Transformer based multi-head attention network for aspect-based sentiment classification

Abhinandan Shirahatti[1], Vijay Rajpurohit[2], Sanjeev Sannakki[2]
[1]Department of Computer Science, KLE Society's Dr. M. S. Sheshgiri College of Engineering and Technology, Visvesvaraya Technological University, Belagavi, Karnataka, India
[2]Department of Computer Science, KLS Gogte Institue of Technology, Visvesvaraya Technological University, Belagavi, India

Article Info

Article history:

Received Oct 28, 2021
Revised Feb 10, 2022
Accepted Feb 18, 2022

Keywords:

Aspect
Attention
Parts-of-speech
Sentiment
Transformer

ABSTRACT

Aspect-based sentiment classification is vital in helping manufacturers identify the pros and cons of their products and features. In the latest days, there has been a tremendous surge of interest in aspect-based sentiment classification (ABSC). Since it predicts an aspect term sentiment polarity in a sentence rather than the whole sentence. Most of the existing methods have used recurrent neural networks and attention mechanisms which fail to capture global dependencies of the input sequence and it leads to some information loss and some of the existing methods used sequence models for this task, but training these models is a bit tedious. Here, we propose the multi-head attention transformation (MHAT) network the MHAT utilizes a transformer encoder in order to minimize training time for ABSC tasks. First, we used a pre-trained Global vectors for word representation (GloVe) for word and aspect term embeddings. Second, part-of-speech (POS) features are fused with MHAT to extract grammatical aspects of an input sentence. Whereas most of the existing methods have neglected this. Using the SemEval 2014 dataset, the proposed model consistently outperforms the state-of-the-art methods on aspect-based sentiment classification tasks.

Corresponding Author:

Abhinandan P. Shirahatti
Department of CSE, KLE DR MSSCET, affiliated to VTU Belagavi
R.C Nagar 2nd Stage Belagavi Karnataka, India
Email: abhinandans2010@gmail.com

1. INTRODUCTION

Natural language processing is a domain at the intersection of artificial intelligence, computer science, and linguistics; here the main aim is to understand the natural language to perform the tasks such as questioning and answering language translations and review analysis. Sentiment analysis is nothing but contextual text mining that identifies the subjective information and extracts this information in the source material [1]. This source material helps in improving business through understanding the social sentiment of a particular service, product, or brand. Sentiment analysis (SA) allows the brand to make use of unstructured data, it possesses various advantages such as real-time analysis, and for instance, product review analysis can help increase the business [2].

Aspect-based sentiment classification (ABSC) is a fine-grained sentiment analysis type and is used to determine the sentiment (e.g. Positive, Negative, Neutral) of an aspect term that is specifically mentioned in the context [3], [4]. An example, 'Laptop's processing speed is incredible! However, the battery life is limited’, the sentiment polarity of the aspects 'processing speed is positive, but for 'battery life' is negative. Aspect-based sentiment classification solves the limitation of sentence-level sentiment classification in that the sentiment polarity of each aspect may fluctuate when a sentence has multiple aspects. ABSC is split into two phases: aspect extraction [5], [6] and sentiment classification [7], [8]. This research focuses solely on the ABSC task. In this field of study, most of the researchers used machine learning and neural network models. Natural language processing (NLP) features like parts-of-speech, and lexical units are used to train the traditional sentiment classification models [9]. For example, ABSC could be performed using a support vector machine (SVM) with well-designed handcrafted features [10]. In recent years, recurrent neural network models such as long short-term memory (LSTM) [11] and gated recurrent units (GRU) [12] have been widely used in aspect-based sentiment classification [13], [14]. Regardless of how effective methods these are, layer models encode words independently, which is tedious. To address this, [15] presented a similar technique based on convolutional neural networks (CNN). Even though CNNs are quite good at lowering training time, they are incapable of capturing long-term dependencies in sentences. Furthermore, aspect-level sentiment polarity is highly reliant on both the aspect and the review context. To incorporate aspect information, several models used an attention mechanism [16], [17]. When an aspect contains multiple words, these approaches overlook the distinct relevance of the words in the aspect phrase, resulting in information loss. We propose an efficient approach, called the multi-head attention transformation network (MHAT), in this study to resolve the difficulties raised above for ABSC tasks. MHAT first embeds context words and their associated parts-of-speech information in word embeddings and then generates contextualized word representations.

The primary contributions of this paper are as shown in:

- We introduce a new model (MHAT) that analyses words in sentences in parallel using a multi-head attention mechanism. MHAT is capable of accurately capturing the global interdependence of the words in a sentence.
- Employed global vectors (GloVe) as an input embedding to amplify the effect of subsequent tasks.

2. RELATED WORK

Aspect-based sentiment classification has gained significant popularity in recent years. The bulk of established approaches relies on classical classifiers (for example, SVM) that are highly dependent on large-scale well-crafted features [9]. Such as bag-of-words, lexemes [18], [19]. However, there is a substantial impact on outcomes produced by these methods depending on the features quality. For example, [20] through huge number of instantaneous features target-specific sentiment classification is accomplished [21]. The supervised machine-learning approach for determining the aspect terms sentiment and aspect groups is described. These models, however, are heavily reliant on the quality of their features [22]. Proposed a framework that focuses on adjectives that appear before or after aspect phrases inside a specified context window [23]. Construct a sentiment lexicon in which word polarities are based on topics or domains using a hierarchical supervision topic model. The Stanford NLP tool is being used to recognize the part of speech of each word and to generate a syntax tree of the input sequences. Secondly, certain studies depend on statistical methods [24]. Employed supervised machine learning models, including naïve Bayes, K-Nearest neighbor, decision trees, and support vector machines (SVM) with syntactic, morphological, and semantic features [25].

To integrate aspect information, attention techniques were applied to sequential models such as LSTM. Ma *et al.* [17] used an attention method to capture interactive information between the context and the aspect [26]. Attention-based LSTM models were used for classification task which improves the model performance. These models learn aspects and context relationships on a wider scale, which results in knowledge loss. To solve this, proposed a sentiment classification [27]. They used a close-grained attention model for word-level learning and coarse-grained attention to obtaining collective information of a sentence. GCSE [28] is a CNN and gating mechanism model, can recognize synchronization while learning. To improve the model's impact, and [29] used Graph convolutional networks (GCN) to model long-term word dependencies, extract grammatical features and create a dependency tree structure.

The proposed approach is based on a transformer that incorporates multi-head attention levels to gather information and rely on residual connections [30], as well as a normalization layer [31]. In this case, the attention mechanism enables our system to parse embedded sequences concurrently. It receives information between every two words, allowing it to understand both word-level information and long-term dependencies in an input sentence.

3. PROPOSED METHODOLOGY

3.1. Task definition

The structure of the proposed model MHAT is presented in this section. Figure 1, depicts the overall architecture. We are given a review sentence $S = \{s_1, s_2, s_3, \dots s_n\}$ and the aspect term $T = \{\ t_1 t_2 t_3, \dots\ t_n\}$. In a given review sentence S, the sentiment type may be *positive*, *negative*, *or neutral*.

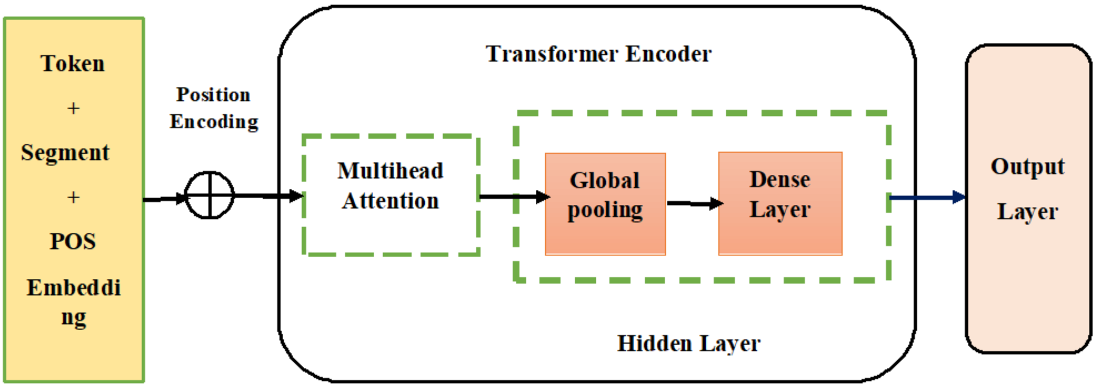

Figure 1. MHAT architecture

3.2. Embedding layer

In this paper we used an efficient word embedding encoder. That is GloVe has been extensively employed in several neural networks for natural language processing problems [32]. We derive word embedding and POS of each word is $P = \{p_1, p_2, p_3, \ldots p_n\} \in R^{n \times dim_p}$. Where$dim_x$ and represents the dimension of both word and Parts-of-Speech embedding respectively. In order to attain the input representation of a word $W = \{w_1, w_2, w_3, \ldots w_n\} \in R^{n \times dim_w}$, we concatenated X and P. Where $dim_w = dim_x + dim_p$. Then, word embedding and aspect term embedding are input into the corresponding transformer encoder. Embeddings are especially instrumental for aspect-based sentiment classification tasks and productively enhance the performance of the downstream task.

3.3. Position encoding

Position encoding is used to feed the model about the tokens relative positions. The embedding vector gains the positional encoding vector. Tokens with the same significance are closer together in d-dimensional space. The position encoding is calculated as (1) and (2).

$$PE(pos, 2_i) = sin(\frac{pos}{1000^{\frac{2i}{dh}}}) \tag{1}$$

$$PE(pos, 2_i + 1) = cos(\frac{pos}{1000^{\frac{2i}{dh}}}) \tag{2}$$

Here, pos is the position, i is the dimension of the vector and distance between two tokens in the sentence would be described by the cosine functions.

3.4. Transformer encoder

Typically, a transformer encoder has two levels: a multi-head attention mechanism and a fully connected layer. Here, multihead attention mechanism (MHA) is used to capture to hidden states of inputs. Without employing sequence-aligned RNNs or convolution, it relies only on self-attention to construct representations of its input and output.

3.4.1. Multi-head attention mechanism (MHA)

Context-aspect word interaction is achieved through multi-head attention. Here, we represent a sequence of task-related Query vectors. Normally; a query q and set of key and value pairs are the inputs for an attention function. Where keys are represented as $k = \{k_1, k_2, k_3, \ldots k_n\}$. Presently in natural language processing tasks key and value are often the same, that is, key= value. In scaled dot-product attention, the weights are computed by the dot products of the query q and the keys as follows. The corresponding dissemination of attention is calculated by a method shown in (3).

$$Attention(k, q) = Softmax\big(s(k, q)\big) \tag{3}$$

Here, $\boldsymbol{s}$ is the semantic gain and it calculates the semantic relevance between context and aspect word. This is also called a scoring mechanism, whose value is determined by a certain procedure and is depicted in (4).

$$s = ktanh\left(\left[k_i, q_j\right].W_a\right) \tag{4}$$

Here, MHA parallelizes the calculation of the input data. MHA acknowledges collaborative learning across delineation batches. The parameters are not going to be shared between the heads and here we are using eight heads. The reason behind this is that k and q values are constantly changing here heads are considered as a parallel scaled dot product mechanism and here the information from different subspaces is learned using linear projections. Finally, multi-head attention is computed as.

$$head_i = Attention(k, q) \tag{5}$$

$$MHA = (head_1 \oplus head_2 \oplus head_3 \dots \oplus head_h).W_O \tag{6}$$

Where, $h \in [1,8]$ $W_O \in R^{(dim_{hidden}) \times (dim_{hidden})}$ $h \in [1,8]$ is the respected weight, dim_{hidden} represents the dimension that is hidden. Attention alleviates the vanishing gradient problem by giving a direct path to the inputs. Here, attention is provided by adding context vector to previous blocks output and hidden state, and context vector is C_i is shown in (7).

$$C_i = \sum_{j=i}^{T_x} a_{ij}\, h_j \tag{7}$$

where a_{ij} denotes attention of i^{th} output should pay to j^{th}input, and is defined by (8) and (9).

$$a_{ij} = softmax\left(e_{ij}\right) = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})} \tag{8}$$

$$e_{ij} = f(S_{i-1},\ h_i) \tag{9}$$

Here the alignment model f, scores how well the inputs around position j and the output at position i match, and S_{i-1}, is the hidden state derived in the preceding time step. The alignment model can be computed as a basic dot product, multiplicative and additive way.

3.5. Model training and regularization

The proposed model is trained from beginning to end by reducing loss as much as possible with L_2 regularization, and is calculated as (10).

$$Loss = -\sum_{i=1}^{C} \quad \hat{y}_i log(y_i) + \lambda L_2(\Theta) \tag{10}$$

In our work$\widehat{y_i}$, indicates the sentiment polarity that was accurately anticipated, y_i indicates the sentiment polarity that was successfully predicted for the supplied input sentence, where i, represents sentence index. L_2Is the regularization, and $\Theta$is parameters list.

4. RESULTS AND DISCUSSION

4.1. Dataset

We evaluate MHAT on Laptop14 dataset, is from “SemEval 2014 Task4” dataset which is related aspect-based sentiment classification task [33], [34]. Table 1, shows the statistics of Laptop14 dataset. Figures 2-4 show the detail analysis of dataset.

Table 1. Dataset details

Dataset	Positive	Negative	Neutral
	Train Test	Train Test	Train Test
Laptop	994 341	870 128	464 169

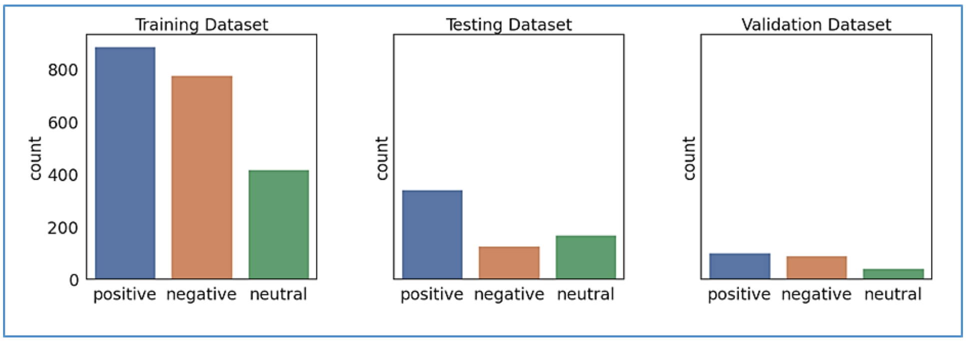

Figure 2. Bar graph representing the distribution of samples for each sentiment category in training, testing, and validation dataset

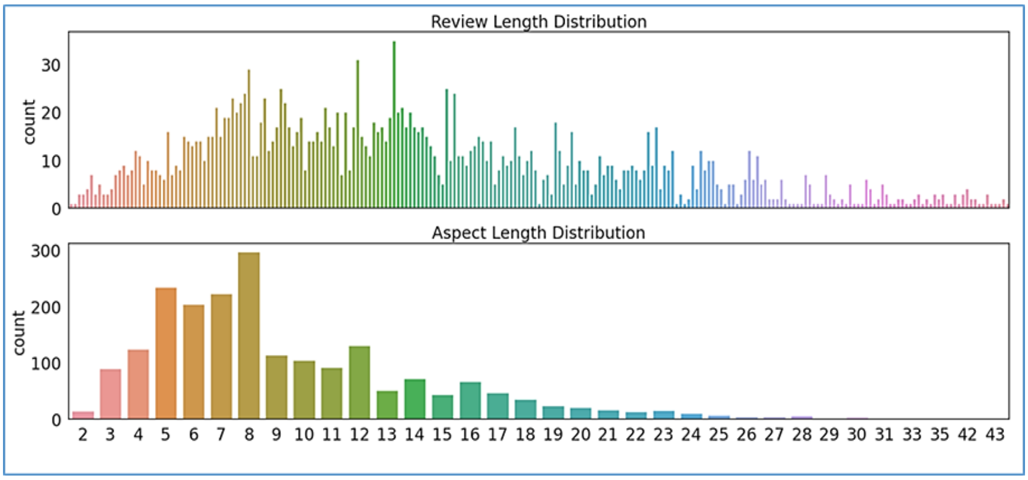

Figure 3. Character length of review and aspect for the training dataset

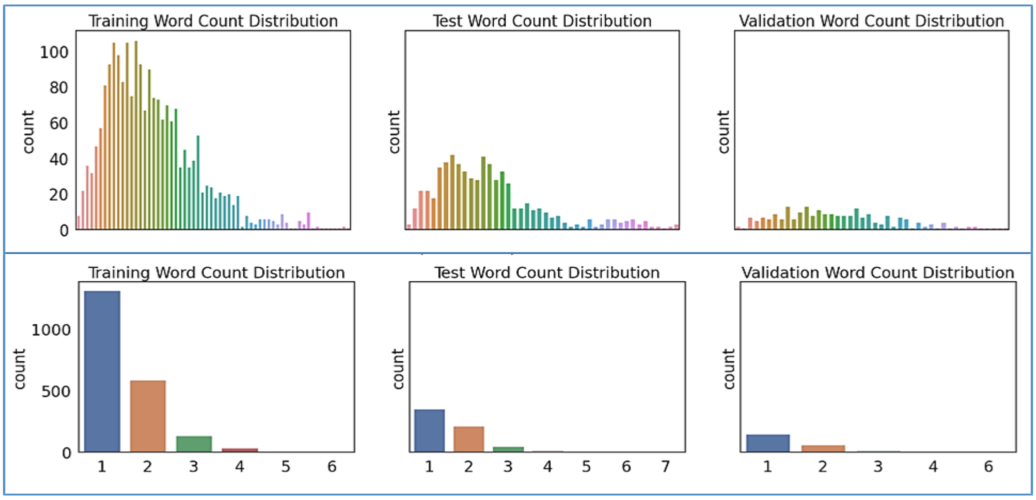

Figure 4. Word count distribution in training, testing, and validation dataset. The top row is for review content and the bottom row is for aspect

4.1.1. Evaluation measure

A common problem in all tasks is determining how to fairly evaluate the model's performance. Different research employs various evaluation metrics. As a result, performing a horizontal comparison becomes quite difficult. We employ the currently common metric, accuracy value, to compare the performance of all models on ABSC. Accuracy is a measurement of how well a classification model performs. This quantity is known as the classification percentage, and it represents the ratio of a number of correct predictions to the total number of input samples.

$$Accuracy = \frac{True_{positve} + True_{negative}}{Total_{no\_of\_samples}} \tag{11}$$

4.2. Baseline models

In our experiments, the proposed model's performance is compared to other models. To ensure fairness and impartiality, the 300d GloVe word embeddings and batch size is 64 used in all baseline methods. All the baseline methods are listed in Table 2.

Table 2. Displays the MHAT comparison findings with other baseline methods. Best result are in bold

Model	Laptop 2014 Accuracy (%)
ATAE-LSTM [13]	68.70
GCAE [15]	69.46
TD-LSTM [14]	71.48
IAN [18]	72.10
MemNet [16]	72.34
RAM [35]	74.51
MGAN [27]	75.39
Tnet-LF [8]	76.32
MAN [36]	78.13
MHAT	81.10

Accuracy metrics were used as an evaluation metric for the proposed model. Table 2, displays the experimental outcomes and the proposed model outperforms as compared to baseline methods. Its accuracy, in particular, demonstrates an improvement of around 2.97% when compared to MAN, the current top model on this dataset. This is due to the fact that the multi-head-attention mechanism determines the relationship between every two words, including contextual information and long-term relationship dependencies, and also we have incorporated the POS features of words into our model, which may make our proposed model more successful at learning the shifting significance of words.

4.3. Model training and analysis

The proposed model is trained using the Laptop14 dataset by considering embedding dimension of 300, with pre-trained GloVe-300 vectors, maximum aspect length as 8 and maximum review content length as 32, maximum tokens as 32000. The model has been trained 30 epochs and training results and model loss for each epoch are shown in Figure 5. The proposed model is able to achieve a training accuracy of about 93% but whereas validation accuracy is about 73%. Whereas on testing dataset proposed model able to achieve accuracy of about 81.10% it shows in the Table 2.

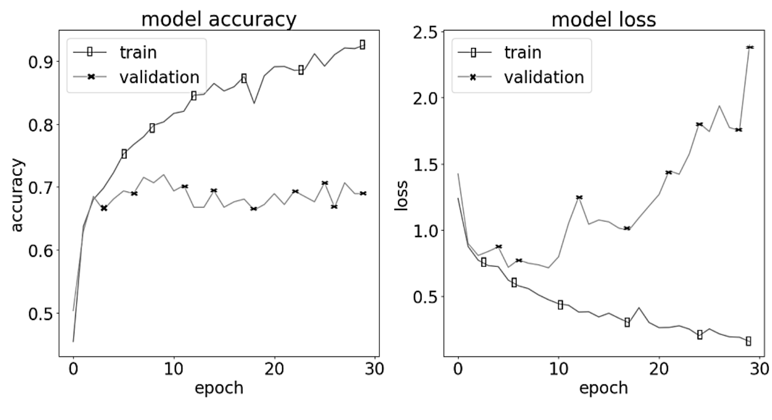

Figure 5. Training and validation accuracy and loss for each epoch

4.4. Model analysis

Here, we examine the impact of MHAT modules, like a stack of layers and attention modules. The transformer encoder includes four attention layers in total, which have an effect on the proposed model's performance. The stacking attention layers are used to manage complicated sentiment connections in the input sequence. However, we evaluate the performance of the proposed model with one to four attention layers. The transformer encoder has 4 layers with 8-head multi-head attention and its model has 4* 8=32 heads, so the model learns the relations of tokens of the input on 32 different standards. Figures 6-9, show the attentions heat maps from the first network layer to the fourth network layer of the network. We have chosen random sentences from the Laptop2014 dataset to demonstrate how attentions vary from layer to layer and also in heads.

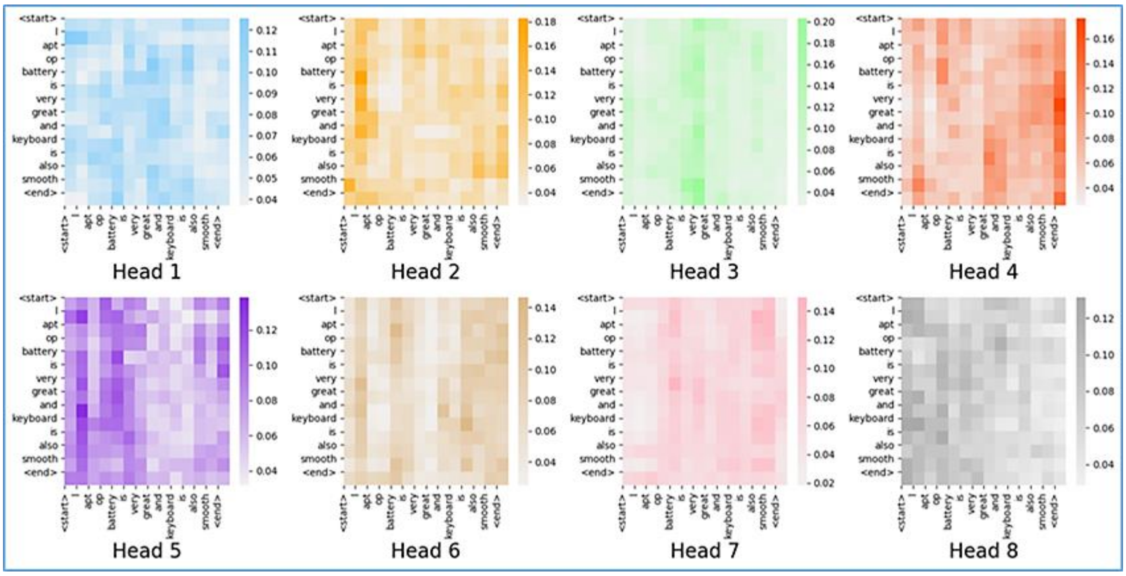

Figure 6. Attentions heat map for the first layer for all eight heads for review sentence "Laptop battery is very good and keyboard is also smooth"

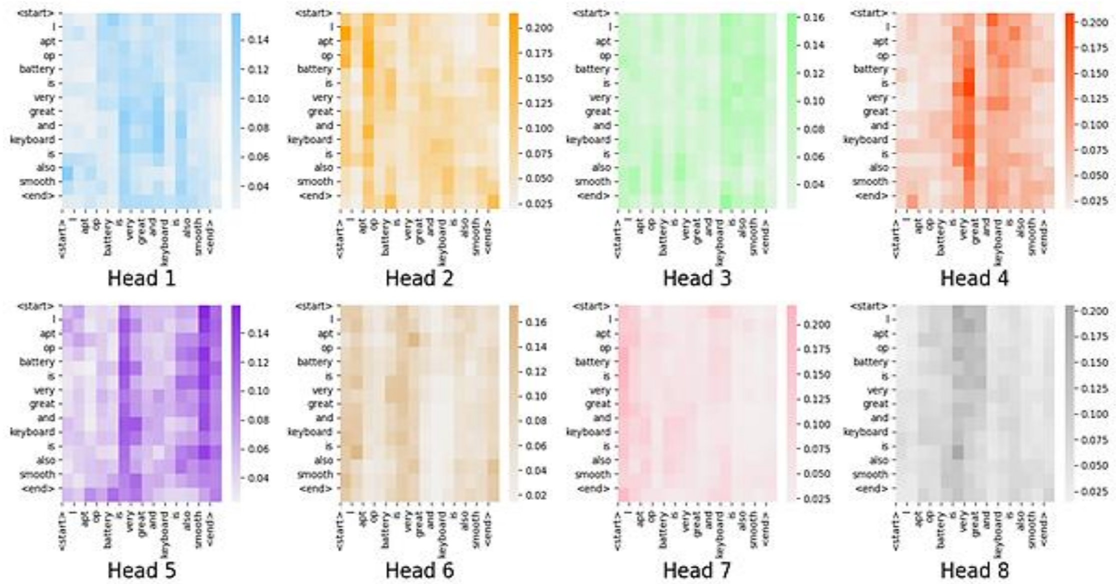

Figure 7. Attentions heat map for the second layer for all eight heads for review sentence "Laptop battery is very good and keyboard is also smooth"

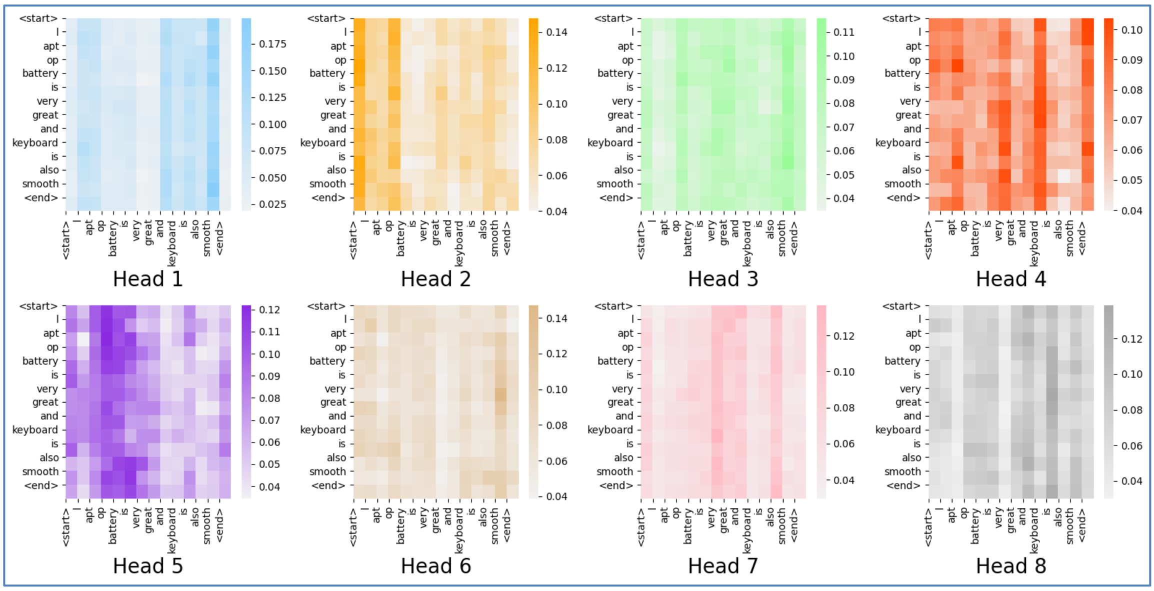

Figure 8. Attentions heat map for the third layer for all eight heads for review sentence "Laptop battery is very good and keyboard is also smooth"

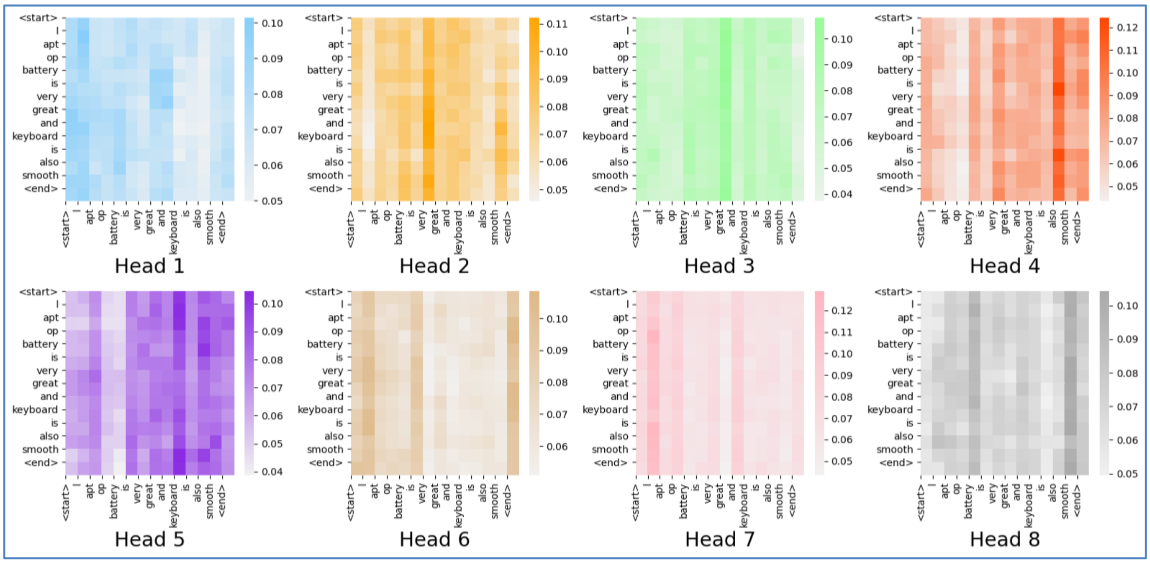

Figure 9. Attentions heat map for the fourth layer for all eight heads for review sentence "Laptop battery is very good and keyboard is also smooth"

5. CONCLUSION

In this work, we created a distinctive approach for tackling the aspect-based sentiment classification problem. We begin by emphasizing the limitations of existing approaches for tackling ABSC through an in-depth analysis. Thus, we developed the MHAT model; it uses an attention strategy to express the context-aspect relationship in a task. As the initial embedding, we start with pre-trained GloVe word vectors, which serve as the foundation for generating cutting-edge results in the next layers. The multi-head attention technique obtains hidden representations in this model. However, to increase the model's performance even further, we integrate POS features into the model, and that might be significant in capturing grammatical features of sentences, as well as providing crucial information about words and their adjacent components. Our model is more effective than other state-of-art methods. ABSC is a complicated and fine-grained work; hence there are still many unsolved challenges in this discipline. For example, some opinion expression plays two roles that is indicating sentiment and implying an (implicit) aspect (target). This will be investigated more in our future study.

REFERENCES

[1] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, Mar. 2016, doi: 10.1109/MIS.2016.31.

[2] W. Zhao *et al*., "Weakly-supervised deep embedding for product review sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 1, pp. 185–197, Jan. 2018, doi: 10.1109/TKDE.2017.2756658.

[3] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008, doi: 10.1561/1500000011.

[4] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–184, May 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[5] L. Shu, H. Xu, and B. Liu, "Lifelong learning CRF for supervised aspect extraction," in *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2017, vol. 2, pp. 148–154, doi: 10.18653/v1/P17-2023.

[6] H. Xu, B. Liu, L. Shu, and P. S. Yu, "Double embeddings and cnn-based sequence labeling for aspect extraction," in *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2018, vol. 2, pp. 592–598, doi: 10.18653/v1/p18-2094.

[7] J. Wang *et al*., "Aspect sentiment classification with both word-level and clause-level attention networks," in *IJCAI International Joint Conference on Artificial Intelligence*, Jul. 2018, vol. 2018-July, pp. 4439–4445, doi: 10.24963/ijcai.2018/617.

[8] X. Li, L. Bing, W. Lam, and B. Shi, "Transformation networks for target-oriented sentiment classification," *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, vol. 1, pp. 946–956, May 2018, doi: 10.18653/v1/p18-1087.

[9] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent Twitter sentiment classification," *ACL-HLT 2011 - Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 151–160, 2011, [Online]. Available: https://aclanthology.org/P11-1016.

[10] V. Perez-Rosas, C. Banea, and R. Mihalcea, "Learning sentiment lexicons in Spanish," *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, 2012, pp. 3077–3081.

[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[12] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint*, Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.3555

[13] Y. Wang, M. Huang, L. Zhao, and X. Zhu, "Attention-based LSTM for aspect-level sentiment classification," in *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2016, pp. 606–615, doi: 10.18653/v1/d16-1058.

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

[15] W. Xue and T. Li, "Aspect based sentiment analysis with gated convolutional networks," in *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2018, vol. 1, pp. 2514–2523, doi: 10.18653/v1/p18-1234.

[16] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," in *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2016, pp. 214–224, doi: 10.18653/v1/d16-1021.

[17] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 0, pp. 4068–4074, 2017, doi: 10.24963/ijcai.2017/568.

[18] Y. Cui, Z. Chen, S. Wei, S. Wang, T. Liu, and G. Hu, "Attention-over-attention neural networks for reading comprehension," *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, vol. 1, 2017, pp. 593–602, doi: 10.18653/v1/P17-1055.

[19] D. Rao and D. Ravichandran, "Semi-supervised polarity lexicon induction," *EACL 2009 - 12th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings*, 2009, pp. 675–682, doi: 10.3115/1609067.1609142.

[20] D. T. Vo and Y. Zhang, "Target-dependent twitter sentiment classification with rich automatic features," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2015-January, pp. 1347–1353, 2015.

[21] S. Kiritchenko, X. Zhu, C. Cherry, and S. M. Mohammad, "NRC-Canada-2014: Detecting aspects and sentiment in customer reviews," in *8th International Workshop on Semantic Evaluation, SemEval 2014 - co-located with the 25th International Conference on Computational Linguistics, COLING 2014*, Proceedings, 2014, pp. 437–442, doi: 10.3115/v1/s14-2076.

[22] V. K. Singh, R. Piryani, A. Uddin, and P. Waila, "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification," in *Proceedings - 2013 IEEE International Multi Conference on Automation, Computing, Control, Communication and Compressed Sensing, iMac4s 2013*, Mar. 2013, pp. 712–717, doi: 10.1109/iMac4s.2013.6526500.

[23] D. Deng, L. Jing, J. Yu, S. Sun, and M. K. Ng, "Sentiment lexicon construction with hierarchical supervision topic model," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 27, no. 4, pp. 704–718, Apr. 2019, doi: 10.1109/TASLP.2019.2892232.

[24] M. Federici and M. Dragoni, "A knowledge-based approach for aspect-based opinion mining," in *Communications in Computer and Information Science*, vol. 641, pp. 141–152, 2016.

[25] M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh, and O. Qawasmeh, "Enhancing Aspect-Based Sentiment Analysis of Arabic Hotels' reviews using morphological, syntactic and semantic features," *Information Processing and Management*, vol. 56, no. 2, pp. 308–319, Mar. 2019, doi: 10.1016/j.ipm.2018.01.006.

[26] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "Exploiting document knowledge for aspect-level sentiment classification," in *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2018, vol. 2, pp. 579–585, doi: 10.18653/v1/p18-2092.

[27] F. Fan, Y. Feng, and D. Zhao, "Multi-grained attention network for aspect-level sentiment classification," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, 2020, pp. 3433–3442, doi: 10.18653/v1/d18-1380.

[28] C. Wu, Q. Xiong, M. Gao, Q. Li, Y. Yu, and K. Wang, "A relative position attention network for aspect-based sentiment analysis," *Knowledge and Information Systems*, vol. 63, no. 2, pp. 333–347, Feb. 2021, doi: 10.1007/s10115-020-01512-w.

[29] B. Zhang, X. Li, X. Xu, K. C. Leung, Z. Chen, and Y. Ye, "Knowledge guided capsule attention network for aspect-based sentiment analysis," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 28, pp. 2538–2551, 2020, doi: 10.1109/TASLP.2020.3017093.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[31] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer Normalization," *arXiv preprint*, Jul. 2016, [Online]. Available: http://arxiv.org/abs/1607.06450.
[32] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2014, pp. 1532–1543, doi: 10.3115/v1/d14-1162.
[33] J. Zhou, J. X. Huang, Q. Chen, Q. V. Hu, T. Wang, and L. He, "Deep learning for aspect-level sentiment classification: Survey, vision, and challenges," *IEEE Access*, vol. 7, pp. 78454–78483, 2019, doi: 10.1109/ACCESS.2019.2920075.
[34] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect based sentiment analysis," in 8*th International Workshop on Semantic Evaluation, SemEval 2014 - co-located with the 25th International Conference on Computational Linguistics, COLING 2014*, Proceedings, 2014, pp. 27–35, doi: 10.3115/v1/s14-2004.
[35] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2017, pp. 452–461, doi: 10.18653/v1/d17-1047.
[36] Q. Xu, L. Zhu, T. Dai, and C. Yan, "Aspect-based sentiment classification with multi-attention network," *Neurocomputing*, vol. 388, pp. 135–143, May 2020, doi: 10.1016/j.neucom.2020.01.024.

BIOGRAPHIES OF AUTHORS

Abhinandan Shirahatti is Assistant Professor in Computer Science and Engineering at KLE DR MSSCET, Belagavi, India and he is currently pursuing PhD studies at the Visvesvaraya Technological University, Belagavi, India in the Department of Computer Science and Engineering. He received his Master and Bachelor of Engineering degrees from the Visvesvaraya Technological University, Belagavi, India in 2012 and 2010, respectively. He can be contacted at email: abhinandans2010@gmail.com.

Dr. Vijay Rajpurohit Professor and Head, Department of CSE at GIT, Belgaum, Karnataka, India. Received B.E. in Computer Science from KUD Dharwad, M.Tech from N.I.T.K Surathkal, and PhD from Manipal University His research interest include Image Processing, Cloud Computing, and Data Analytics. He has good number of publications. He can be contacted at email: vijaysr2k@yahoo.com.

Dr. Sanjeev Sannakki Professor in the department of CSE at GIT, Belgaum, and Karnataka, India. He received his B.E. in ECE from KUD Dharwad in 2009, M.Tech and PhD from VTU, Belagavi. His research interest include Image Processing, Cloud Computing, Computer Networks, and Data Analytics. He has good number of publications. He is the reviewerfor a few international journals. He can be contacted at email: sannakkisanjeev@gmail.com.