

Learning Implicit Sentiment in Aspect-based Sentiment Analysis with Supervised Contrastive Pre-Training

Zhengyan Li¹, Yicheng Zou¹, Chong Zhang¹, Qi Zhang^{1*} and Zhongyu Wei²

¹Shanghai Key Laboratory of Intelligent Information Processing,
School of Computer Science, Fudan University, Shanghai, China

²School of Data Science, Fudan University

{lizy19, yczou18, chongzhang20, qz, zywei}@fudan.edu.cn

Abstract

Aspect-based sentiment analysis aims to identify the sentiment polarity of a specific aspect in product reviews. We notice that about 30% of reviews do not contain obvious opinion words, but still convey clear human-aware sentiment orientation, which is known as implicit sentiment. However, recent neural network-based approaches paid little attention to implicit sentiment entailed in the reviews. To overcome this issue, we adopt Supervised Contrastive Pre-training on large-scale sentiment-annotated corpora retrieved from in-domain language resources. By aligning the representation of implicit sentiment expressions to those with the same sentiment label, the pre-training process leads to better capture of both implicit and explicit sentiment orientation towards aspects in reviews. Experimental results show that our method achieves state-of-the-art performance on SemEval2014 benchmarks, and comprehensive analysis validates its effectiveness on learning implicit sentiment.

1 Introduction

Aspect-level sentiment analysis (ABSA) is a fine-grained variant aiming to identify the sentiment polarity of one or more mentioned aspects in product reviews. Recent studies tackle the task by either employing attention mechanisms (Wang et al., 2016b; Ma et al., 2017) or incorporating syntax-aware graph structures (He et al., 2018; Tang et al., 2020; Zhang et al., 2019; Sun et al., 2019; Wang et al., 2020). Both methodologies aim to capture the corresponding sentiment expression towards a particular aspect, which is usually an opinion word that explicitly expresses sentiment polarity. For instance, given the review on a restaurant “*Great food but the service is dreadful*”, current models attempt to find “*great*” for aspect “*food*” to determine the positive sentiment polarity towards it.

* Corresponding author.

Reviews contain implicit sentiment
The waiter poured water on my hand and walked away
The bartender continued to pour champagne from his reserve
10 hours of battery life ...
The battery life is probably an hour

Table 1: Examples of reviews contain implicit sentiment where aspects are marked to bold. In the above examples, “*pour*” expresses opposite emotions in different contexts. In the below examples, people determine the sentiment orientations towards “*battery*” by referring to a common lifetime.

However, implicit sentiment expressions widely exist in the recognition of aspect-based sentiment. Implicit sentiment expressions indicate sentiment expressions that contain no polarity markers but still convey clear human-aware sentiment polarity in context (Russo et al., 2015). As illustrated in Table 1, the comment “*The waiter poured water on my hand and walked away*” towards aspect “*waiter*” contains no opinion words, but can be clearly interpreted to be negative. According to Table 2 (as seen in Section 4), 27.47% and 30.09% of reviews contain implicit sentiment among Restaurant and Laptop datasets. However, most of the previous methods generally pay little attention on modeling implicit sentiment expressions. This motivates us to better solve the task of ABSA by capturing implicit sentiment in an advanced way.

To equip current models with the ability to capture implicit sentiment, inadequate ABSA datasets are the main challenge. With only a few thousand labeled data, models could hardly recognize comprehensive patterns of sentiment expressions, and are unable to capture enough commonsense knowledge, which is required in sentiment identification. It reveals that external sentiment knowledge should be introduced to solve the problem.

Therefore, we adopt Supervised ContrAstive Pre-Training(SCAPT) on external large-scale

sentiment-annotated corpora to learn sentiment knowledge. Supervised contrastive learning gives an aligned representation of sentiment expressions with the same sentiment label. In embedding space, explicit and implicit sentiment expressions with the same sentiment orientation are pulled together, and those with different sentiment labels are pushed apart. Considering the sentiment annotations of retrieved corpora are noisy, supervised contrastive learning enhances noise immunity of the pre-training process. Also, SCAPT contains review reconstruction and masked aspect predication objectives. The former requires representation encoding review context besides sentiment polarity, and the latter adds the model’s ability to capture the sentiment target. Overall, the pre-training process captures both implicit and explicit sentiment orientation towards aspects in reviews.

Experimental evaluations conducted on SemEval-2014 (Pontiki et al., 2014) and MAMS (Jiang et al., 2019) datasets show that proposed SCAPT outperforms baseline models by a large margin. The results on partitioned datasets demonstrate the effectiveness of both implicit sentiment expression and explicit sentiment expression. Moreover, the ablation study verifies that SCAPT efficiently learns implicit sentiment expression on the external noisy corpora. Codes and datasets are publicly available¹.

The contributions of this work include:

- We reveal that ABSA was only marginally tackled by previous studies since they paid little attention to implicit sentiment.
- We propose Supervised Contrastive Pre-training to learn sentiment knowledge from large-scale sentiment-annotated corpora.
- Experimental results show that our proposed model achieves state-of-the-art performance, and is effective to learn implicit sentiment.

2 Implicit Sentiment

As sentiment that can only be inferred within the context of reviews, many researches address the presence of implicit sentiment in sentiment analysis. Toprak et al. (2010); Russo et al. (2015) proposed similar terminologies (as *implicit polarity* or *polar facts*), and provided corpora containing implicit sentiment. Deng and Wiebe (2014) detected implicit sentiment via inference over explicit sentiment expressions and so-called

goodFor/badFor events. Choi and Wiebe (2014) used +/-EffectWordNet lexicon to identify implicit sentiment, by assuming sentiment expressions are often related to states and events which have positive/negative/null effects on entities.

To investigate the ubiquitous of implicit sentiment in ABSA, we split SemEval-2014 Restaurant and Laptop benchmarks into Explicit Sentiment Expression (ESE) slice and Implicit Sentiment Expression (ISE) slice, based on the presence of opinion words. Fan et al. (2019) have annotated opinion words for target aspects on SemEval benchmarks. We notice that the provided datasets do not keep the original order and have some differences in texts. Thus, we first match the annotations to the original datasets, and then manually pick the reviews including opinion words towards the aspect from the remaining part. As results shown in Table 2 (as seen in Section 4), 27.47% and 30.09% of reviews are divide into ISE part among Restaurant and Laptop, revealing that implicit sentiment exists widely in ABSA and is worthy to be explored.

3 Methodology

In this section, we introduce the pre-training and fine-tuning scheme of our models. In pre-training, we introduce Supervised ContrAstive Pre-Training (SCAPT) for ABSA, which learns the polarity of sentiment expressions by leveraging retrieved review corpus. In fine-tuning, aspect-aware fine-tuning is adopted to enhance the ability of models on aspect-based sentiment identification.

3.1 Supervised Contrastive Pre-training

Three objectives are included in SCAPT: supervised contrastive learning, masked aspect prediction, and review reconstruction. The details of SCAPT’s procedure are shown in Figure 1.

Transformer Encoder Backbone The pre-training scheme is built on Transformer encoder (Vaswani et al., 2017). We denote the retrieved review corpus used in SCAPT as $D = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ including n sentences. The i -th sentence \mathbf{x}_i is labeled with y_i . For each input sentence \mathbf{x}_i , following Devlin et al. (2019), we format the input sentence as $I_i = [\text{CLS}] + \mathbf{x}_i + [\text{SEP}]$ to feed into the model. The output vector of [CLS] token encodes the sentence representation \bar{h}_i :

$$\bar{h}_i, \dots = \text{TransEnc}(I_i) \quad (1)$$

¹<https://github.com/Tribleave/SCAPT-ABSA>

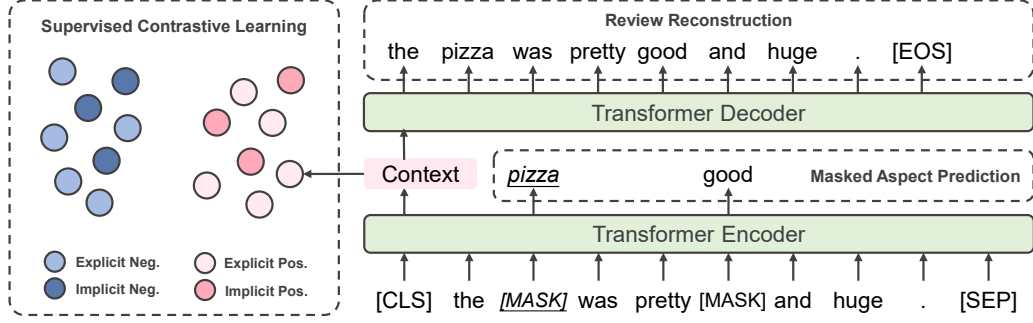


Figure 1: An overview of SCAPT on ABSA. SCAPT consists of three objectives, in which Supervised Contrastive Learning aligns the representations with the same sentiment label.

Supervised Contrastive Learning Inspired by Khosla et al. (2020), we adopt supervised contrastive learning objective in SCAPT to align the representation of explicit and implicit sentiment expressions with the same emotion. Supervised contrastive learning encourages the model to capture the entailed sentiment orientation in context and incorporate it in sentiment representation.

Specifically, for (\mathbf{x}_i, y_i) within a batch B , we first extract sentiment representation $s_i = \mathbf{W}_s \bar{h}_i$ from sentence representation \bar{h}_i of \mathbf{x}_i . \mathbf{W}_s could be seen as a trainable sentiment perceptron for sentences. The supervised contrastive loss on the batch B is defined as:

$$P_B^{sup}(i, c) = \frac{\exp(\text{sim}(s_i, s_c)/\tau)}{\sum_{b \in B, b \neq i} \exp(\text{sim}(s_i, s_b)/\tau)} \quad (2)$$

$$\mathcal{L}_B^{sup} = \sum_{i \in B} -\log \frac{1}{C_i} \sum_{y_i = y_c, c \neq i} P_B^{sup}(i, c) \quad (3)$$

Here, $P_B^{sup}(i, c)$ indicates the likelihood that s_c is most similar to s_i and τ is the temperature of softmax. Here we simply use $\text{sim}(s_i, s_c) = s_i \cdot s_c$ for similarity metric. Supervised contrastive loss \mathcal{L}_B^{sup} is calculated for every sentence s_i among B , where $C_i = |\{c | y_c = y_i, c \neq i\}|$ is the number of samples in the same category y_i in B . Notably, we do not directly use sentence representation in the supervised contrastive pre-training process. Instead, we use sentiment representation to make full use of document-level labeled corpora in mining the inherent sentiment perception.

Review Reconstruction Motivated by the power of denoising auto-encoder (Vincent et al., 2008) and its success in pre-training models (Lewis et al., 2020), we further propose review reconstruction task to enhance the sentence representation on context semantic modeling. With solely pre-trained on the supervised contrastive learning task

which only focuses on sentiment regularization, the essential semantic information is not completely preserved in the sentence representations. Thus, we additionally employ review reconstruction in SCAPT to capture comprehensive context information in sentence representations.

Generally, this objective reconstructs the whole sentence \mathbf{x}_i with the sentence representation \bar{h}_i . After encoding \mathbf{x}_i to the sentence representation \bar{h}_i , the latter is fed to Transformer decoder for autoregressive generation:

$$P^{rec}(\tilde{\mathbf{x}}_i) = \text{TransDec}(\bar{h}_i) \quad (4)$$

$\tilde{\mathbf{x}}_i$ is the recovered sentence. \bar{h}_i acts as a beginning-of-sentence input embedding in the decoding process to control the whole generation. We use the original sentence \mathbf{x}_i without masking as the gold reference of review reconstruction objective:

$$\mathcal{L}_i^{rec} = -\log P^{rec}(\tilde{\mathbf{x}}_i) \quad (5)$$

Masked Aspect Prediction In masked aspect prediction, the model learns to predict the masked aspect from a corrupted version for each review. The masking strategy of input reviews consists of following two steps:

1. **Aspect Span Masking.** Since all inputs are from our retrieved corpora, we ensure that each review contains at least one aspect. For each input, the tokens of aspect spans are replaced with [MASK] with 80% probability, or replaced with a random token with 10% probability, otherwise kept unchanged. Aspect span masking provides a better capture of aspect words.
2. **Random Masking.** After aspect span masking, if the proportion of masked tokens is less than 15%, we randomly mask extra tokens from the rest ones to reach the proportion.

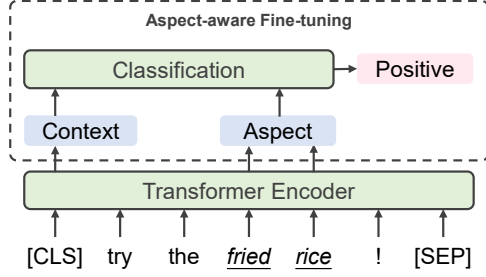


Figure 2: Aspect-aware fine-tuning on Transformer encoder based models. Sentiment representation and aspect-based representation are taken into account in sentiment classification.

We denote the input token of [MASK] as w_{MASK} . For each masked input token at k -th position, its contextualized hidden representation h_{ik} is fed into a softmax layer to predict the original word:

$$P^{map}(k) = \text{softmax}(\mathbf{W}_o h_{ik}) \quad (6)$$

Specific to the above equation, h_{ik} is the output of Transformer encoder at k -th position, \mathbf{W}_o is a trainable parameter matrix, and $P^{map}(k)$ indicates the predict probability of the original word at k -th position. The masked aspect prediction loss is an accumulation of log-likelihood on predictions of each masked position:

$$\mathcal{L}_i^{map} = \sum_{x_{ik}=w_{MASK}} -\log P^{map}(k) \quad (7)$$

Different from MLM (Devlin et al., 2019) or sentiment masking (Tian et al., 2020), masked aspect prediction focuses more on modeling aspect-related context information in aspect-based representations, which complements the other pre-training objectives and purposefully benefits our fine-tuning scheme.

Joint Training The three losses mentioned above are combined and jointly trained in SCAPT. For the overall pre-training loss \mathcal{L}_B^{pre} on batch B , the review reconstruction loss and masked aspect prediction loss are counted on each example $b \in B$, and α and β are coefficients to balance the objectives:

$$\mathcal{L}_B^{pre} = \mathcal{L}_B^{sup} + \alpha \sum_{b \in B} \mathcal{L}_b^{rev} + \beta \sum_{b \in B} \mathcal{L}_b^{map} \quad (8)$$

3.2 Aspect-Aware Fine-tuning

Our proposed models are fine-tuned on ABSA benchmarks by aspect-aware fine-tuning, to fully leverage their ability of sentiment identification. They also learn to capture aspect-related sentiment information during fine-tuning.

Specifically, given a sentence $\mathbf{x}^{ab} = \{w_1, \dots, w_a, \dots, w_n\}$ in ABSA dataset D^{ab} , and w_a is one of the aspects occurring in \mathbf{x}^{ab} . In fine-tuning, models predict aspect-level sentiment orientation y_a^{ab} according to aspect-based representation \bar{h}_a^{ab} and sentiment representation s^{ab} .

Aspect-based Representation The research (Ethayarajh, 2019) on pre-trained contextualized word representation has demonstrated that it can capture context information related to the word. Thus, in spite of using laborious methods to embed the aspect information, we extract aspect-based representation \bar{h}_a^{ab} by collecting final hidden states that correspond to w_a . In fine-tuning, \bar{h}_a^{ab} would focus on aspect-related words in context, which we believe would enhance the perception of aspect-specific opinion words and bring the model with a good view of explicit sentiment. Specifically, let \mathcal{I}_a be the token index in aspect x_a , we average the hidden state h_i for all $i \in \mathcal{I}_a$ to acquire aspect-based representation:

$$\bar{h}_a^{ab} = \text{AveragePooling}_{i \in \mathcal{I}_a}(h_i) \quad (9)$$

Notably, when processing multiple aspects w_{a1}, w_{a2}, \dots in sentence \mathbf{x}^{ab} , we extract aspect-based representation $\bar{h}_{a1}^{ab}, \bar{h}_{a2}^{ab}, \dots$ in a single run, while previous methods embed aspect and encoder whole input for each aspect one-by-one.

Representation Combination For sentiment classification, aspect-based representation and sentiment representation are considered jointly to predict aspect-level sentiment polarity. In that case, fine-tuned model builds the perception of both word-occurrence-related explicit sentiment and semantic-related implicit sentiment. We use the same sentiment perceptron \mathbf{W}_s in pre-training to extract sentiment representation s^{ab} from sentence representation. Then sentiment representation s^{ab} and aspect-based representation \bar{h}_a^{ab} are concatenated for predicting aspect-level sentiment polarity:

$$y_a^{ab} = \text{softmax}(\mathbf{W}_a [s^{ab}; \bar{h}_a^{ab}]) \quad (10)$$

y_a^{ab} is the prediction on aspect x_a and \mathbf{W}_a is trainable parameter matrix. Lastly, our fine-tuning objective is cross-entropy loss for prediction task $\mathcal{L}^{ab} = -\sum_{\mathbf{x}^{ab} \in D^{ab}} \log y_a^{ab}$.

4 Experimental Settings

ABSA Datasets Our experiments are mainly conducted on two benchmarks, Laptop and Restaurant review from SemEval 2014 task 4 (Pontiki et al., 2014). We use ESE and ISE slices of their test parts to evaluate model performance on explicit and implicit sentiment respectively. The process to build these slices is detailed in Section 2. Furthermore, we also use a more challenging dataset, Multi-Aspect Multi-Sentiment (MAMS) (Jiang et al., 2019), which shares the same domain to SemEval2014 Restaurant. All these datasets involve three sentiment categories which are positive, neutral, and negative. The details of these ABSA datasets can be found in Table 2.

Retrieved External Corpora We retrieve large-scale sentiment-annotated corpora from document-level labeled data for pre-training. Specifically, we first extract five-stars-rated/one-star-rated reviews from the Yelp² and Amazon Review (He and McAuley, 2016) datasets, and label them as positive/negative. Such a procedure can mitigate the noise in the 5-way rated document-level sentiment language source. Then we preserve reviews within the topic of restaurant/laptop to make sure that pre-train corpora and ABSA datasets are in the same domain. Later, we split these document-level reviews into sentences and preserve sentences containing the same aspect term as those mentioned in ABSA training sets. The sentiment label of each sentence is determined by the label of its original review. After the retrieving process, we finally acquire about 1.56/0.51 million sentence-level reviews from Yelp/Amazon that are noisy-labeled as positive/negative. After manually checking a small portion of both corpora, we confirm that both implicit and explicit sentiment expressions are available. We pre-train our models on the retrieved corpus that shares the same domain with the downstream ABSA task. Specifically, we adopt Yelp when dealing with Restaurant or MAMS, and Amazon for Laptop.

²<https://www.yelp.com/dataset>

Dataset	Positive	Neutral	Negative	Total	Implicit Sentiment %
Restaurant-train	2164	805	633	3602	28.59
Restaurant-test	728	196	196	1120	23.84
Restaurant	2892	1001	829	4722	27.47
Laptop-train	987	866	460	2313	30.87
Laptop-test	341	128	169	638	27.27
Laptop	1328	994	629	2951	30.09
MAMS	4183	6253	3418	13854	-
YELP	1.17M	-	0.39M	1.56M	-
Amazon	0.38M	-	0.13M	0.51M	-

Table 2: Statistics on three datasets of ABSA and two external corpus for SCAPT.

Models with SCAPT We apply SCAPT to Transformer encoder and BERT, and these models are fine-tuned by aspect-aware fine-tuning. The models are so-called TransEncAsp+SCAPT and BERTAsp+SCAPT respectively. We use a 300-dimensional randomly initialized Transformer encoder with 6 layers and 6 heads and BERT-base-uncased as the basis. The pre-training for Transformer encoder and BERT takes 80 and 8 epochs respectively. We adopt Adam (Kingma and Ba, 2015) with warm-up to optimize our models with learning rate $1e-3$ for Transformer encoder and $5e-5$ for BERT. The pre-trained models are fine-tuned by aspect-aware fine-tuning with $5e-5$ learning rate. The hyper-parameters are set as $\alpha = \beta = 1$ for combining objectives in SCAPT, and $\tau = 0.07$ in supervised contrastive learning.

Baselines We compare the proposed models with baselines from different perspectives to comprehensively evaluate the performance of our approach:

- **Attention-based models:** ATAE-LSTM (Wang et al., 2016b), IAN (Ma et al., 2017), RAM (Chen et al., 2017), and MGAN (Fan et al., 2018).
- **Graph neural networks:** ASGCN (Zhang et al., 2019), BiGCN (Zhang and Qian, 2020), CDT (Sun et al., 2019), and RGAT (Wang et al., 2020).
- **Knowledge-enhanced methods:** TransCap (Chen and Qian, 2019), BERT-SPC³ (Devlin et al., 2019), CapsNet+BERT (Jiang et al., 2019), BERT-PT (Xu et al., 2019), BERT-ADA (Rietzler et al., 2020), and R-GAT+BERT (Wang et al., 2020).

For better analyze the effect of SCAPT and aspect-aware fine-tuning, we further propose the following variants as baselines:

³BERT-SPC denotes fine-tuning BERT for Sentence Pair Classification, in which the input review is transformed to “[CLS] + context + [SEP] + aspect + [SEP]” and fed into BERT for classification (Song et al., 2019).

Method		Restaurant				Laptop			
		Acc.	F1	ESE	ISE	Acc.	F1	ESE	ISE
Attention	ATAE-LSTM (Wang et al., 2016a)	76.90*	62.64*	84.16	53.71	65.37*	62.92*	75.69	37.86
	IAN (Ma et al., 2017)	76.88*	67.71*	86.52	46.07	67.24*	63.72*	75.86	44.25
	RAM (Chen et al., 2017)	80.23	70.80	85.11	55.81	74.49	71.35	75.86	44.25
	MGAN (Fan et al., 2018)	81.25	71.94	85.18	60.04	75.39	72.47	76.16	56.31
GNN	ASGCN (Zhang et al., 2019)	80.77	72.02	84.29	62.91	75.55	71.05	75.46	57.77
	BiGCN (Zhang and Qian, 2020)	81.97	73.48	87.19	59.05	74.59	71.84	79.53	62.64
	CDT (Sun et al., 2019)	82.30	74.02	88.79	65.87	77.19	72.99	77.53	68.90
	RGAT (Wang et al., 2020)	83.30	76.08	89.45	61.05	77.42	73.76	80.17	65.52
Knowledge Enhanced	TransCap (Chen and Qian, 2019)	79.55	71.41	86.52	59.93	73.87	70.10	77.16	60.34
	BERT-SPC (Devlin et al., 2019)	83.57*	77.16*	89.21	65.54	78.22*	73.45*	81.47	69.54
	CapsNet+BERT (Jiang et al., 2019)	85.09*	77.75*	91.68	64.04	78.21*	73.34*	82.33	67.24
	BERT-PT (Xu et al., 2019)	84.95	76.96	92.15	64.79	78.07	75.08	81.47	71.27
	BERT-ADA (Rietzler et al., 2020)	87.14	80.05	94.14	65.92	78.96	74.18	82.76	70.11
	R-GAT+BERT (Wang et al., 2020)	86.60	81.35	92.73	67.79	78.21	74.07	82.44	72.99
Ours	TransEncAsp	77.10	57.92	86.97	48.96	65.83	59.53	74.31	43.20
	BERTAsp	85.80	78.95	92.73	63.67	78.53	74.07	82.33	68.39
	BERTAsp+CEPT	87.50	82.07	93.67	67.79	81.66	78.38	83.84	75.86
	TransEncAsp+SCAPT	83.39	74.53	88.04	68.55	77.17	73.23	78.70	72.82
	BERTAsp+SCAPT	89.11	83.79	94.37	72.28	82.76	79.15	84.70	77.59

Table 3: Overall performance of different methods on Restaurant and Laptop. We rerun the code of baselines and report their accuracy on ESE and ISE slices of the two datasets. For the baselines of which the accuracy or F1-score is missing, we also report the accuracy and F1-score of our rerunning version, and these results are marked with *.

- **TransEncAsp**: Directly apply aspect-aware fine-tuning on randomly initialized Transformer encoder without pre-training.
- **BERTAsp**: Directly apply aspect-aware fine-tuning on BERT-base.
- **BERTAsp+CEPT**: Merely replace the supervised contrastive learning loss with cross-entropy loss in SCAPT. Other settings are the same as BERTAsp+SCAPT.

5 Results and Analysis

This section mainly demonstrates the experiment results. Our model achieves state-of-the-art on three ABSA benchmarks, and we illustrate the representation alignment effect of supervised contrastive learning and the effectiveness of other parts from several perspectives. Moreover, we reveal that our model is capable to identify implicit sentiment, and attributes its effectiveness to supervised contrastive learning in SCAPT.

5.1 Main Results

The performance of baselines and our proposed models are shown in Table 3. Models are evaluated with Accuracy and Macro-F1. According to the results, several observations can be noted.

Our model achieves SOTA performance. BERTAsp+SCAPT outperforms the current SOTA model by 1.97%/3.80% on Restaurant/Laptop. TransEncAsp+SCAPT performs better than most baselines without pre-trained knowledge. Moreover, BERTAsp+SCAPT also achieves the best performance on ESE/ISE slices of the two datasets, revealing the effectiveness of the proposed pre-training scheme.

After pre-trained with SCAPT, models improve significantly on ABSA tasks. Compared with BERTAsp which directly fine-tuned on ABSA datasets, BERTAsp+SCAPT achieves a 3.31%/4.23% performance gain on Restaurant/Laptop, which is a convinced proof that acquiring in-domain knowledge with proper adaptive pre-training is still necessary for knowledge-enhanced models, and SCAPT is an effective approach to be adopted. Moreover, TransEncAsp+SCAPT is 6.29%/11.34% better than TransEncAsp, illustrating that incorporating sentiment knowledge with SCAPT greatly potentiates ABSA models.

SCAPT is good at learning implicit sentiment. This could be verified from several perspectives. First, compared with its performance on ESE, BERTAsp+SCAPT appears to be much better on ISE. Compared with other

Method	MAMS	
	Acc.	F1
ATAE-LSTM	77.05	-
IAN	76.60	-
CapsNet	79.78	-
BERT-SPC	82.22	-
CapsNet+BERT	83.39	-
BERTAsp	84.20	83.82
TransEncAsp+SCAPT	80.54	79.83
BERTAsp+SCAPT	85.63	85.24

Table 4: Model performance on MAMS.

works, BERTAsp+SCAPT is around 0-2% better on ESE slices, but surpasses the previous SOTA model by 4.49%/4.60% on ISE slices. Therefore, the well performance of BERTAsp+SCAPT mainly contributes to its awareness of implicit sentiment. Second, TransEncAsp+SCAPT behaves much better than BERTAsp on ISE slices. With only exposing to million-scale pre-training corpus, TransEncAsp+SCAPT is generally worse than BERTAsp on the whole task, but exceeds BERTAsp by 4.88%/4.43% on ISE slices. This demonstrates that SCAPT is data-effective on learning implicit sentiment. Last, after pre-trained with SCAPT, models attain remarkable performance gain on ISE which is much more significant than ESE. BERTAsp+SCAPT is 2% better than BERTAsp on ESE, but outperforms the latter by 8.61%/9.20% on ISE. As for Transformer encoder based models, the performance gain on ISE after SCAPT goes beyond 20%. We conclude that what models have learned in SCAPT is dominantly the perception of implicit sentiment.

Aspect-aware fine-tuning serves as a complement to SCAPT. We find that models with aspect-aware fine-tuning perform better on ESE slices of the datasets. Specifically, BERTAsp performs worse on ISE but better on ESE compared with BERT-SPC, and is therefore evaluated to be better on the two datasets. The better performance of BERTAsp on ESE slices may mainly due to its use of aspect-based representation, which attends to aspect-related context that may contain sentiment orientation. This characteristic of aspect-aware fine-tuning makes it suitable to enhance the recognition of explicit sentiment of models pre-trained with SCAPT.

Method	Restaurant	
	Acc.	F1
TransEncAsp	77.10	57.92
TransEncAsp+SCAPT	83.39	74.53
BERTAsp	85.71	78.95
BERTAsp+SCAPT	89.11	83.79
(-SCL)	86.73	80.90
(-MAP)	88.13	83.22
(-RR)	87.95	82.38
(-MAP-RR)	87.14	81.10

Table 5: Ablation study on SCAPT.

5.2 Effectiveness on Multi-aspect ABSA

Table 4 shows the performance of baselines and our models in MAMS datasets. Though it is challenging to distinguish the sentiment polarities of multiple aspects in a single sentence, the result shows TransEncAsp+SCAPT outperforms baselines that lack external sentiment knowledge, and BERTAsp+SCAPT achieves state-of-the-art in the multi-aspect scenario. The efficiency of our models can attribute to both SCAPT and aspect-aware fine-tuning since they enhance the learning of implicit and explicit sentiment respectively. Besides, BERTAsp performs much more better than BERT-SPC in MAMS than in Restaurant/Laptop. We suppose the exceeding performance of BERTAsp credits to its modeling of contextual information in aspect-based representation, which is more important in multi-aspect ABSA.

5.3 Implicit Sentiment Learning in SCAPT

We conclude the key aspects of learning implicit sentiment in SCAPT as exposing to sentiment knowledge and using supervised contrastive learning. The results in Table 3 shows that implicit sentiment is more challenging to learn than explicit sentiment, and previous methods based on attention or syntax modeling are not tackling the issue perfectly. The knowledge-enhanced baselines perform slightly better with 5% performance gain on ISE. By pre-training on large-scale sentiment-annotated corpora, our models achieve remarkable performance improvement on implicit sentiment learning, with 19.59%/29.62% relative gain on TransEncAsp. These results prove that in-domain sentiment knowledge is absolutely necessary for implicit sentiment learning, which is provided by our retrieved corpora. Furthermore, the models pre-trained with supervised contrastive learning objective surpasses cross-entropy classification in

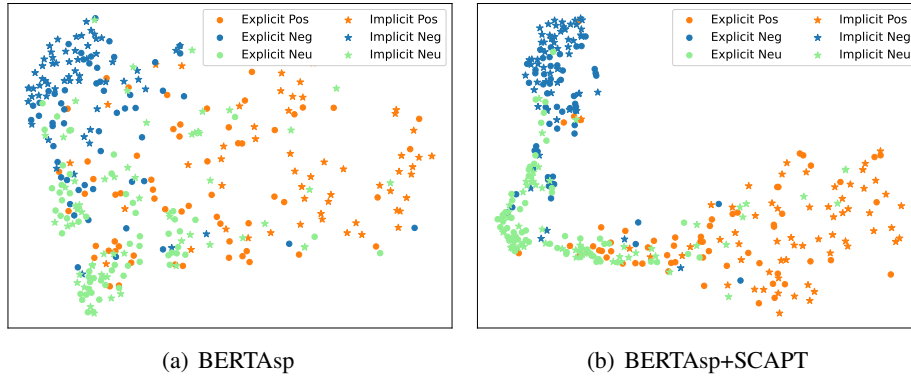


Figure 3: Visualization of the hidden sentiment representations on Restaurant (best to view the colored version). BERTAsp+SCAPT tightly clusters the representations of both explicit and implicit sentiment expressions.

ISE slices. Compared with BERTAsp+CEPT, BERTAsp+SCAPT is 4.49%/1.73% better on ISE, which leads to its better performance on the whole tasks. The deployment of supervised contrastive learning objective enhances noise immunity of the pre-training process, thus the pre-trained models are more effective in learning implicit sentiment.

5.4 Ablation Study on SCAPT

As illustrated in Table 5, we validate the effectiveness of each part by ablation study. First, removing supervised contrastive learning loss (-SCL) leads to a 2.38% performance drop on Restaurant, which is more significant than the occasion of removing the other two objectives (-MAP-RR). This verifies that supervised contrastive learning plays a primary role in SCAPT. Besides, we observe that the removing of masked aspect prediction and review reconstruction objectives also brings about performance drop. This demonstrates that these mechanisms are also indispensable in SCAPT.

5.5 Hidden Sentiment Representations

For better understanding the behavior of our proposed methods, we further perform a visualization of the sentiment representation using t-SNE (Van der Maaten and Hinton, 2008). As seen in Figure 3, models with sentiment pre-training have a strong embedding ability for sentiment expression, while many misclassifications can be found in BERTAsp. The visualization also shows that BERTAsp+SCAPT tightly clusters the representations of both implicit and explicit sentiment expressions.

Method	Restaurant-test		Laptop-test	
	Ori → New	Decline	Ori → New	Decline
LSTM	75.98→14.64	-61.34	67.55→9.87	-57.68
ASGCN	77.86→24.73	-53.13	72.41→19.91	-52.50
CapsNet+BERT	83.48→55.36	-28.12	77.12→25.86	-51.46
BERT	83.04→54.82	-29.22	77.59→50.94	-26.65
BERT-PT	86.70→59.29	-27.41	78.53→53.29	-25.24
TransEncAsp+SCAPT	83.39→67.76	-15.63	76.80→52.52	-24.28
BERTAsp+SCAPT	89.11→80.06	-9.05	82.76→76.13	-6.63

Table 6: Model performance on aspect robustness test sets. We compare the model accuracy on the original and new test sets, and the decline of prediction on new examples are reported.

5.6 Aspect Robustness

We analyze the robustness of our proposed models on aspect robustness test sets. Aspect robustness of ABSA was first emphasized and tested in Xing et al. (2020) by applying several perturbations on reviews from Restaurant and Laptop. TextFlint (Wang et al., 2021) extended these transformations by introducing transformations from various linguistic perspectives. The test sets are designed to probe whether models could distinguish the sentiment of the target aspect from the non-target aspects and unrelated information.

Table 6 lists the performance of tested models, in which the robustness of our proposed models is convincingly proved. Comparing to obvious performance drop in baseline models, BERTAsp+SCAPT performs significantly better than other models with 9.05%/6.63% decline on Restaurant and Laptop. The results show that models pre-trained with SCAPT are more robust for aspect-level perturbations, which attribute to the better modeling for sentiment and context information with the enhancement of in-domain sentiment knowledge.

6 Related Work

Neural Network Methods for ABSA The early neural network methods (Wang et al., 2016b; Ma et al., 2017) in ABSA employed various of attention mechanisms to identify aspect-related context. Memory Network (Tang et al., 2016; Chen et al., 2017; Wang et al., 2018) was further proposed to identify corresponding sentiment expression for aspects. Recent efforts (He et al., 2018; Tang et al., 2020) used syntax information from dependency trees to enhance attention-based models. A lot of works (Zhang et al., 2019; Sun et al., 2019; Wang et al., 2020) make use of graph neural networks to incorporate tree-structured syntactic information and capture aspect-related information in text. Another line in ABSA concentrated on utilizing external corpus and pre-trained knowledge to enhance semantic awareness of models (Xu et al., 2019; Rietzler et al., 2020; Dai et al., 2021).

Contrastive Representation Learning Our work adopts contrastive method in representation learning to acquire discriminating instance representations. Recent work on contrastive representation learning of instances usually based on estimating representation similarities on similar and dissimilar pairs, which are usually composed in a self-supervised manner (Chen et al., 2020; He et al., 2020). Specially, Khosla et al. (2020) illustrated a supervised contrastive method to build positive pairs between instances with same class label, and put their representations together. In this work, our models learn to capture implicit sentiment from informative but noisy language resources in supervised contrastive pre-training.

7 Conclusion

In this paper, we introduce Supervised Contrastive Pre-Training (SCAPT) for ABSA. By noticing that implicit sentiment is not well-handled by current neural network based ABSA models, we argue that more sentiment knowledge is required to solve this issue. We therefore retrieve large-scale in-domain annotated corpora, and propose SCAPT to learn sentiment knowledge from the corpora. Experimental results show that our proposed models with SCAPT achieve SOTA performance. Moreover, SCAPT is proven to be effective in implicit sentiment learning. We hope to inspire future researches on learning and modeling implicit sentiment with knowledge-enhanced methods.

Acknowledgments

The authors wish to thank the anonymous reviewers for their helpful comments. This work was partially funded by China National Key R&D Program (No. 2017YFB1002104), National Natural Science Foundation of China (No. 62076069, 61976056), Shanghai Municipal Science and Technology Major Project (No. 2021SHZDZX0103).

References

- Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. [Recurrent attention network on memory for aspect sentiment analysis](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 452–461, Copenhagen, Denmark. Association for Computational Linguistics.
- Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR.
- Zhuang Chen and Tiejun Qian. 2019. [Transfer capsule network for aspect level sentiment classification](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 547–556, Florence, Italy. Association for Computational Linguistics.
- Yoonjung Choi and Janyce Wiebe. 2014. [+/-EffectWordNet: Sense-level lexicon acquisition for opinion inference](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1181–1191, Doha, Qatar. Association for Computational Linguistics.
- Junqi Dai, Hang Yan, Tianxiang Sun, Pengfei Liu, and Xipeng Qiu. 2021. [Does syntax matter? a strong baseline for aspect-based sentiment analysis with RoBERTa](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1816–1829, Online. Association for Computational Linguistics.
- Lingjia Deng and Janyce Wiebe. 2014. [Sentiment propagation via implicature constraints](#). In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 377–385, Gothenburg, Sweden. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*,

- pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Kawin Ethayarajh. 2019. [How contextual are contextualized word representations? comparing the geometry of BERT, ELMo, and GPT-2 embeddings](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 55–65, Hong Kong, China. Association for Computational Linguistics.
- Feifan Fan, Yansong Feng, and Dongyan Zhao. 2018. [Multi-grained attention network for aspect-level sentiment classification](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3433–3442, Brussels, Belgium. Association for Computational Linguistics.
- 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.
- Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738.
- Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018. [Effective attention modeling for aspect-level sentiment classification](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1121–1131, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web*, pages 507–517.
- Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao, and Min Yang. 2019. [A challenge dataset and effective models for aspect-based sentiment analysis](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6280–6285, Hong Kong, China. Association for Computational Linguistics.
- Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschiot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 33.
- Diederik P. Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. [BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.
- Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. [Interactive attention networks for aspect-level sentiment classification](#). In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 4068–4074.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. [SemEval-2014 task 4: Aspect based sentiment analysis](#). In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.
- Alexander Rietzler, Sebastian Stabinger, Paul Opitz, and Stefan Engl. 2020. [Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4933–4941, Marseille, France. European Language Resources Association.
- Irene Russo, Tommaso Caselli, and Carlo Strapparava. 2015. [SemEval-2015 task 9: CLIPeval implicit polarity of events](#). In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 443–450, Denver, Colorado. Association for Computational Linguistics.
- Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao. 2019. Attentional encoder network for targeted sentiment classification. *arXiv preprint arXiv:1902.09314*.
- Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019. [Aspect-level sentiment analysis via convolution over dependency tree](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5679–5688, Hong Kong, China. Association for Computational Linguistics.
- Duyu Tang, Bing Qin, and Ting Liu. 2016. [Aspect level sentiment classification with deep memory network](#). In *Proceedings of the 2016 Conference on*

- Empirical Methods in Natural Language Processing*, pages 214–224, Austin, Texas. Association for Computational Linguistics.
- Hao Tang, Donghong Ji, Chenliang Li, and Qiji Zhou. 2020. [Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6578–6588, Online. Association for Computational Linguistics.
- Hao Tian, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu. 2020. [SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4067–4076, Online. Association for Computational Linguistics.
- Cigdem Toprak, Niklas Jakob, and Iryna Gurevych. 2010. [Sentence and expression level annotation of opinions in user-generated discourse](#). In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 575–584, Uppsala, Sweden. Association for Computational Linguistics.
- Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6000–6010.
- Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. 2008. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103.
- Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. [Relational graph attention network for aspect-based sentiment analysis](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3229–3238, Online. Association for Computational Linguistics.
- Shuai Wang, Sahisnu Mazumder, Bing Liu, Mianwei Zhou, and Yi Chang. 2018. [Target-sensitive memory networks for aspect sentiment classification](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 957–967, Melbourne, Australia. Association for Computational Linguistics.
- Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016a. [Recursive neural conditional random fields for aspect-based sentiment analysis](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 616–626, Austin, Texas. Association for Computational Linguistics.
- Xiao Wang, Qin Liu, Tao Gui, Qi Zhang, et al. 2021. [Textflint: Unified multilingual robustness evaluation toolkit for natural language processing](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 347–355, Online. Association for Computational Linguistics.
- Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016b. [Attention-based LSTM for aspect-level sentiment classification](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, Austin, Texas. Association for Computational Linguistics.
- Xiaoyu Xing, Zhijing Jin, Di Jin, Bingning Wang, Qi Zhang, and Xuanjing Huang. 2020. [Tasty burgers, soggy fries: Probing aspect robustness in aspect-based sentiment analysis](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3594–3605, Online. Association for Computational Linguistics.
- Hu Xu, Bing Liu, Lei Shu, and Philip Yu. 2019. [BERT post-training for review reading comprehension and aspect-based sentiment analysis](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2324–2335, Minneapolis, Minnesota. Association for Computational Linguistics.
- Chen Zhang, Qiuchi Li, and Dawei Song. 2019. [Aspect-based sentiment classification with aspect-specific graph convolutional networks](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4568–4578, Hong Kong, China. Association for Computational Linguistics.
- Mi Zhang and Tiejun Qian. 2020. [Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3540–3549, Online. Association for Computational Linguistics.