ISSN (e): 2250 – 3005 || Volume, 14 || Issue, 4|| Jul. - Aug. – 2024 ||
International Journal of Computational Engineering Research (IJCER)

A Survey on Deep Learning Methods for Aspect Based Sentiment Analysis

Abani Kumar Bisoyi[1], Sarbasri Sahoo[2], Sandhya Rani Rout[3]
[1,2,3] Department of Computer Science and Engineering, Einstein Academy of Technology & Management, Bhubaneswar, India

Abstract— *The practice of examining, interpreting, drawing conclusions from, and extrapolating sentiment from subjective materials is known as sentiment analysis. Businesses utilize sentiment analysis for market research, customer experience analysis, brand reputation analysis, public opinion analysis, and social media influence research. It can be further classified as document-, sentence-, and aspect-based granularity based on the various needs. An aspect-based sentiment analysis challenge is explained in this article along with the newly suggested solutions. Currently, lexicon-based, conventional machine learning, and deep learning techniques are the three main approaches. We offer a comparison of the most recent deep learning techniques in this survey article. We discuss a number of widely used benchmark data sets, evaluation measures, and the state-of-the-art deep learning techniques.*

Key Words— *Aspect-based sentiment Analysis, Deep Learning, NLP, Data Mining, Sentiment Analysis, Gated Recurrent Unit.*

I. Introduction

SENTIMENT analysis has become a significant researchdirection in NLP. It consists of a combination of information retrieval, NLP, and artificial intelligence. Sentiment analysis is also known as opinion mining or subjectivity analysis.It studies various aspects, such as opinions, sentiments, evaluations, appraisals, attitudes, and emotions [1]. The commonlyused phrase for sentiment analysis is "opinion mining," whichis derived from the data mining and information retrievalcommunity. Its main goal is to determine the opinions ofa group of people on a certain topic. Sentiment analysisis a commonly used term that focuses on identifying the sentiment expressed in a text. It has become a rapidly growingresearch area since 2000 when Pang and Lee [2] createda comprehensive study to determine the sentiment polarityof movie reviews. It has received attention from not onlyacademia but also from the industry because it can providefeedback information of customers through online reviews,help in deciding marketing policies, and detect changes incustomers' opinions about various subjects, e.g., COVID-19'shandling. It is used to identify and extract opinions within texts, sentences, or documents. Its basic task is to classify theexpressed opinion of a given text into positive, negative, and neutral ones.

Nowadays, reviewing online customer comments and ratings before purchasing a product has become a very common and popular trend practice. Studies have shown that consumers trust online reviews or comments from strangers before purchasing a product or service [3]. There have been many statistical surveys and studies conducted in this area [82]. A study conducted in [4] shows that 39% of customers read approximately eight reviews, while 12% of them read 16 or more reviews before deciding on buying a product;98% of the customers admit that their purchasing decision is influenced by customer reviews of previous buyers according to [5]. As stated in [6], statistics show that potential buyers are willing to spend 31% more on a product or service having outstanding reviews.

The customer reviews have become so significant that a study in [7] shows that buyers are not likely going to choose a product that has fewer or no reviews whenever they are confused between two products; 98% of buyers are resistantto buy a product with less or no reviews, as shown in [8], while almost four out of five customers change their minds about buying a particular product recommended by their friends or family because of negative reviews [9].

There have been several investigations being conducted in this area. The survey article by Harrang *et al.* [10] discusses,in detail, the improvements done in the field of prediction of customer reviews and ratings. There have been many scholars who have also compared different types of approachesfor sentiment analysis along with the evaluation of several algorithms to conclude the algorithm that fits best with their respective data sets [11], [12].

II. Related Work

Opinion mining and sentiment analysis (Pang and Lee, 2008) on user-generated reviews can provide valuable information for providers and consumers. Instead of predicting the overall sen1.The code and data is available at https://github. com/wxue004cs/GCAE sentiment polarity, fine-grained aspect-based sentiment analysis (ABSA) (Liu and Zhang, 2012) is proposed to better understand reviews than traditional sentiment analysis. Specifically, we are interested in the sentiment polarity of aspect categories or target entities in the text. Sometimes, it is coupled with aspect term extractions (Xue et al., 2017). A number of models have been developed for ABSA, but there are two different subtasks, namely aspect-category sentiment analysis (ACSA) and aspect-term sentiment analysis (ATSA). The goal of ACSA is to predict the sentiment polarity with regard to the given aspect, which is one of a few predefined categories. On the other hand, the goal of ATSA is to identify the sentiment polarity concerning the target entities that appear in the text instead, which could be a multi-word phrase or a single word. The number of distinct words contributing to aspect terms could be more than a thousand. For example, in the sentence "Average to good Thai food, but terrible delivery.", ATSA would ask the sentiment polarity towards the entity Thai food; while ACSA would ask the sentiment polarity toward the aspect service, even though the word service does not appear in the sentence. Many existing models use LSTM layers (Hochreiter and Schmidhuber, 1997) to distill sentiment information from embedding vectors, and apply attention mechanisms (Bahdanau et al., 2014) to enforce models to focus on the text spans related to the given aspect/entity. Such models include Attention-based LSTM with Aspect Embedding (ATAE-LSTM) (Wang et al., 2016b) for ACSA; Target-Dependent Sentiment Classification (TD-LSTM) (Tang et al., 2016a), Gated Neural Networks (Zhang et al., 2016) and Recurrent Attention Memory Network (RAM) (Chen et al., 2017) for ATSA. Attention mechanisms has been successfully used in many arXiv:1805.07043v1 [cs.CL] 18 May 2018 NLP tasks. It first computes the alignment scores between context vectors and target vector; then carry out a weighted sum with the scores and the context vectors. However, the context vectors have to encode both the aspect and sentiment information, and the alignment scores are applied across all feature dimensions regardless of the differences between these two types of information. Both LSTM and attention layer are very time-consuming during training.

LSTM processes one token in a step. Attention layer involves exponential operation and normalization of all alignment scores of all the words in the sentence (Wang et al., 2016b). Moreover, some models need the positional information between words and targets to produce weighted LSTM (Chen et al., 2017), which can be unreliable in noisy review text. Certainly, it is possible to achieve higher accuracy by building more and more complicated LSTM cells and sophisticated attention mechanisms; but one has to hold more parameters in memory, get more hyper-parameters to tune and spend more time in training. In this paper, we propose a fast and effective neural network for ACSA and ATSA based on convolutions and gating mechanisms, which has much less training time than LSTM based networks, but with better accuracy. For ACSA task, our model has two separate convolutional layers on the top of the embedding layer, whose outputs are combined by novel gating units. Convolutional layers with multiple filters can efficiently extract n-gram features at many granularities on each receptive field. The proposed gating units have two nonlinear gates, each of which is connected to one convolutional layer. With the given aspect information, they can selectively extract aspect-specific sentiment information for sentiment prediction. For example, in the sentence "Average to good Thai food, but terrible delivery.", when the aspect food is provided, the gating units automatically ignore the negative sentiment of aspect delivery from the second clause, and only output the positive sentiment from the first clause. Since each component of the proposed model could be easily parallelized, it has much less training time than the models based on LSTM and attention mechanisms. For ATSA task, where the aspect terms consist of multiple words, we extend our model to include another convolutional layer for the target expressions. We evaluate our models on the SemEval datasets, which contains restaurants and laptops reviews with labels on aspect level. To the best of our knowledge, no CNN based model has been proposed for aspect-based sentiment analysis so far. 2 Related Work We present the relevant studies into following two categories. 2.1 Neural Networks Recently, neural networks have gained much popularity on sentiment analysis or sentence classification task. Tree-based recursive neural networks such as Recursive Neural Tensor Network (Socher et al., 2013) and Tree-LSTM (Tai et al., 2015), make use of syntactic interpretation of the sentence structure, but these methods suffer from time inefficiency and parsing errors on review text. Recurrent Neural Networks (RNNs) such as LSTM (Hochreiter and Schmidhuber, 1997) and GRU (Chung et al., 2014) have been used for sentiment analysis on data instances having variable length (Tang et al., 2015; Xu et al., 2016; Lai et al., 2015). There are also many models that use convolutional neural networks (CNNs) (Collobert et al., 2011; Kalchbrenner et al., 2014; Kim, 2014; Conneau et al., 2016) in NLP, which also prove that convolution operations can capture compositional structure of texts with rich semantic information without laborious feature engineering. 2.2 Aspect based Sentiment Analysis There is abundant research work on aspect-based sentiment analysis. Actually, the name ABSA is used to describe two different subtasks in the literature. We classify the existing work into two main categories based on the descriptions of sentiment analysis tasks in SemEval 2014 Task 4 (Pontiki et al., 2014): Aspect-Term Sentiment Analysis and Aspect-Category Sentiment Analysis. Aspect-Term Sentiment Analysis. In the first category, sentiment analysis is performed toward the aspect terms that are labeled in the given sentence. A large body of literature tries to utilize the relation or position between the target words and the surrounding context words either by using the tree structure of dependency or by simply counting the number of words between them as a relevance information (Chen et al., 2017). Recursive neural networks (Lakkaraju et al., 2014; Dong et al., 2014; Wang et al., 2016a) rely on external syntactic parsers, which can be very inaccurate and slow on noisy texts like tweets and reviews, which may result in inferior performance. Recurrent neural networks are commonly used in many NLP tasks as well as in ABSA problem. TD-LSTM (Tang et al., 2016a) and gated neural networks (Zhang et al., 2016) use two or three LSTM networks to model the left and right contexts of the given target individually. A fully connected layer with gating units predicts the sentiment polarity with the outputs of LSTM layers. Memory network (Weston et al., 2014) coupled with multiple-hop attention attempts to explicitly focus only on the most informative context area to infer the sentiment polarity towards the target word (Tang et al., 2016b; Chen et al., 2017). Nonetheless, memory network simply bases its knowledge bank on the embedding vectors of individual words (Tang et al., 2016b), which makes itself hard to learn the opinion word enclosed in more complicated contexts. The performance is improved by using LSTM, attention layer and feature engineering with word distance between surrounding words and target words to produce target-specific memory (Chen et al., 2017). Aspect-Category Sentiment Analysis. In this category, the model is asked to predict the sentiment polarity toward a predefined aspect category. Attention-based LSTM with Aspect Embedding (Wang et al., 2016b) uses the embedding vectors of aspect words to selectively attend the regions of the representations generated by LSTM

III. Gated Convolutional Network with Aspect Embedding (Gated Mechanism)

The proposed Gated Tanh-ReLU Units control the path through which the sentiment information flows towards the pooling layer. The gating mechanisms have proven to be effective in LSTM. In aspect based sentiment analysis, it is very common that different aspects with different sentiments appear in one sentence. The ReLU gate in Equation 2 does not have upper bound on positive inputs but strictly zero on negative inputs. Therefore, it can output a similarity score according to the relevance between the given aspect information va and the aspect feature ai at position t. If this score is zero, the sentiment features si would be blocked at the gate; otherwise, its magnitude would be amplified accordingly. The max-over-time pooling further removes the sentiment features which are not significant over the whole sentence. In language modeling (Dauphin et al., 2017; Kalchbrenner et al., 2016; van den Oord et al., 2016; Gehring et al., 2017), Gated Tanh Units (GTU) and Gated Linear Units (GLU) have shown effectiveness of gating mechanisms. GTU is represented by $\tanh(X * W + b) \times \sigma(X * V + c)$, in which the sigmoid gates control features for predicting the next word in a stacked convolutional block. To overcome the gradient vanishing problem of GTU, GLU uses $(X*W+b)\times\sigma(X*V+c)$ instead, so that the gradients would not be down scaled to propagate through many stacked convolutional layers. However, a neural network that has only one convolutional layer would not suffer from gradient vanish problem during training. We show that on text classification problem, our GTRU is more effective than these two gating units.

IV. Experiments

Datasets and Experiment Preparation

We conduct experiments on public datasets from SemEval workshops (Pontiki et al., 2014), which consist of customer reviews about restaurants and laptops. Some existing work (Wang et al., 2016b; Ma et al., 2017; Chen et al., 2017) removed “conflict” labels from four sentiment labels, which makes their results incomparable to those from the workshop report (Kiritchenko et al., 2014). We reimplemented the compared methods, and used hyper-parameter settings described in these references. The sentences which have different sentiment labels for different aspects or targets in the sentence are more common in review data than in standard sentiment classification benchmark.

To comprehensively evaluate the performance of GCAE, we compare our model against the following models. NRC-Canada (Kiritchenko et al., 2014) is the top method in SemEval 2014 Task 4 for ACSA and ATSA task. SVM is trained with extensive feature engineering: various types of n-grams, POS tags, and lexicon features. The sentiment lexicons improve the performance significantly, but it requires large scale labeled data: 183 thousand Yelp reviews, 124 thousand Amazon laptop reviews, 56 million tweets, and 3 sentiment lexicons labeled manually. CNN (Kim, 2014) is widely used on text classification task. It cannot directly capture aspectspecific sentiment information on ACSA task, but it provides a very strong baseline for sentiment classification. We set the widths of filters to 3, 4, 5 with 100 features each. TD-LSTM (Tang et al., 2016a) uses two LSTM networks to model the preceding and following contexts of the target to generate target-dependent representation for sentiment prediction. ATAE-LSTM (Wang et al., 2016b) is an attention-based LSTM for ACSA task. It appends the given aspect embedding with each word embedding as the input of LSTM, and has an attention layer above the LSTM layer. IAN (Ma et al., 2017) stands for interactive attention network for ATSA task, which is also based on LSTM and attention mechanisms. RAM (Chen et al., 2017) is a recurrent attention network for ATSA task, which uses LSTM and multiple attention mechanisms. GCN stands for gated convolutional neural network, in which GTRU does not have the aspect embedding as an additional input.

V. Results and Analysis

Following the SemEval workshop, we report the overall accuracy of all competing models over the test datasets of restaurant reviews as well as the hard test datasets. Every experiment is repeated five times. The mean and the standard deviation are reported in Table 4. LSTM based model ATAE-LSTM has the worst performance of all neural networks. Aspect-based sentiment analysis is to extract the sentiment information closely related to the given aspect. It is important to separate aspect information and sentiment information from the extracted information of sentences. The context vectors generated by LSTM have to convey the two kinds of information at the same time. Moreover, the attention scores generated by the similarity scoring function are for the entire context vector. GCAE improves the performance by 1.1% to 2.5% compared with ATAE-LSTM. First, our model incorporates GTRU to control the sentiment information flow according to the given aspect information at each dimension of the context vectors. The element-wise gating mechanism works at fine granularity instead of exerting an alignment score to all the dimensions of the context vectors in the attention layer of other models. Second, GCAE does not generate a single context vector, but two vectors for aspect and sentiment features respectively, so that aspect and sentiment information is unraveled. By comparing the performance on the hard test datasets against CNN, it is easy to see the convolutional layer of GCAE is able to differentiate the sentiments of multiple entities.

	Positive		Negative		Neutral		Conflict	
	Train	Test	Train	Test	Train	Test	Train	Test
Restaurant-Large	2710	1505	1198	680	757	241	-	-
Restaurant-Large-Hard	182	92	178	81	107	61	-	-
Restaurant-2014	2179	657	839	222	500	94	195	52
Restaurant-2014-Hard	139	32	136	26	50	12	40	19

The performance of SVM (Kiritchenko et al., 2014) depends on the availability of the features it can use. Without the large amount of sentiment lexicons, SVM perform worse than neural methods. With multiple sentiment lexicons, the performance is increased by 7.6%. This inspires us to work on leveraging sentiment lexicons in neural networks in the future. The hard test datasets consist of replicated sentences with different sentiments towards different aspects. The models which cannot utilize the given aspect information such as CNN and GCN perform poorly as expected, but GCAE has higher accuracy than another neural network models. GCAE achieves 4% higher accuracy than ATAE-LSTM on Restaurant-Large and 5% higher on SemEval-2014 on ACSA task. However, GCN, which does not have aspect modeling part, has higher score than GCAE on the original restaurant dataset. It suggests that GCN performs better than GCAE when there is only one sentiment label in the given sentence, but not on the hard test dataset.

VI. Conclusions and Future Work

In this paper, we proposed an efficient convolutional neural network with gating mechanisms for ACSA and ATSA tasks. GTRU can effectively control the sentiment flow according to the given aspect information, and two convolutional layers model the aspect and sentiment information separately. We prove the performance improvement compared with other neural models by extensive experiments on SemEval datasets. How to leverage large-scale sentiment lexicons in neural networks would be our future work.

References

[1]. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. In ICLR, pages CoRR abs–1409.0473.
[2]. Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent Attention Network on Memory for Aspect Sentiment Analysis. In EMNLP, pages 463–472.
[3]. Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. In NIPS.
[4]. Ronan Collobert, Jason Weston, Leon Bottou, Michael ´ Karlen, Koray Kavukcuoglu, and Pavel P Kuksa. 2011. Natural Language Processing (Almost) from Scratch. Journal of Machine Learning Research, 12:2493–2537.
[5]. Alexis Conneau, Holger Schwenk, Lo¨ıc Barrault, and Yann LeCun. 2016. Very Deep Convolutional Networks for Text Classification. In EACL, pages 1107–1116.
[6]. Yann N Dauphin, Angela Fan, Michael Auli, and David Grangier. 2017. Language Modeling with Gated Convolutional Networks. In ICML, pages 933–941.
[7]. Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014. Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification. In ACL, pages 49–54.
[8]. John C Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. Journal of Machine Learning Research, pages 2121–2159.
[9]. Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. 2017. Convolutional Sequence to Sequence

Learning. In ICML, pages 1243–1252.
[10]. Sepp Hochreiter and Jurgen Schmidhuber. 1997. ¨ Long Short-Term Memory. Neural computation, 9(8):1735–1780.
[11]. Nal Kalchbrenner, Lasse Espeholt, Karen Simonyan, Aaron van den Oord, Alex Graves, and Koray ¨ Kavukcuoglu. 2016. Neural Machine Translation in Linear Time. CoRR, abs/1610.10099.
[12]. Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. In ACL, pages 655–665.
[13]. Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In EMNLP, pages 1746– 1751. Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif M. Mohammad. 2014.
[14]. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In SemEval@COLING, pages 437–442, Strouds burg, PA, USA. Association for Computational Linguistics.
[15]. Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent Convolutional Neural Networks for Text Classification. AAAI, pages 2267–2273.
[16]. Himabindu Lakkaraju, Richard Socher, and Chris Manning. 2014. Aspect specific sentiment analysis using hierarchical deep learning. In NIPS Workshop on Deep Learning and Representation Learning.
[17]. Hoa T Le, Christophe Cerisara, and Alexandre Denis. 2017. Do Convolutional Networks need to be Deep for Text Classification? CoRR, abs/1707.04108.
[18]. Bing Liu and Lei Zhang. 2012. A Survey of Opinion Mining and Sentiment Analysis. Mining Text Data, (Chapter 13):415–463.
[19]. Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive Attention Networks for Aspect-Level Sentiment Classification. In IJCAI, pages 4068–4074.
[20]. International Joint Conferences on Artificial Intelligence Organization. Aaron van den Oord, Nal Kalchbrenner, Lasse Espe- ¨ holt, Koray Kavukcuoglu, Oriol Vinyals, and Alex Graves. 2016. Conditional Image Generation with PixelCNN Decoders. In NIPS, pages 4790–4798