES16 in terms of accuracy with 89.30% while our method is the best in Macro-F1 with 74.44%. On the laptop domain, our model did not generalize well and performs below some of the compared algorithms. This is because the document-level data comes from general electronics domain most of which are not specific about laptops. It is noteworthy that our method achieves a significant superior Macro-F1 on the RES15 and RES16 datasets with an average difference to trailing results of about 3%.

6. Ablation studies

We conduct ablation experiments to show the effect of different document-level weights on the best among joint, transfer and joint/transfer learning techniques by varying the values of λ from 0.1 to 0.9 compared with the main model which uses 1.0. The results of the experiments are presented in Table 4.

The results show diverse performance with various document-level weights on each of the datasets. However, λ value of 1 remains the optimal value, consistently producing the best performance across the datasets.

7. Conclusion

It is an indisputable fact that inadequate training data is still a major problem affecting the accuracy of aspect sentiment classification. Consequent upon this, we show in this work that aspect sentiment classification can significantly benefit from data augmentation transfer and/or joint learning on document and aspect-level sentiment data. Furthermore, we hypothesize and show that aspect sentiment classification can gain from effective representation of the complex semantic features in opinionated texts through the application of attention mechanism with deep neural networks. Experimental results, including those of ablation studies proves the effectiveness of our method.

References

[1] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. URL: https://www.tensorflow.org/. software available from tensorflow.org.

[2] Ayetiran, E.F., 2020. An index-based joint multilingual/cross-lingual text categorization using topic expansion via babelnet. Turkish Journal Electrical Engineering and Computer Sciences 28, 224–237. URL: https://doi.org/10.3906/elk-1901-140, doi:10.3906/elk-1901-140.

[3] Ayetiran, E.F., Agbele, K.K., 2018. An optimized lesk-based algorithm for word sense disambiguation. Open Computer Science 8, 165–172. URL: https://doi.org/10.1515/comp-2018-0015, doi:10.1515/comp-2018-0015.

[4] Ayetiran, E.F., Boella, G., Caro, L.D., Robaldo, L., 2015. Enhancing Word Sense Disambiguation Using A Hybrid Knowledge-Based Technique. De Gruyter. pp. 15–26. URL: https://doi.org/10.1515/9781501501289.15, doi:doi:10.1515/9781501501289.15.

[5] Ayetiran, E.F., Sojka, P., Novotný, V., 2021. EDS-MEMBED: multi-sense embeddings based on enhanced distributional semantic structures via a graph walk over word senses. Knowledge-Based Systems 219, 106902. URL: https://doi.org/10.1016/j.knosys.2021.106902, doi:10.1016/j.knosys.2021.106902.

[6] Bao, L., Lambert, P., Badia, T., 2019. Attention and lexicon regularized LSTM for aspect-based sentiment analysis, in: Alva-Manchego, F.E., Choi, E., Khashabi, D. (Eds.), Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28 - August 2, 2019, Volume 2: Student Research Workshop, Association for Computational Linguistics. pp. 253–259. URL: https://doi.org/10.18653/v1/p19-2035, doi:10.18653/v1/p19-2035.

[7] Chen, P., Sun, Z., Bing, L., Yang, W., 2017. Recurrent attention network on memory for aspect sentiment analysis, in: Palmer, M., Hwa, R., Riedel, S. (Eds.), Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017, Association for Computational Linguistics. pp. 452–461. URL: https://doi.org/10.18653/v1/d17-1047, doi:10.18653/v1/d17-1047.

[8] Chen, Z., Qian, T., 2019. Transfer capsule network for aspect level sentiment classification, in: Korhonen, A., Traum, D.R., Màrquez, L. (Eds.), Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers, Association for Computational Linguistics. pp. 547–556. URL: https://doi.org/10.18653/v1/p19-1052, doi:10.18653/v1/p19-1052.

[9] Chevalier, J.A., Mayzlin, D., 2006. The effect of word of mouth on sales: Online book reviews. Journal of Marketing Research 43(3), 345–354. doi:10.1509/jmkr.43.3.345.

[10] Chollet, F., et al., 2015. Keras. https://keras.io.

[11] Du, C., Sun, H., Wang, J., Qi, Q., Liao, J., Xu, T., Liu, M., 2019. Capsule network with interactive attention for aspect-level sentiment classification, in: Inui, K., Jiang, J., Ng, V., Wan, X. (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019, Association for Computational Linguistics. pp. 5488–5497. URL: https://doi.org/10.18653/v1/D19-1551, doi:10.18653/v1/D19-1551.

[12] Fan, F., Feng, Y., Zhao, D., 2018. Multi-grained attention network for aspect-level sentiment classification, in: Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J. (Eds.), Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018, Association for Computational Linguistics. pp. 3433–3442. URL: https://doi.org/10.18653/v1/d18-1380, doi:10.18653/v1/d18-1380.

[13] Fellbaum, C., 1998. WordNet: An Electronic Lexical Database. Bradford Books.

[14] Fukushima, K., Miyake, S., 1982. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. Pattern Recognition 15, 455–469. URL: https://doi.org/10.1016/0031-3203(82)90024-3, doi:10.1016/0031-3203(82)90024-3.

[15] Graves, A., Jaitly, N., Mohamed, A., 2013. Hybrid speech recognition with deep bidirectional LSTM, in: 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc, Czech Republic, December 8-12, 2013, IEEE. pp. 273–278. URL: https://doi.org/10.1109/ASRU.2013.6707742, doi:10.1109/ASRU.2013.6707742.

[16] He, R., Lee, W.S., Ng, H.T., Dahlmeier, D., 2018a. Effective attention modeling for aspect-level sentiment classification, in: Bender, E.M., Derczynski, L., Isabelle, P. (Eds.), Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018, Association for Computational Linguistics. pp. 1121–1131. URL: https://www.aclweb.org/anthology/C18-1096/.

[17] He, R., Lee, W.S., Ng, H.T., Dahlmeier, D., 2018b. Exploiting document knowledge for aspect-level sentiment classification, in: Gurevych, I., Miyao, Y. (Eds.), Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers, Association for Computational Linguistics. pp. 579–585. URL: https://www.aclweb.org/anthology/P18-2092/, doi:10.18653/v1/P18-2092.

[18] Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation 9, 1735–1780. URL: https://doi.org/10.1162/neco.1997.9.8.1735, doi:10.1162/neco.1997.9.8.1735.

[19] Huang, B., Ou, Y., Carley, K.M., 2018. Aspect level sentiment classification with attention-over-attention neural networks, in: Thomson, R., Dancy, C.L., Hyder, A., Bisgin, H. (Eds.), Social, Cultural, and Behavioral Modeling - 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings, Springer. pp. 197–206. URL: https://doi.org/10.1007/978-3-319-93372-6_22, doi:10.1007/978-3-319-93372-6_22.

[20] Kazmaier, J., van Vuuren, J.H., 2020. A generic framework for sentiment analysis: Leveraging opinion-bearing data to inform decision making. Decision Support Systems 135, 113304. URL: https://doi.org/10.1016/j.dss.2020.113304, doi:10.1016/j.dss.2020.113304.

[21] Li, W., Shao, W., Ji, S., Cambria, E., 2022. Bieru: Bidirectional emotional recurrent unit for conversational sentiment analysis. Neurocomputing 467, 73–82. URL: https://doi.org/10.1016/j.neucom.2021.09.057, doi:10.1016/j.neucom.2021.09.057.

[22] Li, W., Zhu, L., Shi, Y., Guo, K., Cambria, E., 2020. User reviews: Sentiment analysis using lexicon integrated two-channel cnn-lstm family models. Applied Soft Computing 94, 106435. URL: https://doi.org/10.1016/j.asoc.2020.106435, doi:10.1016/j.asoc.2020.106435.

[23] Liang, B., Su, H., Gui, L., Cambria, E., Xu, R., 2022. Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. Knowledge-Based Systems 235, 107643. URL: https://doi.org/10.1016/j.knosys.2021.107643, doi:10.1016/j.knosys.2021.107643.

[24] Liu, J., Zhang, Y., 2017. Attention modeling for targeted sentiment, in: Proceedings of the 15th Conference of the European Chapter of the

Association for Computational Linguistics: Volume 2, Short Papers, Association for Computational Linguistics, Valencia, Spain. pp. 572–577. URL: https://www.aclweb.org/anthology/E17-2091.

[25] Ma, D., Li, S., Zhang, X., Wang, H., 2017. Interactive attention networks for aspect-level sentiment classification, in: Sierra, C. (Ed.), Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, ijcai.org. pp. 4068–4074. URL: https://doi.org/10.24963/ijcai.2017/568, doi:10.24963/ijcai.2017/568.

[26] McAuley, J.J., Targett, C., Shi, Q., van den Hengel, A., 2015. Image-based recommendations on styles and substitutes, in: Baeza-Yates, R., Lalmas, M., Moffat, A., Ribeiro-Neto, B.A. (Eds.), Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015, ACM. pp. 43–52. URL: https://doi.org/10.1145/2766462.2767755, doi:10.1145/2766462.2767755.

[27] Meskele, D., Frasincar, F., 2020. Aldonar: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model. Information Processing & Management 57, 102211. URL: https://doi.org/10.1016/j.ipm.2020.102211, doi:10.1016/j.ipm.2020.102211.

[28] Novotný, V., Stefánik, M., Ayetiran, E.F., Sojka, P., Rehurek, R., 2022. When fasttext pays attention: Efficient estimation of word representations using constrained positional weighting. Journal Universal Computer Science 28, 181–201. URL: https://doi.org/10.3897/jucs.69619, doi:10.3897/jucs.69619.

[29] Pennington, J., Socher, R., Manning, C.D., 2014. Glove: Global vectors for word representation, in: Moschitti, A., Pang, B., Daelemans, W. (Eds.), Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, ACL. pp. 1532–1543. URL: https://doi.org/10.3115/v1/d14-1162, doi:10.3115/v1/d14-1162.

[30] Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., Clercq, O.D., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N.V., Kotelnikov, E.V., Bel, N., Zafra, S.M.J., Eryigit, G., 2016. Semeval-2016 task 5: Aspect based sentiment analysis, in: Bethard, S., Cer, D.M., Carpuat, M., Jurgens, D., Nakov, P., Zesch, T. (Eds.), Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016, The Association for Computer Linguistics. pp. 19–30. URL: https://doi.org/10.18653/v1/s16-1002, doi:10.18653/v1/s16-1002.

[31] Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., Androutsopoulos, I., 2015. Semeval-2015 task 12: Aspect based sentiment analysis, in: Cer, D.M., Jurgens, D., Nakov, P., Zesch, T. (Eds.), Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, Denver, Colorado, USA, June 4-5, 2015, The Association for Computer Linguistics. pp. 486–495. URL: https://doi.org/10.18653/v1/s15-2082, doi:10.18653/v1/s15-2082.

[32] Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., 2014. Semeval-2014 task 4: Aspect based sentiment analysis, in: Nakov, P., Zesch, T. (Eds.), Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014, Dublin, Ireland, August 23-24, 2014, The Association for Computer Linguistics. pp. 27–35. URL: https://doi.org/10.3115/v1/s14-2004, doi:10.3115/v1/s14-2004.

[33] Schouten, K., Frasincar, F., 2016. Survey on aspect-level sentiment analysis. IEEE Transactions on Knowledge and Data Engineering 28, 813–830. URL: https://doi.org/10.1109/TKDE.2015.2485209, doi:10.1109/TKDE.2015.2485209.

[34] Shuang, K., Ren, X., Yang, Q., Li, R., Loo, J., 2019. Aela-dlstms: Attention-enabled and location-aware double lstms for aspect-level sentiment classification. Neurocomputing 334, 25–34. URL: https://doi.org/10.1016/j.neucom.2018.11.084, doi:10.1016/j.neucom.2018.11.084.

[35] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C., 2013. Recursive deep models for semantic compositionality over a sentiment treebank, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL, ACL. pp. 1631–1642. URL: https://www.aclweb.org/anthology/D13-1170/.

[36] Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. Journal Machine Learning Research 15, 1929–1958. URL: http://dl.acm.org/citation.cfm?id=2670313.

[37] Tang, D., Qin, B., Liu, T., 2015. Learning semantic representations of users and products for document level sentiment classification, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers, The Association for Computer Linguistics. pp. 1014–1023. URL: https://doi.org/10.3115/v1/p15-1098, doi:10.3115/v1/p15-1098.

[38] Tieleman, T., Hinton, G., 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural networks for machine learning 4, 26–31.

[39] Wang, Y., Huang, M., Zhu, X., Zhao, L., 2016. Attention-based LSTM for aspect-level sentiment classification, in: Su, J., Carreras, X., Duh, K. (Eds.), Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016, The Association for Computational Linguistics. pp. 606–615. URL: https://doi.org/10.18653/v1/d16-1058, doi:10.18653/v1/d16-1058.

[40] Wu, S., Xu, Y., Wu, F., Yuan, Z., Huang, Y., Li, X., 2019. Aspect-based sentiment analysis via fusing multiple sources of textual knowledge. Knowledge-Based Systems 183. URL: https://doi.org/10.1016/j.knosys.2019.104868, doi:10.1016/j.knosys.2019.104868.

[41] Yadav, R.K., Jiao, L., Goodwin, M., Granmo, O.C., 2021. Positionless aspect based sentiment analysis using attention mechanism. Knowledge-Based Systems 226, 107136. URL: https://www.sciencedirect.com/science/article/pii/S0950705121003993, doi:https://doi.org/10.1016/j.knosys.2021.107136.

[42] Yang, C., Zhang, H., Jiang, B., Li, K., 2019. Aspect-based sentiment analysis with alternating co-attention networks. Information Processing & Management 56, 463–478. URL: https://doi.org/10.1016/j.ipm.2018.12.004, doi:10.1016/j.ipm.2018.12.004.

[43] Yoo, K.H., Gretzel, U., 2008. What motivates consumers to write online travel reviews? Journal of Information Technology & Tourism 10(4), 283–295. URL: https://doi.org/10.3727/109830508788403114, doi:10.3727/109830508788403114.

[44] Zhou, J., Chen, Q., Huang, J.X., Hu, Q.V., He, L., 2020. Position-aware hierarchical transfer model for aspect-level sentiment classification. Information Sciences 513, 1–16. URL: https://doi.org/10.1016/j.ins.2019.11.048, doi:10.1016/j.ins.2019.11.048.