Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Weighted Words Multi-Domain Model for Aspect-Opinion Pairs Extraction

Atefeh Mohammadi*, *Mohammad Reza Pajoohan**, *Ali Mohammad Zareh Bidoki
Department of Computer Engineering, Yazd University, Yazd, Iran.

Article Info

Article History:
Received 19 February 2025
Reviewed 27 April 2025
Revised 17 May 2025
Accepted 06 June 2025

Keywords:
Multi-domain
Opinion entities
Sentiment analysis
Attention network

*Corresponding Author's Email Address: *pajoohan@yazd.ac.ir*

Abstract

Background and Objectives: In Natural Language Processing (NLP), sentiment analysis is crucial for understanding and extracting aspects and opinions expressed in textual data. Recent methods have emphasized determining polarity in multi-domain sentiment analysis while paying less attention to aspect and opinion extraction. Furthermore, in different domains, the terms conveying aspects and opinions may have different importance, and this difference should be considered to enhance the extraction of aspect-opinion pairs.

Methods: To address these challenges, a Weighted Words Multi-Domain (WWMD) model is proposed for extracting aspect-opinion pairs, consisting of a self-attention mechanism and a dense network. Each word's significance is extracted by the self-attention mechanism and according to the sentence's overall meaning. The dense network is employed for domain prediction. It assigns greater weight to words relevant to each domain, resulting in considering the different significance of terms across various contexts. Adding an attention mechanism to the domain module provides a clearer understanding of different aspects and opinions across various domains. A two-channel approach is utilized; one channel extracts aspects and opinions, while the other extracts the relationships between them. This model's weighted words are simultaneously considered as the input for both channels.

Results: The model output will be improved using weighted words specific to each domain.

Conclusion: Evaluation results on benchmark datasets demonstrate the superiority of the proposed model compared to state-of-the-art techniques.

How to cite this paper:
A. Mohammadi, M. R. Pajoohan, A. M. Zareh Bidoki, "Weighted words multi-domain model for aspect-opinion pairs extraction," J. Electr. Comput. Eng. Innovations, 14(1): 41-54, 2026.

DOI: 10.22061/jecei.2025.11741.832

URL: https://jecei.sru.ac.ir/article_2348.html

Introduction

Sentiment Analysis, or opinion mining, uses natural language processing to identify subjective information from text data [1], [2]. An entity in this context refers to a specific item or subject evaluated for sentiment. It operates at three levels: document level, sentence level, and aspect level [3], [4].

Aspect-opinion pair extraction is a task in natural language processing that entails recognizing and extracting pairs of aspects (features or entities) and opinions (sentiments or evaluations) from textual data. This task is of paramount importance for sentiment analysis, opinion mining, and various text analysis applications [5]. Typically, a sentence includes one or more aspects or opinions. For example, in the sentence "The hotel rooms were very clean, but the staff was not friendly," the aspect-opinion pairs of (rooms, clean) and (staff, not friendly) from opinion entity extraction can be obtained.

Multi-domain sentiment classification refers to a complex area in natural language processing. It uses data from various domains to improve classification performance [6], [7], and [8]. The goal is to train a classifier with a suitable set of labelled data, reducing the need for extensive domain-specific data and addressing data scarcity by leveraging information from other domains [9]. The challenge lies in the different meanings of words across domains; for instance, "bland" can refer to emotional disengagement in book reviews but describe flavorlessness in restaurant reviews [10]. Consequently, a classifier trained in one domain may struggle to interpret sentiments accurately in another due to unfamiliar word meanings.

While the attention mechanism has been proven effective in detecting beneficial aspects, its applicability across multi-domain sentiment analysis remains a critical area that demands further exploration.

Chen et al. [11] introduced the Synchronous Double-channel Recurrent Network (SDRN), comprising primarily an opinion entity extraction unit, a relation detection unit, and a synchronization unit. The opinion entity extraction and relation detection units operate as two channels, enabling the simultaneous extraction of opinion entities (aspects and opinions) and relations. The opinion entity refers to extracting opinions and aspects, while the relation detection unit extracts relations between aspects and opinions in sentences. The challenges of the above model are discussed, examples are provided to illustrate these challenges, and ways to improve them are proposed.

Consider the following two sentences.

"This game is too difficult."

"This microwave is difficult to use."

Upon comparison, we observe that both reviews include "difficult". However, in the first review, "difficult" conveys a meaning of "challenging/attractive," while in the second review, it signifies "inconvenient/not user-friendly". Consequently, these two sentences express entirely different sentiments. A Weighted Words Multi-Domain (WWMD) model for aspect-opinion pairs extraction is employed to first capture the importance of each word in the context of the sentence's overall meaning. It then examines the internal dependencies of words within the same sentence to identify interdependent and collocated words. A fully connected network with Softmax is employed for domain prediction, assigning higher weights to words linked with a specific domain through training. The network learns the relationships between words and domains throughout training through adjusting the connection weights among neurons. According to the training data, the words indicating a specific domain will have increased associated weights. This prioritization empowers the network to give these words a more significant influence in predicting the domain for a given input. Similarly, following the SDRN model [11], two channels are used to extract aspects, opinions, and the relationships between them; however, the weighted words specific to each domain acquired by our proposed model replaced word embeddings from BERT.

Fig. 1 illustrates the aspect-opinion pairs extraction in multi-domain contexts, consisting of five main steps: Encoder Layer, WWMD Layer, Opinion Entities Identification Layer, Relation Detection Layer, and Aspect-Opinion Pairs Extraction Layer. The first step involves the encoding layer of the input sentence. The second one is the weighting layer, highlighting domain-specific words. The third step focuses on identifying opinion entities to extract aspects and opinions. The fourth step involves identifying the relationships between aspects and opinions, while the last step focuses on the extraction of aspect-opinion pairs.

The main contributions of our paper are summarized as follows:

- The weight to each word is assigned by the WWMD layer based on the overall meaning of the sentence and the corresponding domain, ultimately improving the model's performance in extracting aspect-opinion pairs.
- Instead of using the BERT-encoded context, here the domain-weighted context is utilized, leading to more emphasis on relevant features and relationships when identifying a domain's aspect-opinion pairs.
- The proposed model is evaluated on a multi-domain dataset and compared with state-of-the-art models. According to the experimental results, given its ability to capture domain-specific features and leverage broader context, the WWMD model achieves competitive performance.

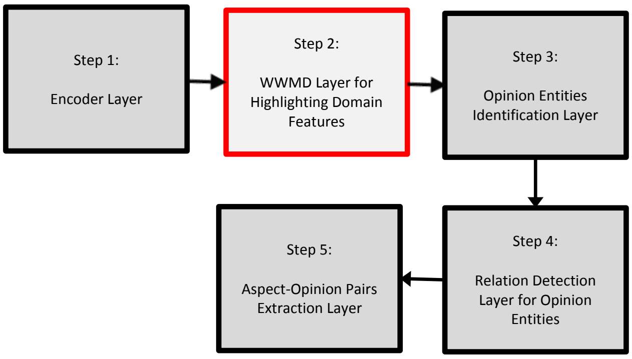

Fig. 1: The illustration of aspect-opinion extraction in multi-domain contexts (The red box is our proposed model).

The rest of this paper is organized as follows: Section 2 provides a brief survey of related work on sentiment analysis. Section 3 describes the proposed model in detail, and Section 4 evaluates the model and compares the results with state-of-the-art methods. Finally, concluding remarks and directions for future work are offered in Section 5.

Related Work

Traditional sentiment analysis methods include lexicon-based approaches and non-neural network classifiers [12]. Using an emotional dictionary and extracting emotional values, sentiment polarity is determined by the lexicon-based method. Aspect extraction primarily relies on association mining to identify frequent nouns and noun phrases as potential aspects [13]-[15].

As data volume grows and diverse tasks become more prevalent, more scholars are turning to neural network models based on deep learning for sentiment analysis tasks, yielding improved outcomes. Recently, multi-layer neural networks have become increasingly crucial in NLP, offering rapid and advanced results. This success can be attributed to the abundance of data and the rise of Graphical Processing Units. Deep learning has emerged as a highly effective technique in machine learning for text-mining tasks such as text classification, sentiment analysis, question-answering systems, and semantic analysis [16].

Chen et al. [11] investigated the aspect-opinion Pair Extraction task, focusing on extracting aspects and opinion expressions in pairs. They introduced the Synchronous Double-channel Recurrent Network, comprising an opinion entity extraction unit, a relation detection unit, and a synchronization unit. The opinion entity extraction and relation detection units act as two channels to simultaneously extract opinion entities and relations. Within the synchronization unit, they devised the Entity Synchronization Mechanism (ESM) and Relation Synchronization Mechanism (RSM) to enhance mutual benefits across the two channels. In order to assess the performance of SDRN, they manually curated three datasets based on the SemEval 2014 and 2015 benchmarks. Nevertheless, its limitation to single-domain scenarios represented a drawback of the approach. This paper only extracts aspect-opinion pairs, without considering the importance of each word within its domain.

Kumar et al. [17] introduced a novel Hierarchical Self-Attention Network demonstrating strong performance while requiring less memory and training time. A hierarchical self-attention mechanism is used by HSAN to assess the significance of each word in the context of the overall sentence meaning. Subsequently, it delves into the internal relationships between words within the same sentence to identify interdependent collocated words. By integrating these two attention mechanisms, HSAN effectively predicts multiple aspect terms within a given sentence, including multi-token aspect terms. This paper only extracts aspects, not aspect-opinion pairs.

Most studies have shown encouraging results in aspect extraction and polarity detection of texts and aspects, while fewer ones have been conducted in the field of multi-domain sentiment analysis [18]-[21]. Multi-domain sentiment analysis focuses on understanding opinions across various categories, such as product reviews and social media posts. This necessitates developing models accurately capturing and interpreting sentiments in different domains, often by extracting relevant aspects tailored to each domain. While attention mechanisms may help identify key features, their effectiveness in multi-domain contexts is still under exploration [22] and [23]. Variations in language use and context often make traditional sentiment analysis models difficult to generalize, causing specialized models for specific domains, which adds complexity and resource demands [24] and [25].

Alyoubi and Sharma [22] introduced a sentiment classification system using a deep bi-directional Recurrent Neural Network with an attention mechanism for multi-domain classifications. Their approach generates domain representations by extracting relevant text aspects, fed into the sentiment classifier through shared hidden layers. However, focusing only on classifying sentences without extracting aspect-opinion pairs, and while extracting the polarity of each aspect, their method does not address aspect-opinion pair extraction.

For analyzing multi-domain sentiments, Ghorbanali and Sohrabi's method utilizes a combination of the pre-trained BERT model, a convolutional neural network (CNN), a bi-directional Long Short-Term Memory (LSTM), and a gated recurrent unit (GRU) within a Capsule Network framework. The model in their study uses pre-trained BERT with CNN and LSTM to extract relevant aspects for CapsuleNet. The effectiveness of this approach is assessed by the Dranziera protocol [9]. This method only extracts polarity and the domain of the document.

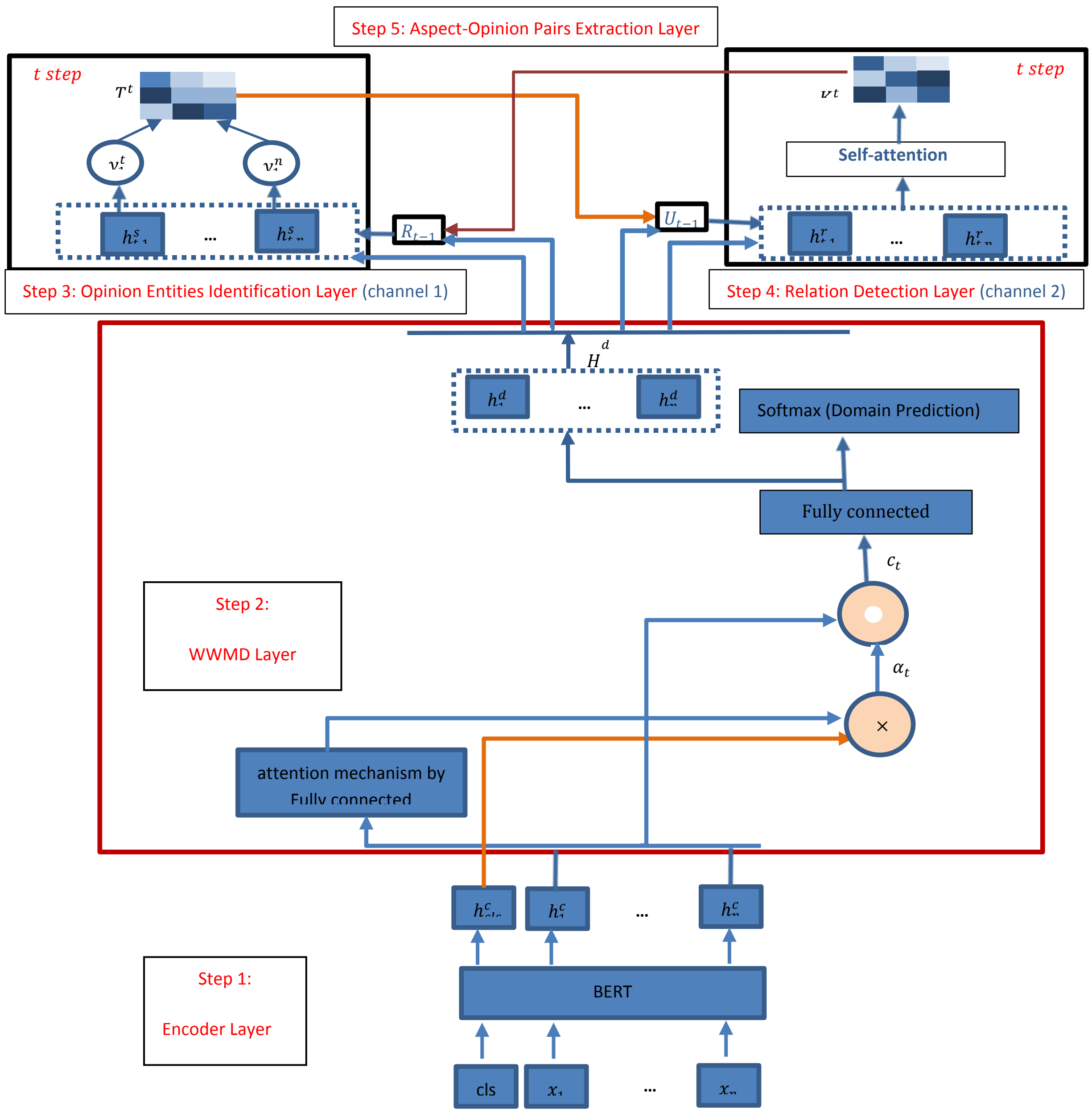

Fig. 2: Proposed model for multi-domain contexts (The red box is our contribution to extending the model).

The challenge of insufficient datasets for training classifiers is addressed by multi-domain sentiment analysis. The aforementioned models either extract aspect-opinion pairs within a single domain or assess the polarity of aspects across multi-domain contexts, although they do not extract aspect-opinion pairs in multi-domain contexts. This research employs the WWMD model to extract aspect-opinion pairs from multi-domain contexts.

Proposed Model

In a sentence X=$\{x_1 .\dots. x_n\}$ containing n tokens, the objective of the aspect-opinion task is to extract all aspects along with their associated opinions $(a.o)$ within the sentence, where a and o define the aspect and opinion, respectively. In this study, BERT is adopted as the encoding layer to learn contextual representations initially. Subsequently, the WWMD layer is proposed, consisting of an attention mechanism and a dense network. The WWMD layer assigns weights to words according to their relevance to the sentence's overall meaning and domain, ensuring that words specific to each domain receive more weight compared to others. Next, the Opinion Entities Identification Layer and the Relation Detection Layer are implemented as dual channels to extract opinion entities and their relationships simultaneously. This approach is inspired by the work of Chen et al., with the distinction that weighted contextual representations, rather than context representations, are input into the dual channels.

The aspect-opinion pairs extraction step is performed to derive the final aspect-opinion pairs.

Fig. 2 illustrates aspect-opinion pair extraction in multi-domain contexts, consisting of five main steps: Encoder Layer, WWMD Layer, Opinion Entities Identification Layer, Relation Detection Layer, and Aspect-Opinion Pairs Extraction Layer. The figure below shows the steps introduced in Fig. 1 in detail. Explanations of each step will be provided in the following sections.

A. Step 1: Encoder Layer

The Encoder Layer in natural language processing is a crucial component that converts input text data into meaningful representations that neural networks can process. In this model, BERT stands for a transformer-based model commonly used for encoding sentences. By leveraging its bidirectional architecture, BERT can capture context from both directions in a sentence, generating high-quality sentence embeddings that encapsulate rich semantic information. In our model, an input sentence X is processed, comprising N words, by first tokenizing the text into a sequence of tokens and then inserting unique tokens such as *[CLS]* and *[SEP]*. Segment embeddings are then added to distinguish between two sentences, followed by positional embeddings to show the position of each token. The input format for BERT includes tokenized text, segment embeddings, positional embeddings, and unique tokens. In BERT, token embeddings, segment embeddings, and positional embeddings are combined through an element-wise sum. Each token w_t is embedded as e_t by summing the respective embeddings. The resulting embedding sequence E is fed into BERT, utilizing a transformer framework to capture bidirectional contextual information. The output from the final transformer block forms the context representation sequence $H = ({h_1}^c . \dots . {h_N}^c)$.

The *[CLS]* token in BERT holds aggregated information about the entire input sequence. Therefore, the output of the *[CLS]* token is used to generate a summary representation of the input sequence. This process is formulated as:

$$S = h_{cls} \tag{1}$$

where S is a summary of the entire input sequence.

B. Step 2: WWMD Layer

Our proposed model, WWMD, is aimed at identifying specific words for each domain. To this end, an attention mechanism is employed to assign appropriate weights to the words in the sentence based on their overall meaning. Subsequently, a fully connected layer and Softmax are utilized to predict the domain. Finally, a cost function will be defined to calculate these weights based on the domain and overall sentence content. The attention mechanism is formulated as follows:

$$\alpha_t = \frac{exp(Sw_e\, {h_t}^T)}{\sum_{t=1}^{n} exp(Sw_e\, {h_t}^T)} \tag{2}$$

where w_e represents a trainable matrix and α_t signifies the likelihood of x_t being the correct word for identifying domain words. α_t is determined by an attention model, considering the encoded-word representation h_t, and the summarized information of the sentence S.

Then, our model develops a word representation aware of global context, denoted as c_t, for every word in the sentence, formulated as:

$$c_t = \alpha_t * h_t \tag{3}$$

where c_t represents a sequence of values calculated by multiplying the attention weight α_t by the hidden state h_t. Then, C represents the global context-aware word representation of all words x_t in the given sentence.

$$C=(c_1 . \dots . c_n) \tag{4}$$

Combining a fully connected layer and Softmax contributes to domain prediction by utilizing the global context representation c_t of each word as input. This process allows for the prediction of the domain based on the contextual information captured by the global representations of individual words c_t, and processed by a fully connected layer to learn domain-specific words through the weight parameter w_d updates. This process is formulated as:

$${h_i}^d = {w_{d_i}}^T * c_t \tag{5}$$

where ${h_i}^d$ refers to the input sequence's domain-specific representation or feature vector for the i^{th} word, the output of the fully connected layer is passed through a Softmax activation function to generate probabilities for various domain classes. The final domain label is predicted based on the class with the highest probability, allowing for accurate forecasting of the input text's domain. This process is formulated as:

$$softmax({h_i}^d) = \frac{exp({h_i}^d)}{\sum_{j=1}^{n} exp({h_j}^d)} \tag{6}$$

where ${h_i}^d$ represents the input to the softmax function for class *I,* and N illustrates the total number of domains.

Then, H^D represents words aware of the domain. H^D is formulated as:

$$H^D = ({h_1}^d . \dots . {h_n}^d) \tag{7}$$

where H^D is the sentence in which words specific to each domain are highlighted. H^D is utilized in the subsequent steps to recognize aspect-opinion pairs.

Evaluation of a model's performance on a dataset needs a loss function since it quantifies the disparity between the model's predictions and the actual target values, thereby measuring the model's error. By comparing predicted outputs to ground truth labels, the loss function provides feedback guiding the model in adjusting its parameters during training. This iterative process of minimizing the loss function enhances the model's accuracy and predictive capabilities, making it crucial for assessing and improving its performance on the dataset. Cross-Entropy loss is used to minimize this function, enhancing the model's accuracy and predictive capabilities. The definition of Cross-Entropy loss is:

$$\mathcal{L}_D = \frac{1}{N_k} \sum_{k=1}^{M} \sum_{i=1}^{N_k} p\left(d_i^k \middle| x_i^k\right) log\left[\tilde{p}\left(d_i^k \middle| x_i^k\right)\right] \quad (8)$$

where d_i^k and x_i^k denote the actual domain value and input data, respectively. M represents the number of domains, while N_k denotes the number of samples for the k^{th} domain. The WWMD model assists in assigning appropriate weights to each word, leading to efficient extraction of opinion entities in subsequent steps.

C. Step 3: Opinion Entities Identification Layer

This step is aimed at extracting the aspects and opinion entities, utilizing the SDRN channel. A Conditional Random Field (CRF) [26] is utilized upon the WWMD layer to perform sequence labelling. CRF is a powerful model for sequence labelling tasks like opinion entity extraction. In this context, CRF is trained to predict the labels of each word in a sequence, indicating whether it belongs to an opinion entity or not. CRF's ability to effectively capture the necessary contextual information for accurate opinion entity extraction by considering dependencies between neighbouring words instills confidence in its capabilities.

At the t^{th} recurrent step, we utilize the score of a sequence of predicted labels $Y^t = (y_1^t. \ldots . y_n^t)$ as:

$$Score(X.Y^t) = \sum_{i=1}^{N} T_{y_{i-1}^t . y_i^t} + \sum_{i=1}^{N} EE_{i.y_i^t}^t \quad (9)$$

$$EE^t = H_t^s W_{ee} + b_{ee} \quad (10)$$

where X is the input sequence and Y^t stands for the label sequence. H_t^s represents the input hidden representation sequence at the t^{th} recurrent step for the opinion entity identification layer. The matrices W_{ee} and b_{ee} are model parameters. $H_t^s = ({h_{t.1}}^s . \ldots . {h_{t.n}}^s)$ is calculated by the context representation sequence H^c replaced by H^d in our WWMD model and the relation synchronization semantics R_{t-1}. The details will be described in the *Aspect-Opinion Pair Extraction Layer* section.

$Score(X.Y^t)$ calculates the score by summing two components:

- The first sum involves transition scores T between consecutive labels y_{i-1}^t and y_i^t in the sequence.
- The second sum includes emission scores *EE* for each word position *i* and its corresponding label y_i^t.

Combining these transition and emission scores, the scoring function evaluates how well the label sequence Y^t fits the input sequence X in the context of opinion entity identification Layer using CRF.

In a CRF, the probability $p(Y^t|X)$ of the predicted sequence Y^t is defined as:

$$p(Y^t|X) = \frac{exp(Score(X.Y^t))}{\sum_{\tilde{Y}^t \in Y_X^t} exp\left(Score\left(X.\tilde{Y}^t\right)\right)} \quad (11)$$

where $Score(X.Y^t)$ represents the score of the input *X* associated with the label sequence Y^t, and Y_X^t denotes the set of all possible label sequences for input *X*.

D. Step 4: Relation Detection Layer

This step is aimed at creating a relation detection system to extract opinion entities and relationships concurrently, utilizing another channel of SDRN [11]. Considering the intricate connections between aspects and opinion expressions, Chen et al. developed a self-attention mechanism as a layer for relation detection. This mechanism allows for the modelling of token-level relations without being constrained by sequence limitations. At the t^{th} recurrent step, the attention matrix K^t is first computed, where the element $k_{i.j}^t$ represents the relationship score between tokens *i* and *j* at time step *t* as follows [11]:

$$k_{i.j}^t = \frac{exp(\gamma(h_{t.i}^r . h_{t.j}^r))}{\sum_{k=1}^{N} exp\left(\gamma(h_{t.i}^r . h_{t.k}^r)\right)} \quad (12)$$

where γ is a score function, and $h_{t.i}^r$ represents the input hidden representation of the i^{th} token for the relation detection layer. This formula suggests an attention mechanism where tokens *i* and *j* attend to each other based on their representations.

The function $\gamma\left(h_{t.i}^r . h_{t.j}^r\right)$ computes the relationship score based on the representations $h_{t.i}^r$ and $h_{t.j}^r$ of tokens *i* and *j*, respectively. This function is defined as:

$$\gamma\left(h_{t.i}^r . h_{t.j}^r\right) = tanh\left(h_{t.i}^r * W_r^1 + h_{t.j}^r * W_r^2\right) * W_r^3 \quad (13)$$

where W_r^1, W_r^2, and W_r^3 are weight matrices. The function utilizes the hyperbolic tangent activation function. $H_t^r = ({h_{t.1}}^r . \ldots . {h_{t.n}}^r)$ is calculated using the context representation sequence H^c replaced by H^d in our WWMD model and the entity synchronization semantics U_{t-1}. The details will be described in the *Aspect-Opinion Pair Extraction Layer* section.

In the final step *T*, the attention matrix K^t is calculated by maximizing the likelihood probability as:

$$p(Z|X) = \prod_{i=1}^{N} \prod_{j=1}^{N} p\left(z_{i.j} \middle| x_i . x_j\right) \quad (14)$$

$$p\left(z_{i.j} \middle| x_i . x_j\right) = \begin{cases} k_{i.j}^T & if\ z_{i.j} = 1. \\ 1 - k_{i.j}^T & if\ z_{i.j} = 0 \end{cases} \quad (15)$$

where $p(Z|X)$ is the probability distribution. Z is the standard relation matrix. It represents the binary relations between tokens, and *X* represents some input features or representations of the tokens. It's a product of probabilities over all token pairs *(i, j)*, indicating whether there is a relationship between them. If $z_{i.j} = 0$, it indicates that there is no relation between the i^{th} and j^{th} tokens, and vice versa.

In summary, these equations describe a method for relation detection between tokens that incorporates attention mechanisms and probability calculations based on learned representations of the tokens involved.

E. Step 5: Aspect-Opinion Pair Extraction Layer

Chen et al. developed the Entity Mechanism (EM) and Relation Mechanism (RM), aiming at facilitating information exchange between the interdependent Opinion Entities Identification Layer and Aspect-Opinion Pair Extraction Layer, ensuring that both channels benefit. This design updates the hidden representation sequences h_t^s and h_t^r through allowing them to share high-level information.

E.1. Entity Mechanism

Initially, the focus is on extracting aspects like "battery life" and opinions like "well-made" in the laptop domain before extracting aspect-opinion pairs. Opinion entities typically consist of phrases. When only token relations are considered, entities like "battery life" or "well-made" may not be extracted. Hence, both opinion entity semantics and token-level interactions are crucial in detecting relations. Chen et al. initially defined $u_{t.i}$ for token *i* in step *t* to acquire entity semantics.

To capture the corresponding entity semantics for each token and integrate them into the hidden representation sequence h_{t+1}^r, the calculation is carried out based on the predicted label sequence Y^t and its associated probabilities obtained from the Opinion Entities Identification Layer. Precisely, the entity semantics $u_{t.i}$ of the i^{th} token at the t^{th} recurrent step can be calculated as [11]:

$$u_{t.i} = \sum_{j=1}^{N} \varphi(T_{i.j}^t) h_j^c \tag{16}$$

$$\varphi(T_{i.j}^t) = \frac{T_{i.j}^t}{\sum_{k=1}^{N} T_{i.k}^t} \tag{17}$$

where $\varphi(\cdot)$ represents a normalization function. Furthermore, $\mathrm{T}_{i.j}^t$ indicates the label probability of the j^{th} token if both the i^{th} and j^{th} tokens belong to the same entity; otherwise, $\mathrm{T}_{i.j}^t$ is zero.

Then, by combining entity semantics $u_{t.i}$ with context representation $\mathrm{h}_\mathrm{i}^\mathrm{c}$ replaced by $\mathrm{h}_\mathrm{i}^\mathrm{d}$ in our WWMD model, $\mathrm{h}_{t+1.i}^r$ is formed, formulated as:

$$h_{t+1.i}^r = \delta(u_{t.i} W_r^4 + h_i^c W_r^5) \tag{18}$$

where W_r^4 and W_r^5 are model parameters. δ is the activation function can be either the tanh or sigmoid function. The hidden representation $\mathrm{h}_{t+1.i}^r$ is calculated by applying a non-linear transformation to $u_{t.i}$ and $\mathrm{h}_\mathrm{i}^\mathrm{c}$, replaced by $\mathrm{h}_\mathrm{i}^\mathrm{d}$, weighted by the corresponding weights W_r^4 and W_r^5. Note that the entity semantics sequence is initialized with a zero matrix: U_0=($u_{0.1}.u_{0.2}.\dots.u_{0.N}$).

E.2. Relation Mechanism

The relationships between opinion entities provide valuable insights. For example, when "great" describes "battery life", it helps extract the aspect "battery life" and the opinion "great". Subsequently, the relations between opinion entities are identified. $r_{t.i}$ is used to calculate and extract the relations between opinion entities as [11]:

$$r_{t.i} = \sum_{j=1}^{N} \varphi(\phi(k_{i.j}^t)) h_j^c \tag{19}$$

$$\phi(k_{i.j}^t) = \begin{cases} k_{i.j}^t & if\ k_{i.j}^t \geq \beta. \\ 0 & if\ k_{i.j}^t < \beta \end{cases} \tag{20}$$

where $\varphi(\cdot)$ is the normalization function. To reduce noise, $\phi(\cdot)$ filters out correlated scores below the threshold β. The relation semantics $r_{t.i}$ is calculated by summing the product of the normalized correlated scores $\varphi(\phi(k_{i.j}^t))$ and the content representations h_j^c replaced by h_j^d for all tokens *j*. $\mathrm{k}_\mathrm{i.j}^\mathrm{t}$ ensures that only correlated scores above the threshold β are considered. Adjusting the threshold β allows for controlling the level of correlation required for a token pair to be considered relevant in relation to semantics calculation. If $\mathrm{k}_\mathrm{i.j}^\mathrm{t} < \beta$ is set to 0, implying that the correlation between the tokens is below the threshold and is considered insignificant for further processing.

Then, the model undergoes a learning process, computing the hidden representation $h_{t+1.i}^o$ for the next step by applying a non-linear transformation to the sum of the relation semantics $r_{t.i}$ and the context representation $\mathrm{h}_\mathrm{i}^\mathrm{c}$, which can be calculated as:

$$h_{t+1.i}^o = \delta(r_{t.i} W_o^1 + h_i^c W_o^2) \tag{21}$$

where W_o^1 and W_o^2 are model parameters. The relation semantics sequence R_0=($r_{0.1}.r_{0.2}.\dots.r_{0.N}$) is initialized to zero. By adjusting these weights during training, the model learns to assign different levels of importance to the relation semantics and the context representation $\mathrm{h}_\mathrm{i}^\mathrm{c}$ replaced by $\mathrm{h}_\mathrm{i}^\mathrm{d}$ in the WWMD model, when updating the hidden representation for relation detection. The weights W_o^1 and W_o^2 allow the model to capture the interaction between relation semantics and context representations, ultimately influencing the extraction of aspect-opinion pairs.

The above-mentioned steps are executed iteratively, each iteration building on the previous one to capture high-level representations. After multiple iterations, the aspect-opinion pairs extraction step is performed to derive the final aspect-opinion pairs. At the final step of the model, aspect-opinion pairs are extracted using the weight matrix K from the *Relation Detection Layer* section. This process leverages the correlations between aspects and opinion expressions to accurately determine and extract the pairs. The weight matrix K^T is crucial in extracting aspect-opinion pairs, ensuring effective capture of relationships between aspects and opinions. The formula is as [11]

$$\delta = \left(\frac{1}{|a|} \sum_{k=i_S^a}^{i_E^a} \sum_{l=i_S^o}^{i_E^o} k_{k.l} + \frac{1}{|o|} \sum_{k=i_S^o}^{i_E^o} \sum_{l=i_S^a}^{i_E^a} k_{l.k} \right) \tag{22}$$

where |a| and |o| represent the lengths of the aspect and opinion expressions, respectively. When analyzing aspect tokens from the position i_S^a to i_E^a and opinion expression tokens from the position i_S^o to i_E^a, the initial calculation aggregates the correlations $k_{k.l}$ within these specified ranges, reflecting the impact of aspect tokens on opinion expression tokens within the specified context. Similarly, the second expression calculates the influence of the opinion expression tokens on the aspect tokens within the specified context. The pair $(a.o)$ is extracted only if the confidence score δ is higher than a specified threshold $\hat{\delta}$.

E.3. Loss Function

Loss functions are essential for training the model to extract opinion entities and detect relations between them accurately. The loss is defined as:

$$\mathcal{L}(\theta) = \mathcal{L}_E + \mathcal{L}_R + \mathcal{L}_D \tag{23}$$

where the losses $\mathcal{L}_E$, $\mathcal{L}_R$, and $\mathcal{L}_D$ represent different components of a loss function related to entity extraction, relation detection, and domain adaptation, respectively.

$$\mathcal{L}_E = log \sum_{\tilde{Y} \in Y_X^T} exp\left(S(X.\tilde{Y})\right) - S(X.Y) \tag{24}$$

where $\mathcal{L}_E$, is the negative log-likelihood loss for opinion entity extraction, is calculated by comparing the predicted scores of all possible entity sequences with the actual entity sequence.

$$\mathcal{L}_R = -\sum_{i=1}^{N}\sum_{j=1}^{N} p(z_{i.j}|x_i.x_j)\, log[\tilde{p}(z_{i.j}|x_i.x_j)] \tag{25}$$

$\mathcal{L}_R$, the cross-entropy loss for relation detection, is computed by evaluating the predicted relation probabilities against the ground truth relations.

$$\mathcal{L}_D = \frac{1}{N_k}\sum_{k=1}^{M}\sum_{i=1}^{N_k} p(d_i^k|x_i^k)\, log[\tilde{p}(d_i^k|x_i^k)] \tag{26}$$

$\mathcal{L}_D$ denotes the domain loss that calculates the average cross-entropy loss over all instances in each domain. It sums the negative log probabilities of the predicted domain labels d_i^k given the input features x_i^k compared to the actual domain labels. This loss function helps train the model to correctly classify instances into their respective domains by penalizing deviations between predicted and actual domain labels.

$\mathcal{L}(\theta)$ combines these three losses to create the overall loss function of the model. These equations are crucial for training the model effectively by optimizing the parameters to minimize the combined loss.

Experiments

A. Dataset

We use two datasets: the dataset in Table 1, proposed by Cai et al. [27], and the SemEval ABSA dataset, proposed by Chen et al. [11] in Table 2. The dataset in Table 1 spans five domains: Book, Clothing, Hotel, Restaurant, and Laptop, comprising approximately 20,000 review sentences, significantly more significant than the SemEval ABSA datasets.

Table 1: Descriptions of the benchmark datasets with five domains: aspect-sentiment (AS), aspect-opinion (AO), aspect-opinion-sentiment (AOS), aspect-category-sentiment (ACS), and aspect-category-opinion-sentiment (ACOS)

Dataset	Sentence	Aspect	Category	Opinion	Sentiment	AS Pair	AO Pair	AOS Triplet	ACS Triplet	ACOS Quadruple
Books	2967	3781	3593	4291	3781	3931	4493	4493	4048	4507
Clothing	2373	2843	2994	3341	2843	2904	3415	3415	3186	3416
Hotel	3526	4700	4886	5781	4700	4735	6014	6014	5284	6017
Restaurant	5152	7056	6307	7958	7056	7250	8484	8484	7436	3496
Laptop	4076	4958	4992	5378	4958	5035	5726	5731	5227	5758
Total	18094	23338	22772	26749	23338	23855	28132	28137	25181	28194

We compare the WWMD model with the FaiMA model in Table 4 and other baseline models in Table 3.

B. Implementation Details

Our approach utilizes the AdamW optimizer with a maximum learning rate of 9 × 10−5 for BERT weights. The batch size is fixed at 15, and the dropout rate is set to 0.1. For cross-validation, the remaining hyperparameters are defined as: β = 0.1 and $\hat{\delta}$= 0.4. To improve the information interaction between the two channels, the recurrent step t to 2 is set. The number of domains for Table 1 and Table 2 is 5 and 3, respectively.

Table 2: Descriptions of the SemEval ABSA dataset (Number of domains is three)

Dataset	Sentence	Aspect	Opinion	AO Pair
SemEval-14 Restaurant	3841	4827	4526	3745
SemEval-14 Laptop	3845	3012	3177	1915
SemEval-15 Restaurant	2000	1747	1733	1747
Total	9686	9586	9436	7407

Table 3: Performance on different datasets is measured using the F-measure (%) metric: Aspect-sentiment (AS), aspect-opinion (AO), aspect-opinion-sentiment (AOS)

Model	14-Res			14-Lap			15-Res		
	AS	AO	AOS	AS	AO	AOS	AS	AO	AOS
SDRN [11]	-	76.48	-	-	67.13	-	-	70.94	-
EMC-GCN [28]			71.78			58.81			61.93
COM-MRC [29]	-	-	72.89	-	-	60.09	-	-	63.65
TAGS [30]			75.05			64.53			67.90
SJCL [31]		80.54			71.96			73.92	
APSCL [32]	81.28	-	-	78.47	-	-	-	-	-
WWMD	-	78.36	-	-	69.87	-	-	72.75	-

Table 4: Performance comparison in F-measure on five datasets; AS and AO stand for aspect-sentiment and aspect-opinion, respectively

Model	Laptop		Restaurant		Books		Clothing		Hotel	
	AS	AO	AS	AO	AS	AO	AS	AO	AS	AO
FaiMA [33]	70.52	-	81.38	-	66.13	-	76.79	-	83.27	-
WWMD	-	69.69	-	80.50	-	64.97	-	73.98	-	81.63

C. Evaluation Metric

In sentiment analysis and opinion mining, precision, recall, and F-measure are fundamental metrics used to assess the effectiveness of aspect-based sentiment analysis systems for aspect-opinion pairs. F-measure is the balanced combination of precision and recall, offering a unified metric that accounts for both aspects. These metrics are calculated as follows:

$$F - measure = \frac{2 \, * \, (Precision \, * \, Recall)}{(Precision \, + Recall)} \tag{27}$$

where *Precision* focuses on the accuracy of positive predictions and *Recall* emphasizes the model's ability to capture all actual ones. The $F - measure$ value is used as the main evaluation metric.

D. Baselines

We selected specific baseline models to thoroughly and comprehensively assess the WWMD model.

Chen et al. [11] proposed the Synchronous Double-channel Recurrent Network, primarily consisting of an opinion entity extraction unit, a relation detection unit, and a synchronization unit. This network extracts aspect and opinion expressions in pairs.

Chen et al. [28] proposed an enhanced multi-channel graph convolutional network (EMC-GCN). EMC-GCN is a graph-based model for aspect-opinion-sentiment triplet extraction utilizing a multi-channel graph convolutional network to capture complex syntactic and semantic relations between word pairs.

It models ten types of linguistic relations using a biaffine attention mechanism to construct rich adjacency tensors. By integrating multiple relational perspectives, EMC-GCN improves the accuracy of aspect-opinion-sentiment extraction through more refined word-level interactions and contextual understanding.

Zhai et al. [29] presented a novel Context-Masked MRC (COM-MRC) framework for Aspect-Based Sentiment Analysis comprising three interconnected components: a context augmentation strategy, a discriminative model, and an inference method, utilizing a two-stage inference method that extracts all aspects and then identifies their corresponding opinions and sentiments by iteratively masking them. To enhance the identification of information from different aspects, the model utilizes the concept of aspect masking for context augmentation. It is suggested by the authors that a sentence containing multiple aspect terms can be treated as multiple training samples due to the challenges in extracting triplets. Consequently, they establish a regular query with various masked contexts to identify each aspect and its associated components. For a sentence with t aspect terms, the number of samples increases from 1 to 2^t, effectively expanding the training corpus.

Luo et al. [30] proposed a novel tagging-assisted generation model with encoder and decoder supervision (TAGS). TAGS introduces a hybrid approach to triplet extraction through combining the strengths of sequence generation and sequence tagging, enhancing the encoder-decoder framework by integrating an auxiliary tagging task to guide the encoder and incorporating semantic alignment between the decoder's hidden states and label representations. Improving both the structural accuracy and semantic fidelity of the generated aspect-opinion-sentiment triplets, this dual supervision achieves strong performance across benchmark datasets.

Yang et al. [31] designed a span-based joint extraction framework with contrastive learning (SJCL) to enhance both term extraction and pairing in PAOTE. For term extraction, they utilized a span-based convolutional neural network (CNN) to merge contextual, syntactic, and semantic features and precisely identify the boundaries of aspect and opinion terms. This was combined with contrastive learning (CL) to enhance the distinctiveness of different types of terms. For term pairing, they pruned the original dependency trees according to part-of-speech (POS) information to remove insignificant noise and leveraged a graph convolutional network (GCN) to learn pairing features. Then, negative sampling and contrastive learning were employed to avoid mismatched aspect-opinion pairs.

Li et al. [32] introduced an Aspect-Pair Supervised Contrastive Learning (APSCL) model with three main components: the aspect-pair embedding module, the aspect-pair representation learning (APRL) module, and the aspect-level classification module.

An aspect-pair embedding technique involves extracting relational grammar-based aspect representations from the Relational Graph Attention Network (RGAT) and constructing aspect-pair embeddings from these representations. This technique enables the aspect-pair, allowing for the capture of category information based on the sentiment polarity of individual aspects. The aspect-pair is then utilized for comparative learning to constrain the sentiment polarity of each aspect.

Yang et al. [33] introduced a novel framework called Feature-aware In-context Learning for Multi-domain Aspect-based Sentiment Analysis (FaiMA). The central insight of FaiMA is its use of in-context learning as a feature-aware mechanism enabling adaptive learning in multi-domain ABSA tasks. A multi-head graph attention network was implemented as a text encoder optimized through heuristic rules to capture linguistic, domain, and sentiment features. By emphasizing these diverse features, they enhanced the optimization of sentence representations.

E. Main Results

The F-measure of different approaches is presented in Table 3 with results pertaining to a single domain. SDRN, SJCL, and WWMD extract aspect-opinion pairs, while EMC-GCN, COM-MRC, and TAGS extract aspect-opinion-sentiment triplets, and APSCL extracts aspect-sentiment pairs. Extracting aspect-opinion-sentiment triplets is relatively more challenging; consequently, EMC-GCN, COM-MRC, and TAGS exhibit lower F-measure scores compared to the other models. In the COM-MRC model, aspects are extracted sequentially from left to right through an iterative masking process. This increases the model's training time, which may limit the applicability of COM-MRC to massive datasets. When compared to COM-MRC and EMC-GCN, accurate and semantically aligned outputs are provided by TAGS; however, it is sensitive to tagging errors that can degrade output quality if not carefully tuned. The advantage of WWMD compared to EMC-GCN, COM-MRC, and TAGS lies in its lower computational complexity and faster inference, while still maintaining strong performance through domain-aware attention, making it more suitable for scalable and multi-domain applications without the overhead of complex graph processing, iterative masking, or generative decoding.

Comparing the F-measure scores indicates that our model outperforms all models except APSCL and SJCL. SJCL achieves superior precision in identifying aspects and opinions, although it is limited to a single domain and has a slower and more complex architecture. The WWMD outperforms APSCL, highlighting its advantage over the existing model due to the limitations of graph neural networks in learning syntactic dependencies in longer sentences. Transformers are better equipped to capture semantic information over long distances compared to RGAT. Nevertheless, the decrease in F-measure in WWMD compared to APSCL is due to the simultaneous extraction of aspects and opinions. Notably, the APSCL model only extracts aspects and their polarity. Moreover, APSCL risks overfitting on limited or noisy data.

SDRN may struggle in scenarios with overlapping or ambiguous aspect-opinion expressions, although it has a relatively simple structure. WWMD extends SDRN and outperforms it by assigning greater attention to domain features. In multi-domain Aspect-Based Sentiment Analysis, texts across different domains may exhibit distinct features and styles, while texts within the same domain tend to share similar background knowledge and emotional contexts. Therefore, it is of paramount importance to consider domain attributes. In this context, the WWMD model stands out as a significant contribution by offering a multi-domain approach addressing the unique challenges of sentiment analysis across various domains.

Table 4 compares the WWMD and FaiMA models in Multi-domain contexts. The WWMD model has two advantages over the FaiMA model. First, their model can only extract aspects and their polarity. Second, their model does not consider the effect of relationships on each other, since considering the relationships between opinion entities can be a clue for extracting them. Hence, encoding the semantic relations and integrating them with the content display makes the model more capable of extracting opinion entities.

The WWMD model can extract aspect-opinion pairs in multi-domain contexts and give more attention to words specific to their related domains. However, it cannot extract implicit features that should be considered in the future. Furthermore, the performance of WWMD can decline if domain information is missing or noisy.

F. Complexity

To evaluate and compare model complexity, each model's architectural components and their impact on time and computational requirements are considered.

SDRN uses a dual-channel design consisting of a BERT encoder, a CRF for aspect-opinion term extraction, and a supervised self-attention mechanism for relation detection with a relatively simple and efficient architecture, resulting in low computational complexity, making it ideal for deployment on standard hardware.

WWMD builds upon SDRN by introducing a domain-aware attention mechanism and a domain classifier, adapting word weighting based on domain context and slightly increasing computational cost but retaining similar time complexity, overall maintaining WWMD's practicality for multi-domain applications.

SJCL implements a more complex span-based framework. It uses CNNs for span generation, BERT for contextual encoding, and GCNs to model syntactic structures. Besides, it incorporates dual contrastive learning for enhanced span boundary precision. These components make SJCL highly accurate but significantly more resource-intensive in both memory and processing time due to span enumeration and graph operations.

APSCL is designed to extract aspect-sentiment pairs. It uses either BERT or LSTM as the base encoder, combined with relational graph attention networks and supervised contrastive learning. Additionally, the use of relational graphs makes it notably more complex compared to WWMD.

FaiMA integrates multi-head graph attention (MGATE), feature-aware contrastive learning, FAISS-based example retrieval, and in-context learning using large language models. While offering high flexibility and domain adaptability, its dependence on retrieval systems and large-scale model inference necessitates substantial infrastructure and computing resources. EMC-GCN features a multi-channel GCN architecture with biaffine attention to model various syntactic and semantic relations between word pairs. Although effective in capturing linguistic structure, the use of multiple graphs and pre-processed linguistic features leads to increased memory usage and complexity during training.

TAGS combines sequence generation and tagging in an encoder-decoder architecture (e.g., T5 or BART), augmented by tagging supervision and semantic alignment. While capable of generating accurate triplets, this approach adds significant computational overhead due to the generation process and alignment modules.

A machine reading comprehension-based pipeline is adopted by COM-MRC, performing aspect, opinion, and sentiment extraction in two stages and improving precision in aspect-rich sentences through sequentially masking and processing each aspect. However, the repeated inference passes make it one of the most complex models. Details of the modules used by each model are presented in Table 5.

Table 5: Modules used in each model

Model	Modules
SDRN	BERT + CRF + Supervised Self-Attention
WWMD	BERT + CRF + Weighted Self-Attention + Domain Classifier + Supervised Self-Attention
APSCL	BERT or BiLSTM + Relational Graph Attention Network (RGAT) + Contrastive Loss
EMC-GCN	BERT + Multi-Channel Graph Convolutional Network (GCN) + Biaffine Attention
SJCL	BERT + CNN (for spans) + GCN + Dual Contrastive Learning
TAGS	T5 or BART (Encoder-Decoder) + Tagging + Alignment Layers
COM-MRC	BERT + Multi-Task Discriminative Modules + Iterative Masking
FaiMA	BERT + MGATE (Multi-head Graph Attention) + FAISS + LLM (LLaMA)

We also compare the average training time of four baselines—WWMD, SDRN, SJCL, and EMC-GCN—on the SemEval dataset. All experiments targeting training time were conducted in a unified environment with 32.0 GB RAM, an NVIDIA GeForce 1070 GPU, and an Intel Core i7 CPU. The results are illustrated in Fig. 3. Average training time on SemEval datasets typically, as the number of parameters and layers increases, the computational workload on the system increases, leading to longer training times. According to the figure, WWMD requires more training time than SDRN but less than SJCL and EMC-GCN.

G. Ablation Study

The domain features were eliminated to examine the impact of domain features on the performance across different datasets, and the changes in the F-measure scores for the five datasets to validate the efficacy of the proposed model were reported.

The overall results are presented in Table 6 According to the results of Table 6, it is observed that extracting domain features enhances the outcome.

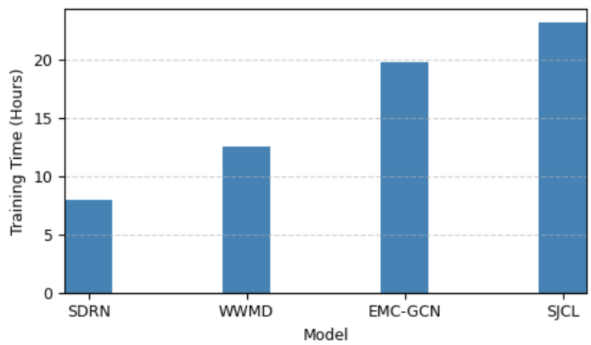

Fig. 3: Average training time on SemEval datasets.

Table 6: F-measure of ablation experiment results on different datasets. Removing the feature domain results in a decrease in performance.

Model	Laptop	Restaurant	Books	Clothing	Hotel
WWMD Without domain	69.18 (-0.51)	79.1518 (-1.35)	63.9618 (-1.01)	73.53(-0.45)	80.79 (-0.84)
WWMD With domain	69.69	80.50	64.97	73.98	81.63

H. *Case Study*

Table 7 presents an example of aspect-opinion pairs to compare this paper's proposed method with baseline approaches.

This example illustrates how multi-domain sentiment analysis can uncover a broader range of aspect-opinion pairs by leveraging the diverse contexts found in various domains, resulting in a more sophisticated understanding of sentiment. In WWMD and FaiMA, both Aspects—"laptop design" and "layout design of the hotel" can be extracted. They belong to different domains: "laptop design" is associated with the laptop domain, while "layout design of the hotel" is related to the hotel domain. Since FaiMA and WWMD are multi-domain, they can better separate aspects related to different domains. Other models labeled the term "layout design of the hotel" simply as "layout design" due to the lack of domain context. EMC-GCN is effective at capturing grammatical structure but misses the full phrase "layout design of the hotel" because of graph pruning.

SJCL is very precise in detecting the correct phrase boundaries, although it mislabels "layout design of the hotel" as "layout design" due to a lack of domain context.

To verify the relation detection capability of our model, we visualize the attention scores in Fig. 4. According to the findings, our proposed model accurately captures the relationships between aspects and opinion expressions across multi-domain reviews.

Table 7: Extraction results comparison of our Model with baseline methods for the ASTE task

Model / Sentence	SJCL	EMC-GCN	FaiMA (multi-domain)	WWMD (multi-domain)
The laptop design is sleek, but the layout design of the hotel feels cramped.	(laptop design, sleek) (layout design, cramped)	(laptop design, sleek, POS) (layout design, cramped, NEG)	(laptop design, POS) (layout design of the hotel, NEG)	(laptop design, sleek) (layout design of the hotel, cramped)

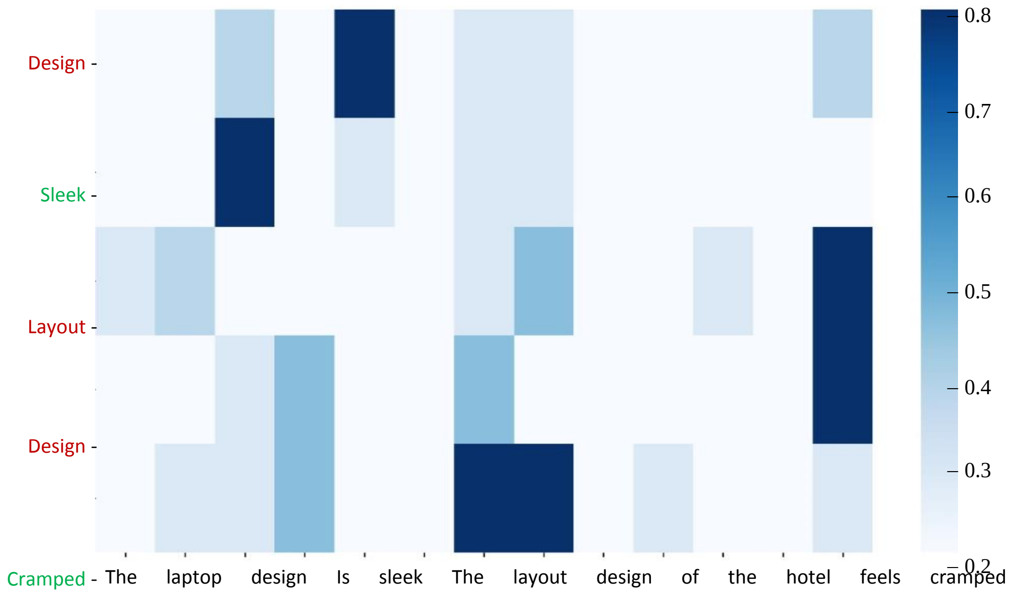

Fig. 4: Attention scores visualization: Aspects are highlighted in red, while opinion expressions are marked in green (Optimal viewing in colour).

Conclusion

Multi-domain approaches can encompass various linguistic features, idiomatic expressions, and contextual nuances across different domains. This diversity enables the model to predict more accurately in ambiguous scenarios. The proposed WWMD model employs an attention mechanism to capture the importance of each word concerning the overall sentence meaning. To extract aspect-opinion pairs, we introduce a domain prediction model that highlights features specific to each domain and combines it with the Synchronous Dual-Channel Recurrent Network method. Experimental results demonstrate the effectiveness of the WWMD model in extracting aspect-opinion pairs, especially across multi-domain contexts. WWMD consistently outperforms several strong baselines, such as COM-MRC, EMC-GCN, and TAGS, regarding F-measure scores on benchmark datasets. The proposed domain-aware attention mechanism captures word importance in a domain-sensitive manner, leading to more precise aspect extraction.

Compared to FaiMA, the multi-domain sentiment analysis model, WWMD demonstrates two distinct advantages. First, WWMD extracts aspect-opinion pairs directly, whereas FaiMA primarily focuses on aspect-sentiment pairs. Second, WWMD captures relationships between opinion entities, which FaiMA does not consider. WWMD exhibits strong efficiency in both time and computational complexity. It is more efficient than graph-based models such as SJCL and EMC-GCN, which incur higher memory and processing costs due to span enumeration and graph convolution operations. Furthermore, WWMD outperforms generative and comprehension-based models, such as TAGS and COM-MRC, by avoiding iterative decoding and multiple inference passes. Generally, WWMD is well-suited for real-world applications, maintaining a practical balance between performance and resource usage.

An ablation study shows that the proposed domain-related features improve overall performance. The F-measure score decreases noticeably when domain-specific weighting is removed, suggesting that domain knowledge is essential for relation detection and attention distribution. Additionally, qualitative analysis indicates that WWMD effectively resolves ambiguities in overlapping aspect-opinion expressions, a challenge frequently encountered by baseline models in cross-domain settings.

The future of sentiment analysis will likely involve integrating multimodal data, such as text, images, and videos, to capture richer sentiments and opinions. Moreover, the field is expected to address challenges related to sarcasm detection, irony, and context-aware sentiment analysis, further enhancing the accuracy and applicability of sentiment analysis models.

Author Contributions

This paper originates from A. Mohammadi’s PhD thesis, conducted under the supervision of M.R. Pajoohan, with A.M. Zareh Bidoki serving as her advisor.

Acknowledgment

We sincerely thank the anonymous reviewers for their valuable comments and constructive suggestions, which significantly improved the quality and clarity of this manuscript. We also thank the editor for their careful guidance during the review process. Additionally, we appreciate the editorial team of JECEI for their dedication and professionalism in managing the submission and publication workflow.

Funding

This research received no external funding.

Conflict of Interest

The authors confirm that there are no conflicts of interest related to the publication of this manuscript.

Abbreviations

SDRN	Synchronous Double-channel Recurrent Network
WWMD	Weighted Words Multi-Domain
ESM	Entity Synchronization Mechanism
RSM	Relation Synchronization Mechanism
EM	Entity Mechanism
RM	Relation Mechanism
APSCL	Aspect-Pair Supervised Contrastive Learning
APRL	Aspect-Pair Representation Learning
RGAT	Relational Graph Attention
FaiMA	Feature-aware In-context Learning for Multi-domain Aspect-based Sentiment Analysis

References

[1] T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, L. Galligan, "Sentiment analysis and opinion mining on educational data: A survey," Nat. Lang. Process. J., 2: 100003, 2023.

[2] H. Zhao, L. Huang, R. Zhang, Q. Lu, H. Xue, "Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction," in Proc. the 58th annual meeting of the association for computational linguistics: 3239-3248, 2020.

[3] H. T. Phan, N. T. Nguyen, D. Hwang, Y. S. Seo, "M2SA: A novel dataset for multi-level and multi-domain sentiment analysis," J. Inf. Telecommun., 7(4): 494-512, 2023.

[4] N. Gupta, R. Agrawal, "Application and techniques of opinion mining," in Hybrid computational intelligence: Elsevier, pp: 1-23, 2020.

[5] S. Chen, Y. Wang, J. Liu, Y. Wang, "Bidirectional machine reading comprehension for aspect sentiment triplet extraction," in Proc. the AAAI Conference on Artificial Intelligence, 35(14): 12666-12674, 2021.

[6] Z. Yuan, S. Wu, F. Wu, J. Liu, Y. Huang, "Domain attention model for multi-domain sentiment classification," Knowledge-Based Syst., 155: 1-10, 2018.

[7] S. Wang, Y. Zhu, W. Gao, M. Cao, M. Li, "Emotion-semantic-enhanced bidirectional LSTM with multi-head attention mechanism for microblog sentiment analysis," Information, 11(5): 280, 2020.

[8] D. Chen, W. Su, P. Wu, B. Hua, "Joint multimodal sentiment analysis based on information relevance," Inf. Process. Manag., 60(2): 103193, 2023.

[9] A. Ghorbanali, M. K. Sohrabi, "Exploiting bi-directional deep neural networks for multi-domain sentiment analysis using capsule network," Multimedia Tools Appl., 82(15): 22943-22960, 2023.

[10] D. Bollegala, D. Weir, J. Carroll, "Cross-domain sentiment classification using a sentiment sensitive thesaurus," IEEE Trans. Knowl. Data Eng., 25(8): 1719-1731, 2012.

[11] S. Chen, J. Liu, Y. Wang, W. Zhang, Z. Chi, "Synchronous double-channel recurrent network for aspect-opinion pair extraction," in Proc. the 58th Annual Meeting of the Association for Computational Linguistics: 6515-6524, 2020.

[12] R. Liu, Y. Shi, C. Ji, M. Jia, "A survey of sentiment analysis based on transfer learning," IEEE access, 7: 85401-85412, 2019.

[13] A. M. M. Yazdian-Dehkordi, M. Nematbakhsh, "Identification of implicit features using persian language rules and sentiments clustering," Tabriz J. Electr. Eng., 50(3): 1395-1404, 2020.

[14] M. Hu, B. Liu, "Mining and summarizing customer reviews," in Proc. the tenth ACM SIGKDD international conference on Knowledge discovery and data mining: 168-177, 2004.

[15] A. M. M-R. Pajoohan, M. Montazeri, "Identification of explicit features in persian comments," J. Comput. Secur., 6: 1-11, 2019.

[16] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explorations Newsletter, 19(1): 22-36, 2017.

[17] A. Kumar, A. S. Veerubhotla, V. T. Narapareddy, V. Aruru, L. B. M. Neti, A. Malapati, "Aspect term extraction for opinion mining using a hierarchical self-attention network," Neurocomputing, 465: 195-204, 2021.

[18] R. Satapathy, S. R. Pardeshi, E. Cambria, "Polarity and subjectivity detection with multitask learning and bert embedding," Future Internet, 14(7): 191, 2022.

[19] J. Shobana, M. Murali, "An improved self attention mechanism based on optimized BERT-BiLSTM model for accurate polarity prediction," Comput. J., 6(5): 1279-1294, 2023.

[20] H. S. Sharaf Al-deen, Z. Zeng, R. Al-sabri, A. Hekmat, "An improved model for analyzing textual sentiment based on a deep neural network using multi-head attention mechanism," Appl. Syst. Innovation, 4(4): 85, 2021.

[21] Y. Fu, J. Liao, Y. Li, S. Wang, D. Li, X. Li, "Multiple perspective attention based on double BiLSTM for aspect and sentiment pair extract," Neurocomputing, 438: 302-311, 2021.

[22] K. H. Alyoubi, A. Sharma, "Deep recurrent neural model for multi domain sentiment analysis with attention mechanism," Wireless Pers. Commun., 130(1): 43-60, 2023.

[23] N. A. Abdullah, A. Feizollah, A. Sulaiman, N. B. Anuar, "Challenges and recommended solutions in multi-source and multi-domain sentiment analysis," IEEE Access, 7: 144957-144971, 2019.

[24] L. P. Hung, S. Alias, "Beyond sentiment analysis: A review of recent trends in text based sentiment analysis and emotion detection," J. Adv. Comput. Intell. Intell. Inf., 27(1): 84-95, 2023.

[25] K. Bhattacharjee, R. Gangadharaiah, "Document-level supervision for multi-aspect sentiment analysis without fine-grained labels," arXiv preprint arXiv:2310.06940, 2023.

[26] J. Lafferty, A. McCallum, F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in Icml, 1(2): 3, Williamstown, MA, 2001.

[27] H. Cai et al., "Memd-absa: a multi-element multi-domain dataset for aspect-based sentiment analysis," arXiv preprint arXiv:2306.16956, 2023.

[28] H. Chen, Z. Zhai, F. Feng, R. Li, X. Wang, "Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction," in Proc. the 60th Annual Meeting of the Association for Computational Linguistics (volume 1: long papers): 2974-2985, 2022.

[29] Z. Zhai, H. Chen, F. Feng, R. Li, X. Wang, "COM-MRC: a context-masked machine reading comprehension framework for aspect sentiment triplet extraction," in Proc. the 2022 Conference on Empirical Methods in Natural Language Processing: 3230-3241, 2022.

[30] L. Xianlong, M. Yang, Y. Wang, "Tagging-assisted generation model with encoder and decoder supervision for aspect sentiment triplet extraction," in Proc. the 2023 Conference on Empirical Methods in Natural Language Processing: 2078-2093, 2023.

[31] J. Yang, F. Dai, F. Li, Y. Xue, "Span-based pair-wise aspect and opinion term joint extraction with contrastive learning," in Proc. CCF International Conference on Natural Language Processing and Chinese Computing: 17-29, 2023.

[32] P. Li, P. Li, X. Xiao, "Aspect-pair supervised contrastive learning for aspect-based sentiment analysis," Knowledge-Based Syst., 274: 110648, 2023.

[33] S. Yang, X. Jiang, H. Zhao, W. Zeng, H. Liu, Y. Jia, "FaiMA: Feature-aware in-context learning for multi-domain aspect-based sentiment analysis," arXiv preprint arXiv:2403.01063, 2024.

Biographies

Atefeh Mohammadi was born in 1984. She received her B.Sc. degree in Computer Software Engineering from Najafabad Branch of Islamic Azad University, and her M.Sc. degree in Computer Software Engineering from Isfahan University. She is currently a Ph.D. student at Yazd University. Her research interests are Natural Language Processing, Deep Learning, and Image Processing.

• Email: atefehmohammadi@stu.yazd.ac.ir
• ORCID: 0000-0001-7811-5877
• Web of Science Researcher ID: NA
• Scopus Author ID: NA
• Homepage: NA

Mohammad Reza Pajoohan is an Associate Professor in the Department of Computer Engineering at Yazd University, Yazd, Iran. He earned his Ph.D. from University Sains Malaysia. His research interests include Data Mining and Knowledge Discovery, Medical Data Processing, Databases and Big Data, Privacy Preservation Data Publication and Business Process Reengineering.

• Email: pajoohan@yazd.ac.ir
• ORCID: 0000-0002-7848-2025
• Web of Science researcher ID: NA
• Scopus Author ID: NA
• Homepage: https://pws.yazd.ac.ir/pajoohan/

Ali Mohammad Zareh Bidoki obtained his Ph.D. in software engineering at the Electrical & Computer Engineering department of the University of Tehran. He also obtained a master's degree from the ECE department of the University of Tehran and a Bachelor's degree from the Computer Engineering department of Isfahan University of Technology. He is currently an associate professor in the CE department of Yazd University and CEO of Parsijoo (a search engine company). His expertise is information retrieval and search engines.

• Email: alizareh@yazd.ac.ir
• ORCID: 0009-0005-4270-0835
• Web of Science researcher ID: NA
• Scopus Author ID: NA
• Homepage: https://yazd.ac.ir/en/people/zareh