DiffuSent: Towards a Unified Diffusion Framework for Aspect-Based Sentiment Analysis

Anonymous ACL submission

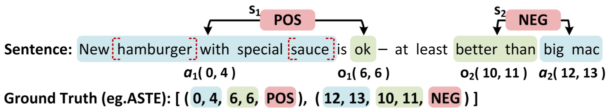
Abstract

Aspect-Based Sentiment Analysis (ABSA) encompasses seven distinct subtasks, each focusing on different extracted elements. Despite the proven success of generative models in unified aspect sentiment analysis, existing approaches often rely on autoregressive token-by-token generation without grasping the whole information of the aspect and opinion terms, resulting in boundary insensitivity, particularly in context of multi-word aspect and opinion terms. To address these issues, we present DiffuSent, a non-autoregressive diffusion framework that systematically formulates all ABSA subtasks as boundary denoising diffusion processes, progressively refining boundaries over noisy states. Furthermore, we introduce a contrastive denoising training strategy which effectively address duplicate predictions with subtle variations introduced by diffusion process. Extensive experiments on four datasets for seven subtasks demonstrate that DiffuSent achieves state-of-the-art performances.[1]

1 Introduction

Aspect-Based Sentiment Analysis (ABSA) stands as a fine-grained branch of sentiment analysis, focusing on evaluating sentiment at the entity level (Pontiki et al., 2016). ABSA comprises three key components: aspect term (a), opinion term (o), and sentiment polarity (s). To illustrate, consider the review sentence in Figure 1: "***New hamburger with special sauce is ok - at least better than big mac.***", "*New hamburger with special sauce*" and "*big mac*" are aspect terms, while "*ok*" and "*better than*" are the corresponding opinion terms linked to "positive" and "negative" sentiment polarities. These elements underlie various ABSA subtasks, each with distinct *extraction* and *classification* goals.

Conventional approaches to ABSA have focused on distinct components such as aspect/opinion term



Subtask	Input	Output	Task Type
Aspect Term Extraction (AE)	S	a_1, a_2	Extraction
Opinion Term Extraction (OE)	S	o_1, o_2	Extraction
Aspect-oriented Opinion Extraction (AOE)	S+a_1	o_1	Extraction
	S+a_2	o_2	
Aspect-Opinion Pair Extraction (AOPE)	S	(a_1,o_1),(a_2,o_2)	Extraction
Aspect-level Sentiment Classification (ALSC)	S+a_1	s_1	Classification
	S+a_2	s_2	
Aspect Extraction and Sentiment Classification (AESC)	S	(a_1,s_1),(a_2,s_2)	Extraction & Classification
Aspect Sentiment Triplet Extraction (ASTE)	S	(a_1,o_1,s_1),(a_2,o_2,s_2)	Extraction & Classification

Figure 1: Illustration of seven ABSA subtasks

extraction (Ma et al., 2019; Dai and Song, 2019; Zhao et al., 2020), sentiment classification for a given aspect (Tang et al., 2015; Liu et al., 2023), or aspect sentiment triplet extraction (Peng et al., 2020; Mukherjee et al., 2021; Zhang et al., 2022; Zhou and Qian, 2023). While these developments have led to successes in individual subtasks, a unified ABSA framework remains an elusive goal.

To bridge this gap, recent research has been shifting towards unified approaches within a pipeline framework (Mao et al., 2021; Fei et al., 2022). However, such paradigms often suffer from error accumulation due to their modular approaches (Fei et al., 2023). Addressing these drawbacks, there is a growing inclination towards employing generative models in ABSA. This shift signifies a move to an end-to-end autoregressive formulation, broadening the scope to include techniques such as word index generation (Yan et al., 2021), label augmented text generation (Zhang et al., 2021), and template filling (Gao et al., 2022; Gou et al., 2023).

However, the autoregressive decoding approach tends to concentrate on individual token during each decoding step. This method restricts the model's ability to holistically process and utilize the full range of context encapsulated within multi-word aspect/opinion terms, impacting its effective-

[1]The source code is anonymous online at: https://anonymous.4open.science/r/DiffuSent-0675/

ness in managing intricate structures and potentially leading to a lack of sensitivity in identifying term boundaries. As illustrated in Figure 1, a model fixated on token-by-token generation might inaccurately label "*hamburger*" or "*new hamburger*" as independent aspect terms, overlooking their contextual role within the broader term "*new hamburger with special sauce*". Furthermore, this autoregressive decoding process can be notably time-intensive (Fei et al., 2023; Xiao et al., 2023), particularly when generating longer target sequences.

Build upon these insights, we propose DiffuSent, a novel unified generative diffusion framework tailored for ABSA. Distinct from traditional token-by-token generation paradigm, DiffuSent is designed to explicitly model boundary indices, and dynamically refines its interpretations based on comprehensive contextual information. Through a non-autoregressive boundary denoising diffusion process, it delivers predictions for boundary indices in a single step. Specifically, we systematically infuse uncertainty via Gaussian noise into the aspect/opinion term boundaries using a forward diffusion process. The subsequent reverse diffusion process then meticulously refines these term boundaries from their initially indeterminate states. Additionally, we introduce a contrastive denoising training strategy designed to systematically differentiate between accurate and inaccurate boundary predictions. It adeptly manages the duplicate predictions with subtle variations in boundary detection, particularly in distinguishing semantically similar terms such as "*hamburger*", "*new hamburger*", and "*new hamburger with special sauce*". We validate DiffuSent on four benchmarks for seven subtasks and DiffuSent yields state-of-the-art performance. In summary, our main contributions are as follows:

- We propose DiffuSent, a novel diffusion-based framework that formulate all ABSA subtasks as boundary denoising diffusion process, offering a unified approach to ABSA. To the best of our knowledge, we are among the first to apply diffusion models in ABSA.
- A novel contrastive denoising training strategy is introduced. This strategy is designed to address duplicate predictions with subtle variations in predicted boundary indices introduced by diffusion process.
- Extensive experiments are conducted on 28 subtasks (7×4 datasets) to evaluate the effectiveness of our approach. Experimental results demonstrate that our model outperforms the state-of-the-art methods.

2 Methodology

2.1 Problem Definition

In this section, we introduce the term boundary denoising diffusion process within the context of the ASTE subtask by default, which can be extended to other subtasks with minor adjustments presented in Table 5. Given a sentence $S = \{w_1, w_2, ..., w_M\}$, the objective of ASTE is to extract the boundary indices of all conceivable aspect terms, associated opinion expression terms, and their corresponding sentiment polarity labels, denoted as $T = \{(a_i^s, a_i^e, o_i^s, o_i^e, s_i)\}_{i=1}^N$. The superscripts s and e denote the start and end indices of aspect or opinion terms within the input text. The sentiment polarity label s_i takes values from $\{\texttt{POS}, \texttt{NEU}, \texttt{NEG}\}$, and N signifies the count of target triples. We define boundary sequences as $T_b = \{(a_i^s, a_i^e, o_i^s, o_i^e)\}_{i=1}^N$ to facilitate the subsequent presentation.

2.2 Boundary Denoising Diffusion Process

As shown in Figure 2, in our boundary denoising diffusion process, the boundary sequences T_b are considered as data samples. During the forward diffusion phase, Gaussian noise is incrementally added to indices in these sequences. Conversely, the reverse diffusion process aims to meticulously restore the original boundary indices.

Boundary Indices Forward Diffusion In this phase, we progressively introduce Gaussian noise to the boundary sequences $T_b \in \mathbb{R}^{N\times4}$, simulating the uncertainty inherent in identifying term boundaries. To facilitate parallel training, we normalize the count N of T_b to N_{train} by duplicating, with normalized sequences represented as $\mathbf{x}_0 \in \mathbb{R}^{N_{train}\times4}$. The noisy sequences at any given timestep t are calculated using a one-step Markov transition as:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon \tag{1}$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ denotes the noise sampled from a standard Gaussian distribution.

Boundary Indices Reverse Diffusion Starting from a noise-perturbed state, the reverse diffusion process employs the non-Markovian denoising strategy DDIM (Song et al., 2021; Shen et al., 2023). DDIM is for precise reconstruction of term boundaries. The process involves selecting

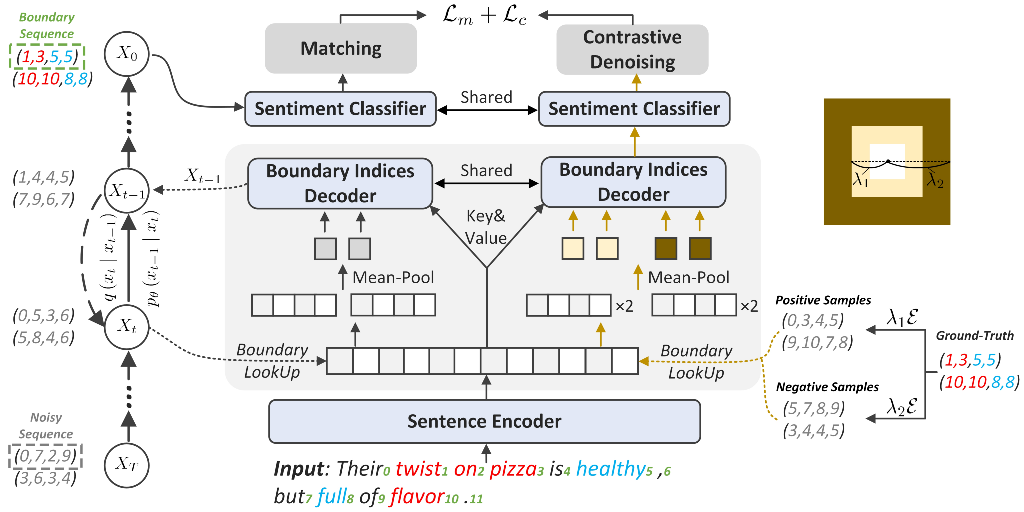


Figure 2: Overview of DiffuSent. "*Boundary LookUp*" denotes get corresponding word embedding with boundary as index. The stream identified with "↑" only occurs in the last reverse process. Noise $\mathcal{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

a subsequence τ from the full timestep sequence $[1, \ldots, T]$, with a length of γ. We iteratively refining the boundary sequences $\mathbf{x}_{\tau_i}$ using the information from the preceding timestep. The iterative refinement process, utilizing a trainable denoising network f_θ conditioned on S at τ_i, as follows:

$$\begin{aligned} \hat{\mathbf{x}}_0 &= f_\theta\left(\mathbf{x}_{\tau_i}, S, \tau_i\right) \\ \hat{\epsilon}_{\tau_i} &= \frac{\mathbf{x}_{\tau_i} - \sqrt{\alpha_{\tau_i}}\hat{\mathbf{x}}_0}{\sqrt{1-\alpha_{\tau_i}}} \end{aligned} \tag{2}$$

where $\hat{\mathbf{x}}_0$ denotes the predicted boundary at timestep τ_i, and $\hat{\epsilon}_{\tau_i}$ denotes the estimated noise. This noise is determined as the normalized difference between the perturbed boundary sequences $\mathbf{x}_{\tau_i}$ and the predicted boundary sequences $\hat{\mathbf{x}}_0$. The refined predictions are then combined with the estimated noise, adjusted by their respective standard deviations. This process is iteratively repeated, as encapsulated in the expression, $\mathbf{x}_{\tau_{i-1}} = \sqrt{\alpha_{\tau_{i-1}}}\hat{\mathbf{x}}_0 + \sqrt{1-\alpha_{\tau_{i-1}}}\hat{\epsilon}_{\tau_i}$. Following γ iterations of the DDIM, the perturbed boundary indices undergo a gradual refinement, converging towards accurate boundary indices.

2.3 Network Architecture

Within our denoising network $f_\theta\left(\mathbf{x}_t, S, t_i\right)$, it takes the perturbed boundary sequences $\mathbf{x}_t$ and the sentence S as input, and subsequently predicts the corresponding term boundary $\hat{\mathbf{x}}_0$ with corresponding sentiment polarity. The architectural design of this denoising network, as illustrated in Figure 2, is parameterized by two key components: a sentence encoder and a boundary indices decoder.

Sentence Encoder The encoder transforms the input sentence $S = \{w_1, w_2, ..., w_M\}$, with a length of M, into a h-dimensional sentence representation $\mathbf{H}_S = \{h_1, h_2, ..., h_M\} \in \mathbb{R}^{M \times h}$. Our implementation involves leveraging pre-trained language models (PLMs) with a bi-directional LSTM.

$$\mathbf{H}_S = BiLSTM(BERT(S)) \tag{3}$$

Boundary Indices Decoder The decoder is tasked with processing the sentence representation $\mathbf{H}_S$ to derive semantic representations for the corrupted sequence of boundary indices $\mathbf{x}_t$, which denote aspect and opinion terms. Initially, the noisy sequences are discretized into word indices through rescaling. Subsequently, the sequence representation $\mathbf{H}_X = \{h_i^X\}_{i=1}^{N_{train}} \in \mathbb{R}^{N_{train} \times h}$ can be computed by mean-pooling over the tokens at the designated start and end indices of aspect and opinion term. Each h_i^X represents the pooled representation of the i-th sequence within boundary sequences, calculated as follows:

$$h_i^X = Pooling(h_{a_i^s}, h_{a_i^e}, h_{o_i^s}, h_{o_i^e}) \tag{4}$$

We further utilize transformer decoder integrated a self-attention and a cross-attention layer to intricately refine sequence representations. The self-attention module fosters increased interactions among sequences by utilizing query, key, and values derived from the sequence representations $\mathbf{H}_X$:

$$\mathbf{H}_{sa} = \mathrm{SelfAttention}(\mathbf{H}_X) \tag{5}$$

where, $\mathbf{H}_{sa} \in \mathbb{R}^{N_{train} \times h}$. In tandem, the cross-attention mechanism further refines the sequence representation by incorporating the broader semantic context of the sentence. This is achieved by utilizing the output of the self-attention module $\mathbf{H}_{sa}$ as a query, with the key and value derived from the sentence representation $\mathbf{H}_S$, denoted as:

$$\mathbf{H}_{ca} = \mathrm{CrossAttention}(\mathbf{H}_{sa}, \mathbf{H}_S) \tag{6}$$

where, $\mathbf{H}_{ca} \in \mathbb{R}^{N_{train} \times h}$. To accommodate the iterative nature of the diffusion process, sinusoidal embeddings $\mathbf{E}_t$ corresponding to each timestep t are integrated into the sequence representations. The final noisy sequence representations $\overline{\mathbf{H}}_X$ are calculated as follows:

$$\overline{\mathbf{H}}_X = \mathbf{H}_{ca} + \mathbf{E}_t \tag{7}$$

Moreover, we employ 4 index pointers to predict boundary indices of aspect and opinion terms, respectively. For each index $\delta \in \{a^s, a^e, o^s, o^e\}$, we create a fused representation $\mathbf{H}^{\delta}_{SX} \in \mathbb{R}^{N_{train} \times M \times h}$, which combines the noisy sequence representation with the sentence representation. The likelihood $\mathbf{P}^{\delta} \in \mathbb{R}^{N_{train} \times M}$ of each index being a boundary of term is as follows:

$$\mathbf{H}^{\delta}_{SX} = \mathbf{W}^{\delta}_S \mathbf{H}_S + \mathbf{W}^{\delta}_X \overline{\mathbf{H}}_X \tag{8}$$

$$\mathbf{P}^{\delta} = FFN\left(\mathbf{H}^{\delta}_{SX} + \mathbf{E}^{\delta}_p\right) \tag{9}$$

where $\mathbf{W}^{\delta}_S, \mathbf{W}^{\delta}_X \in \mathbb{R}^{h \times h}$ are learnable matrices, and $FFN(\cdot)$ denotes a feed-forward network (FFN). $\mathbf{E}^{\delta}_p \in \mathbb{R}^{N_{train} \times M \times h}$ is type embedding to distinguishes between aspect or opinion boundaries.

Sentiment Classifier The sentiment classifier processes the sequence representations $\overline{\mathbf{H}}_X$ through a FFN to output a probability distribution over sentiment categories, denoted as:

$$\mathbf{P}^c = FFN\left(\overline{\mathbf{H}}_X\right) \tag{10}$$

Where, $\mathbf{P}^c \in \mathbb{R}^{N_{train} \times C}$, and C represents the total number of sentiment polarity categories.

Contrastive Denoising Training In the diffusion process of DiffuSent, a certain degree of uncertainty is introduced, leading to duplicate predictions with around the initially predicted boundary indices. It grants the model the flexibility to explore various possible interpretations of where a term might begin or end. However, it is important to note that while this added uncertainty aids in handling multi-word term, it also carries the risk of incorrect predictions of boundary indices due to subtle variations. To further enhance DiffuSent's proficiency in the nuanced delineation of term boundaries and strengthen the sentiment classification process by reducing false triplet generation, we introduce a contrastive denoising training strategy during training phase.

As shown in Figure 2, we generate two types of samples, positive samples and negative samples by adding two different scale of noise λ_1 and λ_2 to N_{train} ground-truth boundary sequences, where $\lambda_1 < \lambda_2$. After diffusion reverse process, the decoder additionally takes the two types of samples as input. Positive samples have a noise scale smaller than λ_1 and are expected to reconstruct their corresponding ground truth. Negative samples have a noise scale larger than λ_1 and smaller than λ_2. They are expected to predict "`Invalid`", denoted as ε. If a sentence has N_{train} ground-truth, contrastive denoising training will have 2 × N_{train} samples with each ground-truth generating a positive and a negative samples.

Similar to previous calculation process, we can obtain the boundary probabilities $\overline{\mathbf{P}}^{\delta}$ of positive samples, classification probabilities $\overline{\mathbf{P}}^{c}$ and $\tilde{\mathbf{P}}^c$ for positive and negative samples, respectively.

2.4 Training Loss

Our training objective consist of a matching loss and a contrastive denoising loss. We discuss each component in detail in following part.

Matching Loss In handling N_{train} predictions and corresponding N_{train} expanded ground-truth values, we leverage the Hungarian algorithm (Kuhn, 1955) to establish an optimal matching $\hat{\psi}$ between the two sets. $\hat{\psi}(i)$ represents the ground-truth corresponding to the i-th noisy sequence. The matching loss encompasses both boundary loss and sentiment classification loss. Subsequently, the reverse process is trained by maximizing the likelihood of the prediction:

$$\begin{aligned}\mathcal{L}_m = -\sum_{i=1}^{N_{train}} \Big(&\sum_{\delta \in \{a^s, a^e, o^s, o^e\}} \log \mathbf{P}^{\delta}_i\left(\hat{\psi}^{\delta}(i)\right) \\ &+ \log \mathbf{P}^c_i\left(\hat{\psi}^c(i)\right)\Big)\end{aligned} \tag{11}$$

Contrastive Denoising Loss The contrastive loss also consists of boundary loss and sentiment classification loss. Specifically, the boundary loss is only calculated according to boundary probabilities $\overline{\mathbf{P}}^{\delta}$ of positive samples. The classification loss is calculated according to classification probabilities $\overline{\mathbf{P}}^{c}$ and $\tilde{\mathbf{P}}^c$ for positive and negative samples, respectively. Consequently, the contrastive loss is computed as follows:

$$\mathcal{L}_c = -\sum_{i=1}^{N_{train}} \Big(\sum_{\delta \in \{a^s, a^e, o^s, o^e\}} \log \overline{\mathbf{P}}_i^{\delta} \left(\hat{Y}_i^{\delta}\right) + \log \overline{\mathbf{P}}_i^{c} \left(\hat{Y}_i^{c}\right) + \log \tilde{\mathbf{P}}_i^{c} \left(\varepsilon\right) \Big) \quad (12)$$

We jointly optimize matching loss $\mathcal{L}_m$ and contrastive denoising loss $\mathcal{L}_c$. The overall training loss can be represented as:

$$\mathcal{L} = \mathcal{L}_m + \mathcal{L}_c \quad (13)$$

2.5 Inference

During the inference stage, DiffuSent initiates by stochastically sampling N_{eval} noisy sequences from a Gaussian distribution. Subsequently, it undertakes iterative denoising with the learned boundary indices reverse diffusion process based on the denoising timestep τ. The predicted probabilities, derived from this denoising process, correspond to the likelihoods associated with various boundary indices and their respective sentiment polarities.

Leveraging these predicted probabilities, the model decodes N_{eval} candidate sentiment triplets $(a_i^s, a_i^e, o_i^s, o_i^e, s_i)_{i=1}^{N_{eval}}$. Following decoding, two essential post-processing steps are employed: deduplication and filtering. For triplets with identical term boundary indices, the algorithm retains the one with the highest polarity probability. Additionally, triplets with a cumulative sum of prediction probabilities falling below the threshold φ are systematically eliminated.

3 Experiments

3.1 Datasets

We evaluate our methods across seven subtasks using four datasets from SemEval Challenges. The D_{17} dataset, annotated by Wang et al. (2017), comprises unpaired opinion terms, while the D_{19} dataset, annotated by Fan et al. (2019), pairs opinion terms with corresponding aspects. Annotated by Peng et al. (2020), the D_{20a} dataset includes aspect labels, corresponding opinion labels, and sentiment polarities. Additionally, the D_{20b} dataset, refined by Xu et al. (2020), eliminates triples with inaccurate sentiments and labels missing triples. We present their statistics in Table 6.

3.2 Baselines

The baselines for evaluating DiffuSent across various datasets are categorized into three groups:

- For AE, OE, ALSC on D_{17}, and AOE on D_{19}: The models considered include: **BART-GEN** (Yan et al., 2021), **SyMux** (Fei et al., 2022), **SK2** (Li et al., 2022a), **MvP** (Gou et al., 2023).
- For AESC, AOPE, ASTE on D_{20a}: The baselines are **Peng-two-stage** (Peng et al., 2020), **Dual-MRC** (Mao et al., 2021), **BART-GEN** (Yan et al., 2021), **LEGO-ABSA** (Gao et al., 2022), **SyMux** (Fei et al., 2022), **SK2** (Li et al., 2022a), **MvP** (Gou et al., 2023).
- For ASTE on D_{20b}: The baselines are **BART-GEN** (Yan et al., 2021), **Span-ASTE** (Xu et al., 2021), **UIE** (Lu et al., 2022), **SK2** (Li et al., 2022a), **SBN** (Chen et al., 2022), **STAGE** (Liang et al., 2023), **SimSTAR** (Li et al., 2023), **SLGM** (Zhou and Qian, 2023), **MvP** (Gou et al., 2023).

3.3 Main Results

We use F1-score as the main evaluation metrics (Gao et al., 2022; Gou et al., 2023). For all ABSA subtasks, a predicted tuple is considered as correct only if all elements are the same as the gold tuple.

We evaluate our method for AESC, AOPE, and ASTE on the D_{20a} and D_{20b} datasets. The comparison results are presented in Table 1 and Table 2, respectively. Our boundary denoising diffusion approach outperforms the state-of-the-art unified baselines, demonstrating significant improvements across all three subtasks, with enhancements ranging from +0.07% to +1.8%. These findings underscore the effectiveness of DiffuSent in accurately locating term boundaries, attributed to the progressive refinement of term boundaries. Additionally, our results validate the capability of DiffuSent in recovering term boundaries from noisy sequences through the boundary denoising diffusion process.

In comparison to the latest ASTE benchmarks, as shown in Table 2, DiffuSent demonstrates superior performance. Specifically, when matched against models based on *Bert-base*, DiffuSent records an average F1-score improvement of +1.04%. In comparison to autoregressive generative models such as UIE, MvP, and SLGM, which utilize *T5-base* with twice the parameters of *Bert-base*, DiffuSent yields improvements of +0.94%, +0.67%, and +0.81% on `Res14`, `Res15`, and `Res16`, respectively. These improvements underscore DiffuSent's capability to refine interpretations dynamically with comprehension of contextual information, moving beyond token-by-token generation. Additionally, we evaluate DiffuSent on D_{17} and D_{19} for AE, OE, ALSC, and AOE, with detailed results in Appendix D.

Table 1: Comparison F1-scores(%) for AESC, AOPE and ASTE on D_{20a} dataset. The best and the second best F1-scores are in **bold** and underlined, respectively. † denotes the reproduced result using the released code. Results marked with "*" indicate a statistically significant improvement with $p < 0.01$ under the bootstrap paired t-test.

Model	PLM	Lap14			Res14			Res15			Res16		
		AESC	AOPE	ASTE	AESC	AOPE	ASTE	AESC	AOPE	ASTE	AESC	AOPE	ASTE
Peng-two-stage	-	62.34	53.85	43.50	74.19	56.10	51.89	65.79	56.23	46.79	71.73	60.04	53.62
Dual-MRC	Bert-base	64.59	63.37	55.58	76.57	74.93	70.32	65.14	64.97	57.21	70.84	75.71	67.40
SyMux	Roberta-base	70.32	67.64	60.11	78.68	79.42	74.84	69.08	69.82	63.13	77.95	78.82	72.76
SK2	Bert-large	69.42	68.12	60.14	78.72	78.19	73.32	73.30	72.05	64.32	77.78	79.89	72.03
BART-GEN	Bart-base	68.17	66.11	57.59	78.47	77.68	72.46	69.95	67.98	60.11	75.69	77.38	69.98
LEGO-ABSA	T5-base	72.3	71.3	62.2	80.6	78.1	73.7	74.2	72.9	64.4	76.1	77.6	71.5
MvP†	T5-base	70.55	71.38	62.42	78.06	77.95	74.6	74.84	74.06	65.25	77.63	80.46	73.28
DiffuSent	Bert-base	**73.74***	**71.67***	**63.31***	**81.13***	**79.86***	**74.91***	**75.85***	**74.19***	**67.05***	**79.16***	**80.9***	**74.14***

Table 2: Comparison F1-scores(%) for ASTE on D_{20b} dataset. Symbols have the same meanings as in Table 1.

Model	PLM	Lap14	Res14	Res15	Res16
Span-ASTE	Bert-base	59.38	71.85	63.27	70.26
SK2	Bert-large	60.56	73.27	65.00	72.19
SBN	Bert-base	62.65	74.34	64.82	72.08
SimSTAR†	Bert-base	59.98	70.15	63.5	70.25
STAGE†	Bert-base	59.58	72.58	63.49	71.06
BART-GEN	Bart-base	58.69	65.25	59.26	67.62
UIE-base	T5-base	62.94	72.55	64.41	72.86
MvP†	T5-base	61.51	73.48	64.65	73.38
SLGM†	T5-base	**63.28**	73.39	65.72	73.41
DiffuSent	Bert-base	63.03*	**74.42***	**66.39***	**74.22***

Table 3: Ablation results (F1-score,%) on Res15 and Res16. The best results are marked in **bold**.

Setting		Res15	Res16
Contrastive Denoising	✗	64.16	71.44
	✓	**66.39**	**74.22**
Duffusion Timestep	1000	**66.39**	**74.22**
	1500	64.42	71.4
	2000	65.57	71.22
Number of Noisy Sequence	30	63.53	72.23
	60	**66.39**	**74.22**
	90	64.61	72.26

3.4 Ablation Study

To further investigate the impact of each component and hyper-parameter in DiffuSent, we conduct comprehensive ablation studies on ASTE task on Res15 and Res16 from D_{20b} in Table 3.

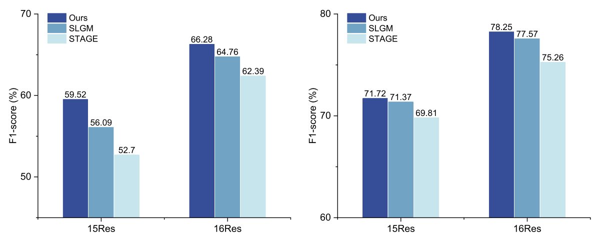
Contastive Denoising We examine the effectiveness of our contrastive denoising training by removing it from our framework. Results indicate a decrease of -2.23% and -2.78% on F1-score for Res15 and Res16, respectively. This substantial drop in performance underscores the importance of contrastive denoising training in managing duplicate predictions with subtle variations in predicted boundary indices, thereby refining predictions and ensuring valid sentiment polarity classification.

Diffusion Timestep The timestep regulates the amount of Gaussian noise introduced during the forward diffusion process. Our analysis indicates that increasing the timestep leads to a noticeable decline in model performance. This trend highlights a trade-off between noise intensity and model accuracy, underscoring the need for balancing noise levels to optimize model performance.

Number of Noisy Sequence The quantity of noisy sequences during both training and inference is indicative of the level of uncertainty. Our experiments investigate how DiffuSent performs across different numbers of noisy sequences. The findings emphasize the importance of selecting an appropriate number of noisy sequences for the model. Insufficient numbers may result in overlooking the ground truth, while an excessive amount can lead to the generation of numerous duplicate predictions with subtle variations, complicating the identification of true targets.

3.5 Performance on Multi-word Triplets

According to statistic data (Zhou and Qian, 2023), multi-word triplets account for roughly one-third of all triplets. To assess DiffuSent's capability with multi-word terms, we focus on triplets containing at least one multi-word aspect or opinion term, contrasting it with single-word triplets. Our evaluation includes comparisons with the latest span-based approach, STAGE (Liang et al., 2023), and a generative method, SLGM (Zhou and Qian, 2023), on the



(a) multi-word triplets (b) single-word triplets

Figure 3: F1-scores of DiffuSent on multi-word and single-word triplets compared with SLGM and STAGE.

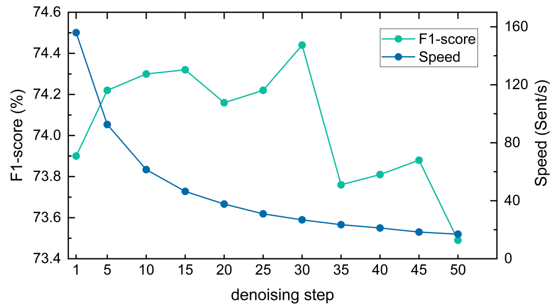


Figure 4: Analysis of denoising timestep γ on Res16

Table 4: Comparison with generative methods on Res16 from D_{20b}. P means the number of parameters. All experiments are conducted on the same setting.

Model	P	F1	Sents/s	SpeedUp
MvP	223M	73.38	0.86	1.00×
SLGM	225M	73.41	24.41	28.38×
DiffuSent$_{[\gamma=1]}$	112M	73.9	**155.98**	**181.37×**
DiffuSent$_{[\gamma=5]}$	112M	74.22	92.61	106.98×
DiffuSent$_{[\gamma=10]}$	112M	**74.3**	61.51	71.52×

Res15 and Res16 datasets from D_{20b}. As shown in Figure 3, our model consistently outperforms others across various metrics. Notably, DiffuSent exhibits a more substantial improvement, achieving an average F1-score increase of 2.48% for multi-word triplets compared to a 0.52% increase for single-word triplets. These results underscore DiffuSent's effectiveness in accurately identifying the boundaries of multi-word terms, consequently enhancing the overall performance.

3.6 Inference Efficiency

To further validate whether our DiffuSent requires more inference computations, we also conduct experiments to compare the inference efficiency between DiffuSent and other generative models: MvP (Gou et al., 2023) and SLGM (Zhou and Qian, 2023). As shown in Table 4, DiffuSent achieves better performance with a faster inference speed and minimal parameter scale. Even with a denoising timestep of $\gamma = 10$, DiffuSent is 71.5× and 2.5× faster than them via generating all triplets in parallel, which avoids generating the linearized sequence in autoregressive manner.

Furthermore, We also conduct experiments to analyze the effect of different denoising timesteps on model performance and inference speed of DiffuSent. As shown in Figure 4, with an increase of denoising steps, the model initially achieves incremental performance improvement while sacrificing inference speed. Subsequently, the model exhibited a significant degradation in performance beyond denoising timesteps $\gamma = 30$, which indicates that preserving a certain level of noise can enhance the diversity of generated triplets. Considering the trade-off between performance and efficiency, we set $\gamma = 5$ as the default setting.

3.7 Case Study

Figure 5 illustrates three distinct case studies from Res15 dataset. In the first example, SLGM wrongly predict "*Smith Street*" as aspect while DiffuSent accurately recovers term boundaries from noisy sequences through boundary denoising diffusion. In the second example with multi-word triplet, SLGM's failure to identify the broader aspect term "*stuff tilapia*" through autoregressive token-by-token generation highlights its limitation in capturing comprehensive context of multi-word term. Notably, the absence of contrastive denoising training strategy in DiffuSent leads to the erroneous prediction of an redundant triplet, highlighting the strategy's importance in mitigating duplicate predictions introduced by diffusion process. This observation is reinforced by the third example, where the lack of contrastive denoising training strategy in DiffuSent leads to the generation of a spurious triplet. Such instances validate the strategy's utility in discerning between precise and imprecise boundary delineations. We conduct additional case studies for further demonstration in Appendix F.

4 Related Work

4.1 Aspect-Based Sentiment Analysis

Aspect-Based Sentiment Analysis (ABSA) encompasses a suite of interrelated subtasks, each focusing on specific components or their combinations within a text as illustrated in Figure 1. Previous studies mainly focus on individual subtasks

Test sentence	*Worst place on Smith Street in Brooklyn.*	*The stuff tilapia was horrid ... tasted like cardboard .*	*never swaying, never a bad meal, never bad service ...*
Gold triplet	(place, Worst, negative)	(stuff tilapia, horrid, negative)	(meal, never a bad, positive) (service, never bad, positive)
SLGM	(Smith Street, Worst, negative)✗	(tilapia, horrid, negative) ✗	~~(meal, never a bad, positive)~~ (service, never bad, positive)✓
DiffuSent w/o CD	(place, Worst, negative)✓	(stuff tilapia, horrid, negative)✓ (stuff tilapia, cardboard, negative)✗	(meal, never swaying, positive)✗ (meal, never a bad, positive)✓ (service, never bad, positive)✓
DiffuSent	(place, Worst, negative)✓	(stuff tilapia, horrid, negative)✓	(meal, never a bad, positive)✓ (service, never bad, positive)✓

Figure 5: Results of case study by different models. *DiffuSent w/o CD* denotes DiffuSent without contrastive denoising. Triplets crossed out by the red line indicate missing predictions.

(Tang et al., 2016; Li and Lam, 2017; Wang et al., 2017), including AE, OE, ALSC. Subsequent research shifted towards integrated models that simultaneously extract aspects, opinions, and their corresponding sentiments (Fan et al., 2019; Gao et al., 2021; Hu et al., 2019), such as AOE, AOPE and AESC. Marking a significant shift in the field, Peng et al. (2020) introduced the Aspect Sentiment Triplet Extraction (ASTE) task, pioneering a unified approach for extracting aspect, opinion, and sentiment triplets. This approach led to the development of advanced techniques in ABSA, such as table filling (Jing et al., 2021; Zhang et al., 2022), sequence tagging (Xu et al., 2020; Li et al., 2023; Zhou and Qian, 2023), and span-based methods (Xu et al., 2021; Chen et al., 2022; Liang et al., 2023). However, these methods focus on individual tasks, rather than a comprehensive solution.

Recent trends in Aspect-Based Sentiment Analysis (ABSA) have seen the emergence of unified methods, such as Mao et al. (2021)'s two-step MRC approach. However, this method suffers from error accumulation due to isolated processing. In response, a shift towards end-to-end generative methods has occurred, addressing all ABSA subtasks more effectively. These include approaches like word index generation (Yan et al., 2021), label augmented text generation (Zhang et al., 2021), and template filling (Gao et al., 2022; Gou et al., 2023; Zhou and Qian, 2023). However, a notable limitation of these generative models is their reliance on autoregressive, token-by-token decoding. This approach, while effective, does not fully capitalize on the information available in multi-word terms and can be inefficient time-wise. In our work, we utilize a diffusion model to facilitate progressive refinements of term boundaries and output all predictions simultaneously in non-autoregressive manner, effectively addressing complex linguistic structures.

4.2 Diffusion Model

Diffusion models (Sohl-Dickstein et al., 2015), primarily used for continuous data like images and audio (Kong et al., 2020; Rombach et al., 2022; Chen et al., 2023), face challenges when applied to the discrete nature of text in NLP. Innovations by Hoogeboom et al. (2021) and Austin et al. (2021) have adapted these models for character-level text generation, while Li et al. (2022b) and Gong et al. (2022) further developed methods to bridge the gap between continuous and discrete domains. Notably, Shen et al. (2023) frame Named Entity Recognition as a boundary denoising process, offering insights into the application of diffusion models in text extraction. Building on this innovation, we have developed DiffuSent, a unified generative diffusion framework designed to address all ABSA subtasks.

5 Conclusion

In this paper, we propose DiffuSent, a novel generative framework for unified aspect-based sentiment analysis (ABSA) that formulate all ABSA subtasks as boundary denoising diffusion process. Different from autoregressive token-by-token generation, DiffuSent explicitly models boundary indices and allows for dynamically refinements in interpreting complex linguistic structures like multi-word terms. In addition, to address duplicate predictions with subtle variations arising from diffusion process uncertainties, we design a contrastive denoising training that further refine aspect and opinion term boundaries. Experimental results demonstrate that DiffuSent yields a new state-of-the-art performance, showcasing superior performance in processing complex linguistic structures efficiently.

Limitations

Despite the strong performance of DiffuSent, its design still has the following limitations. As a latent generative model, DiffuSent relies on sampling from a Gaussian distribution to produce noisy sequences, which leads to a random and uncertain characteristic of generation. Although we propose a contrastive denoising strategy to manage this phenomenon, it inevitably increases some non-negligible computational cost. Additionally, experiments only verified the consistent improvement on ABSA tasks, while intuitively, the idea of DiffuSent can be expanded to any structure prediction tasks, such as information extraction, emotion-cause pair extraction, and stance detection.

Ethics Statement

1. All of the datasets used are collected and annotated in previous studies. The use of these datasets in our work does not involve any interaction or collection of individual privacy data.

2. Our work focuses on methodology studies and experiments. The results and models in our paper will not be used to harm or deceive any individuals or groups.

3. There are no potential conflicts of interest or ethical issues regarding financial support in the sponsors and funds of our research work.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful discussion and feedback. This work was supported by (ANONYMIZED FOR DOUBLE BLIND REVIEW).

References

Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. 2021. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34:17981–17993.

Shoufa Chen, Peize Sun, Yibing Song, and Ping Luo. 2023. Diffusiondet: Diffusion model for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19830–19843.

Yuqi Chen, Chen Keming, Xian Sun, and Zequn Zhang. 2022. A span-level bidirectional network for aspect sentiment triplet extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4300–4309.

Hongliang Dai and Yangqiu Song. 2019. Neural aspect and opinion term extraction with mined rules as weak supervision. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5268–5277.

Zhifang Fan, Zhen Wu, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2019. Target-oriented opinion words extraction with target-fused neural sequence labeling. 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 2509–2518, Minneapolis, Minnesota. Association for Computational Linguistics.

Hao Fei, Fei Li, Chenliang Li, Shengqiong Wu, Jingye Li, and Donghong Ji. 2022. Inheriting the wisdom of predecessors: A multiplex cascade framework for unified aspect-based sentiment analysis. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI*, pages 4096–4103.

Hao Fei, Yafeng Ren, Yue Zhang, and Donghong Ji. 2023. Nonautoregressive encoder–decoder neural framework for end-to-end aspect-based sentiment triplet extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9):5544–5556.

Lei Gao, Yulong Wang, Tongcun Liu, Jingyu Wang, Lei Zhang, and Jianxin Liao. 2021. Question-driven span labeling model for aspect–opinion pair extraction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 12875–12883.

Tianhao Gao, Jun Fang, Hanyu Liu, Zhiyuan Liu, Chao Liu, Pengzhang Liu, Yongjun Bao, and Weipeng Yan. 2022. Lego-absa: A prompt-based task assemblable unified generative framework for multi-task aspect-based sentiment analysis. In *Proceedings of the 29th international conference on computational linguistics*, pages 7002–7012.

Shansan Gong, Mukai Li, Jiangtao Feng, Zhiyong Wu, and LingPeng Kong. 2022. Diffuseq: Sequence to sequence text generation with diffusion models. *arXiv preprint arXiv:2210.08933*.

Zhibin Gou, Qingyan Guo, and Yujiu Yang. 2023. MvP: Multi-view prompting improves aspect sentiment tuple prediction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4380–4397, Toronto, Canada. Association for Computational Linguistics.

Emiel Hoogeboom, Didrik Nielsen, Priyank Jaini, Patrick Forré, and Max Welling. 2021. Argmax flows and multinomial diffusion: Learning categorical distributions. *Advances in Neural Information Processing Systems*, 34:12454–12465.

Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. Open-domain targeted sentiment analysis via span-based extraction and classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 537–546, Florence, Italy. Association for Computational Linguistics.

Hongjiang Jing, Zuchao Li, Hai Zhao, and Shu Jiang. 2021. Seeking common but distinguishing difference, a joint aspect-based sentiment analysis model. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3910–3922, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. 2020. Diffwave: A versatile diffusion model for audio synthesis. In *International Conference on Learning Representations*.

Harold W Kuhn. 1955. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97.

Dongxu Li, Zhihao Yang, Yuquan Lan, Yunqi Zhang, Hui Zhao, and Gang Zhao. 2023. Simple approach for aspect sentiment triplet extraction using span-based segment tagging and dual extractors. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2374–2378.

Jia Li, Yuyuan Zhao, Zhi Jin, Ge Li, Tao Shen, Zhengwei Tao, and Chongyang Tao. 2022a. Sk2: Integrating implicit sentiment knowledge and explicit syntax knowledge for aspect-based sentiment analysis. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 1114–1123.

Xiang Li, John Thickstun, Ishaan Gulrajani, Percy S Liang, and Tatsunori B Hashimoto. 2022b. Diffusion-lm improves controllable text generation. *Advances in Neural Information Processing Systems*, 35:4328–4343.

Xin Li and Wai Lam. 2017. Deep multi-task learning for aspect term extraction with memory interaction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2886–2892, Copenhagen, Denmark. Association for Computational Linguistics.

Shuo Liang, Wei Wei, Xian-Ling Mao, Yuanyuan Fu, Rui Fang, and Dangyang Chen. 2023. Stage: span tagging and greedy inference scheme for aspect sentiment triplet extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13174–13182.

Xueyi Liu, Rui Hou, Yanglei Gan, Da Luo, Changlin Li, Xiaojun Shi, and Qiao Liu. 2023. Aspect-oriented opinion alignment network for aspect-based sentiment classification. In *Proceedings of the 26th European Conference on Artificial Intelligence (ECAI)*, pages 1552–1559.

Yaojie Lu, Qing Liu, Dai Dai, Xinyan Xiao, Hongyu Lin, Xianpei Han, Le Sun, and Hua Wu. 2022. Unified structure generation for universal information extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5755–5772, Dublin, Ireland. Association for Computational Linguistics.

Dehong Ma, Sujian Li, Fangzhao Wu, Xing Xie, and Houfeng Wang. 2019. Exploring sequence-to-sequence learning in aspect term extraction. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 3538–3547.

Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. 2021. A joint training dual-mrc framework for aspect based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 13543–13551.

Rajdeep Mukherjee, Tapas Nayak, Yash Butala, Sourangshu Bhattacharya, and Pawan Goyal. 2021. Paste: A tagging-free decoding framework using pointer networks for aspect sentiment triplet extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9279–9291.

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8600–8607.

Maria Pontiki, Dimitrios Galanis, Harris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, AL-Smadi Mohammad, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphee De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10674–10685.

Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. 2023. DiffusionNER: Boundary diffusion for named entity recognition. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3875–3890, Toronto, Canada. Association for Computational Linguistics.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2256–2265, Lille, France. PMLR.

Jiaming Song, Chenlin Meng, and Stefano Ermon. 2021. Denoising diffusion implicit models. In *International Conference on Learning Representations*.

Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2016. Effective lstms for target-dependent sentiment classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3298–3307.

Duyu Tang, Bing Qin, and Ting Liu. 2015. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 1422–1432.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2017. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Yisheng Xiao, Lijun Wu, Junliang Guo, Juntao Li, Min Zhang, Tao Qin, and Tie-yan Liu. 2023. A survey on non-autoregressive generation for neural machine translation and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Lu Xu, Yew Ken Chia, and Lidong Bing. 2021. Learning span-level interactions for aspect sentiment triplet extraction. 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 4755–4766.

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2339–2349, Online. Association for Computational Linguistics.

Hang Yan, Junqi Dai, Tuo Ji, Xipeng Qiu, and Zheng Zhang. 2021. A unified generative framework 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 2416–2429, Online. Association for Computational Linguistics.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2021. Towards generative 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 2: Short Papers)*, pages 504–510, Online. Association for Computational Linguistics.

Yice Zhang, Yifan Yang, Yihui Li, Bin Liang, Shiwei Chen, Yixue Dang, Min Yang, and Ruifeng Xu. 2022. Boundary-driven table-filling for aspect sentiment triplet extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 6485–6498.

He Zhao, Longtao Huang, Rong Zhang, Quan Lu, and Hui Xue. 2020. Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 3239–3248.

Shen Zhou and Tieyun Qian. 2023. On the strength of sequence labeling and generative models for aspect sentiment triplet extraction. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 12038–12050, Toronto, Canada. Association for Computational Linguistics.

A Denoising Diffusion Implicit Model

Diffusion models are a class of generative models that leverage both forward and reverse processes, which can be likened to Markov chains with Gaussian transitions. The forward process gradually adds Gaussian noise to transform sample data $\mathbf{x}_0$ to a latent noisy sample $\mathbf{x}_t$ for $t \in \{0, 1, \ldots, T\}$, which can be defined as:

$$q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_t \mid \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\,\mathbf{I}\right) \quad (14)$$

where $\bar{\alpha}_t := \prod_{s=0}^{t} \alpha_s = \prod_{s=0}^{t} (1 - \beta_s)$ and β_s represents the predefined variance schedule.

The reverse process then attempts to remove the noise that was added in the forward process and is parameterized by θ as:

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t)) \quad (15)$$

where $\mu_\theta(\cdot)$ and $\Sigma_\theta(\cdot)$ can be implemented by a U-Net or a Transformer. When conditioning also on $\mathbf{x}_0$, $q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0)$ has a closed form so we can manage to minimize the variational lower bound to optimize $\log p_\theta(\mathbf{x}_0)$:

$$\begin{aligned}\mathcal{L}_{\text{vlb}} &= \mathbb{E}_q\left[D_{\text{KL}}\left(q(\mathbf{x}_T \mid \mathbf{x}_0) \| p_\theta(\mathbf{x}_T)\right)\right] + \\ &\mathbb{E}_q\left[\sum_{t=2}^{T} D_{\text{KL}}\left(q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0) \| p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, t)\right)\right] \\ &- \log p_\theta(\mathbf{x}_0 \mid \mathbf{x}_1)\end{aligned} \quad (16)$$

where $\mathbb{E}_q(\cdot)$ denotes the expectation over the joint distribution $q(\mathbf{x}_{0:T})$.

B Optimal Matching

Given a fixed-size set of N_{train} noisy sequences, DiffuSent infers N_{train} predictions, where N_{train} is larger than the number of N ground-truth in a sentence. One of the main difficulties of training is to score the prediction with respect to the ground truth. Thus we utilize an optimal bipartite matching between predicted and ground truth and then optimize the likelihood-based loss.

Assuming $\hat{Y} = \{\hat{Y}_i\}_{i=1}^{N_{train}}$ are the set of N_{train} predictions, where $\hat{Y}_i = \left(\mathbf{P}_i^{a^s}, \mathbf{P}_i^{a^e}, \mathbf{P}_i^{o^s}, \mathbf{P}_i^{o^e}, \mathbf{P}_i^{s}\right)$. We denote the ground truth set of N tuples as $\{(a_i^s, a_i^e, o_i^s, o_i^e, s_i)\}_{i=1}^{N}$, where a_i^s, a_i^e, o_i^s, o_i^e, s_i are the aspect/opinion boundary indices and sentiment for the i-th tuple. Since N_{train} is larger than the number of N ground-truth, we pad Y with $\varnothing$ (invalid). To find a bipartite matching between these two sets we search for a permutation of N_{train} elements $\psi \in \mathfrak{S}_{N_{train}}$ with the lowest cost:

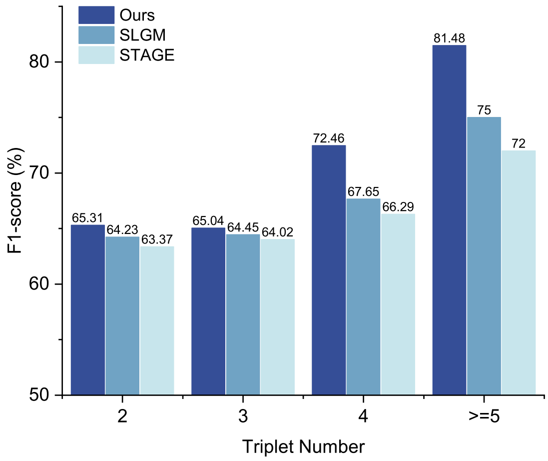


Figure 6: F1-scores of DiffuSent on multi-triplet sentence compared with SLGM and STAGE.

$$\hat{\psi} = \underset{\psi \in \mathfrak{S}_{N_{train}}}{\arg\min} \sum_{i}^{N_{train}} \mathcal{L}_{\text{match}}\left(\hat{Y}_i, Y_{\psi(i)}\right) \quad (17)$$

where $\mathcal{L}_{\text{match}}\left(\hat{Y}_i, Y_{\psi(i)}\right)$ is a pair-wise matching cost between the prediction $\hat{Y}_i$ and ground truth $Y_{\psi(i)}$ with index $\psi(i)$. With these notations we define $\mathcal{L}_{\text{match}}(\hat{Y}_i, Y_{\psi(i)})$ as $-\mathbb{1}(Y_{\psi(i)} \neq \varnothing) \sum_{\sigma \in \{a^s, a^e, o^s, o^e, s\}} \mathbf{P}_i^{\sigma}\left(Y_{\psi(i)}^{\sigma}\right)$, where $\mathbb{1}(\cdot)$ denotes an indicator function. This optimal assignment is computed efficiently with the Hungarian algorithm, following prior work (Shen et al., 2023).

C Implement Details

Our DiffuSent is trained on the NVIDIA A100 Tensor Core GPU. Following previous works (Liang et al., 2023), We employ *bert-base-uncased*[2] as the pre-trained model. We train our model using Adam optimizer with a linear warmup and linear decay learning rate schedule. The initial learning rate is $2e^{-5}$ for AE, OE, ALSC and AOE, $5e^{-5}$ for AESC, AOPE and ASTE. The filtering threshold φ is 0.6 for ALSC, 1.5 for AE, OE and AOE, 2.5 for AESC, 3.5 for AOPE, 4.5 for ASTE. We set dropout as 0.1 and batch size as 16. For diffusion process, the number of noisy sequences N_{train} and N_{eval} are set as 60, the timestep T is 1000, and the sampling timestep γ is 5. The scale factor λ_1 and λ_2 for contrastive denoising training is 1.0 and 2.0, respectively.

[2] https://huggingface.co/bert-base-uncased

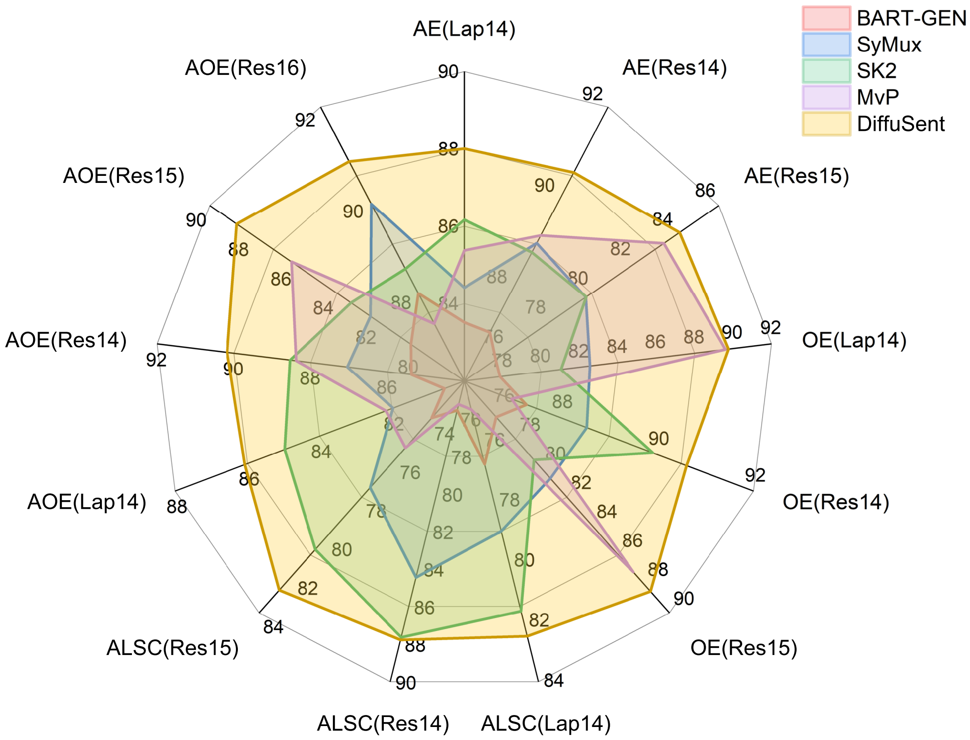


Figure 7: Comparison F1-scores for AE, OE, ALSC on the D_{17} dataset, and AOE on the D_{19} dataset. The results of **MvP** (Gou et al., 2023) are reproduced by us using the released code.

D Additional Result

We extensively evaluate the capabilities of DiffuSent model on D_{17} dataset (Wang et al., 2017) for AE, OE, ALSC and on D_{19} dataset (Fan et al., 2019) for AOE. Figure 7 summarizes the performance on key benchmarks of DiffuSent compared to other state-of-the-art unified ABSA methods. DiffuSent sets new state-of-the-art results on both extraction and classification ABSA subtasks.

E Performance on Multi-triplet

To verify the effectiveness of our framework in handling sentences with multiple triplets, we conduct a comprehensive evaluation on the ASTE task, comparing our model's performance against STAGE and SLGM. Figure 6 showcases our results derived from a meticulous analysis using the Res15 test set, which was segregated into sentences with varying numbers of multi-triplets. In the category of sentences contain two or three triplets, our model exhibited outstanding performance, achieving F1-scores of 65.31% and 65.04%, outperforming the two baseline models. The efficacy of our model becomes even more pronounced in sentences containing four or more triplets. In these instances, our model's scores significantly surpassed those of the leading baseline models. This significant lead underscores the effectiveness of our model's greater flexibility in identifying term boundaries, proving its adeptness in more challenging sentences with intricate structures.

F Additional Case Study

We present extended intances from Res15 dataset analyzed by DiffuSent with and without contrastive denosing training in Figure 8. As illustrated in the first instance, DiffuSent without contrastive denosing training strategy falls short in handling duplicate predictions with subtle variations introduced by diffusion process in boundary identification as it wrongly predicts "*bathroom*" as the aspect term. Furthermore, we observe that DiffuSent without contrastive denoising training strategy typically predicts extra incorrect triplet which does not exist in the given sentence. These cases indicate that DiffuSent is adept at distinguishing between accurate and inaccurate boundary predictions by managing the inherent uncertainty in language interpretation with the help of boundary denoising diffusion process and contrastive denosing training.

Table 5: Experiment settings on each subtask. The underlined tokens are given during inference in subtask that depend on a specific aspect term. Noise $\mathcal{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Subtask	$\mathbb{X}_0$ (Boundary Sequence)	$\mathbb{X}_T$ (Noisy State)	Sentiment Classifier	Contrastive Deniosing
AE/OE	$(a^s/o^s, a^e/o^e)$	$(\mathcal{E}_1, \mathcal{E}_2)$	✗	✔
ALSC	$(\underline{a^s}, \underline{a^e})$	$(\mathcal{E}_1, \mathcal{E}_2)$	✔	✗
AOE	$(\underline{a^s}, \underline{a^e}, o^s, o^e)$	$(a^s, a^e, \mathcal{E}_1, \mathcal{E}_2)$	✗	✔
AESC	(a^s, a^e)	$(\mathcal{E}_1, \mathcal{E}_2)$	✔	✔
AOPE	(a^s, a^e, o^s, o^e)	$(\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4)$	✗	✔
ASTE	(a^s, a^e, o^s, o^e)	$(\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4)$	✔	✔

Table 6: Statistics of the four datasets used in our experiments.

Datasets		Lap14				Res14				Res15				Res16				Tasks
		#s	#a	#o	#p	#s	#a	#o	#p	#s	#a	#o	#p	#s	#a	#o	#p	
D_{17}	train	3048	2373	2504	-	3044	3699	3484	-	1315	1199	1210	-	-	-	-	-	AE,OE,
	test	800	654	674	-	800	1134	1008	-	685	542	510	-	-	-	-	-	ALSC
D_{19}	train	1158	1634	-	-	1627	2643	-	-	754	1076	-	-	1079	1512	-	-	AOE
	test	343	482	-	-	500	865	-	-	325	436	-	-	329	457	-	-	
D_{20a}	train	920	-	-	1265	1300	-	-	2145	593	-	-	923	842	-	-	1289	AESC,
	dev	228	-	-	337	323	-	-	524	148	-	-	238	210	-	-	316	AOPE,
	test	339	-	-	490	496	-	-	862	318	-	-	455	320	-	-	465	ASTE
D_{20b}	train	906	-	-	1460	1266	-	-	2338	605	-	-	1013	857	-	-	1394	
	dev	219	-	-	346	310	-	-	577	148	-	-	249	210	-	-	339	ASTE
	test	328	-	-	543	492	-	-	994	322	-	-	485	326	-	-	514	

-Test sentence: oh speaking of bathroom , the mens bathroom was disgusting
Gold triplet: (mens bathroom, disgusting, negative)
DiffuSent: (mens bathroom, disgusting, negative) ✔
DiffuSent w/o Contrastive Denoising: (bathroom, disgusting, negative),✗ (mens bathroom, disgusting, negative)✔

-Test sentence: Paul , the maitre d ' , was totally professional and always on top of things .
Gold triplet: (Paul, professional, positive)
DiffuSent: (Paul, professional, positive) ✔
DiffuSent w/o Contrastive Denoising: (Paul, professional, positive),✔ (maitre d ', professional, positive)✗

-Test sentence: THE SERVICE IS AMAZING , i 've had different waiters and they were all nice , which is a rare thing in NYC .
Gold triplet: (SERVICE, AMAZING, positive), (waiters, nice, positive)
DiffuSent: (SERVICE, AMAZING, positive),✔ (waiters, nice, positive)✔
DiffuSent w/o Contrastive Denoising: (SERVICE, AMAZING, positive),✔ (waiters, nice, positive),✔ (waiters, rare, positive)✗

-Test sentence: Shame on this place for the horrible rude staff and non-existent customer service .
Gold triplet: (stuff, rude, negative), (customer service, non-existent, negative)
DiffuSent: (stuff, rude, negative),✔ (customer service, non-existent, negative) ✔
DiffuSent w/o Contrastive Denoising: (stuff, rude, negative),✔ (stuff, Shame on this palce for the horrible rude, negative),✗
(stuff, Shame, negative),✗ (customer service, non-existent, negative)✔

-Test sentence: Food was amazing - I love Indian food and eat it quite regularly , but I can say this is one of the best I 've had .
Gold triplet: (Food, amazing, positive)
DiffuSent: (Food, amazing, postive),✔(Indian food, best, positive)✗
DiffuSent w/o Contrastive Denoising: (Food, amazing, postive),✔ (Indian food, best, positive),✗ (Indian food, love, positive),✗
(Food, love, positive),✗ (Food was amazing - I love Indian food, love, positive)✗

Figure 8: Additional case study.