Extensible Multi-Granularity Fusion Network for Aspect-based Sentiment Analysis

Anonymous ACL submission

Abstract

Aspect-based Sentiment Analysis (ABSA) evaluates sentiment expressions within a text to comprehend sentiment information. Previous studies integrated external knowledge, such as knowledge graphs, to enhance the semantic features in ABSA models. Recent research has examined the use of Graph Neural Networks (GNNs) on dependency and constituent trees for syntactic analysis. With the ongoing development of ABSA, more innovative linguistic and structural features are being incorporated (e.g. latent graph), but this also introduces complexity and confusion. As of now, a scalable framework for integrating diverse linguistic and structural features into ABSA does not exist. This paper presents the Extensible Multi-Granularity Fusion (EMGF) network, which integrates information from dependency and constituent syntactic, attention semantic , and external knowledge graphs. EMGF, equipped with multi-anchor triplet learning and orthogonal projection, efficiently harnesses the combined potential of each granularity feature and their synergistic interactions, resulting in a cumulative effect without additional computational expenses. Experimental findings on SemEval 2014 and Twitter datasets confirm EMGF's superiority over existing ABSA methods [1].

1 Introduction

The primary objective of the Aspect-Based Sentiment Analysis(ABSA) task is to assess the sentiment polarity associated with specific aspects or entities in a text, enabling a more comprehensive understanding of the text's sentiment information. For example, give a laptops review "***Looks*** *nice , but has a horribly cheap* ***feel*** ." and the sentiment polarity of the two aspects "*Looks*" and "*feel*" are positive and negative, respectively. Therefore, ABSA accurately identifies the sentiment orientation for individual aspects, rather than assigning a general sentiment label to a whole sentence in sentence-level sentiment analysis. The main challenge of ABSA is to model the relationship between aspects and their associated opinions.

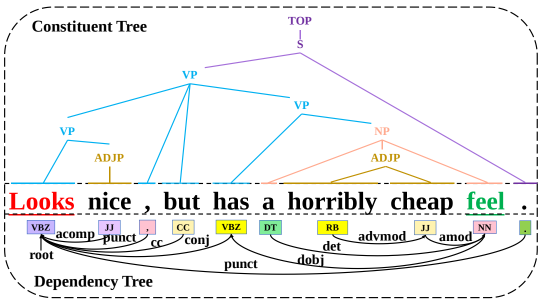

Figure 1: An example sentence with its dependency tree and constituent tree. This sentence from the laptops reviews, contains two aspects but with opposite sentiment polarities.

To this end, previous studies (Ma et al., 2018; Zhou et al., 2020; Zhong et al., 2023) leveraged external knowledge to enhance semantic features in ABSA models. For example, Zhou et al. (2020) employed words related to knowledge graphs to build subgraphs as seed nodes. Subgraph-based approaches yielded remarkable outcomes but may entail complexity, particularly when dealing with a large number of aspect terms. Zhong et al. (2023) incorporated external knowledge graphs into low-dimensional embeddings to efficiently represent aspect-specific knowledge.

More recent studies (Zhang et al., 2019; Sun et al., 2019; Chen et al., 2020; Liang et al., 2020; Wang et al., 2020; Li et al., 2021; Liang et al., 2022) have extensively investigated the use of Graph Neural Networks (GNNs) on dependency trees (Dep.Tree) and constituent trees (Con.Tree) to explicitly leverage sentence syntactic structures. While constituency and dependency trees share common sentential syntactic information, they capture syntactic details from distinct perspectives

[1] Code and datasets are available at `https://anonymous.4open.science/r/EMGF-E7A6`

(Dong et al., 2022).

Dependency trees (Dep.Trees) can establish connections among words in a sentence (Li et al., 2021), while constituent trees (Con.Trees) provide precise phrase segmentation and hierarchical structures, which facilitate precise alignment of aspects with sentiment-indicative words (Liang et al., 2022). We illustrate this with an example in Figure 1: (1) A dependency relation exists between the aspect term "*Looks*" and the opinion term "*nice*"; (2) The phrase segmentation term "*but*" segments "*Looks nice*" from "*has a horribly cheap feel*".

Most of the previous work has already established the effectiveness of single-granularity information for the ABSA task. However, single-granularity features are insufficient to fully capture the rich information contained in the raw data. Li et al. (2021) incorporating SynGCN and SemGCN networks through a Mutual BiAffine module, demonstrating the effectiveness of integrating these two granularity levels for the ABSA task.

However, most current methods use complex and inefficient techniques to integrate diverse types of knowledge. Currently, there is no scalable framework capable of combining various multi-granularity features (*e.g.,* syntactic, semantics, external knowledge graphs information, and so on) to enhance model performance. In this context, a fundamental question arises: **How can we ensure that the combination of multiple granularity features achieves a cumulative effect [2] and addresses the problem of model scalability?**

In this paper, we introduce a novel architecture called the ***E**xtensible **M**ulti-**G**ranularity **F**usion Network* model (EMGF) to address the aforementioned challenges. **Firstly**, we enhance the acquisition of affective representations in ABSA tasks by integrating information from dependency syntax, constituent syntax, semantic attention, and external knowledge graphs. **Secondly**, we have developed an Extensible Multi-Stage Fusion (EMSF) module designed to capture profound and intricate interactions among features at various granularities. Moreover, it can integrate information of multiple granularities at an extremely low computational cost, thereby achieving scalability. To elaborate, our module comprises two stages: the "preprocessing stage" and the "fusion stage." In the "preprocessing stage," we employ a multi-anchor triplet learning approach to combine dependency and constituent syntactic information, enhancing their mutual complementarity. We also utilize an orthogonal projection layer to acquire refined syntactic and semantic discriminative features. **Finally**, external knowledge graphs offer supplementary information support during the "fusion stage."

Our contributions are highlighted as follows:

1) For the ABSA task, we present an Extensible Multi-Granularity Fusion Network designed to capture intricate interactions among features at various granularities, thus achieving the cumulative effect.

2) This network can fuse an arbitrary number of features of different granularities in an expandable manner, at an extremely low computational cost.

3) We present multi-anchor triplet learning to enable mutual learning between dependency syntax and constituent syntax, and employ orthogonal projection techniques to obtain refined syntactic and semantic features.

4) Our experimental findings establish that our EMGF model surpasses the current state-of-the-art methods when evaluated on the SemEval 2014 and Twitter datasets, demonstrating the effectiveness of our EMGF model.

2 Related Work

ABSA is an entity-level and fine-grained task for sentiment analysis (Li et al., 2021; Ma et al., 2023). Early research in ABSA makes use of attention-based neural models for the purpose of capturing semantic interactions (Wang et al., 2016; Ma et al., 2017; Xu et al., 2019).

Dependency with GNNs: Another emerging trend is the effective incorporation of dependency trees with Graph Neural Networks (GNNs). Xu et al. (2020) introduce a GCN model with a heterogeneous graph, merging sentence and aspect nodes via four relationship types, Liang et al. (2021) propose a novel dependency syntactic knowledge augmented interactive architecture with multi-task learning, Zhang et al. (2022) enhance attention score matrices with syntactic mask matrices for integrating syntax and semantics, Zhao et al. (2023) introduce RDGCN to better calculate dependency importance, tackling syntactic ambiguities in aspect-opinion analysis.

Constituent with GNNs: Structural syntax knowledge has been proven effective for seman-

[2] Combining multiple features from various granularity levels results in incremental effects. Specifically, with each additional feature included, the effect improves compared to the previous combination.

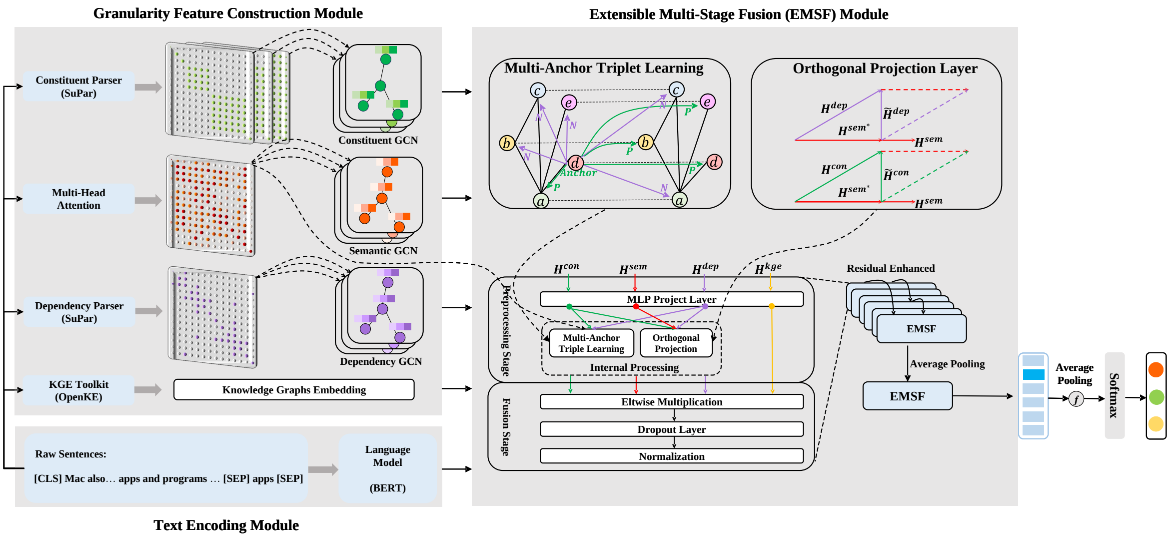

Figure 2: The overall architecture of our EMGF model.

tic role labeling (SRL) (Marcheggiani and Titov, 2020; Fei et al., 2021). Marcheggiani and Titov (2020) showcases the utilization of GCNs to encode constituent structures in an SRL system, Fei et al. (2021) jointly learns phrasal boundaries extracted from constituency and semantic relations from dependency to explore the integration of diverse syntactic representations for SRL. For ABSA, Liang et al. (2022) first focus on effectively harnessing syntactic information from the sentence's constituent tree to model the sentiment context of individual aspects for learning.

3 Methodology

In this section, we provide a detailed explanation of EMGF. The overview of the EMGF framework is shown in Figure 2. The system comprises three components: 1) The Text Encoding Module. 2) The Granularity Feature Construction Module. 3) The Extensible Multi-Stage Fusion Module.

3.1 Text Encoding Module

In the ABSA task, give a n-word sentence $s = \{w_1, w_2, \ldots, w_n\}$, along with a specific aspect represented as $a = \{a_1, a_2, \ldots, a_m\}$, to determine its corresponding sentiment polarity class, c_a. Here, a is a sub-sequence of s, and $c_a \in \{Positive, Neutral, Negative\}$. To obtain contextualized representations, we utilize BERT (Devlin et al., 2019). In the BERT encoder, we construct a sentence-aspect pair as input, represented as $x = ([\text{CLS}]s[\text{SEP}]a[\text{SEP}])$. The output provides contextualized representations, denoted as $H^{\text{bert}} = \text{BERT}(x)$. In this representation, $H^{\text{bert}} = \left[h_1^{\text{bert}}, h_2^{\text{bert}}, \cdots, h_n^{\text{bert}}\right] \in \mathbb{R}^{n \times d}$, where d represents the dimensionality of the last hidden layer of BERT, and h_i^{bert} corresponds to the contextual representation of the i-th word.

3.2 Granularity Feature Construction Module

Dependency GCN The dependency graph convolutional networks (DepGCN) module takes syntactic encoding as input and utilizes the probability matrix of all dependency arcs from a dependency parser to encode syntax information. The dependency graph is embodied as an adjacency matrix $A^{\text{dep}} \in \mathbb{R}^{n \times n}$, which is defined as follows:

$$A_{ij}^{\text{dep}} = \begin{cases} 1, & \text{if link}(i,j) = 1 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $\text{link}(i,j)$ shows that i-th and j-th words have a dependence link. The dependency graph representation $H^{\text{dep}} = \{h_1^{\text{dep}}, h_2^{\text{dep}}, \ldots, h_n^{\text{dep}}\}$ is then obtained from the DepGCN module using the following formula:

$$h_i^l = \sigma\left(\sum_{j=1}^{n} A_{ij} W^l h_j^{l-1} + b^l\right) \tag{2}$$

here, W^l represents a weight matrix, b^l denotes a bias term, and σ is an activation function, such as ReLU.

Constituent GCN We follow the syntax structure of the Con.Tree in a bottom-up manner, inspired by BiSyn-GAT+ (Liang et al., 2022). The Con.Tree is composed of multiple phrases (ph_u^l)

that make up the input text, and we create corresponding graphs based on these phrases ph_u^m.

Given the substantial depth of the constituent tree, we choose a total of m layers with alternating intervals [3]. We make this choice because the variation in phrase hierarchical information between adjacent layers is minimal, and excessive alignment would be an inefficient use of computational resources. Additionally, the chosen value of m aligns with the number of ConGCN layers.

The constituent graph is embodied as an adjacency matrix $A^{\text{con}} \in \mathbb{R}^{l_c \times n \times n}$, which is defined as follows:

$$\mathbf{A}_{i,j}^{\text{con(m)}} = \begin{cases} 1 & \text{if } w_i \text{ and } w_j \text{ are in same } ph_u^m, \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where m denotes the level of the phrase within the selected l_c layers, while u denotes the constituent label associated with the phrase, such as S, NP, VP, and so on. Subsequently yields the output hidden representation $H^{\text{con}} = \{h_1^{\text{con}}, h_2^{\text{con}}, \ldots, h_n^{\text{con}}\}$ is then obtained from the ConGCN module using Eq. (2).

Semantic GCN To construct the attention score matrix A^{sem}, we employ the Multi-Head Attention (MHA) mechanism on the hidden state features H^{bert} derived from the BERT encoder. The MHA computes attention scores among words, and the formulation of the attention score matrix $A^{\text{sem}} \in \mathbb{R}^{n \times n}$ is as follows:

$$A_{ij}^{\text{sem}} = Softmax(\text{MHA}(h_i^{\text{bert}}, h_j^{\text{bert}})) \tag{4}$$

Subsequently yields the output hidden representation $H^{\text{sem}} = \{h_1^{\text{sem}}, h_2^{\text{sem}}, \ldots, h_n^{\text{sem}}\}$ is then obtained from the SemGCN module using Eq. (2).

External Knowledge Zhong et al. (2023) synergistically combine contextual and knowledge information to achieve more comprehensive feature representations. We introduce the external knowledge as proposed by them, represented as $H^{\text{kge}} = \{h_1^{\text{kge}}, h_2^{\text{kge}}, \ldots, h_n^{\text{kge}}\}$.

3.3 Extensible Multi-Stage Fusion Module

In previous studies, it is common to combine only two granularity features, so when trying to combine additional features, the current model is no longer applicable. This means that a model needs to be redesigned that can be compatible with multiple features at the same time, but this will increase the complexity and computational cost of the model.

To address this challenge and capture intricate interactions among features at different granular levels while efficiently integrating diverse granular information, we introduce the extensible multi-granularity fusion (EMGF) module. This innovative approach allows for the expansion and effective exploration of interrelationships among multi-granular features. It achieves this by cascading multiple Extensible Multi-Stage Fusion (EMSF) blocks, each comprising a "preprocessing stage" and a "fusion stage." During the preprocessing stage of EMSF, four features from different levels serve as inputs, namely H^{con}, H^{dep}, H^{sem}, and H^{kge}.

3.3.1 Preprocessing Stage

Con.Tree and Dep.Tree share syntactic information from different viewpoints (Dong et al., 2022). (Ata et al., 2021; Dong et al., 2022) use multi-view learning to study three relationship categories: intra-node intra-view, intra-node inter-view, and inter-node inter-view. We collectively label nodes in these scenarios as "important nodes." However, there is currently no research addressing how to handle "non-important nodes," which could potentially disrupt the complementary learning of "important nodes." Moreover, to handle these three types of collaboration, it's necessary to design three distinct loss functions, adding complexity to the model. To this end, we propose Multi-Anchor Triplet Learning to address the two categories of issues mentioned above.

Additionally, inspired by Qin et al. (2020), we utilize orthogonal projection techniques to encourage the DepGCN and ConGCN networks to acquire distinct syntactic features from the semantic features generated by the SemGCN network. This results in refined and more discriminative syntactic and semantic features.granularity levels.

Within this stage, we combine Multi-Anchor Triplet Learning and Orthogonal Projection Techniques to effectively capture the complementary and discriminative aspects of features across various granularity levels.

Multi-Anchor Triplet Learning We choose a node from the con-view graph as the "Anchor" node and consider three scenarios: **1)** In the con-view, all nodes connected to the Anchor are marked

[3] For instance, you can choose layer 1, skip one layer, pick layer 3, and continue this pattern.

as "pos" nodes, **2)** Nodes in the dep-view that share homologous [4] properties with the "Anchor" node in the con-view are likewise designated as "pos" nodes, **3)** All "pos" nodes in the con-view correspond to homologous nodes in the dep-view. If these nodes are not connected to homologous nodes corresponding to the Anchor, they are still labeled as "pos" nodes. All other cases are labeled as "neg" nodes. The same procedure is applied in the dep-view when the Anchor node is located there.

It is vital to stress that nodes do not possess equal significance. Designating all graph nodes as Anchor nodes would undermine differentiation and precision. Additionally, drawing inspiration from the work of MP-GCN (Zhao et al., 2022), we employ the Multi-Head S-Pool to select Anchor nodes. Specifically, we use the attention matrix A^{sem} to conduct both average and maximum pooling from two distinct perspectives, resulting in the selection of the Top-K important nodes with the highest scores.

Our goal is to have the "Anchor" node stay close to the "pos" nodes to acquire complementary knowledge, while minimizing interference from "neg" nodes. Specifically, we accomplish this goal by minimizing the following loss function:

$$\begin{aligned}\mathcal{L}_{\text{triplet}} = \sum_{i\in\text{Anchor}} \sigma\Big(&\sum_{j\in\text{pos}} f_a(||h_i^z - h_j^{z'}||_2) \\ &- \sum_{j\in\text{neg}} f_a(||h_i^z - h_j^{z'}||_2) + \text{margin}\Big)\end{aligned} \tag{5}$$

$$\text{Anchor} = TopK\left(f_a\left(A^{\text{sem}}\right) + f_m\left(A^{\text{sem}}\right)\right) \tag{6}$$

where z and z' belong to the set $\{dep, con\}$, we determine the size of the anchor set k based on Bourgain's Theorem-1 (You et al., 2019). Here, k is expressed as $k = c\log^2 n$, with c representing a constant, and n denoting the total number of nodes in the graph. Our approach employs functions f_a for average pooling and f_m for maximum pooling. The "margin" hyperparameter in controlling the boundary of the triplet loss function, and σ corresponds to the non-linear activation function ReLU.

Orthogonal Projection Techniques Mathematically, we first project dependency syntax feature H^{dep} onto semantic feature H^{sem}:

$$H^{\text{dep}^*} = Proj\left(H^{\text{dep}}, H^{\text{sem}}\right) \tag{7}$$

[4]A homologous node refers to a node that corresponds to the same entity in different data views.

where $Proj$ is a projection function.

$$Proj(x, y) = \frac{x \cdot y}{|y|}\frac{y}{|y|} \tag{8}$$

where x and y are vectors. Next, we perform the projection in the orthogonal direction of the projected feature H^{dep} to obtain a purer classification feature vector.

$$\widetilde{H^{\text{dep}}} = Proj\left(H^{\text{dep}}, (H^{\text{dep}} - H^{\text{dep}^*})\right) \tag{9}$$

Correspondingly, the terms $\widetilde{H^{\text{con}}}$ in the formula can be expressed as follows:

$$\widetilde{H^{\text{con}}} = Proj\left(H^{\text{con}}, (H^{\text{con}} - H^{\text{con}^*})\right) \tag{10}$$

3.3.2 Fusion Stage

Building on the preprocessing stage, we utilize the purified dependency syntatic $\widetilde{H^{\text{dep}}}$, the purified constituent syntactic $\widetilde{H^{\text{con}}}$, the semantic feature H^{sem}, and the extra knowledge feature H^{kge} as inputs during the fusion stage. Furthermore, inspired by the multimodal fusion method MAMN (Xue et al., 2023a,b), we adopt the extended multimodal factorized bilinear pooling mechanism from MAMN in fusion stage to fuse $\widetilde{H^{\text{dep}}}$, $\widetilde{H^{\text{con}}}$, H^{sem}, and external knowledge feature H^{kge}. The Fusion Stage is calculated as:

$$\begin{aligned}\mathcal{Z}_m^i = Norm\Big(&\tilde{\text{U}}_{\text{dep}}^T \widetilde{H^{\text{dep}}} \circ \tilde{\text{U}}_{\text{con}}^T \widetilde{H^{\text{con}}} \\ &\circ \tilde{\text{U}}_{\text{sem}}^T H^{\text{sem}} \circ \tilde{\text{U}}_{\text{kge}}^T H^{\text{kge}}\Big)\end{aligned} \tag{11}$$

where all $\tilde{\text{U}}$ represent learnable weight parameters, $Norm$ denotes the normalization layer, and $\mathcal{Z}_m^i$ represents the outputs of the fusion stages within the i-th EMSF block. Additionally, we have introduced residual connections between different blocks. Subsequently, we calculate the average of the outputs from these l_e EMSF blocks (where l_e indicates the number of EMSF blocks) to obtain the feature r with four distinct granularity fusions. The specific formula is as follows:

$$\begin{aligned}\mathcal{Z}_m^{i+1} &= \mathcal{Z}_m^i + \text{EMSF}(\\ &\mathcal{Z}_m^i, \widetilde{H^{\text{dep}}}, \widetilde{H^{\text{con}}}, H^{\text{sem}}, H^{\text{kge}})\end{aligned} \tag{12}$$

To obtain the final output, denoted as r for the EMGF, we concatenate the output features from the l_m EMSF blocks and apply average pooling.

$$r = \text{Mean}\left(\mathcal{Z}_m^1, \mathcal{Z}_m^2, \ldots, \mathcal{Z}_m^{l_m}\right) \tag{13}$$

Dataset	#Positve			#Negative			#Neutral		
	Train	Dev	Test	Train	Dev	Test	Train	Dev	Test
Laptop	976	-	337	851	-	128	455	-	167
Restaurant	2164	-	727	807	-	196	637	-	196
Twitter	1507	-	172	1528	-	169	3016	-	336
MAMS	3380	403	400	2764	325	329	5042	604	607

Table 1: Satistics of four datasets.

3.4 Model Training

Softmax Classifier Subsequently, the fusion feature r, obtained from the granularity fusion module, is used to calculate the sentiment probability distribution $\hat{y}_{(s,a)}$ via a linear layer equipped with a softmax function:

$$\hat{y}_{(s,a)} = Softmax\left(W_p r + b_p\right) \tag{14}$$

where (s, a) is a sentence-aspect pair.

Our training goal is to minimize the following overall objective function:

$$\mathcal{L}(\Theta) = \mathcal{L}_c + \beta \mathcal{L}_{\text{triplet}} \tag{15}$$

where Θ denotes all the trainable model parameters, while β are hyperparameters. The cross-entropy loss $\mathcal{L}_c$ for the primary classification task is defined as follows:

$$\mathcal{L}_c = \sum_{(s,a)\in\mathcal{D}} y_{(s,a)} \log \hat{y}_{(s,a)} \tag{16}$$

where $\mathcal{D}$ contains all sentence-aspect pairs and $y_{(s,a)}$ is the real distribution of sentiment.

4 Experiments

4.1 Datasets

Our model was evaluated using four benchmark datasets: Laptop and Restaurant from SemEval2014 Task 4 (Pontiki et al., 2014), Twitter (Dong et al., 2014), and the large-scale multi-aspect MAMS dataset (Jiang et al., 2019). Consistent with prior studies (Chen et al., 2017; Li et al., 2021; Tang et al., 2022) and others, we excluded instances labeled as "conflict." The statistics of these datasets are presented in Table 1.

4.2 Implementation Details

We utilized SuPar [5] as our parser to acquire both the dependency and constituent tree. For constructing our model, we employed the uncased base version of BERT [6] with a dropout rate of 0.3. The training process was conducted with a batch size of 16, utilizing the Adam optimizer with a learning rate of 2e-5. For the four datasets, we set the ConGCN, DepGCN, and SemGCN layers to (6, 3, 6, 6), (3, 3, 9, 3), and (3, 3, 1, 3), respectively, with β coefficients of (0.12, 0.12, 0.07, 0.12). We selected 3 layers (l_c) for the constituent tree and optimized its performance. Additionally, we determined that 6 layers (l_e) are optimal for EMSF blocks. The hyper-parameter margin was set to 0.2. Each experiment is replicated three times, with the results then averaged for consistency. Our primary evaluation metrics include accuracy (Acc.) and macro-f1 (F1).

4.3 Baseline Methods

We compare our EMGF with state-of-the-art baselines, described as follows:
1) **BERT-SRC** (Devlin et al., 2019) represents the fine-tuning of BERT to incorporate aspect-specific representations. 2) **CDT** (Sun et al., 2019) investigate combining dependency trees and neural networks for representation learning. 3) **DualGCN** (Li et al., 2021) simultaneously considers the complementarity of syntax structures and semantic correlations. 4) **SSEGCN** (Zhang et al., 2022) integrates aspect-aware and self-attention mechanisms to enhance the precision of ABSA. 5) **MGFN** (Tang et al., 2022) utilize a latent graph to leverage dependency relation and semantic information. 6) **TF-BERT** (Zhang et al., 2023) examines span-level consistency in multi-word opinion expressions. 7) **HyCxG** (Xu et al., 2023) introduce construction grammar (CxG) to enrich language representation.

4.4 Main Results

Table 2 summarizes the primary experimental results. The EMGF model exceeds the current state-of-the-art (SOTA) baseline, HyCxG (Xu et al., 2023), in both the Laptop and Restaurant benchmarks. Models that incorporate syntactic dependency information tend to outperform those that do not, but relying solely on syntactic information may lead to subpar performance, particularly with informal or complex sentences. Leveraging richer syntax dependency labels and incorporating affective semantic information, as demonstrated by models such as (Li et al., 2021; Tang et al., 2022), generally outperforms syntax-only models, highlighting the effectiveness of integrating diverse

[5] https://github.com/yzhangcs/parser

[6] https://github.com/huggingface/transformers

Model	Laptop		Restaurant		Twitter		MAMS	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
BERT-SRC (Devlin et al., 2019)	78.99	75.03	84.46	76.98	73.55	72.14	82.34	81.94
CDT (Sun et al., 2019)	79.70	75.61	86.36	80.16	77.50	76.54	-	-
DualGCN (Li et al., 2021)	81.80	78.10	87.13	81.16	77.40	76.02	-	-
SSEGCN (Zhang et al., 2022)	81.01	77.96	87.31	81.09	77.40	76.02	-	-
MGFN (Tang et al., 2022)	81.83	78.26	87.31	82.37	78.29	77.27	-	-
TF-BERT (Zhang et al., 2023)	81.49	78.30	86.95	81.43	77.84	76.23	-	-
HyCxG (Xu et al., 2023)	**82.29**	79.11	87.32	82.24	-	-	85.03	84.40
Our EMGF	82.11	**79.24**	**88.42**	**83.20**	**78.87**	**78.06**	**85.48**	**84.73**

Table 2: Experimental results on ABSA datasets with BERT encoder. The best result on each dataset is in **bold.**

Model	Laptop		Restaurant		Twitter		MAMS	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
Our EMGF (M4)	**82.11**	**79.24**	**88.42**	**83.20**	**78.87**	**78.06**	**85.48**	**84.73**
EMGF-M3	81.26	78.24	87.78	82.23	77.53	76.27	85.11	84.13
EMGF-M2	80.79	77.61	87.33	81.59	76.64	76.12	84.32	83.75
EMGF-M1	80.15	77.07	86.24	80.12	76.49	75.05	83.34	82.73
W/O $\mathcal{L}_{\text{triplet}}$	80.84	76.83	86.97	81.06	76.93	75.61	83.54	83.21
W/O Orthogonal Projection	79.41	75.24	86.15	80.22	77.83	76.53	84.44	84.13
W/O Dep Project Sem	80.52	76.92	86.15	79.96	76.04	75.20	84.44	83.87
W/O Con Project Sem	80.37	76.47	85.70	79.66	76.19	74.98	83.99	83.48

Table 3: Ablation study experimental results.

feature information. Experimental results indicate that our EMGF effectively integrates information from four different granularities.

4.5 Ablation Study

We evaluated the extensibility of EMGF and the effectiveness of its fusion approach by investigating how the number of granularity features affects EMGF's performance, the results are shown in Table 3. M4 indicates using all granularity features, M3 represents selecting three out of four granularity features (selected through combinatorial permutations) and averaging all possibilities. M2 and M1 follow a similar pattern. As we reduced the number of granularity features, we observed a decrease in performance, highlighting the extensibility of EMGF and the effectiveness of our fusion approach, which cumulative effects. *W/O* $\mathcal{L}_{\text{triplet}}$ result in reduced performance of EMGF, this shows that multi-anchor triplet learning can gather complementary knowledge from various syntactic feature information, thereby improving the model's performance. The expression "Dep Project Sem (Con Project Sem)" denotes the projection of syntactic features related to dependency (constituent) onto orthogonal spaces associated with semantic features. *W/O* Dep Project Sem, *W/O* Con Project Sem, and *W/O* Orthogonal Projection Techniques, all lead to a decrease in EMGF performance. This implies that omitting the feature projections hinders the model's ability to accurately differentiate between syntactic and semantic information, causing interference from redundant data during the fusion stage.

4.6 Case Study

Table 4 illustrates our model through four examples. Identifying neutral sentiment is challenging due to a lack of strong sentiment words in neutral texts and data imbalance, with more data available for positive and negative sentiments. In the third sentence, MGFN incorrectly predicted the emotional polarity of "*chef*." This can be attributed to MGFN's inability to capture its specific opinion words associated with "*chef*," it incorrectly treated the opinion words from "*food*" and "*service*" as its own. The fourth sentence is particularly challenging, as MGFN, like many models, assigns positive sentiment to an aspect word without strong emotional cues, causing three out of four EMGF predictions to be incorrect. Drawing from our analysis, MGFN combines syntactic features derived from the latent graph with semantic features. However, similar to other models, MGFN does not fully capitalize on the potential offered by a variety of granularity features. In juxtaposition, our EMGF effectively leverages these features and their synergistic effects through multi-anchor triplet learning

Sentence	MGFN (Tang et al., 2022)	EMGF(Ours)
I know real [**Indian food**]$_{neg}$ and this wasn't it.	(neu ✗)	(neg ✔)
Our [**waiter**]$_{pos}$ was friendly and it is a shame that he didnt have a supportive [**staff**]$_{neg}$ to work with.	(pos ✔, pos ✗)	(pos ✔, neg ✔)
Even when the [**chef**]$_{neu}$ is not in the house, the [**food**]$_{pos}$ and [**service**]$_{pos}$ are right on target.	(pos ✗, pos ✔, pos ✔)	(neu ✔, pos ✔, pos ✔)
We started with the [**scallops**]$_{neu}$ and [**asparagus**]$_{neu}$ and also had the [**soft shell crab**]$_{neu}$ as well as the [**cheese plate**]$_{neu}$.	(pos ✗, pos ✗, pos ✗, neu ✔)	(neu ✔, neu ✔, neu ✔, neu ✔)

Table 4: Case study experimental results of MGFN and EMGF.

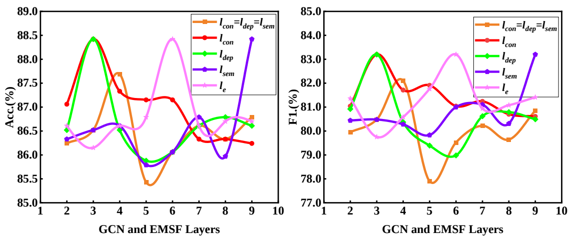

Figure 3: The impact of the number of GCN and EMSF block on Restaurant dataset.

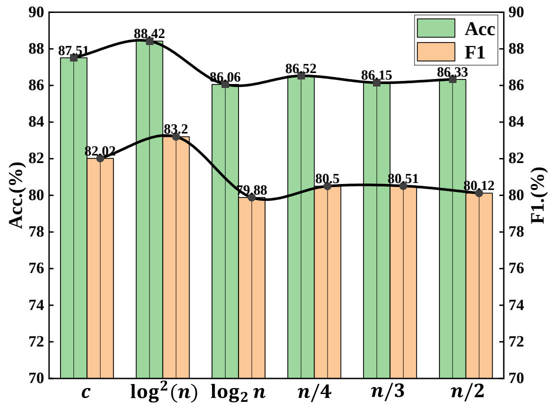

Figure 4: The impact of different k on Restaurant dataset.

and orthogonal projection.

4.7 Impact of Number of GCN and EMSF Blocks

We varied the number of layers, l_{con}, l_{dep}, and l_{sem} from 2 to 9 for ConGCN, DepGCN, and SemGCN to assess their impact on the model's performance on Restaurant dataset. Based on experimental results, we set l_{con}, l_{dep}, and l_{sem} to 3, 3, and 9, respectively. Interestingly, maintaining consistent layer numbers for l_{con}, l_{dep}, and l_{sem} does not necessarily result in optimal performance. We observed that considering the layer count separately for each of the three GCN types tends to enhance performance. The number of cascaded EMFB blocks (denoted as l_e) affects prediction accuracy and F1 score. Through experiments, we determined that the optimal number of modules is 6, as depicted in Figure 3.

4.8 Hype-parameter Analysis

We will investigate the impact of a crucial parameter, k, in EMGF. This relates to selecting the number of crucial nodes in each view. We have conducted experiments with various k values, such as $c, \log^2(n), \log_2 n$, $\frac{n}{4}$, $\frac{n}{3}$, $\frac{n}{2}$, where c is a constant, and n represents the number of view nodes. The value of c varies from 1 to 5, and we calculate the average performance. We can see from Figure 4 that the average performance reaches its peak when k equals $\log^2(n)$.

5 Conclusion

Through efficient integration of diverse granularity features, including dependency and constituent syntactic, attention semantic, and external knowledge graphs, EMGF demonstrates superior performance compared to existing ABSA methods. This study has tackled the persistent challenge of fully leveraging the combined potential of diverse granularity features in the ABSA framework. EMGF effectively captures complex interactions among these features by employing multi-anchor triplet learning and orthogonal projection techniques, yielding a cumulative effect without incurring additional computational expenses. EMGF offers a scalable and flexible framework for integrating a variety of multi-granularity features in ABSA, thereby enhancing model performance.

Limitations

Although our research has achieved commendable results, there are limitations worth acknowledging.

These limitations underscore areas for future improvement and exploration. In this experiment, due to limited computational resources, we selected the top-k nodes as Anchor nodes in multi-anchor triplet learning. However, when we attempted to set the value of k to $\{\log_2 n, \frac{n}{4}, \frac{n}{3}, \frac{n}{2}\}$ magnitude, we observed that the model training was excessively slow, and we had to adjust the magnitude of k to a smaller scale for experimentation. Finally, due to constraints in computational power and time, we were unable to explore larger model architectures or conduct extensive hyperparameter tuning. We hope that future research can address these limitations to enhance the reliability and applicability of the method we propose.

Ethics Statement

Our research adhered to rigorous ethical guidelines and principles throughout its execution. All participants were granted informed consent, with clear and comprehensive information regarding the study's objectives and procedures. We are committed to reporting the study's findings and results objectively and accurately, without any form of manipulation or misrepresentation. Our dedication to upholding the highest ethical standards in research ensures the integrity and validity of our discoveries. None of the authors of this article conducted studies involving human participants or animals. Furthermore, we affirm that none of the authors have any known competing financial interests or personal relationships that might potentially influence the work presented in this paper.

References

Sezin Kircali Ata, Yuan Fang, Min Wu, Jiaqi Shi, Chee Keong Kwoh, and Xiaoli Li. 2021. Multi-view collaborative network embedding. *ACM Trans. Knowl. Discov. Data*, 15(3):39:1–39:18.

Chenhua Chen, Zhiyang Teng, and Yue Zhang. 2020. Inducing target-specific latent structures for aspect sentiment classification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5596–5607, Online. Association for Computational Linguistics.

Peng Chen, Zhongqian Sun, Lidong Bing, 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, Copenhagen, Denmark. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Kuicai Dong, Aixin Sun, Jung-Jae Kim, and Xiaoli Li. 2022. Syntactic multi-view learning for open information extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4072–4083, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

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 (Volume 2: Short Papers)*, pages 49–54, Baltimore, Maryland. Association for Computational Linguistics.

Hao Fei, Shengqiong Wu, Yafeng Ren, Fei Li, and Donghong Ji. 2021. Better combine them together! integrating syntactic constituency and dependency representations for semantic role labeling. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 549–559, Online. Association for Computational Linguistics.

Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao, and Min Yang. 2019. A challenge dataset and effective models for aspect-based sentiment analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6280–6285, Hong Kong, China. Association for Computational Linguistics.

Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard Hovy. 2021. Dual graph convolutional networks for aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6319–6329, Online. Association for Computational Linguistics.

Bin Liang, Rongdi Yin, Lin Gui, Jiachen Du, and Ruifeng Xu. 2020. Jointly learning aspect-focused and inter-aspect relations with graph convolutional networks for aspect sentiment analysis. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 150–161, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Shuo Liang, Wei Wei, Xian-Ling Mao, Fei Wang, and Zhiyong He. 2022. BiSyn-GAT+: Bi-syntax aware graph attention network for aspect-based sentiment analysis. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1835–1848, Dublin, Ireland. Association for Computational Linguistics.

Yunlong Liang, Fandong Meng, Jinchao Zhang, Yufeng Chen, Jinan Xu, and Jie Zhou. 2021. A dependency syntactic knowledge augmented interactive architecture for end-to-end aspect-based sentiment analysis. *Neurocomputing*, 454:291–302.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 4068–4074. ijcai.org.

Fukun Ma, Xuming Hu, Aiwei Liu, Yawen Yang, Shuang Li, Philip S. Yu, and Lijie Wen. 2023. AMR-based network for aspect-based sentiment analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 322–337, Toronto, Canada. Association for Computational Linguistics.

Yukun Ma, Haiyun Peng, and Erik Cambria. 2018. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 5876–5883. AAAI Press.

Diego Marcheggiani and Ivan Titov. 2020. Graph convolutions over constituent trees for syntax-aware semantic role labeling. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3915–3928, Online. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

Qi Qin, Wenpeng Hu, and Bing Liu. 2020. Feature projection for improved text classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8161–8171, Online. Association for Computational Linguistics.

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 (EMNLP-IJCNLP)*, pages 5679–5688, Hong Kong, China. Association for Computational Linguistics.

Siyu Tang, Heyan Chai, Ziyi Yao, Ye Ding, Cuiyun Gao, Binxing Fang, and Qing Liao. 2022. Affective knowledge enhanced multiple-graph fusion networks for aspect-based sentiment analysis. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 5352–5362, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

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*, pages 3229–3238, Online. Association for Computational Linguistics.

Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 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, Austin, Texas. Association for Computational Linguistics.

Hu Xu, Bing Liu, Lei Shu, and Philip Yu. 2019. BERT post-training for review reading comprehension and aspect-based sentiment analysis. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2324–2335, Minneapolis, Minnesota. Association for Computational Linguistics.

Kuanhong Xu, Hui Zhao, and Tianwen Liu. 2020. Aspect-specific heterogeneous graph convolutional network for aspect-based sentiment classification. *IEEE Access*, 8:139346–139355.

Lvxiaowei Xu, Jianwang Wu, Jiawei Peng, Zhilin Gong, Ming Cai, and Tianxiang Wang. 2023. Enhancing language representation with constructional information for natural language understanding. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4685–4705, Toronto, Canada. Association for Computational Linguistics.

Xiaojun Xue, Chunxia Zhang, Zhendong Niu, and Xindong Wu. 2023a. Multi-level attention map network for multimodal sentiment analysis. *IEEE Trans. Knowl. Data Eng.*, 35(5):5105–5118.

Xiaojun Xue, Chunxia Zhang, Tianxiang Xu, and Zhendong Niu. 2023b. Constrained tuple extraction with interaction-aware network. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11430–11444, Toronto, Canada. Association for Computational Linguistics.

Jiaxuan You, Rex Ying, and Jure Leskovec. 2019. Position-aware graph neural networks. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 7134–7143. PMLR.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4568–4578, Hong Kong, China. Association for Computational Linguistics.

Mao Zhang, Yongxin Zhu, Zhen Liu, Zhimin Bao, Yunfei Wu, Xing Sun, and Linli Xu. 2023. Span-level aspect-based sentiment analysis via table filling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9273–9284, Toronto, Canada. Association for Computational Linguistics.

Zheng Zhang, Zili Zhou, and Yanna Wang. 2022. SSEGCN: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4916–4925, Seattle, United States. Association for Computational Linguistics.

Hongyu Zhao, Jiazhi Xie, and Hongbin Wang. 2022. Graph convolutional network based on multi-head pooling for short text classification. *IEEE Access*, 10:11947–11956.

Xusheng Zhao, Hao Peng, Qiong Dai, Xu Bai, Huailiang Peng, Yanbing Liu, Qinglang Guo, and Philip S. Yu. 2023. RDGCN: reinforced dependency graph convolutional network for aspect-based sentiment analysis. *CoRR*, abs/2311.04467.

Qihuang Zhong, Liang Ding, Juhua Liu, Bo Du, Hua Jin, and Dacheng Tao. 2023. Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis. *IEEE Trans. Knowl. Data Eng.*, 35(10):10098–10111.

Jie Zhou, Jimmy Xiangji Huang, Qinmin Vivian Hu, and Liang He. 2020. SK-GCN: modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification. *Knowl. Based Syst.*, 205:106292.