

Self-Attention Networks and Adaptive Support Vector Machine for Aspect-Level Sentiment Classification

Meizhan Liu

Shandong University

Fengyu Zhou (✉ zhoufengyu@sdu.edu.cn)

Shandong University

JiaKai He

Shandong University

Ke Chen

Shandong University

Yang Zhao

Qilu University of Technology

HongChang Sun

Shandong University

Research Article

Keywords: Aspect-level sentiment classification, Long-term Dependencies, Multi-heads Self Attention mechanism, Adaptive Support Vector Machine

Posted Date: January 4th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1221812/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Soft Computing on February 21st, 2022. See the published version at <https://doi.org/10.1007/s00500-022-06793-7>.

Self-Attention Networks and Adaptive Support Vector Machine for Aspect-Level Sentiment Classification

Meizhen Liu¹, FengYu Zhou*¹,
JiaKai He¹, Ke Chen¹, Yang Zhao², HongChang Sun¹,

¹ School of Control Science and Engineering, Shandong University, JiNan, China

² School of Electrical Engineering and Automation, Qilu University of Technology (Shandong academy of sciences), Jinan, China

* Corresponding author

{mzliu94@163.com, zhoufengyu@sdu.edu.cn, nyxhjk@126.com, chenke_zixf@163.com, zdh1136@126.com, sunhongch@163.com}

Abstract. Aspect-level sentiment classification aims to integrating the context to predict the sentiment polarity of aspect-specific in a text, which has been quite useful and popular, e.g. opinion survey and products' recommending in e-commerce. Many recent studies exploit a Long Short-Term Memory (LSTM) networks to perform aspect-level sentiment classification, but the limitation of long-term dependencies is not solved well, so that the semantic correlations between each two words of the text are ignored. In addition, traditional classification model adopts SoftMax function based on probability statistics as classifier, but ignores the words' features in the semantic space. Support Vector Machine (SVM) can fully use the information of characteristics and it is appropriate to make classification in the high dimension space, however which just considers the maximum distance between different classes and ignores the similarities between different features of the same classes. To address these defects, we propose the two-stages novel architecture named Self Attention Networks and Adaptive SVM (SAN-ASVM) for aspect-level sentiment classification. In the first-stage, in order to overcome the long-term dependencies, Multi-Heads Self Attention (MHSA) mechanism is applied to extract the semantic relationships between each two words, furthermore 1-hop attention mechanism is designed to pay more attention on some important words related to aspect-specific. In the second-stage, ASVM is designed to substitute the SoftMax function to perform sentiment classification, which can effectively make multi-classifications in high dimensional space. Extensive experiments on SemEval2014, SemEval2016 and Twitter datasets are conducted, compared experiments prove that SAN-ASVM model can obtains better performance.

Keywords: Aspect-level sentiment classification; Long-term Dependencies; Multi-heads Self Attention mechanism; Adaptive Support Vector Machine.

1 Introduction

With the advancement of electronic informational technology, more and more people share their comments or experiences on the internet, which can reflect their attitudes on a product or service, so how to mine these contexts and extract the main information has been one of hot researching topics in the field of Natural Language Process (NLP). For example, in the process of question survey, sentiment classification can recognize people's attitudes toward some policy or services. Therefore, sentiment analysis is a very useful method to mine the sentiment information (Pang et al.2008; Mohammad et al.2015).

In general, sentiment analysis of aspect-level can excavate semantic correlation with context, but a text consists of several words and the same word may have different semantics in different context. For example, in this text "The speed of computer is so fast that a lot of time is saved", "fast" is attribute word which means the positive polarity. However, another text, "Time has passed so fast, I haven't had time to read this book.", "fast" implies regret. All in all, if semantic environment and context of a word are ignored, the feature representation of the word would lack of contextual information which might lead to the deviation on sentiment orientation of aspect-specific. Therefore, a method based on aspect-level sentiment analysis should be proposed.

LSTM and its' varieties have been a prevalent and effective method to address the problem of large-scale sequential modeling process, however, when the text is too long, the performance of LSTM on long-term dependencies would be decreased greatly (Tang et al.2016). Besides that, LSTM models can deal with the sequential information, but the interactive relationships between each two words are not well considered. Therefore, in order to analyze effectively the semantic correlations between each two words, attention mechanism is developed to solve the above problems which can analyze the correlations between each two words also pay more attention on some important words of the context and decrease the influences of less-related words. Many previous works focus on the design of attention mechanism, to capture more important words and generate the contexts' representation based on different aspects. **Topic-Dependent Attention Model (TDAM)** (Tang et al.2016) employed the bidirectional **Gate Recurrent Unit (GRU)** to model the input sequence, and the hierarchical and multi-tasks learning architecture were designed to calculate the weights but which ignored the relationships between topics and context, therefore the important information would be lost. **ATtention-based Long Short-Term Memory for AspEct-level (ATAE-LSTM)** was proposed to (Gabriele et al.2019) integrate aspect embedding with each word's embedding, and then analyze the mutual influences between the hidden layers' outputs obtained by LSTM, yet this model didn't learn the correlations of every word with context detailly. In order to analyze the influences of aspects on every word of context, Song et al. (Wang et al.2019) designed attentional encoder layer and target-specific attention layer to capture more valuable information between them. Above of all, in order to solve the long-term dependencies and consider the interactive relationships between each two words, LSTM networks or attention mechanism is a good choice to achieve it.

Because SoftMax function is based on probability model, which ignores some valuable information brought by the feature vectors, for example the differences and

similarities of features. Therefore, in order to make full use of features information, SVM is applied to reflect feature vectors of samples to high dimensional space and generate the separate plate by evaluating the maximum distance of different samples (Alyami et al.2020; Han et al. 2020). However, the decision function of SVM just only considers the differences between classes, ignores the valuable information of similarities between classes. Therefore, Fisher method (Belhumeur et al.1996) is introduced, which can project the samples of the high-dimensional space into the low-dimensional space, then take the smallest differences in-cluster and the largest differences between-clusters into account to classify projected samples in subspace. This method makes the classified samples possess the smallest intra-class distance and the largest inter-class distance. Therefore, combining the principle of structural risk minimization and the determinant principle of Fisher classifier to design decision function of Adaptive Support Vector Machine (ASVM) classifier, the goal of this proposal is to balance the differences and similarities of features to improve the classification accuracy.

Above of all, in order to enhance the semantic correlations between each two words in text and improve the accuracy of classification, we put forward two-stages novel architecture. In the first-stage, Self-Attention Network (SAN) is designed to generate feature representations. More concretely, Multi-Heads Self Attention (MHSA) is utilized to extract the semantic correlations between each two words which can solve the long-term dependencies well by reflecting the representation of every word to the same context semantic space. Besides, 1-hop attention mechanism is applied to calculate the correlations between aspect-specific and every word of context, which can help us assign a bigger weigh to the word with strong correlation to aspect-specific and smaller weigh to the word with weak correlation. After that, the cosine similarity is further employed to evaluate the sentimental influences on aspect-specific to generate the final representation integrating the semantics of aspect-specific and context. The architecture of first-stage aims to address problems of long-term dependencies and semantic correlations between each two words. We design the ASVM classifier to conduct sentiment polarity predictions in the second-stage to balance the distance of inter-class and intra-class. Specially, we adopt GloVe pre-trained model to train word embeddings and the whole experiments are trained on public datasets deriving form SemEval2014 (Kiritchenko et al.2014), SemEval2016 (Pontiki et al.2016) and Twitter (Dong et al.2014) to make sentiment classification.

In comparison with previous approaches of sentiment analysis based on aspect-level, the main contribution of this paper can be described as follows:

- We introduce the two-stages novel architecture ASN-ASVM to solve the problem of aspect-level sentiment analysis, in which the sentence's representation integrating with semantic correlations is generated by ASN and sentiment classification is conducted by ASVM.
- In the first-stage, MHSA mechanism is designed to fully learn semantic correlations of context and 1-hop attention mechanism is utilized to capture the vital words with greater influences on sentiment orientation, which significantly enhances the representation ability on dealing with context feature excavation.
- In the second-stage, ASVM classifier integrating with the Fisher regulation is designed, which can balance the feature vectors' distances between inter-class and intra-class, and our approach obtains satisfied results on these four datasets.

This paper is organized as follows. We elaborate the related work in Section 2. The two-stages novel architecture about sentiment analysis on aspect-level proposed here is present in detail in Section 3. And the 10 groups of compared experimental analysis and results are demonstrated in Section 4. At last, conclusion is concluded and an outlook of future work is given in Section 5.

2 Related Work

In this section, we will illustrate the recent researches from the following four aspects: the first part introduces research approaches about aspect-level sentiment classification; the second part discusses the methods of solving long-term dependencies problem; and the third part elaborates the attention mechanism and the last part states SVM classifier to make classification.

2.1 Aspect-Level Sentiment Classification

Aspect-level sentiment classification is a fine-grained classification tasks, which aims to extract the semantic features of context corresponded to aspect-specific and predict the sentiment polarity on the aspect-specific. Traditional methods (Shang et al. 2016) usually extract the context’s feature by using n-gram, bag-of-words, or word order (Atrio et al. 2019), and predict sentiment polarity by using Naïve Bayes [13], Support Vector Machine (SVM) (Akshay et al. 2018) or decision tree (Chauhan et al. 2018; Sridharan et al. 2020). These methods can achieve satisfactory classification results, but the methods depending on feature engineering would lead to a lot of labor and time-waste. Recent researches focus on **Embedding from Language Model (ELMo)** (Ma et al. 2018), **Generative Pre-trained Transformer (GPT)** (Xue et al. 2018), and **Bidirectional Encoder Representations from Transformers (BERT)** (Yi et al. 2018) to generate word representations and using deep networks to conduct sentiment classification.

Tang (Tang et al. 2018) designed two LSTM networks to capture the left and right texts’ features of an aspect and concatenate the succeeding and proceeding outputs as the final feature representation. Meanwhile, Zhang (Zhang et al. 2016) employed the **Gated Neural Networks (GNN)** to extract the sentiment influences on target from the left and right text. However, all these methods lack of semantic logicity and fluency for context divided by aspect-specific, hence the representation of context is continuous so that the semantic information is not fluent. Yunseok Noh (Noh et al. 2019) applied CNN as **Aspect-Map Extraction Network (AMEN)** to extract the representation of aspect to constitute an aspect map, then regarded the representations of aspect and context as input to conduct sentiment classification. Attention over Attention mechanism was introduced to model the interactive relationships between aspect-specific and every word of text, which can assign the greater weights to some

important words on sentiment classification (Huang et al. 2018). **TNet** utilizing **Transformation Networks** architecture with CNN layer was put forward to combining the bidirectional LSTM network's outputs, and generated the next word representation by convolution layer (Li et al. 2018).

Above of all, this part summarizes recent works on aspect-level sentiment analysis, we find that the effectiveness of feature representation is very important to sentiment classification. However, the biggest defect is that these works don't effectively analyze the semantic correlations of context, such as which words in the context is very important to determine the polarity of aspect.

2.2 Methods on Solving Long-Term Dependencies

Recently, LSTM are widely applied to process the sequence data (Bengio et al. 1994; Liu et al.2018; Rehman; Mohammad et al.2019; Tai et al. 2015), however, when the text is very long, the long-term dependencies doesn't be solved efficiently, so the performances of LSTM and LSTM's variants are dissatisfied (Li et al. 2018).

Generally, some researchers adopted LSTM or its variants to solve long-term dependencies problem. In addition, Liu proposed a flexible forget mechanism to decide when and how to forget the state information for LSTM (Liu et al. 2017). But this method just focused on processing units of LSTM, ignored the key word's weight to the whole context and correlations of information transmission between layers. Wang (Wang et al. 2018) integrated the mind of deep residual learning (He et al. 2016) into LSTM to propose a stacked residual LSTM model to predicate the sentiment polarity, which acquired further features between processing units effectively and tackled the degradation problem easily. The novel architecture (Liu et al. 2019) combining the bidirectional LSTM, attention mechanism and convolutional layer was proposed to extract the higher-level semantic features and set the sentiment weights on the output from hidden layers, so that the contexts' representations could be obtained. However, when adapting LSTM model to generate the word embeddings, the long-term dependencies of sequential text would be degraded gradually with the length of text increasing, so that the correlations between each two words cannot be acquired accurately. All in all, it is very significant to consider the semantic correlations between each two words without limitation of text length, therefore the attention mechanism should be introduced to calculate the semantic weights.

2.3 Methods for Attention Mechanism

Attention mechanism can help us draw more attention on significant information and excavate in-depth semantic correlations. Hence, attention mechanism based on feature election or feature representation has been more and more popular. Interactive

Attention Networks (**IAN**) (Ma et al. 2017) was proposed to interactively learn attentions between the context and target. Although this proposal considered the interactive influences of target and context, still used the average vectors to obtain the attention score. Based on that, a Feature-Enhanced Attention Network (Yang et al. 2018) was proposed, which aimed to develop a multi-views co-attention mechanism to generate the final representation integrating the context. Cheng et.al. (Cheng et al. 2017) came up with the hierarchical attention mechanism for aspect-level sentiment analysis. The alternative co-attention was proposed, which could focus on those key words to obtain the context representation according to the attention weights of target to context (Yang et al. 2019). Wang et.al. (Wang et al. 2017) put forward dependency-attention mechanism to solve the long-term dependencies of LSTM and made the LSTM networks capture the key information of targets. Multiple-attention mechanism was also proposed to solve long-term dependencies for sentiment analysis (Chen et al. 2017). All these mechanisms can achieve better performance by analyzing the sentiment relevance between aspects and context (Bahdanau et al. 2019), but don't consider the semantic correlations from the perspective of word-level, which will lead to the word representation integrating context are not fully excavated.

When the BERT (Wang et al.2019; Alexander et al.2019; Devlin et al.2019) model was proposed, many language models applying the BERT could achieve an outstanding effects. Song et al. (Wang et al.2019) adopted GloVe embedding and BERT embedding to acquire word embeddings. Rietzler et al. (Alexander et al.2019) performed a cross-domain evaluation mechanism with an adapted BERT language model for aspect-level sentiment classification, which achieved state-of-the-art performance. Xu et al. (Liu et al. 2019) explored a novel post-training method on the language model BERT to enhance the performances on aspect sentiment classification. DA-BERT (Pei et al. 2019) was proposed, which adopted transformer as the feature extraction to parallel calculate the correlation values of each word to all other words in the context, and experiment results indicated that this model could achieve better performance. Why all of these models of BERT on sentiment analysis can achieve the best, the key reason is that the bidirectional self-attention mechanism which can effectively catch the semantic correlations from the perspective of word-level. Inspired by this mind, we aim to adopt this idea to design multi-heads self-attention mechanism, so as to solve the long-term dependencies and analyze the semantic correlations between each two words of context.

2.4 Methods for SVM Classifier

Comparing with Naive Bayes, Random Forest and K-Nearest Neighbor, SVM has higher classification accuracy in terms of small sample and high dimensional space (Sujata et al. 2019; Sarkar et al. 2019). Kalarani. P. (Kalarani et al. 2019) employed the part-of-speech tags and sentiment features to extract the context feature, and then utilized the SVM and Artificial Neural Network (ANN) to realize the classification of tasks. Wang et.al. (Wang et al. 2019) designed Char Convolutional Neural Networks (**Char CNN**) with SVM to mine the customers' sentiment orientation of purchasing,

and proved that the performance of SVM classifier was better than others. The N-grams or **TF-IDF** (Term Frequency–Inverse Document Frequency) was applied to produce the sentiment knowledge and SVM was used as a classifier (Laoh et al. 2019; de Godoi Brandao et al. 2019). However, there is a common phenomenon, the decision function of SVM which doesn't consider characteristics of inter-class and intra-class, so that the classification accuracy would be decreased greatly. Consequently, we design the adaptive mechanism for SVM to improve the accuracy of sentiment classification.

3 The proposed method

In this section, we will adopt GloVe pre-trained model to simply train every word's representation and then introduce the **Multi-Heads Self-Attention (MHSA)** network to focus on semantic dependencies of every word with context. After the semantic correlations between every word and the aspect are calculated by 1-hop attention and cosine similarity function, the representation of context is obtained by weighted calculation. In the second-stage, we design the ASVM classifier to make sentiment classification. The general flow chart of the model, namely **Self-Attention Network and Adaptive Support Vector Machine (SAN-ASVM)** designed by us is shown as follows.

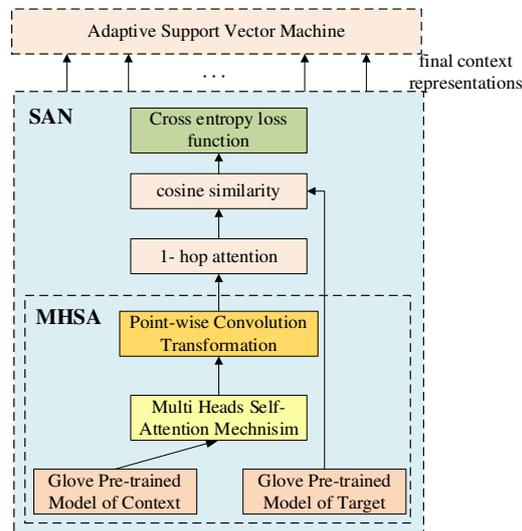


Fig. 1. The flow chart of SAN-ASVM unit.

3.1 Self-Attention Network

In the first-stage of SAN-ASVM, we design the multi-heads self-attention network and 1-hop attention mechanism to capture the semantic dependencies and correlations between each two words of text, and cosine similarity function is employed to pay more attention on some words related to sentiment polarity of aspect-specific. Because, MHSA doesn't need to use the position information, part-of-speech, which can help us simplify the computing progress. Hence combining MHSA network and cosine similarity function, more contextual information related to the aspect can be captured. Next, we present the structure of MHSA mechanism, the simple structure containing GloVe pre-trained model, tanh function and SoftMax function is shown in the Fig.2.

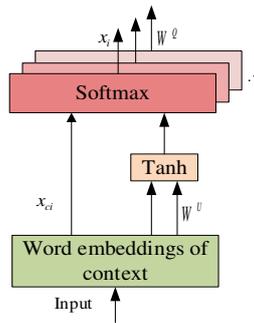


Fig. 2. The structure of MHAS mechanism.

3.1.1 Input Layer

In the field of NLP, because one-hot representation possesses the disadvantages of dimensional disaster and semantics generation gap, in stead of word distributed embedding with advantages of capturing more sematic and syntactic information, so we usually adopt the distributed representations of word embedding. GloVe is selected in our proposed algorithm to train the contexts' feature vectors. We introduce some datasets including SemEval 2014 (including Restaurant and Laptop) and Twitter dataset. Assuming that the size of context is n , namely this context is composed of n words, so the context can be described as $c = (w_1, w_2, w_3, \dots, w_n)$, in which an aspect with m words is contained in a context and the aspect can be depicted as $a = (w_i, w_{i+1}, w_{i+2}, \dots, w_{i+m-1})$. Here, we just consider one situation that one context just contains one aspect, but one context probably takes up several aspects generally.

Therefore, we assume that if the context includes several aspects, we regard every different aspect and the same context as one unclassified example. GloVe pre-trained model is utilized to draw the word embeddings, thence the context representation can be expressed as $hc = (x_{c1}, x_{c2}, x_{c3}, \dots, x_{cn})$, meanwhile the aspect's representation is $ha = (x_{ci}, x_{ci+1}, x_{ci+2}, \dots, x_{ci+m-1})$.

3.1.2 Multi-heads Self-Attention Mechanism

In the part of MHSA mechanism, we introduce two learned matrixes, query matrix $W^Q \in \mathbb{R}^{(n \times d_{dim}) \times d_{dim}}$ and uniform matrix $W^U \in \mathbb{R}^{2 \times d_{dim}}$, and d_{dim} represents the dimension of word embedding.

Firstly, in order to analyze the semantic correlation between each two words, we develop the uniform matrix to learn this dependencies relationship. Thence, the weights between each two words denoting the degree of semantic consistency can be acquired by tanh activate function.

$$U(x_{ci}, x_{cj}) = \tanh(\text{concatenate}[x_{ci}, x_{cj}] \cdot W^U) \quad (1)$$

where $U(x_{ci}, x_{cj})$ denotes the sematic correlation between x_{ci} and x_{cj} . When we select x_{ci} as the central word's representation and calculate the other words' influences on it, the representation of x_{ci} integrating contextual semantics can be deduced by the following equation.

$$x_i = \sum_{j=1}^n \text{softmax}(U(x_{ci}, x_{cj})) \cdot x_{cj} \quad (2)$$

Therefore, the weighted representation of every word can be gained by the same means. The output of single head self-attention mechanism is $hc^1 = (x_1^1, x_2^1, x_3^1, \dots, x_n^1)$.

Moreover, multi-heads self-attention mechanism is adopted to parallelly calculate the feature representations with contextual semantics, which can learn the different semantic information from the various perspective. Concatenating the k heads output representations, and the final output of multi-heads self-attention mechanism is obtained.

$$hc^{mh} = \text{concatenate}[hc^1, hc^2, hc^3, \dots, hc^k] \cdot W^Q \quad (3)$$

where, W^Q is query matrix which can project to a specific semantic space.

In addition, Position-wise Convolution Transformation (PCT) is employed to transform contextual information between the outputs of MHSA, and the equation is shown as follows.

$$hc^* = PCT(hc^{mh}) = \sigma(hc^{mh} * W_1 + b_1) * W_2 + b_2 \quad (4)$$

where, σ is the RELU activate function, and $*$ is the convolution operator. $W_1, W_2 \in \mathbb{R}^{d_{dim} \times d_{dim}}$ are the convolution kernels respectively, and $b_1, b_2 \in \mathbb{R}^{d_{dim}}$ are the bias of kernel functions respectively.

3.1.3 1-Hop Attention Mechanism

In this part, we introduce 1-hop attention mechanism to analyze the influences of aspects on the context from the global perspective.

More concretely, in order to analyze aspect's sentimental influences on every word of text, we simply take the average pooling result of aspects' feature vectors by the pre-trained model, and the context's embedding is original output from PCT models, so the aspect's representation after average pooling operation can be obtained.

$$ha^{ave} = \frac{1}{m} \sum_i^m x_{ci+m-1} \quad (5)$$

where, m is the number of words in an aspect.

In general, we regard the average pooling result $ha^{ave} \in \mathbb{R}^{d_{dim}}$ as an entirety to pay more attention the influences of aspect on other words. Therefore, we adopt the score function to denote the importance, and the score function can be expressed by the following equation.

$$f(ha^{ave}, hc_i^*) = \tanh(hc_i^* \cdot W_3 \cdot ha^{ave} + b_3) \quad (6)$$

where, $hc_i^* \in \mathbb{R}^{d_{dim}}$ denotes i -th word's representation in text; and $W_3 \in \mathbb{R}^{d_{dim} \times d_{dim}}$ is weight matrix and b_3 is bias.

Then we adopt SoftMax function to calculate the normalized weights of 1-hop attention mechanism.

$$\alpha_i = \frac{\exp(f(ha^{ave}, hc_i^*))}{\sum_{i=1}^n \exp(f(ha^{ave}, hc_i^*))} \quad (7)$$

After 1-hop attention mechanism, we obtain the context's representation considering the influences of aspect, so the context's representation is expressed.

$$hc_i^{att} = \alpha_i \cdot hc_i^* \quad i \in [1, n] \quad (8)$$

After the word embeddings fusing the context semantics are obtained, the cosine similarity between aspect-specific and every word in the same contextual semantic space can be calculated to evaluate the influences of every word on aspect-specific. The correlations can be obtained.

$$\beta_i = \frac{\exp(\text{cosine}(ha^{ave}, hc_i^{att}))}{\sum_{i=1}^n \exp(\text{cosine}(ha^{ave}, hc_i^{att}))} \quad (9)$$

Consequently, the final context representation is deduced by the following equation.

$$hc_c = \sum_{i=1}^n \beta_i \cdot hc_i^* \quad (10)$$

3.1.4 Traditional SoftMax Function

Generally, we use SoftMax function to predict the sentiment polarity of context, so we can get the probability distribution by the following formula.

$$y = W \cdot hc_c + b \quad (11)$$

where, y is a 3-dimensional vector, and every dimension indicates the different sentiment, including the *positive*, *neutral*, and *negative*. W and b denotes the weighs and bias of predictive layer.

Moreover, the probability about the polarity of context is calculated by the followings.

$$p_i = \frac{\exp(y_i)}{\sum_{i=1}^3 \exp(y_i)} \quad (12)$$

in which, p_i denotes probability belonging to one of three kinds of sentiment polarity, and i denotes different element of y .

Lastly, the cross-entropy function is utilized to evaluate predictive performance, as the following equation shown.

$$Loss = -\sum_{i=1}^3 y_i \log(p_i) + \lambda \sum_{\theta \in \Phi} \theta^2 \quad (13)$$

where, λ is the coefficient for regulation term, and Φ is the parameters of model.

3.2 Adaptive SVM Classifier

We adopt the first-stage structure to acquire the final context's representation and gather these representations and their actual sentiment labels as dataset of classifier. Given SVM doesn't balance the characteristics of inter-class and intra-class, we design the adaptive mechanism. In detail, we fuse the Fisher regulation to design the classification decision function whose benefit exists in balancing in-cluster-distance and between-cluster-distance of the samples' features, to achieve a better classification performance.

The classification decision function can be expressed.

$$\hat{y} = \text{sign}\left(\sum_{i=1}^Q \alpha_i^* y_i^* K(h, hc_{c_i}) + b^*\right) \quad (14)$$

where, the number of contexts in the whole dataset is Q ; y_i^* is sentimental label of the i -th context's representation; $K(h, hc_c)$ is kernel function, here Gaussian function is selected as kernel function; α_i^* denotes the constant coefficient of support vector; b^* is the bias.

Based on Fisher method and regulation of structural risk minimization, we add the middle parts to the original optimal function of SVM. The middle parts request the ratio of the smallest intra-class divergence and the largest inter-class divergence. In other words, this method requires the contexts' representations have the larger inter-class distance and the smallest intra-class distance in semantic space. The classification plate can be adaptively adjusted by calculating the distances continuously, to produce a more accurate classification result.

The optimal function of ASVM is defined.

$$\min_{w, b, \lambda, \xi} \frac{1}{2} \|w\|_2 + r \frac{w^T S_w w}{w^T S_b w} + C \sum_{i=1}^Q \xi_i \quad (15)$$

$$\begin{aligned} \text{s.t.} \quad & y_i^* (w \cdot hc_{c_i} + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, Q \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, Q \end{aligned}$$

where, $w = \sum_{i=1}^Q \alpha_i^* y_i^* hc_{c_i}$, and $\|w\|_2$ is L_2 regularization; r is constant; C is the penalty parameter; $\xi_i \geq 0$ is the elastic variable. S_w is the sum of intra-class divergence of the same sentiment polarity' samples; S_b is the sum of inter-class divergence of different sentiment polarities' samples.

In our model, by calculating intra-class differences and inter-class differences respectively, S_w and S_b can be obtained, as the following formulas shown.

$$\begin{aligned}
S_w &= \sum_{j=1}^3 \sum_{hc_{c-i} \in T^j, i=1}^{k_j} (hc_{c-i}^j - \mu^j) \cdot (hc_{c-i}^j - \mu^j)^T \\
S_{b1} &= (\mu^1 - \mu^2) \cdot (\mu^1 - \mu^2)^T \\
S_{b2} &= (\mu^2 - \mu^3) \cdot (\mu^2 - \mu^3)^T \\
S_{b3} &= (\mu^3 - \mu^1) \cdot (\mu^3 - \mu^1)^T \\
S_b &= S_{b1} + S_{b2} + S_{b3}
\end{aligned} \tag{16}$$

where, T^j is the sample of different sentiment polarities classified by ASVM; and hc_{c-i}^j is the i -th context's representation corresponding to sentiment polarity $p_j (j=1,2,3)$; μ^j is mean vector of T^j ; k_j is the size of T^j .

Above of all, the architecture of SAN-ASVM model is shown as Fig.3. We utilize the GloVe pre-trained model to acquire word embeddings; then more semantics and sentimental information are obtained by SAN module; at last, ASVM classifier is designed to predicate the sentiment polarity of context

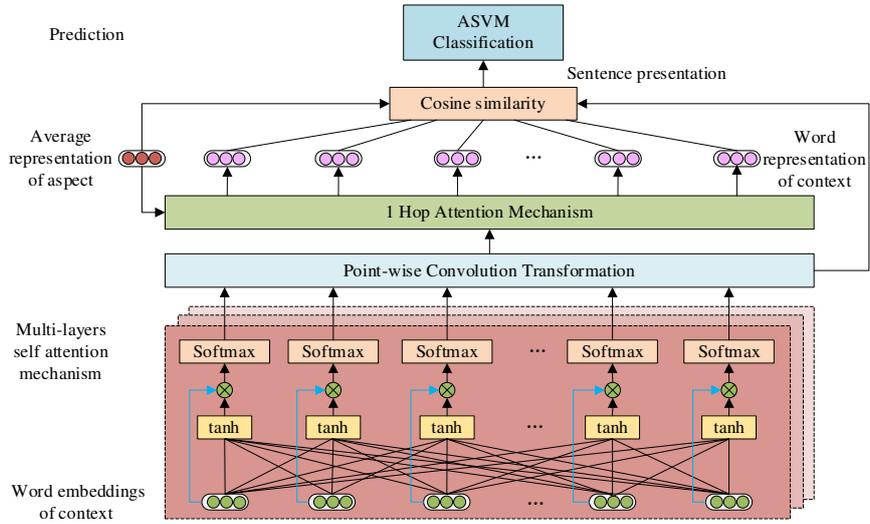


Fig. 2. The structure of MHAS mechanism.

4 Experiments

In this section, a set of experiments and ablation studies of sentiment analysis are conducted on SemEval-2014 Task 4, Restaurant of SemEval-2016 and Twitter datasets to investigate the effectiveness of the proposed method by evaluating Accuracy (Acc) and Macro F1 (MF1).

We mainly design three experiments to prove our proposal: (1) 10 group of experiments on sentiment analysis are conducted to valid the effectiveness of our model; (2) multi-hops attention mechanisms are employed to demonstrate the election of 1-hop is more rational than multi-hops attention mechanism; (3) the compared experiment on classification function about ASVM is designed to achieve the improvement of accuracy.

4.1 Dataset and Evaluation

SemEval-2014 Task4 is selected as experimental dataset, mainly including the Restaurant and Laptop domains. The domain of Restaurant in SemEval-2016 is also selected, for it provides a fine-grained aspect term but Laptop dataset is a coarse-grained term. We also employ the other datasets, Twitter dataset. Their languages are English, all the four domains are composed of users' comments, and each comment contains several sentences which may have one or more aspects and sentiment labels, namely positive, negative, and neutral. The concrete statistics of experimental data is shown as Table 1.

Table 1. The statistics of the dataset

Dataset	Positive		Neural		Negative	
	Train	Test	Train	Test	Train	Test
Restaurant14	2164	341	807	196	637	196
Laptop14	994	728	870	128	464	169
Restaurant16	1620	597	88	38	709	190
Twitter	1561	173	173	346	1560	173

The indicators of sentiment classification, namely Acc and MF1, Acc denotes the ratio of all correct predictions to all samples which can be deduced by the predictive results. However, MF1 is the weighted average value of precision and recall, in which precision and recall also can be calculated by the predictive results. In order to calculate two indicators, the definitions should be listed.

Table 2. The definitions of elements about indicators

Predictive result	Actually	
	Positive	Neural
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

According to the table shown, Acc can be expressed by using the formula: $Acc=(TP+TN)/(TP+FP+FN+TN)$; and the precision refers to the ratio of correct predictions that are actually positive and also predicted to be positive to all predictions that are positive including the actually positive and negative, thence which can be deduce by the equation: $precision=TP/(TP+FP)$. In addition, recall means that the ratio of correct predictions that are positive (TP) to all actually positive samples (TP+FN), which is expressed by the formula: $recall=TP/(TP+FN)$. Therefore, MF1 is expressed as $MF1=2*precision*recall/(precision+recall)$.

4.2 Baseline Models

10 groups of models about sentiment analysis on aspect-specific are chosen to make comparison with our methods.

Feature-based SVM: This model applied traditional feature engineering methods to extract the context feature and then employed the SVM to make sentiment classification (Kiritchenko et al. 2014).

MemNet: This model adopted **Memory Network** to calculate the importance of every word and then generated the context's representation to infer the sentiment polarity of aspect-specific (Chen et al. 2017).

TD-LSTM: Tang et al. employed **Target Dependent LSTM** networks to model the left and right context respectively, and then concatenate hidden state outputs of left and right context to produce the final representation (Tang et al. 2016).

ATAE-LSTM: This model named **ATtention-based LSTM** for **AspECT**-level sentiment classification, which mainly considered the influences of aspect-specific on every word of text, and then calculated the sum of weighted outputs to denote the final representations (Wang et al. 2016).

MGAN: **Multi-Grained Attention Network** was proposed to capture the word-level interaction between aspect and context, especially an aspect alignment loss was designed to fully use the interactions between aspect and context (Fan et al. 2018).

AOA-LSTM: This model introduced **Attention over Attention** mechanism on **LSTM** to model the mutual influences between aspects and context, which could capture the most important word of context influenced by aspect (Huang et al. 2018).

TNet: This model utilized **Transformation Networks** with CNN layer to generate context representations, and in every CPT module, the representations of target and context were combined to generate the next word representation (Li et al. 2018).

RAM: **Recurrent Attention Network** based on **MemNet** introduced the bidirectional LSTM and a gated recurrent unit to produce the context representation (Chen et al. 2017).

CABASC: Content Attention Based Aspect based Sentiment Classification model is proposed, which could capture the important information about given aspects from a global perspective, and considered the order of words and their correlations (Liu et al. 2018).

IAN: Interactive Attention Network was proposed, which analyzed the influences between aspects and context interactively by calculating attention values, to generate the representation of context (Ma et al. 2017).

4.3 Experimental Setting

In this section, we need to pretrain the words from SemEval-2014 Task 4, Restaurant dataset of SemEval-2016 and Twitter dataset. In these four datasets, the Stanford CoreNLP (Manning et al. 2016) is applied as tokenizer to split context. We utilize the GloVe pre-trained model to train the word vectors in pre-trained process. Here we adopt the public pre-trained word embeddings of 300 dimensions. And the batch size of the training data is 32. During training, we usually initialize all matrices with sampling from uniform distribution of $[-0.1, 0.1]$; in order to learn the depth-hyperparameters effectively, the Adam optimizer is adopted to make the errors be backward and update the parameters; and the learning rate is $5e-5$. To avoid the overfitting, the dropout rate is 0.1.

4.4 Analysis of Experimental Results

The results of compared experiments are shown as Table 2. Feature-based SVM based on feature engineering adopted SVM to make sentiment classification, exhibiting a good performance on classification. MemNet model ignored the latent semantics of embeddings, and the performance of MemNet model was worse than Feature-based SVM on the dataset of Restaurant and Laptop. In addition, based on RNN baseline model, TD-LSTM simply analyzed the embeddings of left context and right context, and utilized SoftMax function to predict the sentiment polarity, so that the performance was not good since it did not analyze the influences of target on every word of context. ATAE-LSTM just adopts aspect and every word of context as input of LSTM to calculate the weights, which just considered the sequential influences. IAN mainly took further step to highlight the importance of aspect by learning the interactions between aspect and context in order to generate the concatenated representation. Compared with ATAE-LSTM, 1.4% would be improved in the Restaurant14 dataset and 3.4% improvement occurs in Laptop14 dataset. RAM, TNet and CABASC all have better performances, but on the Twitter dataset, RAM is worse than TNet and CABASC, which illustrates interactive mechanism employed by MGAN and AOA-LSTM could effectively capture the correlation in the small and ungrammatical text.

Table 3. The classification performance of different methods on aspect-level sentiment analysis datasets

Models	Restaurant14		Laptop14		Restaurant16		Twitter	
	Acc	MF1	Acc	MF1	Acc	MF1	Acc	MF1
Feature-based SVM	0.8016	-	0.7049	-	-	-	0.6340	0.6330
MemNet	0.7816	0.6583	0.7033	0.6409	0.8046	0.4979	0.6850	0.6691
TD-LSTM	0.7563	-	0.6813	-	0.8077	0.4912	0.7080	0.6900
ATAE-LSTM	0.7720	-	0.6870	-	0.7923	0.4755	-	-
MGAN	0.8125	0.7194	0.7539	0.7247	0.8123	0.4976	0.7254	0.7081
AOA-LSTM	0.8120	-	0.7450	-	0.8000	0.4718	-	-
TNet-AS	0.8069	0.7127	0.7654	0.7175	0.8015	0.4895	0.7497	0.7360
RAM	0.8023	0.7080	0.7449	0.7135	0.8062	0.5058	0.6939	0.6730
CABASC	0.8089	-	0.7507	-	0.7754	0.4378	0.7153	-
IAN	0.7860	-	0.7210	-	0.7877	0.4657	-	-
GloVe-MHSA/SM	0.7929	0.6820	0.7351	0.6935	0.8001	0.4873	0.7092	0.6548
GloVe-SAN/SM	0.8092	0.7185	0.7532	0.7205	0.8128	0.5091	0.7221	0.6904
GloVe-SAN-ASVM	0.8178	0.7283	0.7583	0.7294	0.8201	0.5182	0.7289	0.7112

In Table 3. GloVe-MHSA/SM model consists of **Multi-Heads Self-Attention** mechanism and **SoftMax** function. And GloVe-SAN/SM model is composed of a whole **Self-Attention Networks** and **SoftMax** function. The last one, utilizing **Adaptive Support Vector Machine** to substitute **SoftMax** function, and to constitute GloVe-SAN-ASVM.

Above of all, we aim at the two issues, long-term dependencies in the sequential model and interactive influences between aspect and context, therefore GloVe-SAN-ASVM sentiment analysis model was designed. According to the compared results, the indicators Acc and MF1 are all improved in various degree on the four datasets. Concretely, selecting GloVe-MHSA/SM as baseline, this model adopted MHSA to calculate the correlations between each two words in context, and compared with IAN, the result proves that the problem of long-term dependences can be solved, moreover compared with TD-LSTM and ATAE-LSTM, the improvement was remarkable. This phenomenon demonstrated MHSA mechanism was more effective to capture the long-term dependences of context than LSTM networks. Besides that, GloVe-SAN/SM solved both issues, compared with GloVe-MHSA/SM, 1.63% improvement occurs on the Restaurant14, 1.81% improvement on Laptop14 dataset, 1.27% improvement on Restaurant16 dataset and 1.29% improvement on Twitter dataset. In addition, on four datasets the performances of GloVe-SAN/SM are superior to AOA-LSTM, RAM, and CABASC respectively, but compared with AOA-LSTM and MGAN the accuracy on Restaurant14 dataset is lower, and the reason is probably these two models analyze the interactive influences between aspect and context by building interactive matrix from the perspective of fine-grained. This is in

line with our main idea about our model, which further illustrated the effectiveness of our method about SAN. Moreover, when adopting ASVM to substitute SoftMax to make classification, the experimental results were shown in Table 3. In comparison with GloVe-SAN/SM and GloVe-MHSA/SM, the classification accuracy was improved in different degrees on each dataset when applying GloVe-SAN-ASVM model, which comprehensively demonstrated ASVM is effective to improve the classification accuracy. Even though, the performance on the first three datasets was superior to other models, but on Twitter dataset, the performance was not better than TNet-AS, the reason was probably that SAN model was not more appropriate to the characteristics (short and unigram) of the Twitter dataset.

4.5 Ablation Studies

In this section, ablation experiments were performed to comprehensively analyze the importance of different mechanisms. Therefore, multiple experiments are designed to verify the effectiveness of every structure.

4.5.1 Analysis of SAN

In order to pay more attention on some vital words related to aspect, we design 1-hop attention mechanism to capture the influences on some key words. By comparing the performances of GloVe-MHSA/SM and GloVe-SAN/SM from Table 2, we valid the effectiveness of 1-hop attention mechanism. Therefore, to study the positive influences on the number of multi-hops mechanism, we design 1-hop and 2-hops mechanism based on SAN model to conduct the experiment.

As the Table 4 shows, we can conclude that SAN adopted 1-hop attention mechanism have more excellent performances than 2-hops attention. In detail, the model SAN/1h achieves better results which outperforms SAN/2h over 2.35%, 6.67%, 14.06% and 3.17% on Restaurant14, Laptop14, Restaurant16 and Twitter datasets respectively in terms of Acc. We think the reason is probably that we adopt the pooling operation of aspect to calculate the influences on context, which discards of some useful information to some extent. Moreover, the variances are smaller and smoother than SAN/2h, which indicates that SAN/1h is more robust.

Table 4. The performance comparison of SAN in respect of multi-hops attention

Models	Indicators	SAN/1h	SAN/2h
Restaurant14	Acc	0.8092 \pm 0.0003	0.7857 \pm 0.0027
	MF1	0.7185 \pm 0.0014	0.6608 \pm 0.0043
Laptop14	Acc	0.7532 \pm 0.0009	0.6865 \pm 0.0018
	MF1	0.7205 \pm 0.0022	0.6289 \pm 0.0024
Restaurant16	Acc	0.8201 \pm 0.0006	0.6795 \pm 0.0003
	MF1	0.5091 \pm 0.0017	0.4398 \pm 0.0017
Twitter	Acc	0.7221 \pm 0.0001	0.6850 \pm 0.0021
	MF1	0.6904 \pm 0.0002	0.6564 \pm 0.0056

In **Table 4**, SAN/1h means this model adopts SAN with 1-hop attention mechanism, and SAN/2h includes 2-hops attention mechanism.

4.5.2 Analysis of ASVM

In addition, based on the results of previous experiments on proving the effectiveness of 1-hop attention mechanism, we further conduct the compared experiments to valid the necessities of ASVM classifier.

Regarding SAN/SM model with SoftMax function as baseline, meanwhile we apply SAN model with standard SVM classifier to realize three classifications of sentiment, and the results manifest SAN-SVM exceeds SAN/SM 0.31% on Restaurant14 dataset, 0.24 % on Laptop14 dataset, 0.48% on Restaurant16 dataset and 0.28% on Twitter dataset in terms of Acc, at the same time MF1 also exhibits a steady increasements. Further, we develop ASVM integrating Fisher regulation to achieve sentiment classification, and compared with SAN/SM model the steady increasement of performances have occurred on the four datasets no matter Acc and MF1, which is a strong argument to valid the effectiveness of ASVM. Therefore, the Fisher regulation inserted into SVM function is very useful to make sentiment classification.

Table 5. The performance comparison of SAN in respect of sentiment classification function

Models	Indicators	SAN/SM	SAN-SVM	SAN-ASVM
Restaurant14	Acc	0.8092 \pm 0.0003	0.8123 \pm 0.0004	0.8178 \pm 0.0002
	MF1	0.7185 \pm 0.0014	0.7215 \pm 0.0011	0.7283 \pm 0.0026
Laptop14	Acc	0.7532 \pm 0.0009	0.7556 \pm 0.0010	0.7583 \pm 0.0007
	MF1	0.7205 \pm 0.0022	0.7262 \pm 0.0021	0.7294 \pm 0.0030
Restaurant16	Acc	0.8128 \pm 0.0003	0.8176 \pm 0.0011	0.8201 \pm 0.0034
	MF1	0.5091 \pm 0.0012	0.5102 \pm 0.0006	0.5182 \pm 0.0010
Twitter	Acc	0.7221 \pm 0.0001	0.7249 \pm 0.0003	0.7289 \pm 0.0002
	MF1	0.6904 \pm 0.0002	0.6998 \pm 0.0004	0.7112 \pm 0.0003

4.5.3 Case Study

In order to further understand the effectiveness of 1-hop attention mechanism and 2-hops attention mechanism, we pitch some examples to visualize the weights of attention mechanism. We design the ablation analysis on SAN with 1-hop or 2-hops attention mechanism to demonstrate the effectiveness. We select one context from test dataset “Everything is so easy to use, Mac software is just so much simpler than Microsoft software”, and analyze which word contributes the biggest influences on the aspects’ sentiment polarities. Therefore, we visualize two groups of weights obtained by SAN with 1-hop or 2-hops attention mechanism, the compared results are shown as the figure 4, in which the color depth indicates the different importance of a word in this context. The first aspect is “use” whose weights distribution is depicted in the first two examples, and the second is “Mac software” whose weights distribution is drawn in the last two examples. As the figure shown, SAN with 1-hop attention mechanism captures the key words “easy” and “simpler” more obviously, and assign the relatively small weights to some adverbs than 2-hops attention, e.g. “so”, and “so much”. Above of all, we can indicate that the SAN with 1-hop attention mechanism on sentiment classification of aspect-specific can strength greatly the ability of extracting the important sentimental words.

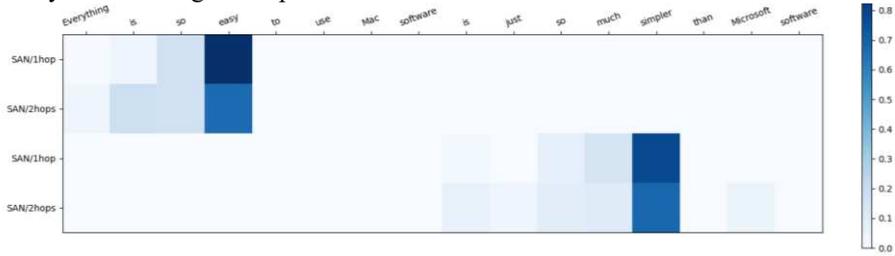


Fig. 4 Examples of attention distribution for sentiment classification on aspect-specific. The color bar denotes the sentimental importance of key words.

5 Conclusions and Future Work

The goal of this paper was to develop an effective sentiment analysis method based on aspect-level. The goal has been successfully achieved by proposing a novelty two-stages architecture namely GloVe-SAN-ASVM for sentiment classification that could effectively extract the context’s feature related to aspect-specific and the performance of sentiment classification was also improved by ASVM.

The effectiveness of the proposed method which adopted two-stages model to conduct sentiment classification was evaluated by comparing with other methods in terms of the Acc and MF1. The experimental results showed that the proposed method could effectively achieve higher classification accuracy than other models on almost all datasets. This is because the MHSA mechanism can effectively mine the latent

semantic dependencies between each two words of context, moreover 1-hop attention mechanism also can excavate the semantic correlations between aspect-specific and vital words of context. Therefore, combining these two mechanisms the more effective feature representations can be learned, besides that ASVM classifier is designed in the second-stage to improve the classification performances. Finally, this paper applied compared experiments to prove the proposal, which can successfully boost the Acc and MF1 of GloVe-SAN-ASVM model based on aspect-specific, and achieve a good result for sentiment classification problems.

However, existing dataset is labeled by human, which is very limited and inefficient, therefore in the future works we aim to research the semi-supervise networks and few shot learning models combined prior language knowledge for sentiment analysis problems.

Compliance with Ethical Standards:

Funding: The authors are grateful to the anonymous reviewers and the editor for their valuable comments and suggestions. This study was funded by The National Key R & D Program of China (Grant No. 2017YFB1302400), National Natural Science Foundation of China (Grant No. 61773242, 61803227), Major Agricultural Applied Technological Innovation Projects of Shandong Province (SD2019NJ014), Shandong Natural Science Foundation (ZR2019MF064).

Conflict of Interest: We confirm that there are no known conflicts of interest associated with this publication.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Pang B, Lee L (2008) Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*. 2(1-2): 1-135
- Mohammad SM (2015) Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text. *Emotion Measurement*. pp: 201-237.
- Tang DY, Bing Q, Feng XC, Liu T (2016) Effective LSTMs for Target-Dependent Sentiment Classification. In: *Proceedings of COLING, the 26th International Conference on Computational Linguistics: Technical Papers*. pp:3298-3307.
- Gabriele P, Lin G, He YL (2019) TDMA: A Topic-Dependent Attention Model for Sentiment Analysis. *Information Processing and Management*. 56(6), DOI: 10.1016/j.ipm.2019.102084.
- Wang YQ, Huang ML, Zhu X (2016) Attention-Based LSTM for Aspect-Level Sentiment Classification. In: *Proceeding of Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. November, pp: 606-615.

- Song YW, Wang JH, Jiang T, (2019) Attentional Encoder Network for Targeted Sentiment Classification. In: Proceedings of 28th International Conference on Artificial Neural Networks. pp:93-103.
- Alyami SN, Olatunji SO (2020) Application of Support Vector Machine for Arabic Sentiment Classification Using Twitter-Based Dataset. Journal of Information and Knowledge Management. 19(1), doi.org/10.1142/S0219649220400183.
- Han KX, Chien W, Chiu CC, Cheng YT (2020) Application of Support Vector Machine (SVM) in the Sentiment Analysis of Twitter Data Set. Applied Science-Basel. 10(3), doi.org/10.3390/app10175944
- Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection. Proceedings of Fourth European Conference on Computer Vision. Cambridge, UK .1996.
- Pontiki M , Galanis D, Pavlopoulos J, Papageorgiou H (2014) SemEval-2014 Task 4: Aspect Based Sentiment Analysis. Proceedings of the 8th International Workshop on Semantic Evaluation. Dublin, Ireland, August, pp: 27-35.
- Dong L, Wei F, Tan C, Tang D, Xu K (2014) Adaptive Recursive Neural Network for Target-Dependent Twitter Sentiment Classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistic. pp: 49-54.
- Shang L,Zhou Z,Liu X (2016) Particle Swarm Optimization-based Feature Selection in Sentiment Classification. Soft Computing. 20, pp:3821-3834.
- Atrio AR, Toni B, Barnes J (2019) On the Effect of Word Order on Cross-lingual Sentiment Analysis. SEPLN. 2(63), pp: 23-30.
- Jamilu A, Bakar A, Yaakubnani A, Ridzwan M (2019) Hybrid N-Gram Model using Naïve Bayes for Classification of Political Sentiments on Twitter. Neural computing & application. 31(12), pp:9207-9220.
- Akshay K, Akhilesh V, Aich A, Hegde Chetana (2018) Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques. Emerging Research in Electronics Computer Science and Technology. ICERECT. 545, pp: 687-696.
- Chauhan C, Sehgal Smriti (2018) Sentiment Classification for Mobile Reviews using KNIME. 2018 International conference on computing, power, and communication technologies. Proceeding Paper. pp:548-553.
- Sridharan K, Komarasamy G (2020) Sentiment classification using harmony random forest and harmony gradient boosting machine. Soft Computing 24, pp:7451-7458.
- Ma Y, Peng H, Cambria E (2018) Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM. In: Proceeding of 32AAAI conference on artificial intelligence. pp: 5876-5883.
- Xue W, Li T (2018) Aspect Based Sentiment Analysis with Gated Convolutional Networks. In: Proceeding of 56th Annual Meeting of the Association for Computational Linguistics (ACL). pp:2514-2523.

- Yi T, Tuan LA, Hui SC (2018) Learning to Attend via Word-Aspect Associative Fusion for Aspect-based Sentiment Analysis, In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. pp: 5956-5963.
- Tang DY, Bing Q, Feng XC, Liu T (2016) Effective LSTMs for TargetDependent Sentiment Classification. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. pp:3298-3307.
- Zhang MS, Zhang Y, Vo DT (2016) Gated Neural Networks for Targeted Sentiment Analysis. In: Proceedings of 30th Association-for-the-Advancement-of-Artificial-Intelligence (AAAI) Conference on Artificial Intelligence. pp: 3087-3093.
- Noh Y, Park S, Park SB (2019) Aspect-Based Sentiment Analysis Using Aspect Map. Applied Science-basel. 9(16), pp:3239-3255.
- Huang BX, Ou YL, Carley KM (2018) Aspect Level Sentiment Classification with Attention-over-Attention Neural Network. In: 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC).pp:197-206.
- Li Xin, Bing L, Lam W, Shi B (2018) Transformation Networks for Target-Oriented Sentiment Classification. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL). 1, pp: 946-956.
- Bengio Y, Simard P, Frasconi P (1994) Learning Long-Term Dependencies with Gradient Descent Is Difficult. IEEE Transactions on Neural Networks. 5(2), pp:157-166.
- Liu Q, Liang B, Xu J, Zhou Q. (2018) A Deep Hierarchical Network Model for Aspect-based Sentiment Analysis. *Chinese Journal of Computers*. 41(12), pp: 2638-2652.
- Mohammad AS, Bashar T, Mahmoud AA, Jararweh Yaser (2019) Using Long Short-Term Memory Deep Neural Networks for Aspect-Based Sentiment Analysis of Arabic Reviews. *International Journal of Machine Learning and Cybernetics*. 10(8), pp:2163-2175.
- Tai KS, Socher R, Manning CD (2015) Improved Semantic Representations from Tree-Structured Long Short-Term Memory Networks. In the proceedings of ACL. pp:1555-1565.
- Li S, Wang QL, Chris C (2018) Independently Recurrent Neural Network (IndRNN): Building A Longer and Deeper RNN. doi: 10.1109/CVPR.2018.00572.
- Liu JJ, Luo B, Yan PF, Wang D, Liu D (2017) Long Short-term Memory based on a Reward/punishment Strategy for Recurrent Neural Networks. 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE. Pp: 327-332.
- Wang J, Peng B, Zhang XJ (2018) Using a Stacked Residual LSTM Model for Sentiment Intensity Prediction. *Neurocomputing*. 332, pp:93-101.

- He KM, Zhang XY, Ren SQ, Sun J (2016) Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition. pp: 770-778.
- Liu G, Guo JB (2019) Bidirectional LSTM with Attention Mechanism and Convolutional Layer for Text Classification. *Neurocomputing*. 337, pp:325-328.
- Ma D, Li S, Zhang X, Wang H (2017) Interactive attention networks for aspect-level sentiment classification. In: Proceedings of the IJCAI. pp:4068-4074.
- Yang M, Qu Q, Chen XJ, Guo C, Shen Y, Lei K (2018) Feature-enhanced Attention Network for Target-Dependent Sentiment Classification. *Neurocomputing*. 307, pp:91-97.
- Cheng J, Zhao S, Zhang J, Zhang JN, King I (2017) Aspect-level Sentiment Classification with Heat (Hierarchical Attention) Network. In: Proceedings of the Conference on Information and Knowledge Management. *ACM*. pp: 91-106.
- Yang C, Zhang HF, Jiang B, Yang W (2019) Aspect-Based Sentiment Analysis with Alternating Coattention Networks. *Information Processing and Management*. 56, pp:463-578
- Wang XB, Chen G, Dependency-Attention-Based LSTM for Target-Dependent Sentiment Analysis. *Social Media Processing*. 2017, 774, 206-217
- Chen P, Sun Z, Bing L (2017) Recurrent Attention Network on Memory for Aspect Sentiment Analysis. In: Proceedings of the EMNLP. pp:452-461.
- Bahdanau D, Cho K, Bengio Y, Neural Machine Translation by Jointly Learning to Align and Translate. arXiv:1409.0473 [cs.CL].
- Alexander R, Sebastian S, Paul O, Stefan Q, Adapt or Get Left Behind: Domain Adaptation through BERT Language Model Finetuning for Aspect-Target Sentiment Classification, 2019, arXiv:1908.11860v2.
- Devlin J, Chang M W, Lee K, Kristina T, BERT: Pre-trained of Deep Bidirectional Transformers for Language Understanding, 2019, arXiv:1810.04805v2.
- Xu H, Liu B, Shu L, Yu P S, BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis, 2019, arXiv:1904.02232v2.
- Pei SW, Wang LL, Shen TM (2019) DA-BERT: Enhancing Part-of-Speech Tagging of Aspect Sentiment Analysis Using BERT. APPT, pp:86-95. DOI:10.1007/978-3-030-29611-7_7.
- Sujata R, Parteek K, A Sentiment Analysis System for Social Media using Machine Learning Techniques: Social enablement. *Digital Scholarship in The Humanities*. 2019, 34(3), 569-581.
- Sarkar S, Ejaz N, Kumar M (2020) Root Cause Analysis of Incidents Using Text Clustering and Classification Algorithms. Proceedings of ICETIT 2019. Emerging Trends in Information Technology. Lecture Notes in Electrical Engineering (LNEE 605). DOI: 10.1007/978-3-030-30577-2_63.
- Kalarani P, Brunda S, Selva (2019) Sentiment Analysis by POS and Joint Sentiment Topic Features using SVM and ANN. *Soft Computing*. 23(16), pp:7067-7079.

- Wang XK, Sheng YQ, Deng HJ, Zhao Z (2019) Char CNN-SVM for Chinese Text Datasets Sentiment Classification with Data Augmentation. *International Journal of Innovation Computing Information and Control*. 15(1), pp: 227-246.
- Laoh E, Surjandari I, Prabaningtyas NI (2019) Enhancing Hospitality Sentiment Reviews Analysis Performance using SVM N-Grams Method. 2019 16th International Conference on Service Systems and Service Management (ICSSSM). 2019.
- de Godoi Brandao, J. Pacheco Calixto, W. (2019) N-Gram and TF-IDF for Feature Extraction on Opinion Mining of Tweets with SVM Classifier. 2019 International Artificial Intelligence and Data Processing Symposium (IDAP). Malatya, Turkey.
- Kiritchenko S, Zhu XD, Cherry C, Mohammad S (2014) Nrc-canada-2014: Detecting Aspects and Sentiment in Customer Reviews. In: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. pp:437-442.
- M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, G. Eryigit, *SemEval-2016 Task 5: Aspect Based Sentiment Analysis*. International Workshop on Semantic Evaluation, 2016 pp.16-17.
- Cheng L, Guo XX, Mei QZ (2017) Deep Memory Networks for Attitude Identification. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. pp: 671-680.
- Fan FF, Feng YS, Zhao DY (2018) Multi-Grained Network for Aspect-level Sentiment Classification. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. pp: 3433-3442.
- Liu Q, Zhang HB, Zeng YF, Huang Z, Wu Z (2018) Content Attention Model for Aspect Based Sentiment Analysis. In: *Proceedings of the 2018 World Wide Web Conference*. doi:10.1145/3178876.3186001
- Manning CD, Surdeanu M, Bauer J, Finkel J (2014) The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. pp:55-60.

Appendix

Abbreviation	Full name
Acc	Accuracy
AMEN	Aspect-Map Extraction Network
ANN	Artificial Neural Network
AOA-LSTM	Attention over Attention mechanism on LSTM
ASVM	Adaptive SVM
ATAE-LSTM	Attention-based Long Short-Term Memory for Aspect-Level
BERT	Bidirectional Encoder Representations from Transformers
CABASC	Content Attention Model for aspect-level sentiment analysis
ELMo	Embedding from Language Model
GNN	Gated Neural Networks
GPT	Generative Pre-trained Transformer
GRU	Gate Recurrent Unit
LSTM	Long Short-Term Memory
IAN	Interactive Attention Networks
MemNet	Memory Network
MF1	Macro F1
MGAN	Multi-Grained Attention Network
MHSA	Multi Heads Self Attention
MHSA/SM	Multi Heads Self-Attention mechanism and SoftMax function
PCT	Position-wise Convolution Transformation
RAM	Recurrent Attention Network based on MemNet
SAN/SM	Self-Attention Networks and SoftMax function
SAN-ASVM	Self-Attention Networks and Adaptive SVM
SVM	Support Vector Machine
TD-LSTM	Target Dependent LSTM
TF-IDF	Term Frequency–Inverse Document Frequency

