Enhancing Aspect-Based Sentiment Analysis with Supervised Contrastive Learning

Bin Liang∗
Wangda Luo∗
bin.liang@stu.hit.edu.cn
huowangda_hitsz@163.com
Harbin Institute of Technology
Shenzhen, China

Xiang Li
Harbin Institute of Technology,
Shenzhen, Joint lab of CMS-HITSZ
Shenzhen, China
xiangli@stu.hit.edu.cn

Min Yang
SIAT, Chinese Academy of Sciences
Shenzhen, China
min.yang@siat.ac.cn

Xiaoqi Yu
China Merchants Securities Co., Ltd.
Shenzhen, China
yuxiaoqi@cmschina.com.cn

Lin Gui
University of Warwick
Coventry, UK
lin.gui@warwick.ac.uk

Ruifeng Xu†
Harbin Institute of Technology,
Shenzhen, Peng Cheng Lab
xuruifeng@hit.edu.cn

ABSTRACT
Most existing aspect-based sentiment analysis (ABSA) research efforts are devoted to extracting the aspect-dependent sentiment features from the sentence towards the given aspect. However, it is observed that about 60% of the testing aspects in commonly used public datasets are unknown to the training set. That is, some sentiment features carry the same polarity regardless of the aspects they are associated with (aspect-invariant sentiment), which props up the high accuracy of existing ABSA models when inevitably inferring sentiment polarities for those unknown testing aspects. Therefore, in this paper, we revisit ABSA from a novel perspective by deploying a novel supervised contrastive learning framework to leverage the correlation and difference among different sentiment polarities and between different sentiment patterns (aspect-invariant/dependent). This allows improving the sentiment prediction for (unknown) testing aspects in the light of distinguishing the roles of valuable sentiment features. Experimental results on 5 benchmark datasets show that our proposed approach substantially outperforms state-of-the-art baselines in ABSA. We further extend existing neural network-based ABSA models with our proposed framework and achieve improved performance.

CCS CONCEPTS
- Information systems → Sentiment analysis; Clustering and classification.

KEYWORDS
aspect sentiment analysis, sentiment analysis, contrastive learning

† Corresponding Author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM Reference Format:

1 INTRODUCTION
Aspect-based sentiment analysis (ABSA) aims to detect the sentiment polarity (e.g. positive, negative, or neutral) towards a specific aspect from a sentence. For example, given a sentence “great food but dreadful service” and aspects “food” and “service”, the sentiment polarity of “food” is positive, while of aspect “service” is negative.

Recently, many deep learning methods achieve promising performance in ABSA task [14, 22, 27, 32, 35]. Such as attention-based models [1, 5, 7, 13, 28], graph network models [6, 12, 18, 21, 23, 26], and BERT-based models [3, 9, 19, 20, 30, 31], etc. Figure 1 shows the ratio of coverage of testing aspects in the training set on five commonly used ABSA datasets. Note that the best ratio is only 50.89%, which indicates that there is a considerable amount of testing aspects are unknown to the training set. This shows that the high accuracy of existing ABSA models is supported by some sentiment features carrying the same polarity regardless of the aspects they are associated with (aspect-invariant sentiment features).

Despite remarkable progress made in ABSA, existing research efforts largely focused on how to extract aspect-dependent contextual sentiment features for specific aspects. This leads to the ignorance of separating the representations between aspect-invariant
sentiment features and aspect-dependent ones, which potentially limit the sentiment learning for ABSA [11]. As shown in Figure 2, the sentiment expressions are quite different between different sentiment patterns (aspect-invariant/-dependent). In which, the aspect-invariant sentiment expression carries the same polarity regardless of the aspects they are associated with (Figure 2 (a)), while the aspect-dependent sentiment expression only takes effect on the specific aspect(s) (Figure 2 (b)). That is, aspect-invariant and aspect-dependent sentiment expressions play different roles in ABSA.

Motivated with these observations, we propose a novel BERT-based supervised contrastive learning (BERT-SCon) framework to leverage the correlation and difference among both different sentiment polarities and patterns. Specifically, in addition to the sentiment polarity, we construct augmentation data by supplying a sentiment pattern label (aspect-invariant/-dependent) for each instance. Then, we advance contrastive learning of both sentiment polarity and sentiment pattern to pull together representations belonging to the same class (sentiment polarity/pattern) in embedding space, while simultaneously push apart representations from different classes. This allows the discrimination of different sentiment expressions to be preferably leveraged to predict sentiment polarities for testing aspects, for those unknown aspects in particular.

The main contributions of our work can be summarized as follows:

- The ABSA task is approached from a new scenario that leveraging the roles of sentiment features to focus on the ABSA improvement for unknown testing aspects.
- A novel BERT-based supervised contrastive learning (BERT-SCon) framework is deployed to discriminate sentiment features from both sentiment polarity and pattern perspectives.
- Experimental results on 5 benchmark datasets show that the proposed framework achieves state-of-the-art performance.

## 2 PROPOSED APPROACH

As demonstrated in Figure 3, the architecture of the proposed BERT-SCon framework mainly contains four components:

1. **Data augmentation**, which derives the augmentation data by supplying a label of "aspect-invariant/-dependent" for each training instance.
2. **Feature extractor**, which encodes the input instance as hidden vectors.
3. **Sentiment classification**, which trains the cross-entropy loss $L^{cla}$ of the input sample following a classifier layer.
4. **Contrastive learning**, which trains the contrastive losses $L^D$ and $L^A$ from sentiment polarity and sentiment pattern perspectives.

### 2.1 Data Augmentation

Inspired by [11], in this work, we attempt to discriminate the aspect-invariant and aspect-dependent sentiment features by automatically supplying a sentiment pattern label (aspect-invariant/-dependent) for each training instance. Specifically, for an instance that has been adequately learned by a well-trained ABSA model, if we modify some aspect-related descriptors and still get the same prediction result with the well-trained model, then this instance potentially expresses aspect-invariant sentiment. By contrary, it is aspect-dependent.

We first train a BERT-based ABSA model $M$ with the training data $D$, here the training accuracy is close to 100% for each dataset. Then, we deploy two ways to construct augmentation data for each training instance. Assuming that there is a sentence consists of $n$ words $s_i = \{w_1, w_2, \ldots, a_i, \ldots, w_n\}$, where $w_j$ denotes the $i$-th contextual word, $a_i$ is the corresponding aspect, which consists of one or multiple words. 1) We replace the aspect $a_i$ with a special token "[MASK]" and derive a masked synthetic instance $s_i^m = \{w_1, w_2, \ldots, [MASK], \ldots, w_n\}$. 2) We randomly select another aspect $a_j$ to replace the aspect $a_i$ and derive an aspect-based synthetic instance $s_i^A = \{w_1, w_2, \ldots, a_j, \ldots, w_n\}$. Then $s_i^m$ and $s_i^A$ are fed as testing instances into $M$ to acquire the prediction results. If both the results are the same as the ground-truth label $y_i$, we supply a sentiment pattern label $z_i$ for the instance with "aspect-invariant", or else with "aspect-dependent".

### 2.2 Feature Extractor

For each instance $s_i$, we adopt pre-trained BERT [4] as the Feature Extractor to acquire a $d_m$-dimensional hidden vector $h_i \in \mathbb{R}^{d_m}$:

$$h_i = \text{BERT}([\text{CLS}]s_i[\text{SEP}]a_i[\text{SEP}])$$

(1)

For a mini-batch sample set $B$, the hidden vectors of the samples can be defined as: $B = \{h_i\}_{i=1}^{N_B}$, $N_B$ is the size of mini-batch.

### 2.3 Sentiment Classification

We feed the hidden vectors of the mini-batch $B$ into a classifier with a softmax function to produce the predicted sentiment distribution:

$$p_i = \text{softmax}(Wh_i + b)$$

(2)

where $p_i \in \mathbb{R}^{dp}$ is the predicted sentiment distribution of $h_i$, $d_p$ is the dimensionality of sentiment polarities, $W \in \mathbb{R}^{dp \times d_m}$ and $b \in \mathbb{R}^{dp}$ are trainable parameters.

Based on the predicted sentiment probability, we employ a cross-entropy loss between predicted and ground-truth distribution $y_i$ to train the classifier:

$$L^{cla} = -\sum_{i=1}^{N_B} dp \sum_{j=1}^{dp} y_{ij} \log p_{ij}$$

(3)

### 2.4 Contrastive Learning

Inspired by [2, 10], for each mini-batch $B$, we explore two contrastive losses to pull together the representations belonging to the same sentiment polarity or pattern, while simultaneously push apart samples from different sentiment polarities or patterns. For
We conduct experiments on five benchmark datasets from SemEval. We train the proposed framework by jointly minimizing the sum of the aforementioned three losses $L^{cla}$, $L^{p}$, and $L^{A}$:

$$L = \sum_{h_i \in B} t^p(h_i) + \lambda \left( \sum_{h_i \in B} t^a(h_i) \right)$$

where $\sum_{h_i \in B} t^p(h_i)$ denotes all trainable parameters of the framework, $\lambda$ is set to 0.00002. The coefficient of $L_2$-regularization $\lambda$ is set to 0.00001. Adam is utilized as the optimizer. The mini-batch is set to 16. For contrastive learning, the temperature parameter $\tau$ is set to 0.1. The reported results are averaged scores of 10 runs. Following previous ABSA works, we perform Accuracy metric to measure the performance of the models. For REST14, LAP14, REST15, and REST16, we randomly select 10% of the training set as the development data.

### 3.3 Main Experimental Results

Table 2 shows the experimental results on 5 benchmark datasets. Note that our proposed BERT-SCon consistently outperforms all compared baselines on all datasets, including the BERT-based models. To be specific, the best improved result is 5.35% compared with the vanilla BERT. This verifies that our proposed BERT-SCon, which deploying contrastive learning to discriminate aspect-invariant/dependent sentiment representations, is effective in ABSA task.

### 3.4 Analysis of Unknown Aspect-based Data

To investigate the performance of the testing instances whose aspects are unknown to the training set, we conduct experiments only testing on the unknown aspect-based set and report the results in Table 3. Note that the performance of unknown aspect-based data is considerably poorer than the overall performance on all datasets, which denotes that detecting sentiment for the unknown aspects is more challenging. Compared with the strong baseline models, our proposed BERT-SCon achieves significantly better performance.

---

### Table 1: Statistics of the experimental datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Positive</th>
<th>Neutral</th>
<th>Negative</th>
<th>Unk.</th>
</tr>
</thead>
<tbody>
<tr>
<td>REST14</td>
<td>2164</td>
<td>728</td>
<td>637</td>
<td>196</td>
</tr>
<tr>
<td>LAP14</td>
<td>994</td>
<td>34</td>
<td>464</td>
<td>169</td>
</tr>
<tr>
<td>REST15</td>
<td>912</td>
<td>326</td>
<td>36</td>
<td>34</td>
</tr>
<tr>
<td>REST16</td>
<td>1240</td>
<td>469</td>
<td>69</td>
<td>30</td>
</tr>
<tr>
<td>MAMS</td>
<td>3380</td>
<td>403</td>
<td>504</td>
<td>607</td>
</tr>
</tbody>
</table>

Unk. denotes the testing instances with unknown aspects.

### Table 2: Main experimental results. Best results are in bold face. Results with * are retrieved from the original papers, with † are retrieved from [8].

<table>
<thead>
<tr>
<th>Model</th>
<th>REST14</th>
<th>LAP14</th>
<th>REST15</th>
<th>REST16</th>
<th>MAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATAE-LSTM [29]</td>
<td>76.80*</td>
<td>68.88*</td>
<td>78.48*</td>
<td>83.77*</td>
<td>77.05*</td>
</tr>
<tr>
<td>MGAN [5]</td>
<td>81.25*</td>
<td>75.39*</td>
<td>79.36*</td>
<td>87.06*</td>
<td>75.98</td>
</tr>
<tr>
<td>CapsNet [8]</td>
<td>80.79*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>79.78*</td>
</tr>
<tr>
<td>ASGCN [33]</td>
<td>80.77*</td>
<td>75.55*</td>
<td>79.89*</td>
<td>88.99*</td>
<td>76.27</td>
</tr>
<tr>
<td>R-GAT [26]</td>
<td>83.30*</td>
<td>77.42*</td>
<td>80.57*</td>
<td>88.79*</td>
<td>77.50</td>
</tr>
<tr>
<td>BiGCN [34]</td>
<td>81.97*</td>
<td>74.59*</td>
<td>81.16*</td>
<td>88.96*</td>
<td>80.07</td>
</tr>
<tr>
<td>InterGCN [12]</td>
<td>82.23*</td>
<td>77.86*</td>
<td>81.76*</td>
<td>89.77*</td>
<td>79.25</td>
</tr>
<tr>
<td>BERT [4]</td>
<td>84.11*</td>
<td>77.59*</td>
<td>83.48*</td>
<td>90.10*</td>
<td>82.22*</td>
</tr>
<tr>
<td>CapsNet-BERT [8]</td>
<td>85.93*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>83.39*</td>
</tr>
<tr>
<td>R-GAT-BERT [26]</td>
<td>86.60*</td>
<td>78.21*</td>
<td>83.27*</td>
<td>89.91*</td>
<td>82.07</td>
</tr>
<tr>
<td>BERT-KVMN [24]</td>
<td>85.98*</td>
<td>79.78*</td>
<td>84.14*</td>
<td>90.52*</td>
<td>-</td>
</tr>
<tr>
<td>Ours BERT-SCon</td>
<td>87.62</td>
<td>82.94</td>
<td>85.42</td>
<td>92.53</td>
<td>85.78</td>
</tr>
</tbody>
</table>

### Table 3: Experimental results of unknown aspect data.

<table>
<thead>
<tr>
<th>Model</th>
<th>REST14</th>
<th>LAP14</th>
<th>REST15</th>
<th>REST16</th>
<th>MAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD-LSTM [22]</td>
<td>73.25</td>
<td>60.37</td>
<td>74.79</td>
<td>80.70</td>
<td>61.08</td>
</tr>
<tr>
<td>MGAN [5]</td>
<td>75.28</td>
<td>61.34</td>
<td>76.32</td>
<td>81.21</td>
<td>60.23</td>
</tr>
<tr>
<td>InterGCN [12]</td>
<td>75.99</td>
<td>67.78</td>
<td>79.70</td>
<td>88.09</td>
<td>65.71</td>
</tr>
<tr>
<td>BERT [4]</td>
<td>77.98</td>
<td>71.05</td>
<td>79.78</td>
<td>84.14</td>
<td>68.66</td>
</tr>
<tr>
<td>BERT-SCon</td>
<td>82.26*</td>
<td>78.88</td>
<td>84.09*</td>
<td>92.60*</td>
<td>77.62</td>
</tr>
</tbody>
</table>

* denotes without sentiment pattern contrastive learning. ** denotes without sentiment polarity contrastive learning. 

### Comparison Models

We compare our proposed BERT-SCon with various ABSA models. Including 1) attention-based models: ATAE-LSTM [29], MGAN [5], and CapsNet [8]; 2) graph-based models: ASGCN [33], R-GAT [26], BiGCN [34], and InterGCN [12]; 3) BERT-based models: BERT [4], CapsNet-BERT [8], R-GAT-BERT [26], and BERT-KVMN [24].

We also design several variants of our proposed BERT-SCon in the ablation study. BERT-Multi” denotes regarding sentiment polarities and patterns classification as a pure multi-task learning. w/o $L^{p}$ denotes without sentiment polarity contrastive learning. w/o $L^{A}$ denotes without sentiment pattern contrastive learning.
Table 4: Experimental results of ablation study.

<table>
<thead>
<tr>
<th>Model</th>
<th>Rest14</th>
<th>Lap14</th>
<th>Rest15</th>
<th>Rest16</th>
<th>MAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT-Multi</td>
<td>83.96</td>
<td>77.62</td>
<td>82.81</td>
<td>90.57</td>
<td>82.53</td>
</tr>
<tr>
<td>w/o $L^A$</td>
<td>85.39</td>
<td>79.10</td>
<td>84.10</td>
<td>91.17</td>
<td>83.61</td>
</tr>
<tr>
<td>w/o $L^A$</td>
<td>85.07</td>
<td>79.22</td>
<td>83.93</td>
<td>91.05</td>
<td>83.45</td>
</tr>
</tbody>
</table>

Table 5: Comparison results of combining our proposed contrastive learning framework with different ABSA models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Rest14</th>
<th>Lap14</th>
<th>Rest15</th>
<th>Rest16</th>
<th>MAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD-LSTM [22]</td>
<td>78.00</td>
<td>71.83</td>
<td>76.39</td>
<td>82.16</td>
<td>74.70</td>
</tr>
<tr>
<td>TD-LSTM-SCon (ours)</td>
<td>79.75</td>
<td>72.63</td>
<td>78.04</td>
<td>85.71</td>
<td>76.18</td>
</tr>
<tr>
<td>MGAN [5]</td>
<td>81.25</td>
<td>75.39</td>
<td>79.36</td>
<td>87.06</td>
<td>75.98</td>
</tr>
<tr>
<td>MGAN-SCon (ours)</td>
<td>82.96</td>
<td>77.85</td>
<td>81.75</td>
<td>89.42</td>
<td>78.33</td>
</tr>
<tr>
<td>InterGCN [12]</td>
<td>82.25</td>
<td>77.86</td>
<td>81.76</td>
<td>89.77</td>
<td>79.25</td>
</tr>
<tr>
<td>InterGCN-SCon (ours)</td>
<td>83.25</td>
<td>79.33</td>
<td>83.05</td>
<td>91.07</td>
<td>81.22</td>
</tr>
</tbody>
</table>

Figure 4: Visualization of intermediate vector representations. Blue=Positive, green=Neutral, red=Negative.

3.7 Visualization

To qualitatively demonstrate how contrastive learning enhances the features learning of ABSA, we show the t-SNE [25] visualization of intermediate vectors learned by BERT and our proposed BERT-SCon in Figure 4 on test set of Rest14 dataset. We can observe that the separations of representations derived from BERT-SCon are significantly clearer than that learned by the vanilla BERT among different sentiment polarities. This verifies that the proposed contrastive learning scenario can derive more definite correlation and clearer difference of representations among different sentiment polarities, so as to improve the performance of ABSA.

3.8 Case Study

To investigate how the proposed BERT-SCon works in improving ABSA performance by discriminating the aspect-invariant sentiment representations in the light of contrastive learning, we look into the Cosine Similarities of an unknown aspect-based testing instance "The teas are great," paired with different kinds of training instances in Figure 5. Note that our proposed contrastive learning framework allows the representation of the testing instance to be more similar as the relevant aspect-invariant training instance "Great food," so as to capture a correct prediction result.

4 CONCLUSION

In this paper, we propose to deploy a novel contrastive learning framework for ABSA based on BERT, called BERT-SCon, in which the aspect-invariant and aspect-dependent sentiment representations can be discriminated to improve the sentiment learning for the testing instances, for those testing aspects are unknown to the training set in particular. Experimental results on 5 benchmark datasets show that the proposed BERT-SCon achieves state-of-the-art performance in ABSA.

ACKNOWLEDGMENTS

This work was partially supported by the National Natural Science Foundation of China (61632011, 61876053, 62006062), the Guangdong Province Covid-19 Pandemic Control Research Funding (2020KZDZX1224), the Shenzhen Foundational Research Funding (JCYJ20180507183527919 and JCYJ20200109113441941), China Postdoctoral Science Foundation (2020M670912), Joint Lab of HITSZ and China Merchants Securities, the UK EPSRC (grant no. EP/T017112/1, EP/V048597/1), Youth Innovation Promotion Association of CAS China (No. 2020357), and Shenzhen Science and Technology Innovation Program (Grant No. KQTD20190929172835662).

