

# FaiMA: Feature-aware In-context Learning for Multi-domain Aspect-based Sentiment Analysis

Songhua Yang<sup>1</sup>, Xinke Jiang<sup>2\*</sup>, Hanjie Zhao<sup>1</sup>  
Wenxuan Zeng<sup>2</sup>, Hongde Liu<sup>1</sup>, Yuxiang Jia<sup>1†</sup>

<sup>1</sup> Zhengzhou University, Henan, China

<sup>2</sup> Peking University, Beijing, China

{suprit,thinkerjiang}@foxmail.com, hjzhao\_zzu@163.com

zwx.andy@stu.pku.edu.cn, lhd\_1013@gs.zzu.edu.cn, ieyxjia@zzu.edu.cn

## Abstract

Multi-domain aspect-based sentiment analysis (ABSA) seeks to capture fine-grained sentiment across diverse domains. While existing research narrowly focuses on single-domain applications constrained by methodological limitations and data scarcity, the reality is that sentiment naturally traverses multiple domains. Although large language models (LLMs) offer a promising solution for ABSA, it is difficult to integrate effectively with established techniques, including graph-based models and linguistics, because modifying their internal architecture is not easy. To alleviate this problem, we propose a novel framework, **Feature-aware In-context Learning for Multi-domain ABSA** (FaiMA). The core insight of FaiMA is to utilize in-context learning (ICL) as a feature-aware mechanism that facilitates adaptive learning in multi-domain ABSA tasks. Specifically, we employ a multi-head graph attention network as a text encoder optimized by heuristic rules for linguistic, domain, and sentiment features. Through contrastive learning, we optimize sentence representations by focusing on these diverse features. Additionally, we construct an efficient indexing mechanism, allowing FaiMA to stably retrieve highly relevant examples across multiple dimensions for any given input. To evaluate the efficacy of FaiMA, we build the first multi-domain ABSA benchmark dataset. Extensive experimental results demonstrate that FaiMA achieves significant performance improvements in multiple domains compared to baselines, increasing F1 by 2.07% on average. Source code and data sets are available at <https://github.com/SupritYoung/FaiMA>.

**Keywords:** Multi-domain Aspect-based Sentiment Analysis, Graph Neural Networks, Large Language Model, In-Context Learning, Linguistics

## 1. Introduction

In the highly interconnected digital era, a myriad of social media platforms are continually emerging (Roccabruna et al., 2022). These platforms generate a vast corpus of user reviews across various domains, providing a rich reservoir of sentiment-related information. For years, aspect-based sentiment analysis (ABSA) has emerged as a long-standing solution to this problem (Pang et al., 2008; Zhang and Liu, 2012; Schouten and Frasincar, 2016). ABSA is a fine-grained sentiment analysis task that can meticulously extract the sentiment polarity of users towards specific aspects. However, the majority existing ABSA methods are confined to single-domain applications, struggling to capture the multifaceted sentiment information prevalent in the real world. Traditional approaches often encounter generalization challenges across multiple domains, limiting the practical and broad-scale applicability of ABSA (Luo et al., 2022). Customizing models and annotating data for each domain is inefficient and costly, especially in resource-limited settings.

Fortunately, the advent of Large Language Models (LLMs) can imbue the multi-domain ABSA with renewed optimism, owing to their remarkable generalization and cross-domain capabilities (Wang et al., 2023; Zhang et al., 2023). Trained on extensive, multi-domain corpora, LLMs assimilate a broad spectrum of common sense and domain-agnostic knowledge, equipping them with the ability to discern nuanced differences and linguistic subtleties across various domains (Zhao et al., 2023; Dillion et al., 2023; Yang et al., 2023b; Luo et al., 2024a). Moreover, emerging in-context learning (ICL) techniques demonstrate that task-specific performance can be significantly amplified by simply incorporating concise, task-relevant instructions, demonstrations, and examples into the prompts (Ye et al., 2023; Jiang et al., 2023). Although initial research has begun to probe the potential of LLMs and associated techniques in ABSA (Fei et al., 2023; Scaria et al., 2023; Varia et al., 2022), empirical investigations explicitly focusing on multi-domain ABSA are still notably scarce.

Another line of ABSA research focuses on graph neural networks (GNNs) and linguistic features (Chen et al., 2022). Linguistic knowledge, epitomized by syntactic and part of speech, is widely regarded as essential for solving ABSA tasks, as they

\*Songhua Yang and Xinke Jiang contributed equally to this research.

† Yuxiang Jia is the corresponding author.

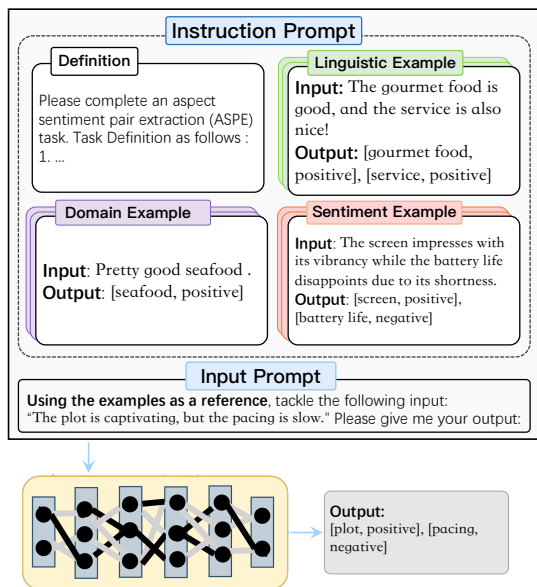


Figure 1: An example of feature-aware in-context learning for ABSA. By selecting one relevant example on each of the three features, sufficient reference is provided for LLM.

share intricate connections with the relationships between sentiment elements (Zhang et al., 2022a; Nazir et al., 2020). Numerous studies demonstrated that leveraging these linguistic features to construct relationships between words and leveraging the unique message-passing mechanism of GNNs can effectively capture complex and latent relationships among sentiment elements (Wu et al., 2021; Chen et al., 2021; Yang et al., 2023a; Shi et al., 2023; Zhong et al., 2023). Features such as domain and sentiment structure are also valuable in multi-domain ABSA (Wu et al., 2020; Gong et al., 2020). In the context of multi-domain ABSA's complex and diverse landscape, general linguistic features can provide substantive support, while specific domain information can serve as unique augmentative features. On the other hand, LLMs are often perceived as inscrutable "black box", making it challenging to directly modify their internal architecture or incorporate additional features (Luo et al., 2023; Zhao et al., 2023). Solely fine-tuning LLMs for ABSA fails to integrate the wealth of domain-specific expertise and the intrinsic relationships between parts of speech and syntax. Seamlessly integrating these well-established traditional methods with cutting-edge LLMs to fully unleash their collective potential remains a pivotal challenge in current research.

Incorporating semantically similar examples into the instructions can significantly enhance the performance of LLMs on specific tasks (Liu et al., 2022). Unlike unsupervised strategies, supervised example retrieval methods have proven to be more ef-

fective (Rubin et al., 2022; Zhang et al., 2022b). In light of this, we propose the following critical hypothesis: **ICL is not only a tool to guide the model but also an efficient feature-aware mechanism.** We further hypothesize that the stable retrieval of representative examples for various features, followed by their precise incorporation into fine-tuning instructions, can give the model a structured and enriched feature context, as shown in 1. This, in turn, substantially enhances its performance on the target task. By undergoing supervised fine-tuning (SFT) on extensive data, LLMs with strong comprehension capabilities can fully grasp, understand, and apply these features, achieving marked performance improvements in ABSA tasks.

In light of the above, we introduce a novel **Feature-aware In-context Learning for Multi-Domain ABSA (FaiMA)** framework. FaiMA ingeniously amalgamates traditional techniques with cutting-edge LLMs, using ICL as the linchpin that coherently integrates these components. Explicitly, we architect a Multi-head Graph Attention Network Encoder (MGATE) to function as the sentence encoder. Employing a multi-headed Graph Attention Network (GAT) architecture, MGATE concentrates on a panoply of linguistic, domain, and sentiment features, thereby engendering a unique sentence encoding paradigm. The essence of MGATE is its ability to wisely choose examples that are highly aligned with any given input across a variety of feature dimensions. To achieve this, the ICL technique is combined with SFT in the training stage to impart LLMs a nuanced, feature-aware understanding and learning capacity. In order to make it easier to retrieve the most relevant examples, we craft a set of heuristic rules that quantify sentence similarity across various feature dimensions. This approach generates a balanced mix of positive and negative samples for the next MGATE contrastive learning training. After training and optimizing sentence representations, the MGATE can achieve a refined understanding of the features critical for multi-domain ABSA tasks, producing high-quality sentence representations. Building on this, we select the most similar examples across features and insert them into instruction prompts during both the training and inference stages, further enhancing performance for multi-domain ABSA tasks.

In response to the lack of specialized multi-domain ABSA datasets, we also constructed a benchmark dataset named MD-ASPE, which combines 16,000 sentences across nine diverse domains. Extensive experiments show that FaiMA performs in all these domains and increases average performance by 2.07% compared to baseline models.

Our contributions can be summarized as follows:

- We introduce FaiMA, a novel framework based

on LLMs for multi-domain ABSA tasks, demonstrating that ICL can be an effective feature-aware tool.

- We propose a sentence encoding model, MGATE, which combines multi-head GAT and contrastive learning. It fully integrates linguistic, domain, and sentiment features, allowing the robust retrieval of highly relevant examples in multiple dimensions.
- We present MD-ASPE, the first benchmark dataset for multi-domain ABSA. Extensive experiments demonstrate that our method achieves state-of-the-art performance across nearly all domains and on average.

## 2. Related Work

Historically, extensive research has demonstrated the universal applicability of specific features for ABSA tasks (Zhang et al., 2022a). For example, dependency parse trees and part-of-speech tagging naturally captured relationships between words and were considered crucial linguistic features for tackling ABSA; they were closely related to underlying sentiment elements (Zhang et al., 2019; Wu et al., 2021; Chen et al., 2021; Shi et al., 2023). Furthermore, Wu et al. (2020); Chen et al. (2022) introduced a grid tagging scheme, formalizing ABSA as a task to predict the types of edge relations between words. Given that ABSA spans multiple domains, domain-specific information is often considered a crucial feature (Gong et al., 2020; jie Tian et al., 2021). Since ABSA can be considered an edge-sensitive task, GNN-based models demonstrated remarkable performance (Zhang et al., 2019; Wang et al., 2022; Zhang et al., 2022c). Significantly, the multi-head GAT model, which can flexibly focus on multiple features, achieved superior performance (Wang et al., 2020; Liang et al., 2022; Yang et al., 2023a).

Recently, LLMs like ChatGPT or LLaMA have achieved groundbreaking success (Touvron et al., 2023a,b). With the increasing scale of LLMs, novel techniques such as ICL (Ye et al., 2023) and Chain of Thought (CoT) (Wei et al., 2022) emerged. ICL demonstrates that adding detailed instructions and examples to task prompts can significantly enhance task performance, whether in zero-shot inference or supervised training. Current research has begun to investigate optimal example selection to further augment ICL’s capabilities. Studies (Liu et al., 2022; Min et al., 2022) found that choosing examples semantically and label-wise closer to the actual input is more effective. Moreover, Rubin et al. (2022); Zhang et al. (2022b) revealed that training a retriever in a supervised way to find more relevant examples is a more practical approach.

Traditional methods based on Small Language Models (SLMs) for multi-domain ABSA showed limited performance (Hu et al., 2019a; Ji et al., 2020; Luo et al., 2022). Previous approaches usually trained models for every domain, leading to computational and resource costs. Recent work has begun to explore the combination of LLMs for ABSA. For example, Varia et al. (2022) demonstrated a few-shot generalizability across various ABSA sub-tasks using SFT and multi-task learning. Scaria et al. (2023) employed ICL with fixed examples to achieve marked performance, while Fei et al. (2023) designed a multi-turn CoT for understanding implicit sentiments and opinions. These studies provide initial evidence of the substantial research potential of LLM in ABSA tasks.

## 3. Methodology

In this section, we introduce our proposed FaiMA framework, depicted in Figure 2. In §3.2, we describe the MGATE, elaborate on the heuristic rules, and contrastive learning. In §3.3, we discuss how to perform feature-aware example retrieval, along with the specific implementation of the ICL strategy.

### 3.1. Problem Definition

As a fine-grained sentiment analysis task, ABSA can be formalized as a hybrid task of extraction and classification. Given a sentence  $L = \{t_1, t_2, \dots, t_n\}$ , where  $t_i$  represents the  $i$ -th word in this sentence, and the multi-domain ABSA refers to the extraction of all sentiment pairs  $P = \{(A_1, S_1), \dots, (A_m, S_m)\}$  in  $L$  jointly for different domains.<sup>1</sup> Formally,  $A$  indicates an entity or phase in the sentence  $S$  that are related to sentiment, defined as  $A = \{a_1, a_2, \dots, a_p\} \subseteq L$  and the sentiment polarity  $S \in \{\text{positive, negative, neutral}\}$ .

### 3.2. Multi-head Graph Attention Network Encoder

For a long time, the linguistic, domain, sentiment features, and the GNN model have been crucial components for ABSA (Zhang et al., 2022a; Chen et al., 2022). In light of this, we propose MGATE, a submodel designed to investigate and understand the intricate interplay of these three complex features between words within sentences. To enhance the training process, we develop a set of sophisticated heuristic rules to generate positive and negative training sentence pairs for each feature, and then employ contrastive learning to train the graph

<sup>1</sup>This task is also referred to as aspect sentiment pair extraction (ASPE) in some literature.

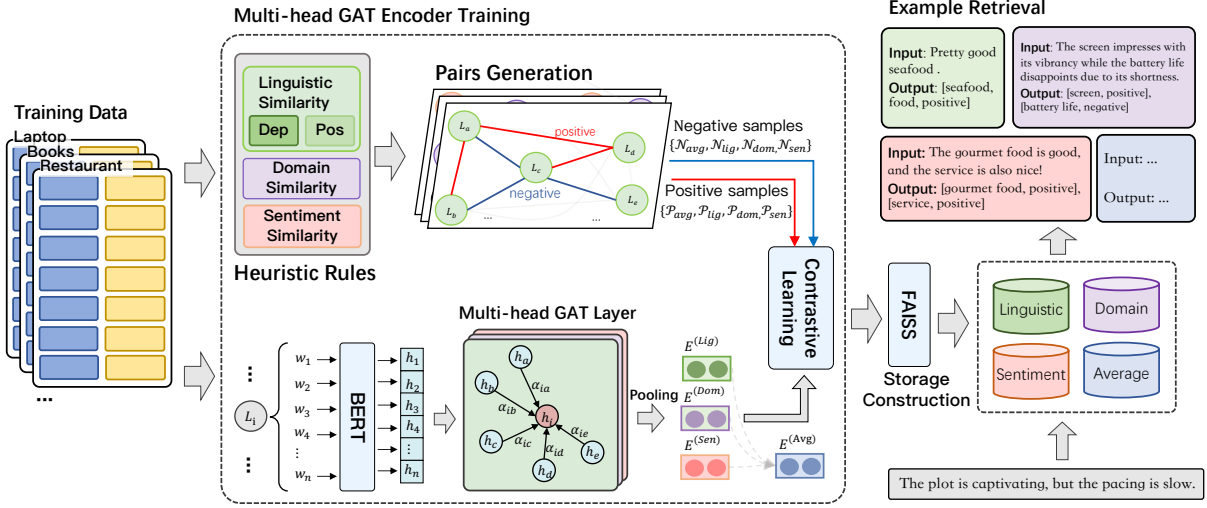


Figure 2: The overall architecture of FaiMA: MGATE training part and example retrieval part. MGATE training involves three steps: heuristic rules for positive/negative pairs generation, multi-head graph attention network to embed sentences upon three features, and contrastive learning. The diagram to the far right illustrates an ICL process that reliably fetches three domain-relevant and global average samples for any input sentence.

neural encoder, optimizing sentence representations from these three perspectives. The detailed implementation is as follows.

### 3.2.1. Feature Selection and Heuristic Rules

To accommodate the properties of the three different features, we devise unique processing rules for each. For linguistic and sentiment features, direct conversion to trainable positive and negative sample pairs presents challenges. Therefore, we design a set of heuristic algorithms to precisely calculate the similarity between two sentences given in the ABSA task.

**Linguistic Similarity** Linguistic knowledge has always been considered an essential resource to solve the ABSA task (Chen et al., 2021; Shi et al., 2023). We select the most representative part-of-speech combinations and syntactic dependency types and define refined feature modeling methods. Initially, for a sentence  $L$ , using a parser to establish part-of-speech combination matrices  $R^{pos} \in \mathbb{R}^{n \times n}$  and syntactic dependency matrices  $R^{dep} \in \mathbb{R}^{n \times n}$ , where each type of relationship corresponds to a unique numerical ID.

In the ABSA task, the aspect is always considered the key to solving this task and is closely related to other elements, so we redefine the aspects and opinions in the sentence as central words  $C = A = \{c_1, c_2, \dots, c_k\}$ . The central word for multi-token phrases is selected based on its highest number of relationships with other words. For each main word  $c_k$ , we assign weights to the other words in the sentence using a Gaussian function, ensuring that terms closer to the main word receive more

significant weight. It is defined as:

$$W(c_k) = \left[ e^{-\frac{(1-k)^2}{2\sigma^2}}, \dots, e^{-\frac{(n-k)^2}{2\sigma^2}} \right] \in \mathbb{R}^n. \quad (1)$$

The similarity calculation uses the weighted Hamming distance  $\text{HM}(\cdot)$ , which can effectively capture minor structural changes in the sentence and amplify the influence of core words nearby, considering comprehensive linguistic structures beyond direct word-to-word connections, defined as the weighted Hamming distance:

$$H(i, j) = W(c_i) \circ \text{HM}([R^{dep}(c_i), R^{pos}(c_i)], [R^{dep}(c_j), R^{pos}(c_j)]), \quad (2)$$

where  $\circ$  denotes the dot product operation, and  $c_i \in C_1$  and  $c_j \in C_2$  are the central word sets of two sentences  $L_1$  and  $L_2$ , respectively. The overall similarity distance is calculated as:

$$D(L_1, L_2) = \frac{1}{|C_1||C_2|} \sum_{i=1}^{|C_1|} \sum_{j=i}^{|C_2|} H(i, j). \quad (3)$$

Finally, the linguistic similarity score between the two sentences is obtained through the  $\text{Sigmoid}(\cdot)$  function:

$$S_{Lig}(L_1, L_2) = \text{Mean}(\text{Sigmoid}(D(L_1, L_2))) \quad (4)$$

where  $\text{Mean}$  refers to the averaging operation,  $D(L_1, L_2) \in \mathbb{R}^{\min(n, m)^2}$  and the final output is a scalar. This strategy combines part-of-speech, syntactic dependencies, and core word concepts, providing an effective quantitative measure of sentence linguistic similarity for the ABSA task.

**Domain Similarity** In Multi-domain ABSA, texts from different domains may possess entirely different features and styles, while texts from the same domain share similar background knowledge and emotional objects. Therefore, taking into account domain similarity becomes a critical factor. We define a simple binary metric to measure this. Given two sentences  $L_1$  and  $L_2$  that belong to domains  $D_1$  and  $D_2$  respectively, the domain similarity  $S_{\text{Dom}} = \mathbb{1}_{D_1=D_2}$ , where  $\mathbb{1}(\cdot)$  is the indicator function, taking the value of 1 if the condition is met and 0 otherwise.

**Sentiment Similarity** Sentiment similarity in ABSA is not directly measurable. Review text often contains different sentiment polarities across multiple aspects, especially in long or complex sentences. To capture these nuanced variations, we introduce a sentiment vector representation. For each sentence  $L$ , we define a sentiment vector  $\mathbf{v} = [n_{\text{pos}}, n_{\text{neu}}, n_{\text{neg}}]$ , where  $n_{\text{pos}}, n_{\text{neu}}, n_{\text{neg}}$  represent the count of positive, neutral, and negative sentiments in the text, respectively. For two sentences  $L_1$  and  $L_2$  and their corresponding sentiment vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , their sentiment similarity is calculated as follows:

$$S_{\text{sen}}(L_1, L_2) = \frac{1}{2} \cdot \frac{\mathbf{v}_1 \circ \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|} + \frac{1}{2}. \quad (5)$$

Here,  $\circ$  denotes the dot product between the two vectors, and  $\|\mathbf{v}\|$  represents the Euclidean norm of the vector.

Through the aforementioned method, we obtain a quantified inter-sentence similarity measure as follows:

$$S(L_1, L_2) = [S_{\text{Lig}}(L_1, L_2), S_{\text{Dom}}(L_1, L_2), S_{\text{Sen}}(L_1, L_2)] \quad (6)$$

which integrates the three feature dimensions of linguistics, domain, and sentiment. By further setting three thresholds  $\theta_{\text{Lig}}, \theta_{\text{Dom}}, \theta_{\text{Sen}}$ , these continuous similarity values can be mapped into a three-dimensional 0-1 tensor  $T$ .  $T_{ijk}$  represents the value of sentence  $i$  and  $j$  in dimension  $k$ . The tensor  $T = [\mathbb{1}_{S_{\text{Lig}} \geq \theta_{\text{Lig}}}, \mathbb{1}_{S_{\text{Dom}} \geq \theta_{\text{Dom}}}, \mathbb{1}_{S_{\text{Sen}} \geq \theta_{\text{Sen}}}]$  serves as a multi-dimensional matrix of positive and negative samples, and will be subsequently used for contrastive learning to optimize the training of the graph encoder.

### 3.2.2. Multi-head Graph Attention Network

The multi-head GAT is designed to discern intricate interrelations among linguistic, domain, and sentiment features at the token level. To this end, we deploy three distinct sub-linear layers that serve as encoders for token adjacency matrices, subsequently leveraging multi-head graph attention networks (Zhang et al., 2024; Luo et al., 2024b) for aggregating features. The aggregated features

are globally pooled to produce a graph-level (or sentence-level) representation.

Given a sentence  $L = \{w_1, w_2, \dots, w_n\}$ , we apply a pre-trained BERT model to encode word tokens:

$$(h_1, h_2, \dots, h_n) = \text{BERT}(w_1, w_2, \dots, w_n), \quad (7)$$

we denote sentence’s token vectors  $H = (w_1, w_2, \dots, w_n)$ . Then an Adaptive Adjacency Matrix is employed as the feature propagation matrix for these token vectors. For instance, the Adaptive Adjacency Matrix corresponding to linguistic features denoted as  $A^{(\text{Lig})}$ , is computed through:

$$A^{(\text{Lig})} = \text{Sigmoid}(HW^{(\text{Lig})}H^T). \quad (8)$$

and  $W^{(\text{Lig})}$  is learnable weight to linguistic feature. We also compute  $A^{(\text{Dom})}, A^{(\text{Sen})}$  by learnable weights  $W^{(\text{Dom})}, W^{(\text{Sen})}$  for domain and sentiment features, respectively.

The attention coefficients between the  $i^{\text{th}}$  and  $j^{\text{th}}$  tokens are then calculated as:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\vec{a}^T [W_a h_i \| W_a h_j]))}{\sum_{A_{ik}^{(\text{Lig})} > \delta} \exp(\text{LeakyReLU}(\vec{a}^T [W_a h_i \| W_a h_k]))} \quad (9)$$

where  $\delta$  serves as a threshold to filter out noise in the adjacency matrix.  $\vec{a}, W_a$  are learnable weights.

Then the token-level representation for the  $i^{\text{th}}$  token  $E_i$ , is computed via attention mechanism:

$$E_i = \text{LayerNorm}\left(\sum_{A_{ij}^{(\text{Lig})} > \delta} \alpha_{ij} A_{ij}^{(\text{Lig})} W_a h_j\right). \quad (10)$$

To synthesize the graph-level representation, an average pooling operation is applied across all token-level features:

$$E^{(\text{Lig})} = \text{AVG}(E_1, \dots, E_n). \quad (11)$$

Lastly, we compute the comprehensive graph-level representations for linguistic, domain-specific and sentiment features—denoted  $E^{(\text{Lig})}, E^{(\text{Dom})}, E^{(\text{Sen})}$ , and their average  $E^{(\text{Avg})}$  serves as the global feature representation.

When calculating the multi-head attention mechanism, attention is paid to the knowledge of the “Lig”, “Dom”, and “Sen” three levels. Unlike the traditional multi-head attention method that applies multi-head attention for “Lig”, “Dom”, and “Sen” respectively, we treat the “Lig”, “Dom”, and “Sen” as three heads respectively, as FaiMA mainly focuses on the knowledge of these three levels.

### 3.2.3. Contrastive Learning

Next, we will introduce the graph-level contrastive learning loss (Li et al., 2022), which aims at optimizing the representation of the three aspects after the multi-head graph attention network.

Table 1: The statistics of train and test datasets. #S, #P represent the number of sentences and sentiment pairs in the dataset, and #Pos, #Neg, #Neu refer to the amount of corresponding sentiment polarity.

Dataset	Train set					Test set				
	#S	#P	#Pos	#Neg	#Neu	#S	#P	#Pos	#Neg	#Neu
laptop	1148	1384	745	518	121	339	418	279	93	46
restaurant	1500	2125	1525	452	148	496	726	555	128	43
twitter	1500	1500	353	390	757	500	500	134	112	254
books	1411	1780	1282	445	53	421	538	394	127	17
clothing	1303	1567	1158	381	28	318	369	274	88	7
device	948	1405	905	500	0	482	696	480	216	0
finance	1500	2139	675	608	856	500	593	291	220	82
hotel	1468	1963	1856	100	7	500	678	636	42	0
service	1432	1842	1032	698	112	500	618	350	229	39
<b>Overall</b>	<b>12210</b>	<b>15705</b>	<b>9531</b>	<b>4092</b>	<b>2082</b>	<b>4056</b>	<b>5136</b>	<b>3393</b>	<b>1255</b>	<b>488</b>

In Section 3.2.1, we obtain positive and negative sample pairs from linguistic, domain, and sentiment feature perspectives through heuristic rules. Similarly, taking the linguistic perspective as an example, for any sentence  $L_i$ , we define its positive sample sentence set as  $\mathcal{P}_i$ , and its negative sample sentence imposed as  $\mathcal{N}_i$ . We take the global representation (sentence representation) obtained in Section 3.2.2 as input to maximize the similarity between positive sample pairs and minimize the similarity between negative sample pairs. To this end, we define the contrastive learning formula as follows:

$$\mathcal{L}_{CL}^{(Lig)} = \frac{1}{B} \sum_{L_i} \left[ -\log \frac{\sum_{L_j \in \mathcal{P}_i} \exp(\Gamma(E_i, E_j)/\tau)}{\sum_{L_k \in \mathcal{N}_i} \exp(\Gamma(E_i, E_k)/\tau)} \right] \quad (12)$$

where  $(L_i, L_j), L_j \in \mathcal{P}_i$  is the positive pair and  $(L_i, L_k), L_k \in \mathcal{N}_i$  is the negative pair for sentence  $L_i$ . Moreover, we define the critic function as:  $\Gamma(u, v) = \cos(\text{Linear}(u), \text{Linear}(v))$ .  $\text{Linear}(\cdot)$  represents the projection function implemented with a two-layer perceptron model.  $\cos(\cdot)$  means cosine similarity, and we first normalize the embedding and then calculate point multiplication instead.  $\tau \in (0, 1)$  is the annealing coefficient to avoid smoothing the exp function around 0 to speed up model convergence.

For linguistic features, domain features, sentiment features and global features, we calculate their respective contrastive learning loss functions  $\mathcal{L}_{CL}^{(Lig)}, \mathcal{L}_{CL}^{(Dom)}, \mathcal{L}_{CL}^{(Sen)}, \mathcal{L}_{CL}^{(Avg)}$  and sum them up as the model’s contrastive learning loss:

$$\mathcal{L}_{CL} = \beta_1 \mathcal{L}_{CL}^{(Lig)} + \beta_2 \mathcal{L}_{CL}^{(Dom)} + \beta_3 \mathcal{L}_{CL}^{(Sen)} + \mathcal{L}_{CL}^{(Avg)}, \quad (13)$$

where  $\beta_1, \beta_2, \beta_3$  are the reweighting coefficients. By continuously optimizing  $\mathcal{L}_{CL}$ , we can achieve the goal of optimizing the model.

### 3.3. Feature-aware Example Retrieval

After completing the MGATE training, the last phase of FaiMA involves example retrieval across different feature dimensions. Through the ICL method, the LLM can fully perceive the influence of different feature dimensions in the ABSA task on the output, thereby making more accurate predictions. Given an input sentence, the trained graph encoder  $\mathcal{G}$  can yield three feature vectors  $h_{Lig}, h_{Dom}, h_{Sen}$  and a pooled average vector  $h_{Avg}$ . We only use the training set as the retrieval library and adopt the efficient FAISS (Facebook AI Similarity Search) algorithm<sup>2</sup> (Johnson et al., 2019) for approximate nearest neighbor search:

$$\mathcal{N}_k(h) = \underset{x_1, \dots, x_k \in \mathcal{S}}{\text{argmin}} \|h - x_i\|_2^2 \quad (14)$$

Here,  $k$  is the number of instances to be retrieved. For each feature dimension, we retrieve at least one nearest neighbour instance (i.e.,  $k \geq 3$ ), and the retrieved instances are strictly de-duplicated.

For multi-domain ABSA tasks, we carefully design multiple different templates. The retrieved examples will be directly inserted into these templates to fine-tune the LLM in an ICL manner further.

## 4. Experiments

### 4.1. Dataset

To bridge the absence of multi-domain benchmark datasets, we combine nine high-quality datasets from various domains into a comprehensive dataset, named MD-ASPE, including 14Restaurant (Pontiki et al., 2014), 14Laptop (Pontiki et al., 2014), Device (Hu and Liu, 2004), Service (Toprak et al., 2010), Books, Clothing, Hotel (Luo et al., 2022), Twitter (Dong et al., 2014), and Financial News Headlines (Sinha et al., 2022). MD-ASPE

<sup>2</sup>[github.com/facebookresearch/faiss](https://github.com/facebookresearch/faiss)

Table 2: Performances over five different runs with Macro-F1 score (%) metric. The best performance is in bold and the second best results are underlined.

	laptop	restaurant	twitter	books	clothing	device	finance	hotel	service	Overall
BERT-CRF	55.32	68.15	57.85	42.15	61.41	55.78	55.37	61.44	54.77	58.03
SpanABSA-joint	59.12	72.65	61.05	45.55	65.61	59.38	59.47	65.74	58.47	61.89
BART-Index	65.57	76.71	66.89	60.93	72.96	67.78	69.29	80.21	67.52	69.57
T5-Index	68.05	79.44	67.85	64.12	<u>78.59</u>	<u>71.13</u>	75.81	82.05	70.96	73.18
T5-Paraphrase	68.29	<u>80.77</u>	<b>69.52</b>	64.25	<b>79.13</b>	70.82	75.77	82.17	<u>71.47</u>	<u>73.55</u>
LLaMA-SFT	68.50	76.26	62.80	60.29	73.71	<b>71.21</b>	73.74	<u>82.87</u>	67.13	70.54
LLaMA-Random	<u>69.54</u>	78.23	63.60	64.07	74.07	66.82	<u>76.56</u>	82.45	71.29	72.40
LLaMA-SBERT	68.41	76.80	60.90	62.32	73.64	67.58	74.29	81.04	68.89	70.99
LLaMA-FaiMA	<b>70.58</b>	<b>81.39</b>	<u>68.00</u>	<b>66.33</b>	77.08	70.85	<b>77.60</b>	<b>83.45</b>	<b>72.24</b>	<b>75.62</b>

incorporates annotations from diverse teams and draws from rich data sources, effectively mimicking real-world multi-domain scenarios. We ensure data balance by employing random sampling strategies and standardizing the selected data by rectifying punctuation errors and addressing whitespace inconsistencies. Train and Test datasets statistics are summarized in Table 1.

## 4.2. Baselines

To rigorously and comprehensively evaluate our proposed approach, we chose a range of baseline models, from SLMs and Generative SLMs to LLMs. **1)** Within the SLM category, we employ two cross-domain competent models based on the BERT framework (Devlin et al., 2018): SpanABSA-joint (a span-level focused model) (Hu et al., 2019b) and BERT-CRF (a BERT-based model augmented with a CRF layer). **2)** Generative SLMs include BART-Index based on BART (Raffel et al., 2020), its T5 variant T5-Index, and the variance model T5-Paraphrase (labels transduced into sequences using text templates) (Zhang et al., 2021). **3)** We also incorporate three LLM-based methods. The first is to conduct SFT directly based on LLaMA while keeping the instruction unchanged and only removing examples (LLaMA-SFT). Other ICL methods involved randomly selecting an equal number  $k$  of examples (LLaMA-Random), and utilizing SentenceBERT<sup>3</sup> as sentence encoder to index and retrieve the most similar instances using the Euclidean algorithm (LLaMA-SBERT).

## 4.3. Experimental Settings

Our FaiMA comprises multiple stages, including the generation of pairs using heuristic rules, MGATE training, and SFT with ICL. For the heuristic rules, we set  $\theta_{Lig} = 0.43$ ,  $\theta_{Dom} = 0.5$  and  $\theta_{Sen} = 0.8$  to differentiate feature similarity. Within the

MGATE training phase, we employ the BERT-base-uncased<sup>4</sup> as a token encoder with an initial learning rate of  $2 \times 10^{-4}$  running for 10 epochs. In the ICL and SFT stage, we insert the 5 ( $k = 5$ ) most relevant examples into the prompt ordered by similarity (Liu et al., 2022), including 2 average, 1 linguistic, 1 domain and 1 sentiment samples, respectively. We use LLaMA2-7b<sup>5</sup> as the backbone model and leverage low-rank adaptation (LoRA) (Hu et al., 2021) for efficient parameter tuning, coupled with gradient accumulation and mixed-precision training. The learning rate and epochs are set to  $8 \times 10^{-5}$  and 7, respectively. All methods use AdamW optimizer (Loshchilov and Hutter, 2017) with gradient decay, dynamic learning rate, and gradient clipping technique. The batch size  $B$  is set to 128, and  $\tau = 0.1$ ,  $\delta = 0.2$ ,  $\beta_1 = \beta_2 = \beta_3 = 1$ . All experiments are conducted on an Ubuntu 18.04.5 LTS server with an A800-80G GPU. We randomly divide 10% of the validation set from the training set, select the best-performing model on it, and employ the Macro-F1 value as the principal evaluation metric. We repeat experiments with different random seeds five times and report the average results.

## 4.4. Main Results

Table 2 shows the main experimental results. Our proposed LLaMA-FaiMA outperforms all baseline models in most domains, demonstrating an average performance gain of 3.22% over the best previous method. This underscores the efficacy of the Feature-Aware ICL strategy in multi-domain ABSA scenarios. Among various SLMs, generative models such as BART, T5, and LLaMA evidently outclass BERT-based models. This superiority may stem from the generative models' more efficient pretraining methodology, which enables them to undergo large-scale unsupervised training on massive corpora, thereby acquiring richer and more diverse domain knowledge. Intriguingly,

<sup>3</sup>[huggingface.co/sentence-transformers/all-MiniLM-L6-v2](https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2)

<sup>4</sup>[huggingface.co/bert-base-uncased](https://huggingface.co/bert-base-uncased)

<sup>5</sup>[huggingface.co/meta-llama/Llama-2-7b](https://huggingface.co/meta-llama/Llama-2-7b)

Table 3: Macro-F1 score of ablation experiment results on different datasets. Values in green indicate the drop in performance after removing a feature.

Model	laptop	restaurant	twitter	books	clothing	avg.
All	70.58	81.39	68.00	66.33	77.08	73.94
w/o Lig.	(-1.35)	(-2.10)	(-0.95)	(-1.80)	(-1.22)	(-1.75)
w/o Dom.	(-1.78)	(-1.87)	(-1.52)	(-1.09)	(-1.67)	(-1.68)
w/o Sen.	(-0.37)	(-0.92)	(-0.68)	(-0.53)	(-0.86)	(-0.71)

although the other LLaMA-based methods (LLaMA-SFT, LLaMA-Random, and LLaMA-SBERT) have larger model sizes than T5, their performance is somewhat lacking. We speculate that this could be due to the excessive size of LLM models, resulting in difficulties in learning transfer and adaptability. That efficient parameter fine-tuning alone may not be sufficient for optimal training (Wang et al., 2023). Despite employing a more advanced sentence encoder, for example, retrieval, LLaMA-SBERT experiences a decline in performance, indicating that conventional sentence encoding models struggle to adapt to the complexities of multi-domain ABSA tasks. In contrast, FaiMA provides stable examples from similar tasks, allowing the model to grasp the essence of the task at hand more rapidly. This demonstrates the effectiveness of our proposed approach and provides a robust new framework for the multi-domain ABSA task.

#### 4.5. Ablation Study

To investigate the impact of different features on the performance of different domains, we sequentially remove three features (linguistic, domain, and sentiment) and then report the changes in the Macro-F1 score in the top five domains to validate the efficacy of the three features. The overall results are demonstrated in Table 3. Taking the results of the average drop, linguistic features have the most significant reduction to  $-1.75$  in performance, followed by domain at  $-1.68$  and sentiment feature at  $-0.71$ , substantiating the crucial role of linguistic features in ABSA tasks. Additionally, in some domains, such as Twitter, due to its unique characteristics, the impact of domain features is especially notable compared to linguistic features. In contrast, linguistic characteristics have the most significant impact in the restaurant and book domains. Text data in linguistic domains are generally more structured, making them more susceptible to the influence of linguistic features. Meanwhile, the clothing and restaurant domains show a more pronounced dependency on sentiment features due to the high diversity in aspects and sentiments. The variation in the impact of linguistic features across domains is a reflection of unique language usage and contextual factors inherent to each domain. Typically, the lack of any variation leads to a decrease in per-

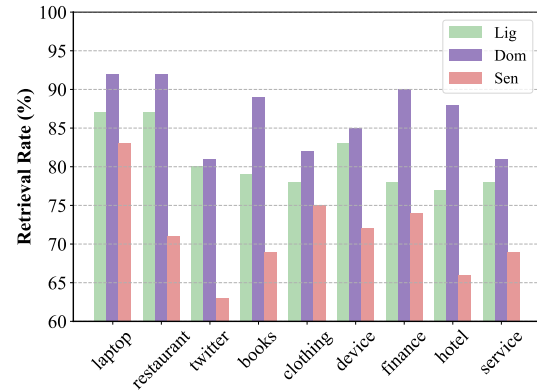


Figure 3: The retrieval success rate of the three relevant feature examples retrieved for each domain.

formance when compared to the complete model.

#### 4.6. Effectiveness Analysis of MGATE

To validate the effect of MGATE (cf. Section 3.2) for Feature-aware ICL components, we employ gpt-3.5-turbo<sup>6</sup> as an adjudicator to determine whether the examples retrieved by MGATE are similar in the validation set<sup>7</sup>. The retrieval rates for three features are illustrated in Figure 3 in various domains, indicating that all three features achieve a relatively high success rate (over 50%), proving the effectiveness of MGATE for multi-domain ABSA sentence encoding. Domain features exhibit the most explicit retrieval rate. Sentiment features are inferior to Linguistic features, and we attribute that Sentiment features are more multi-component and complex, leading to relatively low retrieval rates.

#### 4.7. Case Study

To provide an insightful understanding of the efficacy of MGATE, we conduct case studies to detail retrieval and predictive results. 1) For the correctly predicted Case #1, we observe that all three examples show a high similarity to the input sentence in the corresponding feature dimensions. They share very similar syntactic structures linguistically, while

<sup>6</sup>[openai/api/openai/chat-completion](https://openai.com/api/openai/chat-completion)

<sup>7</sup>Through testing, we found GPT can achieve human-like judgment due to excellent understanding ability.



Table 4: Case study reports two representative samples, including the retrieved most relevant examples on three features using MGATE and the prediction of FaiMA.

	Case #1	Case #2
<b>Sample</b>	Input: The food was great - sushi was good, but the cooked food amazed us. Output: [food, <b>positive</b> ], [sushi, <b>positive</b> ], [cooked food, <b>positive</b> ] Predict: [food, <b>positive</b> ], [sushi, <b>positive</b> ], [cooked food, <b>positive</b> ]	Well, it happened because of a graceless manager and a rude bartender who had we waiting 20 minutes for drinks, and then tells us to chill out. Output: [manager, <b>negative</b> ], [bartender, <b>negative</b> ], [drinks, <b>neutral</b> ] Predict: [manager, <b>negative</b> ], [bartender, <b>negative</b> ]
<b>Lig.</b>	The service was excellent, the food was excellent, but the entire experience was very cool. Output: [service, <b>positive</b> ], [food, <b>positive</b> ], [experience, <b>positive</b> ]	The whole setup is truly unprofessional and I wish Cafe Noir would get some good stuff, because despite the current one this is a great place. Output: [staff, <b>negative</b> ]
<b>Dom.</b>	The food was very expensive (we spent \$160 for lunch for two) but extremely tasty. Output: [food, <b>positive</b> ]	One would think we'd get an apology or complimentary drinks - instead, we got a snobby waiter who wouldn't even take our order for 15 minutes and gave us a lip when we asked him to do so. Output: [waiter, <b>positive</b> ]
<b>Sen.</b>	The spicy tuna roll was unusually good and the rock shrimp tempura was awesome, great appetizer to share! Output: [spicy tuna roll, <b>positive</b> ], [rock shrimp tempura, <b>positive</b> ], [appetizer, <b>positive</b> ]	We actually gave 10% tip (which we have never done despite mediocre food and service), because we felt totally ripped off. Output: [food, <b>neutral</b> ]

also being from the same domain and possessing consistent sentiment polarity and quantity. These well-matched examples enable the model to fully apprehend each feature's role and effectiveness. 2) On the other hand, for Case #2, the sentence composition and sentiment are somewhat complex, and only the domain feature successfully matching. The linguistic examples focus on partial similarity ("bartender," "manager," and "staff"), while the sentiment examples, possibly due to limited sample size, only offer support for the "neutral" label. Despite these limitations, the model still delivers accurate predictions, only overlooking the less frequent "neutral" label.

## 5. Conclusion and Future Direction

In this paper, we introduce FaiMA, a novel framework tailored to address the challenges of multi-domain Aspect-Based Sentiment Analysis (ABSA). The core insight of FaiMA is to utilize in-context learning (ICL) as a feature-aware tool in LLM. Moreover, FaiMA leverages GNNs and proposes MGATE, which captures the intricate interplay between linguistic, domain-specific, and sentiment features. Together with contrastive learning, MGATE empowers the model to retrieve highly analogous examples for any given input. Comprehensive experiments carried out in several domains demonstrate the effectiveness of FaiMA.

In summary, our research reveals the potential

of LLMs in advancing ABSA studies, especially in multi-domain and cross-domain intricacies, providing a new insight and solution for integrating traditional GNN-based methods and LLMs, holding promise for broader sentiment analysis applications, i.e. Aspect Sentiment Triplet Extraction (ASTE) and Aspect Sentiment Quad Prediction (ASQP). Despite these successes, FaiMA, as an LLM-based model, needs higher training and deployment costs compared to previous methods. Another limitation of our model is its current focus on extracting binary sentiment elements, we plan to explore the extraction of triplet and quadruple and continue to build the appropriate dataset in future.

## 6. Ethics Statement

There are no ethics-related issues in this paper. We conduct experiments on publicly available datasets. These datasets do not share personal information and do not contain sensitive content that can be harmful to any individual or community.

## 7. Acknowledgments

The authors thank anonymous reviewers for their insightful comments. This work is mainly supported by the Key Program of the Natural Science Foundation of China (NSFC) (Grant No. U23A20316) and Key R&D Project of Hubei Province (Grant No.2021BAA029).

## 8. Bibliographical References

- Hao Chen, Zepeng Zhai, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. 2022. [Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction](#). In *Annual Meeting of the Association for Computational Linguistics*.
- Zhexue Chen, Hong Huang, Bang Liu, Xuanhua Shi, and Hai Jin. 2021. Semantic and syntactic enhanced aspect sentiment triplet extraction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1474–1483.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Danica Dillion, Niket Tandon, Yuling Gu, and Kurt Gray. 2023. Can ai language models replace human participants? *Trends in Cognitive Sciences*.
- Hao Fei, Bobo Li, Qianchu Liu, Lidong Bing, Fei Li, and Tat-Seng Chua. 2023. [Reasoning implicit sentiment with chain-of-thought prompting](#). In *Annual Meeting of the Association for Computational Linguistics*.
- Chenggong Gong, Jianfei Yu, and Rui Xia. 2020. [Unified feature and instance based domain adaptation for end-to-end aspect-based sentiment analysis](#). In *Conference on Empirical Methods in Natural Language Processing*.
- Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. [Lora: Low-rank adaptation of large language models](#).
- Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019a. Open-domain targeted sentiment analysis via span-based extraction and classification. *arXiv preprint arXiv:1906.03820*.
- Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019b. Open-domain targeted sentiment analysis via span-based extraction and classification. *arXiv preprint arXiv:1906.03820*.
- Qian Ji, Xiang Lin, Yinghua Ma, Gongshen Liu, and Shilin Wang. 2020. [A unified labeling model for open-domain aspect-based sentiment analysis](#). *2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC)*, pages 186–189.
- Xinke Jiang, Ruizhe Zhang, Yongxin Xu, Rihong Qiu, Yue Fang, Zhiyuan Wang, Jinyi Tang, Hongxin Ding, Xu Chu, Junfeng Zhao, et al. 2023. Think and retrieval: A hypothesis knowledge graph enhanced medical large language models. *arXiv preprint arXiv:2312.15883*.
- Ying jie Tian, LinRui Yang, Yunchuan Sun, and Dalian Liu. 2021. [Cross-domain end-to-end aspect-based sentiment analysis with domain-dependent embeddings](#). *Complex*, 2021:5529312:1–5529312:11.
- Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2019. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3):535–547.
- Rongfan Li, Ting Zhong, Xinke Jiang, Goce Trajcevski, Jin Wu, and Fan Zhou. 2022. Mining spatio-temporal relations via self-paced graph contrastive learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 936–944.
- 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 799–810.
- Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, and Weizhu Chen. 2022. [What makes good in-context examples for GPT-3?](#) In *Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, pages 100–114, Dublin, Ireland and Online. Association for Computational Linguistics.
- Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- Jiayuan Luo, Songhua Yang, Xiaoling Qiu, Panyu Chen, Yufei Nai, Wenxuan Zeng, Wentao Zhang, and Xinke Jiang. 2024a. Kuaiji: the first chinese accounting large language model. *arXiv preprint arXiv:2402.13866*.
- Jiayuan Luo, Wentao Zhang, Yuchen Fang, Xiaowei Gao, Dingyi Zhuang, Hao Chen, and Xinke Jiang. 2024b. Timeseries suppliers allocation risk optimization via deep black litterman model. *arXiv preprint arXiv:2401.17350*.
- Yun Luo, Hongjie Cai, Linyi Yang, Yanxia Qin, Rui Xia, and Yue Zhang. 2022. Challenges for open-domain targeted sentiment analysis. *arXiv preprint arXiv:2204.06893*.

- Ziyang Luo, Can Xu, Pu Zhao, Xiubo Geng, Chongyang Tao, Jing Ma, Qingwei Lin, and Daxin Jiang. 2023. Augmented large language models with parametric knowledge guiding. *arXiv preprint arXiv:2305.04757*.
- Sewon Min, Xinxi Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2022. Rethinking the role of demonstrations: What makes in-context learning work? *arXiv preprint arXiv:2202.12837*.
- Ambreen Nazir, Yuan Rao, Lianwei Wu, and Ling Sun. 2020. Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Transactions on Affective Computing*, 13(2):845–863.
- Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2):1–135.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.
- Gabriel Roccabruna, Steve Azzolin, and Giuseppe Riccardi. 2022. Multi-source multi-domain sentiment analysis with bert-based models. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 581–589.
- Ohad Rubín, Jonathan Herzig, and Jonathan Berant. 2022. Learning to retrieve prompts for in-context learning. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2655–2671.
- Kevin Scaria, Himanshu Gupta, Siddharth Goyal, Saurabh Arjun Sawant, Swaroop Mishra, and Chitta Baral. 2023. Instructabsa: Instruction learning for aspect based sentiment analysis. *arXiv preprint arXiv:2302.08624*.
- Kim Schouten and Flavius Frasincar. 2016. [Survey on aspect-level sentiment analysis](#). *IEEE Transactions on Knowledge and Data Engineering*, page 813–830.
- Jingli Shi, Weihua Li, Quan Bai, Yi Yang, and Jianhua Jiang. 2023. Syntax-enhanced aspect-based sentiment analysis with multi-layer attention. *Neurocomputing*, 557:126730.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023a. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.
- Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023b. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.
- Siddharth Varia, Shuai Wang, Kishaloy Halder, Robert Vacareanu, Miguel Ballesteros, Yassine Benajiba, Neha Ann John, Rishita Anubhai, Smaranda Muresan, and Dan Roth. 2022. [Instruction tuning for few-shot aspect-based sentiment analysis](#). In *Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.
- Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. [Relational graph attention network for aspect-based sentiment analysis](#). In *Annual Meeting of the Association for Computational Linguistics*.
- Yadong Wang, Chen Liu, Jinge Xie, Songhua Yang, Yuxiang Jia, and Hongying Zan. 2022. [Aspect-based sentiment analysis with dependency relation graph convolutional network](#). *2022 International Conference on Asian Language Processing (IALP)*, pages 63–68.
- Zengzhi Wang, Qiming Xie, Zixiang Ding, Yi Feng, and Rui Xia. 2023. Is chatgpt a good sentiment analyzer? a preliminary study. *arXiv preprint arXiv:2304.04339*.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837.
- Shengqiong Wu, Hao Fei, Yafeng Ren, Donghong Ji, and Jingye Li. 2021. [Learn from syntax: Improving pair-wise aspect and opinion terms extraction with rich syntactic knowledge](#). In *International Joint Conference on Artificial Intelligence*.
- Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid tagging scheme for aspect-oriented fine-grained opinion extraction. *arXiv preprint arXiv:2010.04640*.
- Songhua Yang, Tengxun Zhang, Hongfei Xu, and Yuxiang Jia. 2023a. Improving aspect sentiment triplet extraction with perturbed masking and edge-enhanced sentiment graph attention network. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

- Songhua Yang, Hanjia Zhao, Senbin Zhu, Guangyu Zhou, Hongfei Xu, Yuxiang Jia, and Hongying Zan. 2023b. Zhongjing: Enhancing the chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue. *arXiv preprint arXiv:2308.03549*.
- Seonghyeon Ye, Hyeonbin Hwang, Sohee Yang, Hyeongu Yun, Yireun Kim, and Minjoon Seo. 2023. In-context instruction learning. *arXiv preprint arXiv:2302.14691*.
- Chen Zhang, Qiuchi Li, and Dawei Song. 2019. [Syntax-aware aspect-level sentiment classification with proximity-weighted convolution network](#). *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- Lei Zhang and Bing Liu. 2012. [Sentiment analysis and opinion mining](#). In *Encyclopedia of Machine Learning and Data Mining*.
- Ruizhe Zhang, Xinke Jiang, Yuchen Fang, Jiayuan Luo, Yongxin Xu, Yichen Zhu, Xu Chu, Junfeng Zhao, and Yasha Zhao. 2024. Infinite-horizon graph filters: Leveraging power series to enhance sparse information aggregation. *arXiv preprint arXiv:2401.09943*.
- Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021. Aspect sentiment quad prediction as paraphrase generation. *arXiv preprint arXiv:2110.00796*.
- Wenxuan Zhang, Yue Deng, Bing Liu, Sinno Jialin Pan, and Lidong Bing. 2023. Sentiment analysis in the era of large language models: A reality check. *arXiv preprint arXiv:2305.15005*.
- Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2022a. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*.
- Yiming Zhang, Shi Feng, and Chenhao Tan. 2022b. Active example selection for in-context learning. *arXiv preprint arXiv:2211.04486*.
- Zheng Zhang, Zili Zhou, and Yanna Wang. 2022c. [Ssegcn: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis](#). In *North American Chapter of the Association for Computational Linguistics*.
- Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223*.
- 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 Transactions on Knowledge and Data Engineering*, page 1–14.

## 9. Language Resource References

- 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.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.
- Yun Luo, Hongjie Cai, Linyi Yang, Yanxia Qin, Rui Xia, and Yue Zhang. 2022. Challenges for open-domain targeted sentiment analysis. *arXiv preprint arXiv:2204.06893*.
- 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.
- Ankur Sinha, Satishwar Kedas, Rishu Kumar, and Pekka Malo. 2022. Sentfin 1.0: Entity-aware sentiment analysis for financial news. *Journal of the Association for Information Science and Technology*, 73(9):1314–1335.
- Cigdem Toprak, Niklas Jakob, and Iryna Gurevych. 2010. Sentence and expression level annotation of opinions in user-generated discourse. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 575–584.