A3SN: Amplifying Aspect-Sentence Awareness for Aspect-based Sentiment Analysis

Anonymous ACL submission

Abstract

Aspect-based sentiment analysis (ABSA) is a vital natural language processing task that extracts fine-grained sentiments for specific text aspects, yielding nuanced insights from customer reviews, social media and beyond. Native Sparse Attention (NSA), an efficient alternative to dense attention-based methods, excels at modeling long-context dependencies, local precision and fine-grained features. However, NSA faces three ABSA challenges: (1) Aspect overlap, where proximate aspects trigger selection conflicts; (2) Sparse misses, omitting critical sentiment cues in sparse selections; and (3) Global noise, where token compression dilutes aspect-specific signals. To address these challenges, we introduce a simple yet effective method, Amplifying Aspect-Sentence Awareness (A3SN), a novel method that enhances aspect-sentence interactions by doubling attention weights between aspects and contextual sentences, capturing subtle dependencies precisely. Experimental results on three benchmark datasets demonstrate A3SN's effectiveness, outperforming state-of-the-art (SOTA) baseline models while maintaining simplicity.

1 Introduction

Aspect-based Sentiment Analysis (ABSA), an advanced form of sentiment analysis in Natural Language Processing (NLP), addresses the limitations of traditional sentiment analysis by extracting fine-grained sentiments toward specific aspects or features in text, rather than providing a broad sentiment overview. With the rise of user-generated content on online platforms, ABSA's ability to analyze detailed sentiments in applications like customer feedback and product recommendations offers valuable insights for decision-making and enhanced user experiences.

In the field of ABSA, attention-based (semantic) approaches (Xu et al., 2020; Wang et al., 2016; Peng et al., 2017; He et al., 2018; Cai et al., 2021; Ma et al., 2017; Tang et al., 2016; Zhang et al., 2019; Fan et al., 2018; Wu and Li, 2022; She et al., 2023; Vaswani et al., 2017) have emerged as powerful tools for unraveling the complex sentiments expressed within texts. Work by (Tang et al., 2016) pioneered deep memory networks, highlighting the significance of individual context words in aspect-level sentiment classification. Building upon this, (Wang et al., 2016) introduced attention-based extended short-term memory networks, emphasizing the relationship between sentiment polarity and specific sentence aspects. Further refinement came with interactive attention networks by (Ma et al., 2017), which recognized the importance of modeling targets and contexts separately. Meanwhile, (Peng et al., 2017) proposed a neural framework leveraging multiple attention mechanisms, integrating recurrent neural networks and weighted-memory mechanisms to enhance model capacity. To address information loss with multi-word aspects, (Fan et al., 2018) introduced a multi-grained attention network combining fine- and coarse-grained mechanisms. Advances continued as (He et al., 2018) incorporated syntactic information to refine target representations within attention. Additionally, (Song et al., 2019) presented attentional encoder networks as efficient alternatives to recurrent networks, while (Xu et al., 2020) combined multi-attention networks with global and local attention modules to capture differentially grained interactions between aspects and context. Finally, (Zhao et al., 2024) emphasized robust interactions using multi-head attention networks (MHA).

In parallel, advancements in graph-based (syntactic) approaches (Sun et al., 2019; Wang et al., 2020; Xiao et al., 2021; Lu et al., 2022; Song et al., 2024; Ouyang et al., 2024; Shang et al., 2024) have reshaped the ABSA landscape. For instance, (Sun et al., 2019) proposed combining convolution over a dependency tree (CDT) with bidirectional LSTM (Bi-LSTM) to model sentence structure, enhancing it further with graph convolutional networks (GCNs). (Wang et al., 2020) tackled the problem of linking aspect and opinion terms using relational graph attention networks for more accurate sentiment prediction. Recent innovations like type-aware GCNs (T-GCN) by (Tian et al., 2021) explicitly consider dependency types, improving performance. To represent multiple aspects effectively, (Lu et al., 2022) proposed a heterogeneous graph neural network framework that integrates syntax, word relations, and external knowledge. Similarly, (Song et al., 2024) introduced Knowledge-guided Heterogeneous GCN (KHGCN), which leverages Bidirectional Encoder Representations from Transformers (BERT) and merges sub-word vectors dynamically. (Shang et al., 2024) proposed Aspect Sentence GCN (ASGCN), capturing both grammatical and semantic dependencies for comprehensive ABSA. To overcome limitations of traditional neural models, (Yuan et al., 2024) introduced (SGAN), a syntactic graph attention network incorporating dependency-type knowledge. Lastly, (Ouyang et al., 2024) developed Aspect-based sentiment classification with aspect-specific hypergraph attention networks (ASHGAT), using a word-level relational hypergraph to enhance syntactic relation modeling for sentiment classification.

Recent advances leverage attention mechanisms to model complex linguistic relationships, with Native Sparse Attention (NSA) emerging as an efficient alternative to traditional attention-based methods. NSA excels at capturing long-context dependencies and fine-grained features with reduced computational complexity, making it well-suited for processing extended sequences in ABSA datasets. However, NSA faces three critical challenges, illustrated by the example sentence *"The service was exceptional, but the staff was unhelpful in the restaurant"* from a restaurant review: (1) **Aspect Overlap**: Proximate or semantically related aspects (e.g., *"service"* and *"staff"* in the above example) trigger selection conflicts, leading to ambiguous attention allocations. (2) **Sparse Misses**: The sparse selection mechanism may omit critical sentiment cues (e.g., sentiment-bearing adjectives like *"excellent"*) that fall outside the selected token subset, thereby reducing accuracy. (3) **Global Noise**: Token compression can aggregate irrelevant context, diluting aspect-specific signals and introducing noise particularly in noisy social media data. These limitations, including aspect overlap, sparse misses, and global noise, hinder NSA's ability to accurately model aspect-specific sentiment interactions, leading to reduced sentiment polarity prediction accuracy in ABSA datasets. This underscores the urgent need for a method that enhances NSA's focus on sentiment-critical aspect-sentence relationships.

To address these challenges, we propose A3SN, a simple yet effective novel method designed to enhance ABSA by strengthening aspect-sentence interactions. A3SN doubles attention weights between aspect and contextual sentence tokens using an amplify matrix, enabling the model to capture subtle dependencies with high precision while mitigating issues like aspect overlap, sparse misses, and global noise. The main contributions of this paper are as follows:

- We introduce amplify aspect-sentence awareness attention, which enhances the MHA mechanism by doubling the attention on aspect-sentence relationships. This enhancement helps the model capture subtle relationships and dependencies more accurately, mitigating aspect overlap, sparse misses, and global noise.

- We present A3SN, a novel framework integrating three NSA branches (token compression, selection, sliding window) and an amplify aspect-sentence awareness attention branch.

- The experimental results on three benchmark datasets (Restaurant14, Laptop, and Twitter) showcase the effectiveness of the A3SN model, surpassing SOTA baseline models that incorporate semantic, syntactic, and common knowledge.

2 Related Work

In ABSA, relevant methods are broadly categorized into attention-based models focusing on semantic relationships and hybrid models that combine attention mechanisms with graph-based syntactic structures.

Attention-based neural networks dominate ABSA by effectively capturing semantic relationships between aspects and context. Deep memory networks (Tang et al., 2016) employ external memory and attention to model the importance of context words. (Wang et al., 2016) utilize attention mechanisms to highlight aspect-specific parts of the sentence. Interactive Attention Networks (IAN) (Ma et al., 2017) separately model targets and contexts, thereby enhancing sentiment classification precision. Multi-Grained Attention Networks (MGAN) (Fan et al., 2018) capture fine-grained word-level interactions between aspects and contexts. Semantic Distance Attention (SDA-BERT) (Cai et al., 2021) leverages BERT to extract high-quality semantic features. Multi-Attention Networks (MAN) (Xu et al., 2020) and models based on MHA (Zhang et al., 2019; Wu and Li, 2022; She et al., 2023) integrate intra- and inter-level attention mechanisms often alongside BERT embeddings to improve aspect-context interactions. Conditional BERT augmentation (Wu et al., 2018) further enriches data diversity and reduces overfitting.

Hybrid approaches combine attention with graph-based syntactic models to better encode sentence structure (Tian et al., 2021; Yuan et al., 2024; Xiao et al., 2022; Feng et al., 2022). Relational Graph Attention Networks (R-GAT) (Wang et al., 2020) and BERT4GCN (Xiao et al., 2021) integrate dependency trees and BERT features to enhance sentiment prediction. Type-aware GCNs (T-GCN) (Tian et al., 2021) and gated GCNs (Xiao et al., 2022) leverage syntactic dependency types to improve graph representation learning. Heterogeneous Granph Neural Networks (GNNs) (Lu et al., 2022) incorporate word relations and opinion lexicons into the modeling process. More advanced models such as KHGCN (Song et al., 2024), AS-GCN (Shang et al., 2024), SGAN (Yuan et al., 2024), and ASHGAT (Ouyang et al., 2024) utilize GCNs or GATs to robustly model the interactions between aspects and contexts. AG-VSR (Feng et al., 2022) exemplifies this trend by combining attention-assisted GCNs with variational sentence representations for more robust classification. However, the expressiveness of these hybrid models comes at the cost of increased computational overhead and architectural complexity, which can limit their scalability in large-scale applications.

3 Overview of our Proposed Model Framework

In ABSA, we define the task as a sequence-to-class problem. Given a sentence-aspect pair (s, a), where $s = \{w_1, w_2, \ldots, w_n\}$ represents the sentence tokens and $a = \{a_1, a_2, \ldots, a_m\}$ denotes a subsequence corresponding to the aspect term, the objective is to predict the sentiment polarity of aspect a within the sentence s.

To address the three key challenges of NSA in ABSA, aspect overlap (conflicts arising from proximate or semantically related aspects), sparse misses (omission of sentiment-bearing cues outside selected tokens), and global noise (signal dilution caused by token compression), we introduce (A3SN), a novel attention-based framework. It incorporates four attention branches: compression (`cmp`) to model long-range dependencies, selection (`slc`) to capture fine-grained features, sliding window (`win`) to retain local contextual cues, and amplification (`amp`) to emphasize aspect-sentence relevance. For each branch $C \in \{\texttt{cmp}, \texttt{slc}, \texttt{win}, \texttt{amp}\}$, attention is computed across h heads to extract diverse relational patterns. A gated fusion mechanism then adaptively integrates these representations, enabling the model to concentrate on essential features. This design empowers the model to capture subtle yet crucial relationships between aspects and their sentence contexts, improving performance on ABSA tasks.

3.1 Embedding Module

In BERT (Devlin et al., 2018) encoding, the sentence-aspect pair is structured as `[CLS]` + sentence + `[SEP]` + aspect + `[SEP]`, forming the input sequence. This format allows for extracting an aspect-aware hidden state vector, denoted as h. This aspect-aware hidden state vector serves as a rich representation that incorporates information from both the input sentence and the associated aspect, enabling deeper understanding and more effective analysis in ABSA tasks.

$$h_1, \ldots, h_n = \text{BERT}([w_1, \ldots, w_n]) \quad (1)$$

3.2 Token Compression

The Token Compression branch is designed to capture the broader semantic context in which an aspect appears. In ABSA tasks, understanding the general sentiment of a review or sentence often requires looking beyond the immediate vicinity of the aspect term. To facilitate this, the branch compresses the key and value token sequences, K_t and V_t, into coarser-grained representations that summarize higher-level abstractions over extended spans of the input sequence.

This compression is achieved by dividing the token sequence into overlapping blocks of length l

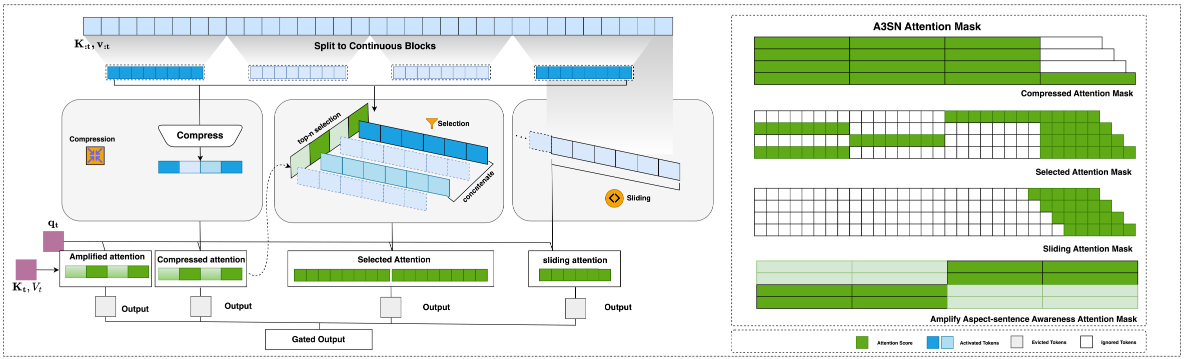

Figure 1: A3SN Architecture. Left: The framework processes through four parallel attention branches: compressed attention for coarse sentiment patterns, selected attention for sentiment-critical tokens, sliding attention for local aspect context, and amplified attention to strengthen aspect-sentiment relationships, Right: Visualization of attention patterns, with green areas indicating computed sentiment scores and white areas showing skipped regions.

with stride d, and applying a learnable transformation ϕ (implemented as an MLP) to each block:

$$q_t = hW_q, \quad K_t = hW_k, \quad V_t = hW_v \tag{2}$$

where W_q, W_k, and W_v are trainable weight matrices.

$$\tilde{K}_t^{\text{cmp}} = \left\{ \phi\left(K_{id+1:id+l}\right) \mid 0 \leq i \leq \left\lfloor \frac{t-l}{d} \right\rfloor \right\} \tag{3}$$

$$\tilde{V}_t^{\text{cmp}} = \left\{ \phi\left(V_{id+1:id+l}\right) \mid 0 \leq i \leq \left\lfloor \frac{t-l}{d} \right\rfloor \right\} \tag{4}$$

The attention weights and final output are then computed as:

$$P_t^{\text{cmp}} = \text{Softmax}(q_t(\tilde{K}_t^{\text{cmp}})^T) \tag{5}$$

$$O_t^{\text{cmp}} = \text{Attn}(q_t, \tilde{K}_t^{\text{cmp}}, \tilde{V}_t^{\text{cmp}}) \tag{6}$$

This reduces the effective sequence length from t to approximately $\lfloor t/l \rfloor$, providing a scalable mechanism for modeling long-range sentiment dependencies in text.

$$\text{Attn}(q_t, \tilde{K}_t^c, \tilde{V}_t^c) = \sum_{i=1}^{t} \frac{\alpha_{t,i} \tilde{v}_i^c}{\sum_{j=1}^{t} \alpha_{t,j}}, \tag{7}$$

$$\alpha_{t,i} = e^{\frac{q_t^\top \tilde{k}_i^c}{d_k}} \tag{8}$$

d_k is the dimension of k

3.3 Token Selection

The Token Selection branch focuses on identifying fine-grained, aspect-specific sentiment cues. Unlike the Token Compression branch, which summarizes the global context, this branch selects the most informative blocks based on attention-derived importance scores.

To ensure consistency, the same block size and stride are used across both branches: $l' = l = d$. The selection branch reuses the attention scores from the compression branch:

$$P_t^{\text{slc}} = P_t^{\text{cmp}} \tag{9}$$

It selects the top-n most relevant blocks:

$$I_t = \{\text{top-}n(P_t^{\text{slc}})\} \tag{10}$$

The corresponding keys and values are concatenated as:

$$\tilde{K}_t^{\text{slc}} = \text{Cat}[K_{il'+1:(i+1)l'} \mid i \in I_t] \tag{11}$$

$$\tilde{V}_t^{\text{slc}} = \text{Cat}[V_{il'+1:(i+1)l'} \mid i \in I_t] \tag{12}$$

The output is then:

$$O_t^{\text{slc}} = \text{Attn}(q_t, \tilde{K}_t^{\text{slc}}, \tilde{V}_t^{\text{slc}}) \tag{13}$$

This branch ensures that specific, sentiment-bearing phrases (e.g., adjectives and opinion terms) are directly considered in the final aspect representation.

3.4 Sliding Window

The Sliding Window branch captures sentiment signals that occur near the aspect. This is especially useful for short texts like tweets or comments, where sentiment expressions are typically located in the immediate neighborhood of the aspect term.

This branch preserves local dependencies by maintaining a fixed-size window of tokens near the current position t:

$$\tilde{K}_t^{\text{win}} = K_{\max(0,t-w):t} \tag{14}$$

$$\tilde{V}_t^{\text{win}} = V_{\max(0,t-w):t} \tag{15}$$

$$O_t^{\text{win}} = \text{Attn}(q_t, \tilde{K}_t^{\text{win}}, \tilde{V}_t^{\text{win}}) \tag{16}$$

The hyperparameter w defines the window size, allowing the model to focus on nearby tokens that may include sentiment-modifying words.

3.5 Amplify Aspect-Sentence Awareness

While the previous branches tackle distinct ABSA challenges, each remains persistently vulnerable to a specific limitation: token selection suffers from *aspect overlap*, where proximate aspects trigger selection conflicts; token compression leads to *global noise*, as it may dilute aspect-specific signals; and sliding window mechanisms often result in *sparse misses*, omitting critical sentiment cues due to fixed-size windowing. To overcome these persistent issues, we propose a fourth attention mechanism A3SN, designed to explicitly enhance interactions between aspect terms and sentence context by *doubling* the attention weights between them. By doing so, we aim to encourage the model to extract and harness richer aspect-sentence information.

This augmentation is achieved by utilizing an amplify matrix, denoted as $\text{Amplify}_{\text{mat}}$, which mirrors the size of the multi-head attention weight matrix, as shown in Figure 1. Given the final input sequence [CLS] w_1, a_1, a_2, w_4 [SEP] a_1, a_2 [SEP], the amplify matrix $\text{Amplify}_{\text{mat}}$ is formulated as:

$$\text{Amplify}_{\text{mat}} = \begin{bmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,N} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N,1} & g_{N,2} & \cdots & g_{N,N} \end{bmatrix} \tag{17}$$

where the elements $g_{i,j}$ are defined as:

$$g_{i,j} = \begin{cases} 2, & \text{if } w_i \in s \;\&\; w_j \in a \text{ or } w_i \in a \;\&\; w_j \in s \\ 1, & \text{otherwise} \end{cases} \tag{18}$$

$$\text{score} = \frac{\text{softmax}(q_t {K_t}^T + \text{Mask})}{\sqrt{d_k}} \tag{19}$$

$$\text{Score}^{\text{amp}} = \text{score} \circ \text{Amplify}_{\text{mat}} \tag{20}$$

$$O_t^{\text{amp}} = \text{score}_{\text{amp}} \cdot V_t \tag{21}$$

where $\circ$ represents the Hadamard product.

This approach enforces stronger interactions between aspect and sentence tokens by magnifying their attention weights, ensuring that even in the presence of overlapping aspects, sparse cues, or noisy compression, the model retains and leverages crucial aspect-specific sentiment signals.

3.6 Attention and Gating Mechanism

To combine the information from different branches: token compression, selection, sliding window, and amplified attention, we use a gating mechanism. Each gate controls the contribution of its respective branch to the final output:

$$h_c = g_t^{\text{cmp}} \cdot O_t^{\text{cmp}} + g_t^{\text{slc}} \cdot O_t^{\text{slc}} + g_t^{\text{win}} \cdot O_t^{\text{win}} + g_t^{\text{amp}} \cdot O_t^{\text{amp}} \tag{22}$$

Here, each gate value $g_t^c \in [0, 1]$ is computed using a learnable function (an MLP followed by a sigmoid activation). These values are dynamically adjusted based on the input context, enabling the model to adaptively weigh local, global, and aspect-specific cues during sentiment inference.

3.7 Training

We employ mean pooling to condense the contextualized embeddings h_c, which assists downstream classification tasks. Following this, we apply a linear classifier to generate logits. Finally, a softmax transformation converts the logits into probabilities, facilitating ABSA. Each component is pivotal in analyzing input text for ABSA tasks, from the embedding layer to the sentiment classification layer.

$$h_{\text{mp}} = \text{MeanPooling}(h_c) \tag{23}$$

$$p(a) = \text{softmax}(W_p h_{\text{mp}} + b_p) \tag{24}$$

Here, W_p and b_p are trainable parameters, consisting of learnable weights and biases.

We utilize the standard cross-entropy loss as our primary objective function:

$$\mathcal{L}(\theta) = - \sum_{(s,a)\in\mathcal{D}} \sum_{c\in\mathcal{C}} \log p(a) \tag{25}$$

computed over all sentence-aspect pairs in the dataset $\mathcal{D}$. For each pair (s, a), representing a sentence s with aspect a, we compute the negative log-likelihood of the predicted sentiment polarity $p(a)$. Here, θ encompasses all trainable parameters, and $\mathcal{C}$ denotes the collection of sentiment polarity classes.

4 Experiment

4.1 Datasets

Our experiments utilize three public sentiment analysis datasets: the *Laptop* and *Restaurant14* review datasets from the SemEval 2014 Task (Pontiki et al., 2014), and the *Twitter* dataset employed by (Dong et al., 2014). For detailed statistics of these datasets, refer to Table 2.

Model	Restaurant14		Laptop14		Twitter	
	Acc.	F1	Acc.	F1	Acc.	F1
ATAE-LSTM (Wang et al., 2016)	77.20	-	68.70	-	-	-
IAN (Ma et al., 2017)	78.60	-	72.10	-	-	-
RAM (Peng et al., 2017)	80.23	70.80	74.49	71.35	69.36	67.30
MGAN (Fan et al., 2018)	81.25	71.94	75.39	72.47	72.54	70.81
BERT (Devlin et al., 2018)	85.79	80.09	79.91	76.00	75.92	75.18
CBERT (Wu et al., 2018)	86.27	80.00	79.83	76.12	76.44	75.35
AEN+BERT (Song et al., 2019)	83.12	73.76	79.93	76.31	74.71	73.13
IMAN+BERT (Zhang et al., 2019)	83.95	75.63	80.53	76.91	75.72	74.50
MAN (Xu et al., 2020)	84.38	71.31	78.13	73.20	76.56	72.19
MAMN_W (Wang et al., 2021)	86.52	81.57	81.35	77.83	76.59	75.27
HN-PMAT+BERT (Wu and Li, 2022)	85.13	76.21	79.71	75.80	75.45	73.30
IMHSACap+BERT (She et al., 2023)	85.00	77.90	81.03	77.62	76.30	75.19
RGAT+BERT (Wang et al., 2020)	86.60	81.35	78.21	74.07	76.15	74.88
BERT4GCN (Xiao et al., 2021)	84.75	77.11	77.49	73.01	74.73	73.76
TGCN+BERT (Tian et al., 2021)	86.16	79.95	80.88	77.03	76.45	75.25
AGVSR+BERT (Feng et al., 2022)	86.34	80.88	79.92	75.85	76.45	75.04
KHGCN+BERT (Song et al., 2024)	-	-	80.87	77.90	-	-
ASHGAT+BERT (Ouyang et al., 2024)	85.49	79.23	79.98	76.58	-	-
A3SN (ours)	**87.32**	**81.27**	**82.05**	**78.92**	**77.97**	**76.33**

Table 1: Experimental results comparison on three publicly available datasets.

Dataset	Division	Pos	Neg	Neu
Rest14	Train	2164	807	637
	Test	727	196	196
Laptop	Train	976	851	455
	Test	337	128	167
Twitter	Train	1507	1528	3016
	Test	172	169	336

Table 2: Statistics of three benchmark datasets

4.2 Implementation

Our A3SN model employs the pre-trained BERT model to extract word representations from the last hidden states (Devlin et al., 2018). We adopt 4 attention heads to enhance representation learning. For the model architecture, we experimented with varying numbers of layers: 3 layers proved optimal for the *Laptop* and *Twitter* datasets, while the *Restaurant* dataset achieved superior performance with a 2 layers. We set the sliding window size to 10, compress block size to 4, and compress block sliding stride to 2. For the selection branch, we use a selection block size of 4 and retain number of selection blocks as 2. To mitigate overfitting, we apply a dropout rate of 0.2. During training, we utilize the Adam optimizer with its default configuration, as outlined in (Kingma and Ba, 2014), to optimize model parameters and promote convergence toward an optimal solution. These representations are fine-tuned during training to adapt to our specific ABSA task. We implement the model using the PyTorch framework, which supports efficient and scalable training.

4.3 Baseline Comparisons

We conduct a comprehensive comparison with state-of-the-art (SOTA) baselines to evaluate the effectiveness of our model against both attention-based models: ATAE-LSTM (Wang et al., 2016), IAN (Ma et al., 2017), RAM (Peng et al., 2017), MGAN (Fan et al., 2018), BERT (Devlin et al., 2018), CBERT (Wu et al., 2018), AEN (Song et al., 2019), IMAN (Zhang et al., 2019), MAN (Xu et al., 2020), MAMN_W (Wang et al., 2021), HN-PMAT (Wu and Li, 2022), IMHSACap (She et al., 2023) and hybrid-based models: RGAT+BERT (Wang et al., 2020), BERT4GCN (Xiao et al., 2021), TGCN (Tian et al., 2021), AG-VSR+BERT (Feng et al., 2022), KHGCN (Song et al., 2024), ASH-

Text	A3SN w/o Amplified Attention	A3SN	Labels
Then the [system]$_{\text{neg}}$ would many times not [power down]$_{\text{neg}}$ without a forced power-off	(N✓, N✓)	(N✓, N✓)	(N, N)
Our [waiter]$_{\text{pos}}$ was friendly and it is a shame that he didn't have a supportive [staff]$_{\text{neg}}$ to work with.	(P✓, P×)	(P✓, N✓)	(P, N)
Both a number of the [appetizer]$_{\text{pos}}$ and [pasta specials]$_{\text{pos}}$ were amazing.	(P✓, P✓)	(P✓, P✓)	(P, P)
Great [food]$_{\text{pos}}$ but the [food]$_{\text{neg}}$ was dreadful!	(P✓, P×)	(P✓, N✓)	(P, N)
It was our only opportunity to visit and wanted an authentic [Italian meal]$_{\text{neu}}$.	(O✓)	(O✓)	(O)

Table 3: Case studies comparing A3SN with and without the amplified attention module. Predictions are shown alongside gold sentiment labels.

GAT+BERT (Ouyang et al., 2024).

The overall performance of all the models is shown in Table 1, from which several observations can be noted. The A3SN model outperforms all baseline models across all three datasets, surpassing those that incorporate semantic, syntactic, and external knowledge information. Notably, the performance of BERT is significantly enhanced when integrated with A3SN, even without relying on syntactic structures or additional knowledge resources, resulting in a simpler yet more effective model. This demonstrates that A3SN improves the model's capacity to understand and utilize the complex relationships between sentences and their corresponding aspects, thereby boosting overall performance. Furthermore, the results reaffirm the effectiveness of the pre-trained BERT model in the ABSA task, as it already outperforms several existing models. Nevertheless, integrating A3SN yields further performance gains, indicating that even strong base models benefit from our proposed method. these findings collectively confirm the efficacy of A3SN in capturing essential interactions between aspect terms and contextual sentences for sentiment classification, while effectively addressing challenges such as sparse misses and aspect overlap.

4.4 Ablation study

We performed ablation experiments on the Restaurant, Laptop14, and Twitter datasets to investigate the impact of each component in the proposed A3SN model. The results are presented in Table 4. Specifically, the variant "w/o amplified attention" removes the amplified aspect-sentence awareness attention branch, which plays a crucial role in addressing challenges such as aspect overlap and sparse aspect-opinion associations. Excluding this component leads to notable performance drops of 2.10%, 1.84%, and 1.86% in accuracy on the Restaurant, Laptop, and Twitter datasets, respectively, highlighting its effectiveness in enhancing fine-grained aspect-context interaction. The "w/o compression attention" setting removes the representation derived from the token compression mechanism. This ablation results in a decline of 1.21%, 0.68%, and 1.19%, indicating that token compression plays a meaningful role in reducing noise and preserving salient contextual information. Similarly, the "w/o selection attention" variant omits the contribution of the token selection module. We observe a performance reduction of 1.29%, 0.72%, and 1.15%, suggesting that selecting high-relevance tokens is important for modeling aspect-aware sentiment signals. Lastly, removing the sliding window attention component in "w/o sliding window" leads to accuracy reductions of 0.68%, 0.44%, and 1.00%, confirming the benefit of localized context aggregation over fixed-length segments. These results highlight that each component contributes to A3SN's ability to model fine-grained aspect-context interactions. The full A3SN model achieves the best performance, demonstrating the effectiveness of jointly integrating compressed, selected, and localized attention with the amplified aspect-sentence awareness mechanism for robust ABSA.

Model	Rest14 Acc.	Lapt14 Acc.	Twit Acc.
A3SN	**87.32**	**82.05**	**77.97**
w/o amplified attention	85.22	80.21	76.11
w/o compression attention	86.11	81.37	76.78
w/o selection attention	86.03	81.33	76.82
w/o sliding window	86.64	81.44	76.97

Table 4: Ablation study results on three benchmark datasets (%).

4.5 Case Study

To further evaluate the effectiveness of A3SN in modeling semantic information and capturing fine-

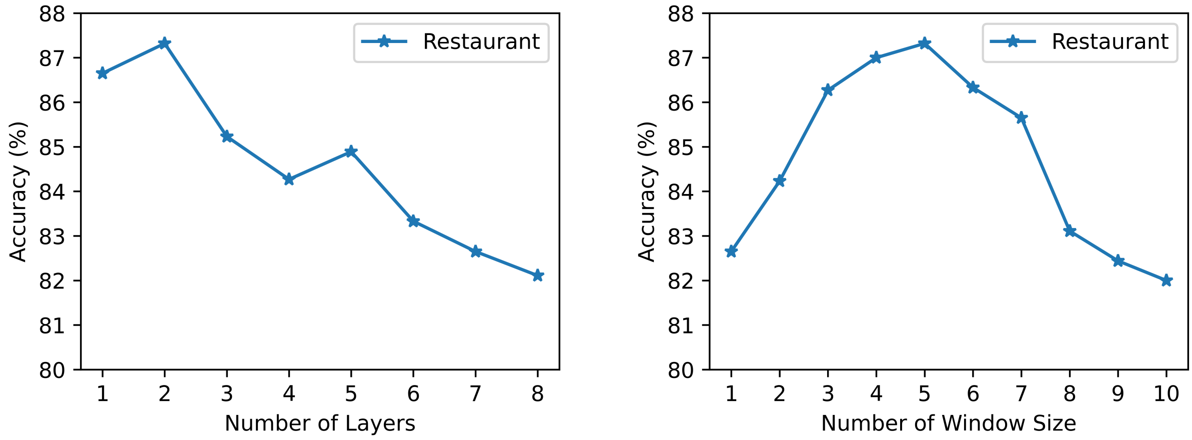

Figure 2: Effect of the number of A3SN layers (l) and the window size (w)

grained relationships between aspect terms and contextual sentences for ABSA, we conducted a qualitative case study using representative sample inputs. Table 3 presents predictions alongside ground truth labels for selected examples.

Consider the sentence: "Our waiter was friendly, and it is a shame that he didn't have a supportive staff to work with." This sentence features two aspect terms, "waiter" (positive sentiment) and "staff" (negative sentiment), presenting challenges for sparse attention models due to aspect overlap and diluted contextual signals. A3SN accurately predicts the sentiment polarity for "waiter" using its four-branch attention framework. The compressed attention for coarse sentiment patterns, selected attention for sentiment-critical tokens (e.g., "friendly"), and sliding attention for local aspect context. The amplified attention branch then strengthens the alignment between "waiter" and its sentiment-bearing context, disambiguating it from "staff." This layered strategy enhances A3SN's ability to resolve aspect overlap, suppress irrelevant noise, and prioritize relevant contextual cues. In contrast, an ablated A3SN without amplified attention incorrectly predicts (waiter, negative) and (staff, positive), likely due to aspect overlap between the semantically similar "waiter" and "staff," underscoring the amplified attention's critical role in precise sentiment alignment. The same case applies when the input aspect is "staff."

4.6 Effect of Number of Layers

In A3SN's evaluation on ABSA datasets, we observed distinct optimal layer configurations for the Laptop and Restaurant datasets, as shown in Figure 2 (on the left hand side). The Laptop dataset achieved the highest sentiment polarity prediction performance with three layers, reflecting its complex sentence-aspect relationships that require deeper modeling. In contrast, the Restaurant dataset performed best with a double layer, leveraging its simpler structure and direct sentiment associations. This variation underscores the importance of tailoring model depth to dataset complexity. Using too few layers for complex datasets like Laptop risks insufficient modeling, while excessive layers for simpler datasets like Restaurant may lead to overfitting. These findings guide A3SN's configuration for precise sentiment analysis across diverse ABSA datasets.

5 Conclusion

In this work, we introduce A3SN to address the issues of aspect overlap, sparse misses, and global noise, presenting a novel framework that significantly advances ABSA. By integrating four attention mechanisms, A3SN captures fine-grained, global, and local sentiment dependencies critical for precise sentiment polarity prediction. The enhanced MHA mechanism amplifies aspect-sentence interactions, effectively modeling complex relationships between aspects and their contextual sentences across benchmark dataset. The gated fusion mechanism further integrates these feature representations. Experimental results on three benchmark datasets confirm A3SN's superior performance over some SOTA baseline models, achieving remarkable effectiveness while maintaining simplicity. These results validate the robustness and efficiency of A3SN, making it a valuable advancement in sentiment analysis.

Limitations

A3SN's evaluation is limited to English-language datasets, and its performance on multilingual or low-resource languages remains untested, potentially restricting its applicability in diverse linguistic contexts. Secondly, A3SN's design primarily targets explicit aspects, but may struggle with implicit aspects. Incorporating aspect detection could improve performance.

Acknowledgments

References

Xiaohong Cai, Hui Cao, Jingang Ma, Ming Li, and Xuqiang Zhuang. 2021. Aspect level sentiment classification with semantic distance attention networks. In *ACM International Conference Proceeding Series*. Association for Computing Machinery.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019*.

Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014. Adaptive recursive neural network for target-dependent twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 49–54. Association for Computational Linguistics.

Feifan Fan, Yansong Feng, and Dongyan Zhao. 2018. Multi-grained attention network for aspect-level sentiment classification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3433–3442.

Shi Feng, Bing Wang, Zhiyao Yang, and Jihong Ouyang. 2022. Aspect-based sentiment analysis with attention-assisted graph and variational sentence representation. *Knowledge-Based Systems*, 258.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018. Effective attention modeling for aspect-level sentiment classification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1121–1131.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. In *ICLR 2015*.

Guangquan Lu, Jiecheng Li, and Jian Wei. 2022. Aspect sentiment analysis with heterogeneous graph neural networks. *Information Processing and Management*, 59.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. In *IJCAI'17: Proceedings of the 26th International Joint Conference on Artificial Intelligence*.

Jihong Ouyang, Chang Xuan, Bing Wang, and Zhiyao Yang. 2024. Aspect-based sentiment classification with aspect-specific hypergraph attention networks. *Expert Systems with Applications*, 248:123412.

Chen Peng, Sun Zhongqian, Bing Lidong, and Wei Yang. 2017. Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 452–461.

Maria Pontiki, Haris Papageorgiou, Dimitrios Galanis, Ion Androutsopoulos, John Pavlopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation*, pages 27–35.

Wenqian Shang, Jiazhao Chai, Jianxiang Cao, Xia Lei, Haibin Zhu, Yongkai Fan, and Weiping Ding. 2024. Aspect-level sentiment analysis based on aspect-sentence graph convolution network. *Information Fusion*, 104.

Lina She, Hongfang Gong, and Siyu Zhang. 2023. An interactive multi-head self-attention capsule network model for aspect sentiment classification. *Journal of Supercomputing*.

Xiangxiang Song, Guang Ling, Wenhui Tu, and Yu Chen. 2024. Knowledge-guided heterogeneous graph convolutional network for aspect-based sentiment analysis. *Electronics (Switzerland)*, 13.

Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao. 2019. Targeted sentiment classification with attentional encoder network. In *Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series*, pages 93–103, Cham. Springer International Publishing.

Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019. Aspect-level sentiment analysis via convolution over dependency tree. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 5679–5688.

Duyu Tang, Bing Qin, and Ting Liu. 2016. Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 214–224.

Yuanhe Tian, Chen Guimin, Yan , and Song. 2021. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2910–2922.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*.

Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. Relational graph attention network for aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Xiaodi Wang, Mingwei Tang, Tian Yang, and Zhen Wang. 2021. A novel network with multiple attention mechanisms for aspect-level sentiment analysis. *Knowledge-Based Systems*, 227.

Yequan Wang, Minlie Huang, Li Zhao, and Xiaoyan Zhu. 2016. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615.

Xing Wu, Shangwen Lv, Liangjun Zang, Jizhong Han, and Songlin Hu. 2018. Conditional bert contextual augmentation.

Yuchen Wu and Weijiang Li. 2022. Aspect-level sentiment classification based on location and hybrid multi attention mechanism. *Applied Intelligence*, 52:11539–11554.

Luwei Xiao, Xiaohui Hu, Yinong Chen, Yun Xue, Bingliang Chen, Donghong Gu, and Bixia Tang. 2022. Multi-head self-attention based gated graph convolutional networks for aspect-based sentiment classification. *Multimedia Tools and Applications*, 81:19051–19070.

Zeguan Xiao, Jiarun Wu, Qingliang Chen, and Congjian Deng. 2021. Bert4gcn: Using bert intermediate layers to augment gcn for aspect-based sentiment classification. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9193–9200.

Qiannan Xu, Li Zhu, Tao Dai, and Chengbing Yan. 2020. Aspect-based sentiment classification with multi-attention network. *Neurocomputing*, 388:135–143.

Li Yuan, Jin Wang, Liang Chih Yu, and Xuejie Zhang. 2024. Syntactic graph attention network for aspect-level sentiment analysis. *IEEE Transactions on Artificial Intelligence*, 5:140–153.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks.

Ying Zhao, Tingyu Xia, Yunqi Jiang, and Yuan Tian. 2024. Enhancing inter-sentence attention for semantic textual similarity. *Information Processing and Management*, 61.

A Effect of Window Size w

To determine the optimal w, we conducted experiments on the Restaurant and Laptop datasets, evaluating window sizes $w = \{1, 2, ..., 10\}$. Performance was measured using the accuracy for sentiment polarity classification. A window size of $w = 5$ consistently yielded the highest accuracy across all datasets, balancing the capture of local sentiment cues with sufficient contextual breadth. Smaller windows (e.g., $w = 3$) missed relevant context, while larger windows (e.g., $w = 10$) included less relevant tokens, diluting sentiment focus. The optimal $w = 5$ ensures that A3SN effectively models short-range dependencies, as illustrated in Figure 2 (on the right hand side).