# Toward Prompt-Enhanced Sentiment Analysis with Mutual Describable Information Between Aspects

## Gaofei Xie, Ning Liu, Xiaojie Hu & Yatian Shen

Published online: 15 Mar 2023.

Submit your article to this journal

Article views: 514

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

# Toward Prompt-Enhanced Sentiment Analysis with Mutual Describable Information Between Aspects

Gaofei Xie, Ning Liu, Xiaojie Hu, and Yatian Shen

School of Computer and Information Engineering, Henan University, Kaifeng, China

**ABSTRACT**

Sentiment analysis aims to attain the sentiment polarity of the text, which is a coarse-grained approach and does not focus on the targets . On the other hand, an aspect-based sentiment analysis (ABSA) has recently gained boosting interest. The ABSA is a fine-grained sentiment assignment to determine the sentiment tendency toward a specific aspect. Most previous methods employ Recurrent Neural Network (RNN) coupled with attention mechanisms to accomplish this task. However, such RNN-based models tend to be complex and require much training. Recently, a growing number of BERT-style models have been emerging and presenting better results in ABSA tasks . Nevertheless, these methods cannot well distinguish the various logical relationships between aspects that exist in the data and thus do not model the relationship between aspects. In the manuscript, a prompt-enhanced sentiment analysis (PESA) is proposed. Hence, a novel and efficient approach could retrieve the training set that is most similar to the input text and represents the mutual information between aspects by utilizing the [MASK] token. Moreover, the proposed model only needs to forward once if a sentiment analysis of multiple aspects is required in the inference stage. The language representation model BERT is employed to boost the performance of the proposed method. Comprehensive experiments and conducted analysis indicate the efficiency of the proposed methodology.

## Introduction

ABSA (Brauwers and Frasincar 2023) is a natural language processing technique that involves identifying the distinct phases or attributes of a commodity, service, or topic which people are discussing and determining the sentiment or opinion associated with each aspect. The goal of ABSA is to go beyond traditional sentiment analysis, which only provides an overall sentiment score for a piece of text. By breaking down the text into different aspects, ABSA provides a more detailed and nuanced understanding of people's opinions and sentiments. For example, consider a product review for a laptop that mentions the aspects of battery life, display quality, and keyboard

**CONTACT** Yatian Shen ✉ ytshen2022@126.com School of Computer and Information Engineering, Henan University, Kaifeng, 475004 China

comfort. ABSA would analyze each of these aspects separately to determine whether the neutral, positive, or negative emotions regarding them are articulated (Zhang et al. 2022).

ABSA can be used in a variety of applications, such as surveilling social media, analyzing feedback from customers, and researching the market. It can help businesses and organizations understand the specific aspects of their products or services that are driving customer satisfaction or dissatisfaction and make data-driven decisions based on that understanding.

The ABSA has recently gained a boosting interest in e-commerce, so a fine-grained sentiment assignment to determine the tendency of sentiments toward a particular phase has been widely researched. For instance, given a remark "Average to good Thai food, but terrible delivery." as presented in Figure 1, the ABSA operation aims to pinpoint those two words (i.e., food and delivery in the written text) having positive and negative sentiments tendencies, respectively.

Various types of deep learning approaches were proposed and utilized for the ABSA operation as deep learning models tend to advance, aiming to efficiently find out helpful hidden representations between the particular aspects and the sentences, and thus achieve promising results. Most previous methods employ Recurrent Neural Networks (RNN) combined with attention mechanisms to accomplish these kinds of tasks. Besides, some recent works on the language models dealing with Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2018), A Robustly Optimized BERT Pretraining Approach (RoBERTa) (Liu et al. 2019), and Generalized Autoregressive Pretraining for Language Comprehension (XLNet) (Yang et al. 2019)) have extensively enhanced the outcomes in these types of tasks. These methods used to pre-train language approaches are constructed to represent deep bidirectional forms by utilizing massive unlabeled corpus collectively based on conditioning on two-way context across all layers.
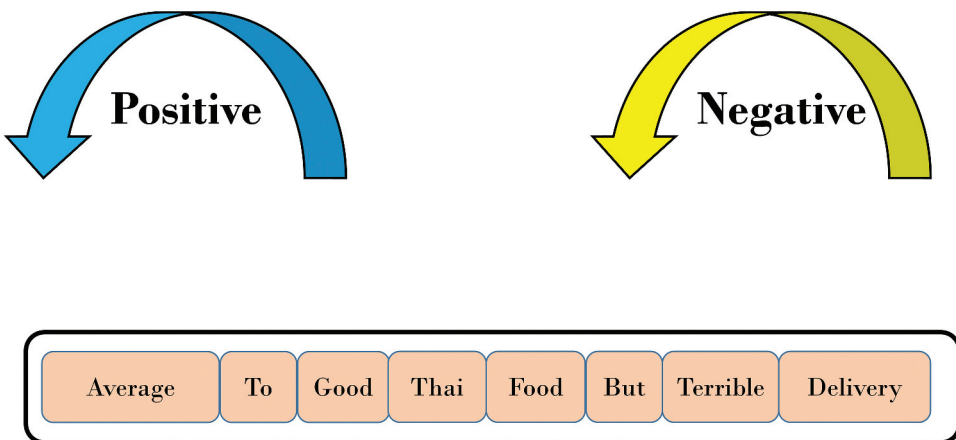


**Figure 1.** An illustration.

They are generally divided into two parts. However, these methods have some limitations. Firstly, RNN-based models become challenging to parallelize, and the backpropagation takes too much time. Besides, as the step size increases, the utilization of long-term memory becomes very difficult to deal with. Moreover, these models ignore grammatical information in sentences, making it challenging to match aspects and corresponding sentiment expressions correctly. Secondly, as scales of the pre-trained models raise, the hardware and data necessities, and costs related to fine-tuning rise rapidly. Furthermore, the rich and distinct downstream assignments also make the construction of the two stages (including pre-training and fine-tuning stages) laborious and complex (e.g.,

answering questions (Rajpurkar et al. 2016), sentiment analysis (Socher et al. 2013), and the identification of named entities (Sang and De Meulder 2003)). Therefore, researchers hope to explore more compact, lightweight, universal, and efficient methods, in which the prompt would be an attempt in this direction. Some recent work is no longer confined to formal prompts, which brings more imagination to related research.

In the manuscript, a novel and effective model was proposed, namely prompt-enhanced sentiment analysis (PESA), which combines the prompt from datasets and mutual information between aspects. Prompt learning employs a cloze-style operation in the fine-tuning operation that functions as a pretraining. Hence, it could employ the information of pretraining more efficiently and greatly improve the few-shot outcome. Besides, the language representation model BERT is leveraged to boost performance (Devlin et al. 2018).

Resemblance matching between written documents and retrieval operations is a substantial direction in the process of retrieving information. (Crammer, Singer, and Singer 2002) implemented a model called BERT to train abstracts/titles and query operations. [CLS], which represents a unique vector, is inserted to the top of the input before transmitting and assigning it to the BERT, and [SEP], which represents a unique tag to differentiate sentences, is inserted as a divider within a text. Afterward, the result of the BERT algorithm (the embedding of sentence pairs) was employed, and [CLS] was employed to finish the computation phase. The result sigmoid was calculated to attain the resemblance between the written document and the query, which was regarded as the resemblance values between the input abstract/title and the query.

Furthermore, various masking blueprints were proposed to take samples. Devlin et al. (2019) casually picked out the input tokens based on using a uniform distribution; Joshi et al. (2020) took a sample set of contiguous spans of a text; Levine et al. (2021) sampled words and spans consisting of increased pointwise mutual information (PMI). These improved sampling

methods were implemented to restrict models to exploit superficial indigenous hints from masking uniformly.

Firstly, to better find out the different emotions corresponding to specific aspects of sentences and decrease the inconsistency between the pre-training and fine-tuning processes to help the algorithm learn the semantic information more sufficiently, employing the classification method based on the [MASK] token is proposed. Specifically, for each sentence to be subjected to sentiment analysis, the same [MASK] token behind each aspect is spliced, and then the sentence is inserted into the pre-trained algorithm to attain the hidden vector representations. Finally, the shared sentiment classification layer is leveraged to attain similar sentiment categories.

Secondly, the training example that is most alike to the input text in the set of training was retrieved, which is then concatenated as the final input and fed into the model to obtain the sentiment category. Even if the model scale increases exponentially, the pre-trained model cannot remember all the patterns in the training data, which in part means that recapturing the relevant training data as the prompt can provide explicit information to improve the model's performance (Wang et al. 2022). The method works well and is easily scalable and computationally inexpensive. The conducted experiments assess the effectiveness of the suggested approach. The outcomes indicate that the suggested approach significantly boosts performance.

## Literature Review

### *Approaches Based on Unsupervised Pretraining*

Mastering word representations for applications has been a hot study field for some decades. Many related studies have appeared one after another. Sarzynska-Wawer et al. (2021) suggested the embeddings from language models (ELMo) based on two unidirectional LSTMs (including left-to-right and right-to-left LSTMs). The architecture of the ELMo was dependent upon the recurrent neural network, which presents difficulty to parallelize, and the backpropagation takes too much time, hindering the large-scale pre-training on the massive corpus. Moreover, as the step size increases, long-term memory becomes difficult.

In recent years, some work has employed an architecture using a transformer-based approach (Vaswani et al. 2017) and designed various pre-training tasks, significantly improving performance. Devlin et al. propose the Masked LM (MLM) and Next Sentence Prediction (NSP) training objectives. Liu et al. introduced the RoBERTa, which uses more data, trains more steps with larger batch sizes, and excludes the NSP task. Some recent work has also

made new attempts and achieved good results (Cui et al. 2020; He et al. 2020; Lan et al. 2019).

## Sentiment Analysis

Pre-training of large language models is conducted using large-scale unlabeled texts having different training goals. Their fine-tuning algorithms have succeeded in good results concerning Natural Language Processing (NLP) assignments such as sentiment analysis (Socher et al. 2013), question responding (Rajpurkar et al. 2016), part-of-speech tagging, named entity identification (Sang and De Meulder 2003), and so on. ABSA, as a sentiment analysis tool, has gained increasing momentum. The ABSA assignment is a fine-grained sentiment operation, determining the sentimental tendency toward a specific phase (Liu 2012). For instance, given the expression "We went for lunch. Although the service was excellent, the taste was not good.," the ABSA operation aims to pinpoint the two phases (i.e., "lunch" and "service" contained in the sentence) that have negative and positive tendencies respectively.

Specifically, Kiritchenko et al. present the feature-based Support Vector Machine (SVM) with conventional feature extraction. Tang et al. devise the Temporal Dependence Long short-term memory (TD-LSTM) that leverages the stacked recursive model of LSTM to encipher the expression and employ the last hidden vector to predict the sentiment categories. A multi-hop attention mechanism equipped with external memory (Mem-Net) (Tang, Qin, and Liu 2016) is introduced to capture the importance of each token by treating context as an external memory and applying a multi-hop attention layer on token embeddings for representation. Ma et al. crunch the terms of the context and aspect with 2 stacked Long- and Short-Term Memory (LSTM) interactively, which output the representations for the context and aspect. Some BERT-style approaches also demonstrate huge benefits. BERT text pair classification model (BERT-SPC) (Song et al. 2019) uses the pre-trained model of BERT as the chipper and uses the result of the [CLS] token at the final layer as the feature concerning sentiment classification.

As machine learning methods are employed, the contribution of the pre-training step brings better performance to these methods (Liu et al. 2015; Singh et al. 2021; Xu, Zhang, and Hong 2022). The BERT model, an approach representing the aspect of pre-trained language, was suggested by (Devlin et al. 2019) and the performance of BERT was determined to be more exceptional than the ones in the literature. Park et al. employed a classifier based on BERT to train retrieved manuscripts and word vectors to characterize manuscripts in medicine. The studies were sorted out based on resemblance values between query semantic items and the manuscript. The outcomes indicated the accuracy was extensively enhanced over available methods. Pan et al. incorporated the health records of patients with manuscripts published in the biomedical

field and utilized three algorithms to supplement the terms utilized in query processes, and the outcomes designated that the suggested algorithm succumbed an encouraging weighted mean accuracy, better reliability, and relevance. Maciej et al. examined the efficacy of a BERT-based rating approach on distinct platforms. The outcomes validated the accuracy of the BERT approach for the implementations of precision medicine too. Bayesian networks into query elongation and probabilistic approaches to supplement query semantic terms to raise the accuracy of the query by Balaneshinkordan and Kotov, 2019. Two kinds of BERT approaches, BERTBASE and BERTLARGE, are readily usable (Pan et al. 2022). Some manuscripts containing several pertinent improvements of BERT could be presented by (Hu et al. 2022; Lee, Son, and Song 2022; Liu et al. 2023; Yang et al. 2021).

## Proposed Methodology

The PESA is introduced in this section, which includes two parts: retrieving the prompt from the training set and mutual information between aspects. These will be presented in subsequent subsections.

### *Prompt Designing*

In the manuscript, a novel and efficient method, namely prompt-enhanced sentiment analysis (PESA), employs BERT-style architecture and combines the prompt from datasets and mutual information between aspects. The prompt can be roughly designed from the following two perspectives: manual design templates (Brown et al. 2020; Petroni et al. 2019; Schick and Schütze 2020) and automatic learning templates (Jiang et al. 2020). The original prompt stemmed from hand-designed templates. Manual design is usually based on human language expression habits to maximize the unambiguous and relatively accurate template of semantic expressions. However, the disadvantages of hand-designed templates are relatively prominent, thus, it requires a lot of experiments, practices, and more professional terminology, and the budget cost is too high. Based on the unchangeable and significant disadvantage of manually designed templates, more researchers have gradually turned directions toward mining to learn and find suitable and accurate prompts automatically.

Even though there are many studies on the prompt, there are still a lot of loopholes in the current research situation (Chen et al. 2021; Jiang et al. 2020; Zhao et al. 2021). To accurately connect the pre-training task and the prompt is formulated as an academic problem worth deep exploration and running various types of practices. The exact mapping of prompts and labels would be the major problem. Besides, the model's expression depends on the conversion

of the prompt and the label concurrently, which leads to solid resistance to how to search or learn the optimal effect of the two illusions simultaneously.

### Retrieving the Prompt from Training Set

Considering that pre-trained models cannot remember all patterns in the training data when the model size even increases, to some extent. Thus, recapturing relevant training data can provide explicit information to improve performance. Therefore, the training dataset that is the most alike to the input text is retrieved, which is then concatenated and supplemented into the model to attain the eventual sentiment category. In our work, a TF-IDF-based retrieval algorithm is used, namely, Okapi Best Matching (BM25) (Robertson et al. 2009), a probabilistic relevance framework, which is mainly composed of 2 modules: improved term frequency (TF) and inverse document frequency (IDF):

$$\text{TF}(q_i) = \frac{f_i \cdot (k_1 + 1)}{f_i + k_1 \cdot (1 - b + b \cdot \frac{d_l}{\text{avgdl}})} \tag{1}$$

where fi denotes the occurrence numbers of the search phrase qi in the sentence related to sentiment analysis, avgdl denotes the sentence mean length in the dataset used for training, dl denotes word numbers in the sentence.

$$IDF(q_i) = \log\left(\frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}\right) \tag{2}$$

where N denotes the sentence numbers in total for the training set, and n(qi) denotes the sentence numbers containing the term qi. In this paper, the general form of the correlation score in BM25 is presented as follows:

$$DM25_{score}(D, Q) = \sum_{i=1}^{n} IDFq_i \cdot TFq_i \tag{3}$$

where D represents the current sentence in the set of training and Q denotes the sentence to be classified.

The TF represents the repetition of a phrase showing up in the presented document, indicating the importance of the term. The IDF characterizes the significance of each phrase in the corpus, indicating the importance of the term in the document and corpus. All the training data with the BM25 algorithm is indexed and the top1 text as the prompt using the sentence is retrieved. As shown in Figure 2, the retrieved sentence is spliced as the prompt in front of the sentence to be classified.
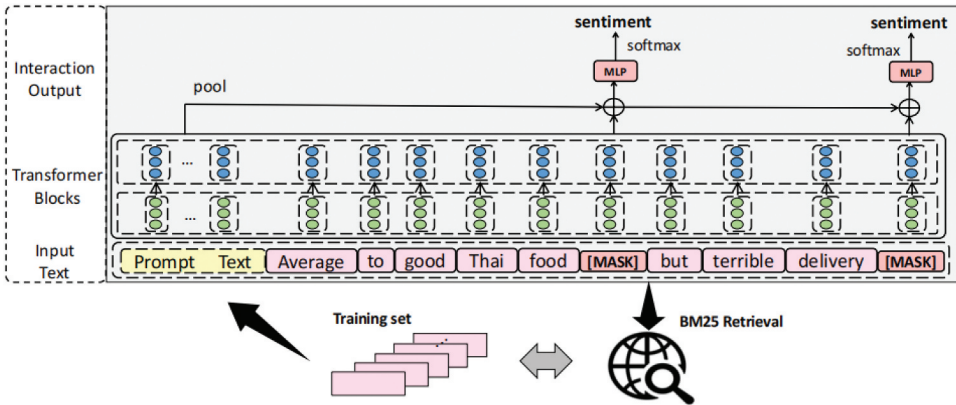
**Figure 2.** The diagram illustrating the model structure.

## Mutual Information Between Aspects

To better learn the different emotions corresponding- ing to specific aspects of the sentence, we propose the classification method to be employed based on the [MASK] token, where the [MASK] token is placed behind each aspect and bi-directional attention is employed to facilitate the representation of different aspects. Moreover, the proposed method only needs to infer once if sentiment analysis of multiple aspects is required in the inference stage.

Specifically, for each sentence to be subjected to sentiment analysis, we splice the same [MASK] to-ken behind each aspect and then input the sentence into the pre-training approach to attain the representation of a hidden vector. Finally, we leverage the shared sentiment classification layer to obtain the corresponding sentiment categories. As shown in Figure 2, we splice the [MASK] token after food and delivery. Then, we input the sentence into transformer blocks to attain the hidden representation and employ the fully connected layer for sentiment analysis so that the inconsistency between the pre-training and fine-tuning could be reduced and the model can learn the semantic information more sufficiently.

## Pre-Training of Language Models

Pre-training language approaches have shown promising results on sentiment tasks. By using as much training data as possible and extracting as many standard features as possible, the learning burden of the pre-trained language approaches for specific tasks could be reduced. In the manuscript, A BERT pre-trained with Masked LM and Next Sentence Prediction operations is employed as the chipper and leverages the representation encoded by the BERT model as guidance information for sentiment classification. Pre-training refers to training a language model (Lee, Son, and Song 2022; Wang

et al. 2022) on a large number of texts, such as news articles, books, and web pages, in an unsupervised manner, i.e., without any labels or annotations. The pre-trained approach could then be adjusted on a smaller dataset with labeled examples concerning a specific assignment, such as sentiment analysis. Pre-training language approaches have shown encouraging outcomes on tasks of analyzing sentiments (Wahba, Madhavji, and Steinbacher).

Pre-training helps the model learn the general patterns and structures of language, which can be useful for a wide range of downstream tasks. In the case of sentiment analysis, pre-training could aid the approach to better to comprehend the association between words and their associated sentiment, as well as the context in which these words appear (H. Yang, Qin, and Deng).

The current studies have shown that pre-trained language models, such as BERT and GPT-2 (Generative Pre-trained Transformer 2), can achieve a pioneering outcome on various sentiment assignments. These models can capture the nuances of language and context. They can be adjusted on distinct kinds of sentiment analysis assignments, such as binary classification (positive or negative sentiment) or fine-grained sentiment analysis (multi-class sentiment analysis with more than two labels).

Apart from the research of Devlin et al., the BERT is implemented to encode each token in the sentence into a vector whose length is fixed and transform the [MASK] token representation to the final score since we want to fully use the [MASK] token that shows in the phase of pre-training and decreases the inconsistency between pre-training and fine-tuning processes. Given the encoder, the sentence was encoded as follows:

$$h_s = W_O^T BERT_{MASK}(s) \tag{4}$$

where $w_o$ denotes the sentiment layer, and s denotes the sentence. We apply several methods to improve performance further and boost the deep bidirectional model. Firstly, different layers in the BERT have different encoding priorities. While the lower layers often learn lexical and syntactic features, the higher layer often learns semantic and other features. By doing so, we proportionally weigh the results of the BERT's singular layer.

Secondly, we employ the adversarial training method (Goodfellow, Shlens, and Szegedy 2014) of the Fast Gradient Method (FGM) to improve the model's generalization ability.

The attribution depiction attained by the BERT is mapped to the identical dimension as the number of sentiment categories through the fully connected layer (FCL). Afterward, Softmax was employed to standardize the result of the FCL to attain the category probability of the approach for each category:

$$p_y = softmax(h_s W_c + b_{c)} \tag{5}$$

where $W_c$ and $b_c$ are learnable parameters. Eq. (6) defines the loss function

$$Loss = \sum\nolimits_{(s,y) \in D} - \log p_y \qquad (6)$$

where D, s, and y represent the training set, the training sample, and the sample's ground truth, respectively.

In the manuscript, a TF-IDF-based retrieval algorithm is used to compute correlation scores in BM25 as presented in Equation (3). While the quantity D characterizes the current sentence in the training set, Q characterizes the sentence to be classified. In Equation (3), the TF and IDF represent the significance of the term and the importance of each term in the corpus. Hence, all the training data with the BM25 algorithm is indexed and the top1 text is employed as the prompt using the sentence that is retrieved. Then, the retrieved sentence is spliced as the prompt in front of the sentence to be classified.

## Experiments

### *Experimental Data*

In the manuscript, the suggested approach has been experimented with by using 3 different datasets, including Semantic Evaluation (SemEval) 2014 Task4 Restaurant, Laptop (Pontiki et al. 2016), and Twitter dataset (Dong et al. 2014). In all experiments, three kinds of samples, namely, positive, neutral, and negative were used. The same split was used in the original research concerning open-source datasets. Table 1 depicts both training and test set numbers when each category is under consideration.

### *Baseline Models*

To comprehensively assess the suggested approach's performance, a comparison concerning baseline models is conducted.

(1) SVM-Support Vector Machine (Kiritchenko et al.): use conventional attribute derivation.
(2) TD-LSTM-Target Dependent Long Short-Term Memory (Tang et al. 2015): leverages the stacked recursive model of LSTM to encode sentences and employ the last hidden vector to predict the sentiment categories.

**Table 1.** Partitions of datasets.

| Dataset | Positive | | Neutral | | Negative | |
|---|---|---|---|---|---|---|
| | Training | Test | Training | Test | Training | Test |
| Restaurant | 2164 | 728 | 637 | 196 | 196 | 196 |
| Laptop | 994 | 341 | 464 | 169 | 169 | 128 |
| Twitter | 1561 | 173 | 3127 | 346 | 346 | 173 |

(3) Mem-Net-Memory Network (Tang, Qin, and Liu 2016): treats context as an external memory and applies a multi-hop attention layer on token embeddings to represent each token's importance.

(4) IAN-Interactive Attention Networks (Ma et al., 2017): represents the context and aspect with two stacked LSTMs interactively, which output the depictions for both context and aspect.

(5) AEN-BERT-Attentional Encoder Network for Targeted Sentiment Classification (Song et al. 2019): employs the multi-head attention mechanism to generate the sentence representations and attributes and utilizes label smoothing to process the neutral samples.

(6) TD-GAT-BERT-Target Dependent Graph Attention Network BERT (Huang and Carley 2019a, 2019b): learns the attribute representations with a multi-layer GCN and employs the LSTM to model the attribution depictions of the multi-layer GCN.

(7) CDT- Convolution Dependency Tree (Sun et al. 2019): employs the bidirectional LSTM to obtain the context-related word representation then uses a layer of graph convolutional neural network, and finally uses the representation of attributes as feature classification.

(8) BERT-SPC-Text Pair Classification Model (Song et al. 2019): uses the pre-trained model of BERT as the encipher and operates the output of the [CLS] token at the last layer as the feature to classify sentiment.

## Experimental Setup

The version of the best-base-uncased approach from the open-source tool 1 is employed, and the hidden state dimension equals 768 with a total of 12 layers. The batch size was assigned to 16, the learning ratio of the BERT algorithm was assigned to 3e-5, and the learning ratio of other modules was assigned to 5e-5. The learning ratio increased linearly as the training step numbers raised, peaked at the first 500 steps, and reduced in a linear form. While weight decay is set to 0.01, a dropout probability is assigned to 0.1. The parameters of the output layer were initialized based on a uniform distribution utilizing the method called Xavier, and then the approach was get trained by employing an optimization method called Adam (Kiritchenko, Zhu, and Mohammad 2014). The experiment averages 5 randomly assigned seeds. Each experiment was trained for three epochs, and the early-stopping approach was implemented. As the accuracy rate of three consecutive verifications did not rise, the training was terminated early.

## Experimental Result

In this work, Accuracy and Macro-F1 are used to assess the performance. The outcomes of seven baseline approaches and the suggested one are presented in

**Table 2.** Outcomes (the outcomes of PESA are written in bold.

| Method | Restaurant | | Laptop | | Twitter | |
|---|---|---|---|---|---|---|
| | Accuracy | Macro-F1 | Accuracy | Macro-F1 | Accuracy | Macro-F1 |
| SVM* | 80.16 | - | 70.49 | - | 63.40 | 63.30 |
| IAN* | 78.60 | - | 72.10 | - | - | - |
| TD-LSTM* | 75.63 | - | 68.13 | - | 70.80 | 69.00 |
| MemNet* | 78.16 | 65.83 | 70.33 | 64.09 | 68.50 | 66.91 |
| TD-GAT-BERT* | 83.00 | - | **79.80** | - | - | - |
| CDT* | 81.78 | 72.32 | 75.73 | 71.42 | 74.00 | 72.19 |
| BERT-SPC | 84.46 | 76.98 | 78.99 | 75.03 | 73.55 | 72.14 |
| AEN-BERT | 83.12 | 73.76 | **79.93** | 76.31 | **74.71** | 73.13 |
| PESA (Ours) | **85.36** | **77.93** | 79.58 | **76.82** | 74.58 | **73.57** |
| **Ablation** | | | | | | |
| PESA | **85.36** | **77.93** | 79.58 | 76.82 | 74.58 | 73.57 |
| PESA w/o MIM | 84.72 | 77.48 | 79.34 | 76.57 | 73.97 | 73.26 |
| PESA w/o SP | 84.53 | 77.02 | 79.16 | 75.68 | 73.61 | 72.54 |

Models designated using "*" indicate that they were quoted from the original paper. Competitive outcomes on all three datasets are presented. The proposed method outperforms all baseline models on the Restaurant dataset, with Accuracy improving by 0.1%. However, it has generated slightly less accuracy in the Twitter and laptop data sets. Alternatively, the suggested algorithm outperforms all once Macro-F1 is a concern. When compared with the three baseline models based on BERT, the proposed model has better results, indicating that using various information, such as the information of the phrase structure and grammatical information in the BERT's intermediate layer, can help solve the aspect-based sentiment analysis problem better. The experiment suggests that the semantic information mastered by the higher layers of BERT is more significant than others in the task because the BERT-SPC model, without using intermediate layers as control information, succeeds in quite good outcomes.

Table 2. "-" denotes that the original research did not have the corresponding outcome.

## Ablation Study

As discussed in previous sections, the improvement of the proposed approach comes from two parts: retrieving the training example that is most similar to the input text in the training data as the prompt (SP) and learning the mutual information between aspects based on the [MASK] token (MIM). For further analysis, we conduct an ablation study as follows: Table 2 depicts that removing any part would have an impact on the suggested approach's performance, and removing the SP module brings a significant reduction, suggesting that retrieving the

**Table 3.** A case study.

| Sentence | Label | Ours |
|---|---|---|
| Although the restaurant [MASK] is nice, I prefer not to go for the **food** [MASK]. | negative | negative |
| I love the drinks, esp the lychee martini, and the **food** [MASK] is also VERY good. | positive | positive |
| I had to wait for my friend at the **bar** [MASK] for a few minutes | neutral | neutral |
| Great food [MASK] but the **service** [MASK] was dreadful! | negative | negative |
| The staff members are extremely friendly and even replaced my **drink [MASK] o**nce when I dropped it outside. | neutral | neutral |
| Average to good Thai **food** [MASK], but terrible [MAKS] delivery | positive | positive |
| The waitstaff is solicitous and friendly and always seems glad to see us, and the **food** [MASK] is wonderful, if not stunningly creative. | positive | positive |

training example that is most similar to the input text in the training data as the prompt (SP) becomes more important.

## Discussion

### Case Analysis

We analyzed several samples that verify the benefits of the suggested approach, which deals with complex contexts and sentiment expressions. Table 3 presents the outcomes. In each sentence, the aspect is written in a bold format, followed by the label of the ground truth and the classification result of the proposed model.

Apart from fine-tuning, prompt learning transforms the training process into the same as pretraining. When the process of pretraining is in use, the input words are casually masked, enabling the approach to recapture these masked words. To be coherent with the process, the input is wrapped by utilizing an assignment-relevant template having one keyword masked. BERT is employed to encode each token in the sentence into a vector whose length is fixed and transform the [MASK] token representation to the final score. By doing so, the [MASK] token is fully utilized in the pre-training state, thus reducing the inconsistency between the pre-training and fine-tuning phases.

In the first sentence, the proposed model correctly classifies the sentiment orientation with the aspect "food" as negative; similarly, in the third sentence, the proposed algorithm classifies the sentiment with the aspect bar for neural. In the following few sentences, the proposed
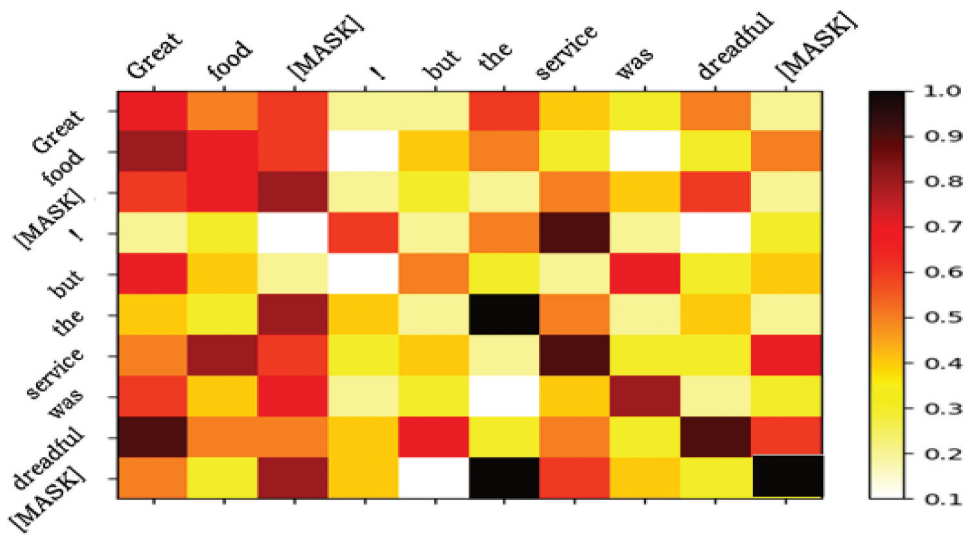


**Figure 3.** Attention scores' heat map.

model also completed the classification, showing that emotional expressions could be matched with the corresponding specific aspects when faced with such complex contexts. For instance, in the sixth sentence, even with the contradictory sentiment tendencies of good and terrible coexisting, the proposed model is not disturbed and finally makes the correct judgment.

### Visualization of Attention Scores

We obtain the attention scores for the last layer of BERT and visualize these results with a heat map. Figure 3 depicts that the first inserted [MASK] token effectively notices the word Great in the sentence that is most nearly pertinent to its aspect of food and the second is aware of service and dreadful, which shows that the suggested approach could efficiently help the approach master the association between aspect and corresponding sentiment.

## Conclusion

In the manuscript, a prompt-enhanced sentiment analysis (PESA) was proposed, which is a novel and efficient approach that combines the mutual information between aspects and prompts from datasets. We employ the language representation model BERT to boost the suggested approach's performance. The outcomes of the comprehensive experimentations indicate that the suggested algorithm could accurately classify the sentiment of a sentence given a specific aspect is a concern. We have also done the case study and constructed a heat map of the attention. So, the suggested model could accurately provide the corresponding sentiment for each aspect. Furthermore, the heat map designates that the suggested algorithm could efficiently construct a link between the aspect and the related sentiment phrases. Future work plans to deal with introducing a dependency syntax three to improve performance further.

## Acknowledgements

## Disclosure Statement

## Funding

# References

Balaneshinkordan, S., and A. Kotov. 2019. Bayesian approach to incorporating different types of biomedical knowledge bases into information retrieval systems for clinical decision support in precision medicine. *Journal of Biomedical Informatics* 98:103238. doi:10.1016/j. jbi.2019.103238.

Brauwers, G., and F. Frasincar. 2022. A survey on aspect-based sentiment classification. *ACM Computing Surveys* 55 (4):1–37.

Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33:1877–901.

Cambria, E., B. Schuller, Y. Xia, and C. Ine Havasi. 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems* 28 (2):15–21. doi:10.1109/MIS.2013.30.

Chen, X., X. Xie, N. Zhang, J. Yan, S. Deng, C. Tan, F. Huang, S. Luo, and H. Chen. 2021. Adaprompt: Adaptive prompt-based finetuning for relation extraction. *arXiv preprint arXiv* 2104:07650.

Crammer, K., Y. Singer, and Y. Singer. 2002. Pranking with ranking. *Advances in Neural Information Processing Systems* 14:641–47.

Cui, Y., W. Che, T. Liu, B. Qin, S.J. Wang, and H. Guoping. 2020. Revisiting pre-trained models for Chinese natural language processing. arXiv preprint arXiv:200413922.

Devlin, J., M. -W. Chang, K. Lee, and K. Toutanova. 2019. "BERT: Pre-training of deep bidirectional transformers for language understanding". In Proceedings of the 17th Conference of the North American chapter of the association for computational linguistics: Human language technologies, (NAACL-HLT'19), Minneapolis, Minnesota, 4171–86. http://arxiv.org/abs/1810.04805

Devlin, J., M.W. Chang, K. Lee, K. Toutanova, W. M. DeCampli, C. A. Caldarone, A. Dodge-Khatami, P. Eghtesady, J. M. Meza, P. J. Gruber, et al. 2018. Intervention for arch obstruction after the Norwood procedure: Prevalence, associated factors, and practice variability. *The Journal of Thoracic and Cardiovascular Surgery*, 157 (2):684–95.e8. arXiv preprint arXiv:1810.04805. doi:10.1016/j.jtcvs.2018.09.130.

Dong, L., F. Wei, C. Tan, D. Tang, M. Zhou, and K. 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), Baltimore, Maryland, 49–54.

Goodfellow, I. J., J. Shlens, and C. Szegedy. 2014. Explaining and harnessing adversarial examples. arXiv preprint arXiv. 1412:6572.

He, X., E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, et al. 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine* 26:672–75. doi:10.1038/s41591-020-0869-5.

Huang, B., and K. M. Carley. 2019a. A hierarchical location prediction neural network for Twitter user geolocation. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Hong Kong, China, 4732–42.

Huang, B., and K. M. Carley. 2019b. Syntax-aware aspect level sentiment classification with graph attention networks. *arXiv preprint arXiv* 1909:02606.

Hu, W., L. Liu, Y. Sun, Y. Wu, Z. Liu, R. Zhang, and T. Peng. 2022. NLIRE: A Natural Language Inference method for Relation Extraction. *Journal of Web Semantics* 72:100686.

Jiang, Z., F. F. Xu, J. Araki, and G. Neubig. 2020. How can we know what language models know? *Transactions of the Association for Computational Linguistics* 8:423–38. doi:10.1162/tacl_a_00324.

Kingma, D. P., and B. Jimmy. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv. 1412:6980.

Kiritchenko, S., X. Zhu, and S. M. Mohammad. 2014. Sentiment analysis of short informal texts. *The Journal of Artificial Intelligence Research* 50 (2014):723–62. doi:10.1613/jair.4272.

Lan, Z., M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv* 1909:11942.

Lee, Y., J. Son, and M. Song. 2022. BertSRC: Transformer-based semantic relation classification. *BMC Medical Informatics and Decision Making* 22 (2022):234. doi:10.1186/s12911-022-01977-5.

Levine, G. N., B. E. Cohen, Y. Commodore-Mensah, J. Fleury, J. C. Huffman, U. Khalid, D. R. Labarthe, H. Lavretsky, E. D. Michos, E. S. Spatz, et al. 2021. Psychological health, well-being, and the mind-heart-body connection: A scientific statement from the American Heart Association. *Circulation* 143 (10):e763–83.

Liu, B. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies* 5 (1):1–167.

Liu, Y., M. Ott, N. Goyal, D. Jingfei, M.D. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv* 1907:11692.

Liu, J.-M., M. You, Z. Wang, G.-Z. Li, X. Xu, and Z. Qiu. 2015. Cough event classification by pre-trained deep neural network. *BMC Medical Informatics and Decision Making* 15 (Suppl 4):S2. doi:10.1186/1472-6947-15-S4-S2.

Liu, Y., R. Zhang, T. Li, J. Jiang, J. Ma, and P. Wang. 2023. MolRoPE-BERT: An enhanced molecular representation with rotary position embedding for molecular property prediction. *Journal of Molecular Graphics & Modelling* 118:108344. doi:10.1016/j.jmgm.2022.108344.

Pan, M., J. Wang, J. X. Huang, A. J. Huang, Q. Chen, and J. Chen. 2022. A probabilistic framework for integrating sentence-level semantics via BERT into pseudo-relevance feedback. *Information Management and Processing* 59 (1):102734. doi:10.1016/j.ipm.2021.102734.

Petroni, F., T. Rocktäschel, P. Lewis, A. Ton Bakhtin, W. Yuxiang, A. H. Miller, and S. Riedel. 2019. Language models as knowledge bases? *arXiv preprint arXiv* 1909:01066.

Pontiki, M., D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M.M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, et al. 2016. Semeval- 2016 task 5: Aspect-based sentiment analysis. International workshop on semantic evaluation, San Diego, California, 19–30.

Rajpurkar, P., J. Zhang, K. Lopyrev, and P. Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250.

Robertson, S., H. Zaragoza, et al. 2009. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends® in Information Retrieval*. 3(4):333–89. doi:10.1561/1500000019.

Sang, E. F., and F. De Meulder. 2003. Introduction to the Conll2003 shared task: Language. Introduction to the Conll2003 shared task: Languageindependent named entity recognition. arXiv preprint cs/0306050.

Sarzynska-Wawer, J., A. Wawer, A. Sandra Pawlak, J. Szymanowska, I. Stefa- Niak, M. Jarkiewicz, and L. Okruszek. 2021. Detecting formal thought disorder by deep contextualized word representations. *Psychiatry Research* 304:114135. doi:10.1016/j.psychres.2021.114135.

Schick, T., and H. Schütze. 2020. Few-shot text generation with pattern-exploiting training. *arXiv preprint arXiv* 2012:11926.

Singh, B., A. Kshatriya, E. Sagheb, W. CIl, J. Yoon, H. Y. Seol, Y. Juhn, and J. Sohn. 2021. Identification of asthma control factor in clinical notes using a hybrid deep learning model. *BMC Med Inform Decis Mak* 21 (Suppl 7):272. doi:10.1186/s12911-021-01633-4.

Socher, R., A. Perelygin, W. Jean, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing, Seattle, Washington, USA, 1631–42.

Song, Y., J. Wang, T. Jiang, Z. Liu, and Y. Rao. 2019. Attentional encoder network for targeted sentiment classification. *arXiv preprint arXiv* 1902:09314.

Sun, K., R. Zhang, S. Mensah, Y. Mao, and X. Liu. 2019. Aspect-level sentiment analysis via convolution over dependency tree. In Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), Hong Kong, China, 5679–88.

Tang, D., B. Qin, X. Feng, and T. Liu. 2015. Effective LSTMs for target-dependent sentiment classification. *arXiv preprint arXiv* 1512:01100.

Tang, D., B. Qin, and T. Liu. 2016. Aspect-level sentiment classification with deep memory network. *arXiv preprint arXiv* 1605:08900.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* 30.

Wang, S., X. Yichong, Y. Fang, Y. Liu, S. Sun, X. Ruochen, C. Zhu, and M. Zeng. 2022. Training data is more valuable than you think: A simple and effective method of retrieving from training data. *arXiv preprint arXiv* 2203:08773.

Xu, H., C. Zhang, and D. Hong. 2022. BERT-based NLP techniques for classification and severity modeling in basic warranty data study. *Insurance, Mathematics & Economics* 107:57–67. doi:10.1016/j.insmatheco.2022.07.013.

Yang, Z., Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems* 32.

Yang, F., X. Wang, H. Maand, and J. Li. 2021. Transformers-sklearn: A toolkit for medical language understanding with transformer-based models. *BMC Medical Informatics and Decision Making* 21 (Suppl 2):90. doi:10.1186/s12911-021-01459-0.

Zhao, Z., E. Wallace, S. Feng, D. Klein, and S. Singh. 2021. Calibrate before use: Improv- ing few-shot performance of language models. In International Conference on Machine Learning, United States, 12697–706. PMLR.