Comparative Study of CNN and LSTM based Attention Neural Networks for Aspect-Level Opinion Mining

Wei Quan*, Zheng Chen*, Jianliang Gao† Xiaohua Tony Hu*
* *College of Computing and Informatics, Drexel University, Philadelphia, USA*
† *School of Information Science and Engineering, Central South University, Changsha, China*
wq34@drexel.edu, zc86@drexel.edu, gaojianliang@csu.edu.cn, xh29@drexel.edu

***Abstract*—Aspect-level opinion mining aims to find and aggregate opinions on opinion targets. Previous work has demonstrated that precise modeling of opinion targets within the surrounding context can improve performances. However, how to effectively and efficiently learn hidden word semantics and better represent targets and the context still needs to be further studied. In this paper, we propose and compare two interactive attention neural networks for aspect-level opinion mining, one employs two bi-directional Long-Short-Term-Memory (BLSTM) and the other employs two Convolutional Neural Networks (CNN). Both frameworks learn opinion targets and the context respectively, followed by an attention mechanism that integrates hidden states learned from both the targets and context. We compare our model with state-of-the-art baselines on two SemEval 2014 datasets[1]. Experiment results show that our models obtain competitive performances against the baselines on both datasets. Our work contributes to the improvement of state-of-the-art aspect-level opinion mining methods and offers a new approach to support human decision-making process based on opinion mining results. The quantitative and qualitative comparisons in our work aim to give basic guidance for neural network selection in similar tasks.**

***Keywords*-Opinion Mining; Sentiment Analysis; Deep Learning;**

I. Introduction

We live in this digital age, which is also referred to as the information society, surrounded by ever growing volumes of data. The past decade has witnessed the rapid growth of online platforms driven by the current generation of Web applications, the nearly limitless connectivity, and the insatiable desire for sharing information [1]. People are provided with new platforms to efficiently create and share information, which has changed the way we communicate with each other [2], and we spend a significant amount of time on social media [3]. We are constantly sharing our opinions, reviews, and preferences with the rest of the world, which leads to the explosion of user-generated contents. Thus, there are valuable, vast and unstructured information about public opinions online [4]. This type of contents is ever more recognized as a source of data that has added value for multiple application domains [1], such as shopping, entertainment, education, etc.

Before making a decision, we often refer to others' opinions. We need fast, accurate and concise information and to integrate diverse experiences and knowledge from other people to make decisions, since "what other people think" has always been an important piece of information during the decision-making process [5]. For instance, we often ask questions like which features of a product are favored or disfavored, why do people prefer one product over another, and what do people think about the new tax policy, etc. Brightlocal's online review survey revealed that every nine out of ten people read online reviews and 88% of them trust online reviews to make their decisions [6]. Opinion not only has a great impact on and provides guidance for individuals, but also for organizations, social communities and governments [7]. For ages, governments, organizations, and companies have been struggling to determine opinions of their target communities and audiences through surveys, questionnaires, and other market research [1]. They want to know how their customers feel about these products, and this information can be acquired by studying opinions from their customers. For example, in commercial airline field, social media has become the preferred method for customers to interact with airlines to acquire latest information, ease frustration, and solve issues, etc. By analyzing customers opinion on social media towards airlines, it offers a new approach to evaluate airline service quality. Now, for the first time, nearly immediate feedback on products, stocks, and policies, etc., and many other desired data is now readily available online, which was hard to obtain in the past [1]. It has been seen that opinions are getting popular day by day, and these opinions represent wealth of information which can be beneficial for the industry as well as for the consumers [8].

However, looking through online contents manually is a mundane and time-consuming job for both individuals and organizations, and unlike factual text, sentiment/opinionated text can often be expressed in a more subtle or arbitrary manner, which makes it difficult to be identified by simply looking at each constituent word in isolation. Thus, the need for opinion mining grows substantially in this era, since

[1]Detailed information of SemEval 2014 datasets can be found here: http://alt.qcri.org/semeval2014/task4/

978-1-5386-5035-6/18/$31.00 ©2018 IEEE

it aims to extract opinions from user generated contents (such as reviews) and present them to the users in a user-friendly manner [8]. Opinion mining is described as a recent discipline encompassing information retrieval, text mining, and computational linguistics, that tries to detect opinions expressed in natural language [5]. In general, opinion mining tasks are categorized into subjectivity and polarity classification, opinion target identification, opinion source identification, and opinion summarization [4], [9]. The goal of opinion mining is to create a knowledge base containing opinions in a more structured and explicit form. More specifically, it is to identify emerging societal trends based on views, dispositions, moods, attitudes and expectations of stakeholder groups or the general public [10]. Opinion mining has now raised growing interests both in the business world and within the scientific community [11]. There are many exciting open challenges from various domains including finance, politics, health, manufacturing, supply chain applications, and human communication comprehension, etc.

Aspect-level opinion mining aims to find and aggregate opinion expression/sentiment on opinion targets or aspects within documents [1]. For example, the following review "I ordered this product a week ago and it is a fantastic product; however, the one that was shipped to me barely works. When it works, it does exactly what its supposed to do and also charges quickly. But it only works maybe 20% of the time that I have tried to use it." evaluates the product from several aspects. The goal of aspect-level opinion mining here is to identify every aspect of the product and the polarity of each aspect. It is a fine-grained method that can be applied in various domains. Aspect-level opinion mining is challenging because it is difficult to model semantic relationships between the opinion targets or aspects and the surrounding context [12], [13], [14]. Previous studies focus on three major tasks: represent the context of a target, generate a target representation, and identify the important sentiment context for the specified target [12]. Recently, researchers have proposed many effective neural networks to deal with aspect-level opinion mining: Tang et al. [15] designed target-dependent LSTM and target-connection LSTM by regarding any given target as a feature and concatenating it with the context features; Ruder et al. [16] proposed a hierarchical and bi-directional LSTM to leverage language independence within a review; Wang et al. [17] designed an attention-based LSTM to enforce the model with attention to the important part of a sentence; Yang et al. [18] proposed two attention based bi-directional LSTMs; Liu et al. [19] differentiated attentions from left and right context; Ma et al. [20] implemented an interactive attention network to integrate attentions from both target and context. They obtained promising results with the development of deep learning techniques.

In this paper, we propose and compare two aspect-level opinion mining frameworks. Specifically, in the first framework bi-directional attention neural networks (BANN), we first implement two bi-directional Long-Short-Term-Memory (BLSTM) to model the opinion targets and the context respectively; then, we employ an attention mechanism to integrate hidden states learned from the targets and context to better represent them, and we concatenate both the target and context representations to predict aspect-level opinion polarity [21]. In the second framework convolutional attention neural networks (CANN), we implement two Convolutional Neural Networks (CNN) to model the opinion targets and the context respectively; then, we employ an attention mechanism to integrate feature maps learned from the targets and context to better represent them, and we concatenate both the target and context representations to predict aspect-level opinion polarity. Finally, we compare our models with state-of-the-art baselines on two SemEval 2014 task datasets. Results show that our models obtain competitive performances against the baselines on both datasets. In particular, our contribution is in three-fold:

1. We propose two interactive attention neural networks for aspect-level opinion mining that improve the state-of-the-art methods;
2. We compare proposed frameworks from quantitative and qualitative perspective to uncover when each framework performs better, aiming to give basic guidance for neural network selection in similar tasks;
3. We offer a new approach to support human decision-making process based on opinion mining results.

II. Related Work

A. Opinion Mining

Opinion mining has raised growing interests both in industry and academia in the past few years. It is described as a recent discipline encompassing information retrieval, text mining, and computational linguistics, that tries to detect opinions expressed in natural language [5]. An opinion is the private state of an individual, and as such. It represents individual's ideas, beliefs, assessments, judgements and evaluations about a specific subject, topics or item [9]. An opinion has three main components: the opinion holder or source of the opinion, the object about which opinion is expressed and the evaluation, view or appraisal [9]. In general, opinion mining tasks are categorized into subjectivity and polarity classification, opinion target identification, opinion source identification, and opinion summarization [4], [9]. The goal of opinion mining is to create a knowledge base containing opinions in a more structured and explicit form. More specifically, it is to identify emerging societal trends based on views, dispositions, moods, attitudes and expectations of stakeholder groups or the general public [10]. The growth of this field has resulted in the emergence of various subfields, each of which addressing a different level of analysis [1], including health, bioinformatics, and medical domains, business, political science, etc.

B. Aspect-level opinion mining

Aspect-level opinion mining aims to find and aggregate opinion expression (sentiment) on opinion targets within documents [1]. An opinion target can be a product aspect or entity and is also referred to as aspect in some literature. For example, if we want to mine opinions about the target "battery" of a smart phone and want to find out all related aspects and opinion terms mentioned in the reviews. We may find aspect words like "battery life", "size of the battery", and "charging time", and opinion related words like "long", "small" and "fast". It is a fine-grained method that can be applied to various domains. Aspect-level opinion mining is challenging because it is difficult to model semantic relationships between the opinion targets and the surrounding context [12]. According to [11], there are three research directions: aspect term extraction, categorizing given aspect terms, and aspect term sentiment classification. Earlier works mostly employed dictionary-based methods, while recent works mostly apply machine learning-based feature engineering and classification.

C. Deep Learning

Deep Learning is a subfield of machine learning which allows machines to learn from experience and understand the world in terms of hierarchy of concepts, with each concept defined in terms of its relation to simpler concepts [22]. It has become more and more useful in recent years as the volume of available data has increased, and it can be applied to various domains, such as image segmentation, object detection, video classification, speech recognition, reinforcement Learning, and robotics, etc. Thus, they attract much research interest in recent years, and achieve state-of-the-art performances in many fields including sentiment classification [23]. In the following section, we will summarize current state-of-the-art deep learning methods for opinion mining in two main categories: CNN and RNN.

Convolutional Neural Networks (CNN) is comprised of one or more convolutional layers, often with a subsampling step, and then followed by one or more fully connected layers as in a standard multilayer neural network. The architecture of a CNN is originally designed to take advantage of the 2D structure of an input image. This is achieved with local connections and tied weights followed by some form of pooling which results in translation invariant features. Another advantage is that CNN is easier to train and has fewer parameters than fully connected networks with the same number of hidden units. It was initially designed to solve problems in image-related area, and it turned out to be the natural choice given the effectiveness in computer vision tasks [24], [25], [26]. However, in recent works, we noticed the growing applications of CNN in text classification [27]. Collobert and Weston [28] are among the first attempts to apply CNN for sentence modeling. They used multi-task learning to output multiple predictions for NLP tasks. In [29], Collobert et al., further extended their work to propose a general CNN-based framework with pre-trained word embeddings to solve a plethora of NLP tasks. Another work that leads to a huge proliferation of CNN-based method is Kim [30]. In his work, a simple CNN with one layer of convolution was trained to produce feature maps. He reported the experiments and showed the model have excellent performance in multi-class text classification, which now is often used as a strong baseline in these tasks. Kalchbrenner et al. [31] also proposed a Dynamic Convolutional Neural Network (DCNN) for the semantic modelling of sentences. More recent work includes: Lei et al. [32] appealed to tensor algebra and used low-rank n-gram tensors to revise the temporal convolution operation in CNNs to better adapt it to text processing; Johnson and Zhang [33] directly applied CNN to high-dimensional text data to exploit the word order of text data for accurate prediction; They [34] further proposed a low-complexity word-level deep CNN architecture that can efficiently represent long-range associations in text; Wang et al. [35] introduced a densely connected CNN with multi-scale feature attention for text classification.

Recurrent neural network (RNN) makes connections between nodes to form a directed graph along a sequence, which allows to exhibit dynamic temporal behavior for a time sequence. Unlike feedforward neural networks, RNN can use the internal state or memory to process sequences of inputs, which makes them applicable to tasks such as speech recognition. Previous studies focus on three major tasks: represent the context of a target, generate a target representation, and identify the sentiment for a specific target [12]. Recently, most state-of-the-art studies use LSTM network and attention mechanism as the basic module in the methods. Tai, Socher, and Manning [36] introduced a tree LSTM for improving semantic representations. Tang et al. [15] designed target-dependent LSTM and target-connection LSTM by regarding any given target as a feature and concatenating it with the context features. Ruder et al. [16] proposed a hierarchical and bi-directional LSTM to leverage language independence within a review. Wang et al. [17] designed an attention-based LSTM to enforce the model with attention to the important part of a sentence. Yang et al. [18] proposed two attention based bi-directional LSTMs. Liu et al. [19] differentiated attentions from left and right context. Ma et al. [20] implemented an interactive attention network to integrate attentions from both target and context. They obtained promising results with the development of deep learning techniques.

Despite these advances on aspect-level opinion mining, previous works still suffer from drawbacks include: assumption of equal importance of multiple instances, some instances are often more related to sentiment; the way how attention mechanism and deep neural network work [37]. Thus, when applying deep neural networks for aspect-level opinion mining, how to effectively and efficiently learn hidden word semantics and better represent targets and the context still needs to be further studied. Qualitative analysis of experimental results is also needed to uncover the black box of deep neural networks.

III. Proposed Method

In this section, we first introduce the architectures of bi-directional attention neural networks (BANN) and convolutional attention neural networks (CANN) for aspect-level opinion mining. Then, we explain the math notations, equations and training details of our models.

A. Overall Framework

The overview of two frameworks are shown in Figure 1 and Figure 2. Both BANN and CANN take word embeddings as inputs. Then, BANN uses two BLSTM to learn hidden word semantics to model opinion targets and context, while CANN uses two CNN to learn hidden word semantics to model opinion targets and context. In both frameworks, attention vectors are generated with an attention mechanism by taking into account the average values of target hidden states and context hidden states respectively. In this way, target representations are obtained by using attention vectors from context to target; and context representations are obtained by using attention vectors from target to context. Finally, we concatenate target representations and context representations as the final representation. Then, we pass it into a softmax layer and predict the opinion polarity.

B. Model Description

This section describes two models in detail with mathematic notations and equations.

1) BANN: First, we use w to denote a specific word. [w_t^1, w_t^2, ..., w_t^m] is an opinion target with m words, and [w_c^1, w_c^2, ..., w_c^n] is the surrounding opinion context with n words. We use word embedding to represent a word. Thus, we can get $w^k \in R^d$, where k is the word index, R is the real numbers, and d is the embedding dimension.

Then, we use the BLSTM to model opinion target and context in the review. After concatenating the forward and backward outputs, we can get hidden states [h_t^1, h_t^2, ..., h_t^m] for target, and [h_c^1, h_c^2, ..., h_c^n] for context, where h denotes the hidden state.

We employ the attention mechanism to emphasize important information that contributes to opinion polarity classification. The inputs are average pooling of target and context hidden states:

$$t_{\text{pool}} = \sum_{i=1}^{m} h_t^i/m \tag{1}$$

$$c_{\text{pool}} = \sum_{i=1}^{n} h_c^i/n \tag{2}$$

We define a score function f that calculates the importance of a hidden state in the target or context:

$$f(h_t^i, c_{\text{pool}}) = tanh(h_t^i \cdot W_t \cdot {c_{\text{pool}}}^T + b_t) \tag{3}$$

$$f(h_c^i, t_{\text{pool}}) = tanh(h_c^i \cdot W_c \cdot {t_{\text{pool}}}^T + b_c) \tag{4}$$

where W_t, W_c, b_t, b_c are weight matrices and bias respectively, and $tanh$ is a non-linear function.

Then, we generate the attention vectors α_i and β_i:

$$\alpha_i = exp(f(h_t^i, c_{\text{pool}}))/\sum_{j=1}^{m} exp(f(h_t^j, c_{\text{pool}})) \tag{5}$$

$$\beta_i = exp(f(h_c^i, t_{\text{pool}}))/\sum_{j=1}^{n} exp(f(h_c^j, t_{\text{pool}})) \tag{6}$$

where α_i is the attention vector for target using average pooling from context representations and β_i is the attention vector for context using average pooling from target representations.

Thus, we can get new weighted target representations t_r and context representations c_r by:

$$t_r = \sum_{i=1}^{m} \alpha_i h_t^i \tag{7}$$

$$c_r = \sum_{i=1}^{n} \beta_i h_c^i \tag{8}$$

Finally, we concatenate t_r and c_r as the final representation vector f_r, and feed it to a softmax function for aspect-level opinion polarity prediction:

$$\hat{y} = softmax(W_f \cdot f_r + b_f) \tag{9}$$

where W_f and b_f are weight matrix and bias of the softmax function.

2) CANN : First, we use w to denote a specific word. [w_t^1, w_t^2, ..., w_t^m] is an opinion target with m words, and [w_c^1, w_c^2, ..., w_c^n] is the surrounding opinion context with n words. We use word embedding to represent a word. Thus, we can get $w^k \in R^d$, where k is the word index, R is the real numbers, and d is the embedding dimension. We can get W_t by simply concatenating opinion target words:

$$W_t = w_t^1 \oplus w_t^2 \oplus ... \oplus w_t^m \tag{10}$$

where $\oplus$ is the concatenation operation. Similarly, we can get W_c by simply concatenating opinion context words:

$$W_c = w_c^1 \oplus w_c^2 \oplus ... \oplus w_c^m \tag{11}$$

Then, we use the CNN to model opinion target and context in the review. The function of convolution layers is to extract higher level features from the input matrix. After

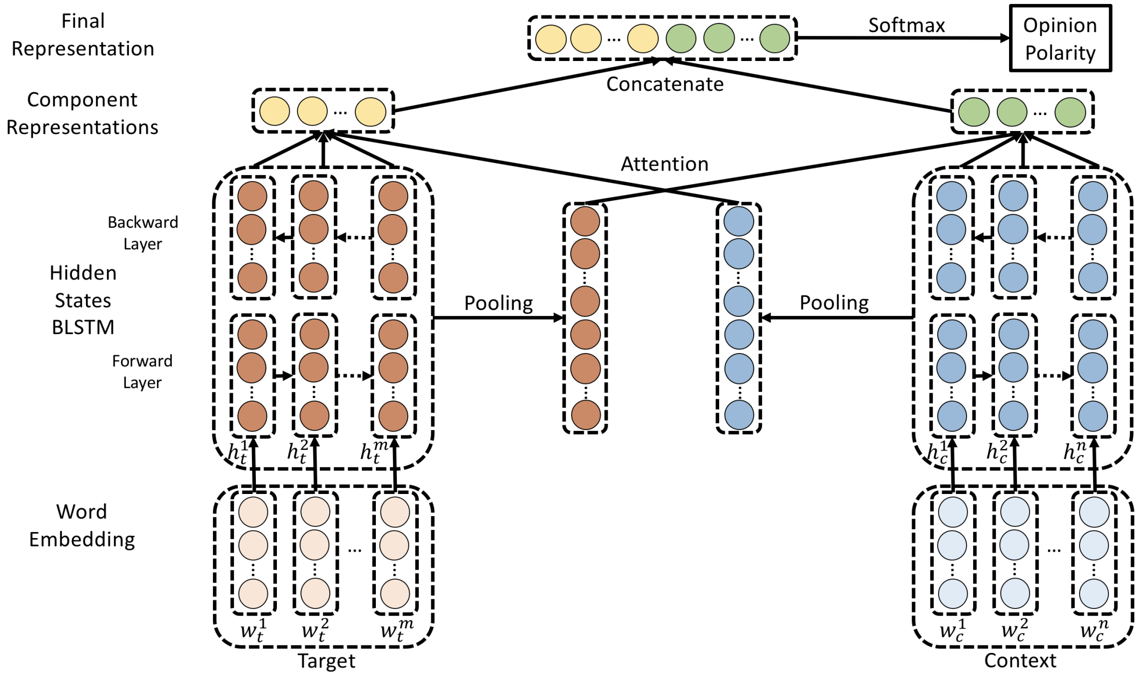

Figure 1. Overview of BANN framework.

a simple CNN with one layer of convolution on top of word embeddings, we can get feature maps [s_t^1, s_t^2, ..., s_t^m] for target, and [s_c^1, s_c^2, ..., s_c^n] for context, where s denotes the feature generated from a window of words h, and h is the size of the filter.

We employ the attention mechanism to emphasize important information that contributes to opinion polarity classification. The inputs are average pooling of target and context hidden states:

$$t_{\text{pool}} = \sum_{i=1}^{m} s_t^i / m \tag{12}$$

$$c_{\text{pool}} = \sum_{i=1}^{n} s_c^i / n \tag{13}$$

We define a score function f that calculates the importance of a hidden state in the target or context:

$$f(s_t^i, c_{\text{pool}}) = tanh(s_t^i \cdot W_t \cdot {c_{\text{pool}}}^T + b_t) \tag{14}$$

$$f(s_c^i, t_{\text{pool}}) = tanh(s_c^i \cdot W_c \cdot {t_{\text{pool}}}^T + b_c) \tag{15}$$

where W_t, W_c, b_t, b_c are weight matrices and bias respectively, and $tanh$ is a non-linear function.

Then, we generate the attention vectors α_i and β_i:

$$\alpha_i = exp(f(s_t^i, c_{\text{pool}})) / \sum_{j=1}^{m} exp(f(s_t^j, c_{\text{pool}})) \tag{16}$$

$$\beta_i = exp(f(s_c^i, t_{\text{pool}})) / \sum_{j=1}^{n} exp(f(s_c^j, t_{\text{pool}})) \tag{17}$$

where α_i is the attention vector for target using average pooling from context representations and β_i is the attention vector for context using average pooling from target representations.

Thus, we can get new weighted target representations t_r and context representations c_r by:

$$t_r = \sum_{i=1}^{m} \alpha_i s_t^i \tag{18}$$

$$c_r = \sum_{i=1}^{n} \beta_i s_c^i \tag{19}$$

Finally, we concatenate t_r and c_r as the final representation vector f_r, and feed it to a softmax function for aspect-level opinion polarity prediction:

$$\hat{y} = softmax(W_f \cdot f_r + b_f) \tag{20}$$

where W_f and b_f are weight matrix and bias of the softmax function.

C. Model Training

We train our model by minimizing the cross entropy error of opinion polarity classification. The loss function is defined as:

$$L = -\sum_{i=1}^{C} y_i log(\hat{y}_i) + \lambda \|\Theta\|^2 \tag{21}$$

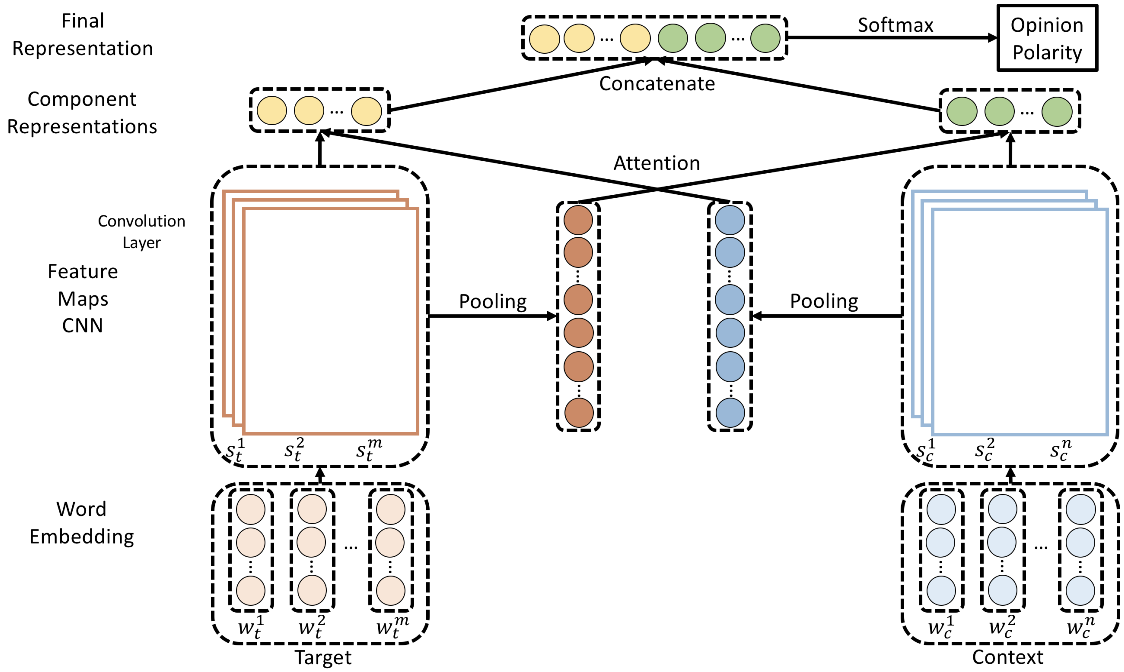

Figure 2. Overview of CANN framework.

where C is the number of class labels, y_i is the one-hot vector for the i-th class ground truth, $\hat{y}_i$ is the predicted probability for the i-th class, λ is the weight of L_2 regularization, Θ is the parameter set from BLSTM networks, attention layer score function f, and softmax layer.

We use back propagation to compute and update Θ, and dropout strategy to avoid overfitting.

IV. Experiments

A. Datasets

We use two datasets from SemEval 2014 Task 4 to evaluate two proposed methods. The first is a laptop review dataset and the second is a restaurant review dataset.

1) Laptop reviews: This dataset consists of over 3000 English sentences extracted from customer reviews of laptops. Experienced human annotators tagged the aspect terms of the sentences and their polarities.

2) Restaurant reviews: This dataset consists of over 3000 English sentences from the restaurant reviews of [38]. The original dataset of [38] included annotations for coarse aspect categories and overall sentence polarities. This dataset was modified to include annotations for aspect terms occurring in the sentences, aspect term polarities, and aspect category-specific polarities for SemEval 2014 Task 4. Experienced human annotators identified the aspect terms of the sentences and their polarities. Additional restaurant reviews, not in the original dataset of [38], were being annotated in the same manner, and used as test data.

Reviews from both datasets are labeled with three opinion polarities: positive, neutral and negative. Table I shows detailed information of two datasets. We use classification accuracy as the metric to evaluate the performance of aspect-level opinion polarity classification.

Table I
Detailed information of two datasets

Dataset		Positive	Neutral	Negative
Laptop	Train	994	464	870
	Test	341	169	128
Restaurant	Train	2164	637	807
	Test	728	196	196

B. Experimental Setup

We use TensorFlow to implement our models. We use GloVe2 [39] to initialize word embedding of opinion targets and context in our experiments. All out-of-vocabulary words and weight matrices are initialized by randomly sampling from the uniform distribution $U(-0.1, 0.1)$, and all bias are set to zeros. The dimension of word embeddings is 300 as in previous studies [17], [20].

In BANN framework, the dimension of BLSTM hidden states is 600 because we concatenate the forward and backward outputs. In CANN framework, the window size is 5 and the feature dimension is 300. We also set stride as 1 and use same padding.

In the parameter training of both models, the learning rate is set to 0.01, the coefficient of $L2$ regularization is set to

10^{-5}, and the dropout rate is set to 0.5. We train both models using Adam optimizer with momentum of 0.9.

C. Experimental Results

We compare our models with seven baselines to evaluate the performance (see Table II). In Table II, we report polarity classification accuracy on two datasets.

Table II
MODEL PERFORMANCE COMPARISON WITH SEVEN BASELINES

	Laptop	Restaurant
Majority	65.00%	53.50%
LSTM	66.50%	74.30%
CNN	69.50%	77.48%
TD-LSTM	68.10%	75.60%
AE-LSTM	68.90%	76.60%
ATAE-LSTM	68.70%	77.20%
IAN	72.10%	78.60%
BANN[2]	**73.51%**	**80.71%**
CANN[3]	69.75%	78.04%

The baselines are introduced as:

- **Majority** assigns the majority polarity in the training set to each sample in the test set;
- **LSTM** employs a standard LSTM for aspect-level classification task [17];
- **CNN** uses a one-layer CNN for text classification proposed by [30];
- **TD-LSTM** uses two LSTMs to model the left and right context with target respectively [15];
- **AE-LSTM** represents target with aspect embeddings to supervise the generation of attention vectors [17], and feeds the embeddings to LSTM;
- **ATAE-LSTM** which was developed on AE-LSTM, appends target embeddings to hidden vectors, using attention mechanism to obtain weights of hidden vectors [17];
- **IAN** applies two LSTMs to learn hidden states of target and context separately, and generates attention vectors from target and context hidden states. In this way, target and context representations are obtained by using attention vectors from context to target and from target to context respectively [20].

D. Results Analysis

The performances of all methods on two datasets are summarized in Table II. Each line lists the accuracy of each method on a specific dataset, where the best score is in bold.

Table II shows our models outperform the seven baseline methods on both datasets. The Majority method has the lowest accuracy, which means 65% and 53.5% of both test sets have the majority polarity from the training sets in the Laptop and Restaurant categories respectively.

[2]BANN achieves the best performances on both datasets

[3]CANN outperforms all baseline models expect IAN.

All LSTM based models have better performances than the Majority method, which indicates that LSTM can learn better representations and improve performances for polarity classification tasks. Among the six LSTM based models, the standard LSTM performs the worst, as opinion target and context are equally treated in this model. When we model left and right context with targets as in TD-LSTM, we can see the accuracy increase in both datasets by 1 to 2%. The use of attention mechanisms in AE-LSTM and ATAE-LSTM brings more improvement. Attention mechanisms help to capture the important information in the context and generate better representations for aspect-level opinion polarity classifation. The IAN model uses two attention networks to model opinion targets and context separately and learn attentions from them interactively, which stably exceeds both AE-LSTM and ATAE-LSTM models. Thus, adopting attention mechanisims to deal with opinion targets and context can improve classification results. On both datasets, our BANN model outperforms the state-of-the-art framework IAN, which reinforces our hypothesis that the capability of capturing information from the past and future states is especially useful for this task.

Compared with other LSTM baselines, we can see that the CNN model outperforms all expect for IAN. In both datasets, it beats the LSTM model by 3.00% to 3.18% in accuracy. Even without adding attention layers, the CNN model performs remarkably well and outperforms most state-of-the-art methods, which is in line with the finding from [30]. Also, it demonstrates that convolution layer can effectively represent the contents and has the ability to acquire richer and complex features compared with other learning approaches. While CANN with additional attention mechanisms performs competitively with CNN and increases the accuracy in both datasets.

The results of two proposed models are shown in Table II. As we expected LSTMs are well suited to encode information and long-range context dependency, while CNNs are considered good at extracting local and position-invariant features. In the Laptop dataset, BANN has 73.51% in accuracy and CANN has 69.75%; in the Restaurant dataset, BANN has 80.71% in accuracy and CANN has 78.04%. This can be explained according to the finding of [40], which type of neural networks performs better in the task depends on how often the comprehension of global/local range semantics is required. We will have further in-depth analysis and qualitative case studies in the next section.

E. Discussions

As IAN[20] is the state-of-the-art framework for aspect-level opinion mining, we compare our frameworks with IAN to verify the effectiveness and advantages in this section.

LSTM has restrictions as the future input information cannot be reached from the current state. While, in BLSTM, the future input information is reachable from the current state.

With this structure, the output layer can get information from the past and future states, which makes BLSTM especially useful when the context of input is needed. From Table II, we can see our model improved classification accuracy by 1.41% in the Laptop dataset and 2.11% in the Restaurant dataset. It indicates that using BLSTM to model opinion targets and context helps to learn hidden word semantics and better represent targets and context, which contributes to aspect-level opinion polarity classification. Thus, our proposed BANN framework can be a better solution in aspect-level opinion polarity classification.

As for CANN model, it performs competitively with CNN and marginally increases the accuracy in both datasets. It also performs competitively with IAN on the restaurant dataset, while slightly fall behind on the laptop dataset. Compared with the CNN baseline model, our CANN with additional attention mechanisms performs competitively with it and marginally increases the accuracy in both datasets. Although, in BANN experiments, attention mechanisms help to capture the important information in the context and generate better representations for aspect-level opinion polarity classification. While in CANN experiments, we cannot see much improvements based on our results. One possible explanation is that our dataset is relatively small, the CANN model did not reach its full potential. Our future work will conduct empirical experiments with larger datasets to further demonstrate the effectiveness of proposed methods. Comparing to BANN which has the similar framework, it also failed to get higher accuracy. This can be explained according to the finding of [40], which type of neural networks performs better in the task depends on how often the comprehension of global/local range semantics is required.

Table III
IAN VS. BANN VS. CANN IN RESTAURANT EXAMPLE

IAN	BANN	CANN	Examples
T	T	F	E1. The **fajita** we tried was tasteless and burned and the mole sauce was way too sweet.
F	F	T	E2. Yes, they use fancy ingredients, but even fancy **ingredients** don't make for good pizza unless someone knows how to get the crust right.
F	T	F	E3. The **waiters** ALWAYS look angry and even ignore their high-tipping regulars.
T	F	F	E4. **Desserts** include flan and sopaipillas.

To further compare BANN and CANN framework, we employ content analysis from the qualitative perspective to uncover the cases where each framework works better aiming to give basic guidance for base neural networks selection. Table III and Table IV show examples E1 to E8 from both datasets in which either IAN, BANN or CANN predicts correctly while the others predicts correctly or falsely or vice versa. T in the column represents the model predicts the sentiment correctly, while F represents the model predicts falsely, where aspect-term is in bold.

Table IV
IAN VS. BANN VS. CANN IN LAPTOP EXAMPLE

IAN	BANN	CANN	Examples
T	T	F	E5. The **battery life** seems to be very good, and have had no issues with it.
F	F	T	E6. The receiver was full of superlatives for the quality and **performance**.
F	T	F	E7. Further, this Mac Mini has a sloppy Bluetooth interface (courtesy of the Mac OS) and the **range** is poor.
T	F	F	E8. I eventually did the migration from my iMac backup disc which uses **USB**.

E1 is labeled as negative, the aspect term is "fajita". It contains the phrases "tasteless" and "burned" that usually appear with negative sentiment. E5 is labeled as positive, the aspect term is "battery life". It contains the phrases "very good" that usually appear with positive sentiment. However, in both examples, they contain phrases like "sweet" (in E1) and "issues" (in E5) that usually appear with the opposite sentiment. Also, in both examples, the sentiment phrases are widely spread throughout the sentence. These might result in the wrong prediction by CANN. Thus, an architecture like IAN or BANN that well suited for encoding information and long-range context dependency is needed to handle long sequences correctly.

E2 is labeled as positive, the aspect term is "ingredients". It contains the phrases "fancy" that usually appear with positive sentiment, but the whole sentence exhibits a negative sentiment. E6 is labeled as positive, the aspect term is "performance". It contains the phrases "full of superlatives" that usually appear with positive sentiment. As we known, CNNs are considered good at extracting local and position-invariant features. Both examples show sentiment phrases closely adhere to aspect terms. Thus, an architecture like CANN is better to handle it correctly.

E3 is labeled as negative, the aspect term is "waiters". It contains the phrases "always", "look angry" and "ignore" that usually appear with negative sentiment. E7 is labeled as negative, the aspect term is "range". It contains the phrases "poor" that usually appear with negative sentiment. In BLSTM, the future input information is reachable from the current state. With this structure, the output layer can get information from the past and future states. Thus, an architecture like BANN is ideal when the future context of input is needed.

E4 is labeled as neutral, the aspect term is "desserts". It does not contain any phrases that usually appear with either positive or negative sentiment. E8 is labeled as neutral, the aspect term is "USB". It does not contain any phrases that usually appear with either positive or negative sentiment. It shows that an architecture like IAN works well when there is no strong sentiment expressed.

In summary, we find that BANN is better when sentiment is determined by a long-range semantic dependency or the entire sentence rather than some local key-phrases is involved, see E1, E3, E5, E7. On the other hand, modeling the whole context sometimes is a burden which can neglect the key phrases. For instance, the first part of E2 seems positive, while the second part seems negative. IAN and BANN encode the entire information in E2, making it hard for the positive key phrase "fancy" to play a main role in the final representation, which might result in the wrong prediction. While CANN good at extracting local and position-invariant features is better to handle it correctly.

V. Conclusions

In this paper, we propose and compare two novel frameworks for aspect-level opinion mining: bi-directional attention neural networks (BANN) and convolutional attention neural networks (CANN). Specifically, BANN framework employs two bi-directional LSTM (BLSTM) to learn hidden word semantics to model opinion target and context, while CANN framework employs two CNN to learn hidden word semantics to model opinion target and context. Then, we implement an attention mechanism to integrate hidden states learned from both the targets and context. Finally, we concatenate both the target and context representations to predict aspect-level opinion polarity. Experiments on two SemEval 2014 task datasets show that two methods obtain competitive performances against the baselines on both datasets.

In particular, our contribution is in three-fold:

1. We propose two interactive attention neural networks for aspect-level opinion mining that improve the state-of-the-art methods;
2. We compare proposed frameworks from quantitative and qualitative perspective to uncover when each framework performs better, aiming to give basic guidance for neural network selection in similar tasks;
3. We offer a new approach to support human decision-making process based on opinion mining results.

The direction of our future work includes conducting empirical experiments with larger datasets to demonstrate the effectiveness of proposed methods. Also, we plan to develop a new framework to analyze aspect-level opinions with a more significant representation of context to better support the multiple opinion targets scenario. In addition, domain adaptation for opinion mining is still a challenging topic due to the fact that supervised classifiers are very sensitive to domain changes. We also hope to build a real-time social media opinion mining system to identify emerging societal trends based on views, dispositions, moods, attitudes and expectations of stakeholder groups or the general public.

Acknowledgment

This work was supported in part by NSF III 1815256, NSF III 1744661, NSF CNS 1650431. The authors would also like to thank the anonymous reviewers for their valuable feedback.

References

[1] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 813–830, 2016.

[2] E. Qualman, *Socialnomics: How social media transforms the way we live and do business*. John Wiley & Sons, 2010.

[3] N. A. Buzzetto-More, "Social networking in undergraduate education," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 7, no. 1, pp. 63–90, 2012.

[4] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013.

[5] B. Pang, L. Lee *et al.*, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[6] K. D. Varathan, A. Giachanou, and F. Crestani, "Comparative opinion mining: a review," *Journal of the Association for Information Science and Technology*, vol. 68, no. 4, pp. 811–829, 2017.

[7] B. Liu, M. Hu, and J. Cheng, "Opinion observer: analyzing and comparing opinions on the web," in *Proceedings of the 14th international conference on World Wide Web*. ACM, 2005, pp. 342–351.

[8] H. Binali, V. Potdar, and C. Wu, "A state of the art opinion mining and its application domains," in *Proceedings of the international conference on industrial technology (ICIT 2009)*. IEEE Computer Society, 2009, pp. 1–6.

[9] K. Khan, B. Baharudin, A. Khan, and A. Ullah, "Mining opinion components from unstructured reviews: A review," *Journal of King Saud University-Computer and Information Sciences*, vol. 26, no. 3, pp. 258–275, 2014.

[10] P. Sobkowicz, M. Kaschesky, and G. Bouchard, "Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web," *Government Information Quarterly*, vol. 29, no. 4, pp. 470–479, 2012.

[11] H. Peng, Y. Ma, Y. Li, and E. Cambria, "Learning multi-grained aspect target sequence for chinese sentiment analysis," *Knowledge-Based Systems*, vol. 148, pp. 167–176, 2018.

[12] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2018.

[13] X. Yu, Z. Chen, W.-S. Yang, X. Hu, E. Yan, and G. Li, "Large-scale joint topic, sentiment & user preference analysis for online reviews," in *Big Data (Big Data), 2017 IEEE International Conference on*. IEEE, 2017, pp. 847–856.

[14] Z. Chen, Y. Zhang, Y. Shang, and X. Hu, "Unifying topic, sentiment & preference in an hdp-based rating regression model for online reviews," in *Asian Conference on Machine Learning*, 2016, pp. 222–237.

[15] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective lstms for target-dependent sentiment classification," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 3298–3307.

[16] S. Ruder, P. Ghaffari, and J. G. Breslin, "A hierarchical model of reviews for aspect-based sentiment analysis," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 999–1005.

[17] Y. Wang, M. Huang, L. Zhao *et al.*, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 606–615.

[18] M. Yang, W. Tu, J. Wang, F. Xu, and X. Chen, "Attention based lstm for target dependent sentiment classification." in *AAAI*, 2017, pp. 5013–5014.

[19] J. Liu and Y. Zhang, "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*, vol. 2, 2017, pp. 572–577.

[20] X. Z. H. W. Dehong Ma, Sujian Li, "Interactive attention networks for aspect-level sentiment classification," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017, pp. 4068–4074.

[21] W. Quan, Z. Chen, and X. Hu, "Bann: A framework for aspect-level opinion mining," in *Proceedings of the 2018 International Conference on Information Technology*. ACM, 2018.

[22] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

[23] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using bilstm-crf and cnn," *Expert Systems with Applications*, vol. 72, pp. 221–230, 2017.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[25] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "Cnn features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806–813.

[26] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[27] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *arXiv preprint arXiv:1708.02709*, 2017.

[28] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 160–167.

[29] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, no. Aug, pp. 2493–2537, 2011.

[30] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[31] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[32] T. Lei, R. Barzilay, and T. Jaakkola, "Molding cnns for text: non-linear, non-consecutive convolutions," *arXiv preprint arXiv:1508.04112*, 2015.

[33] R. Johnson and T. Zhang, "Effective use of word order for text categorization with convolutional neural networks," *arXiv preprint arXiv:1412.1058*, 2014.

[34] ——, "Deep pyramid convolutional neural networks for text categorization," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2017, pp. 562–570.

[35] S. Wang, M. Huang, and Z. Deng, "Densely connected cnn with multi-scale feature attention for text classification." in *IJCAI*, 2018, pp. 4468–4474.

[36] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *arXiv preprint arXiv:1503.00075*, 2015.

[37] Y. Ma, H. Peng, and E. Cambria, "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm," in *Proceedings of AAAI*, 2018.

[38] G. Ganu, N. Elhadad, and A. Marian, "Beyond the stars: improving rating predictions using review text content." in *WebDB*, vol. 9. Citeseer, 2009, pp. 1–6.

[39] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[40] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of cnn and rnn for natural language processing," *arXiv preprint arXiv:1702.01923*, 2017.