# Enhancing Aspect-Based Sentiment Analysis with Supervised Contrastive Learning

Bin Liang[1*], Wangda Luo[1*], Xiang Li[1], Lin Gui[2], Min Yang[3], Xiaoqi Yu[4], Ruifeng Xu[1†]
[1]School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), China
[2]Department of Computer Science, University of Warwick, Coventry, U.K
[3]SIAT, Chinese Academy of Sciences, [4]China Merchants Securities Co ., Ltd. Shenzhen, China

@HITSZ

CIKM 2021 1-5 NOVEMBER

## INTRODUCTION

**Problem Definition:** 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 polarity of "*food*" is *positive*, while of aspect "*service*" is *negative*.

**Main Challenge:**

- There are about 60% of the testing aspects in commonly used public datasets are unknown to the training set.
- Some sentiment features carry the same polarity regardless of the aspects, which props up the high accuracy of existing ABSA models.

## MOTIVATION & CONTRIBUTION

**Motivation:** In Example (a), the aspect-invariant sentiment expression carries the same polarity regardless of the aspects. In Example (b), the aspect-dependent sentiment expression only takes effect on the specific aspect. So 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.

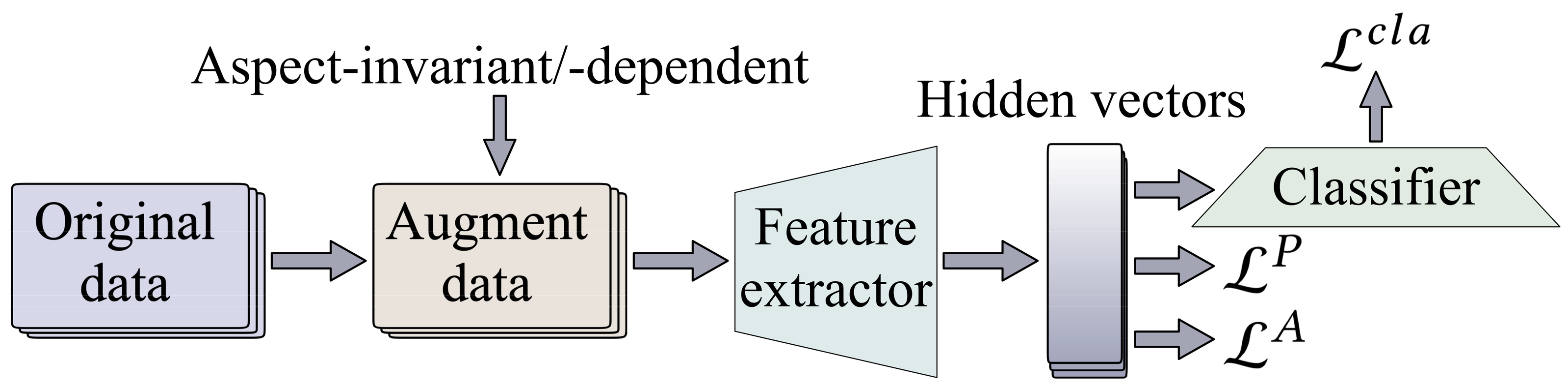

(a) aspect-invariant example (b) aspect-dependent example

**Key Contributions:**

- 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.

## Method

**Framework Architecture of Supervised Contrastive Learning(BERT-SCon)**




The architecture of the proposed **BERT-SCon** framework mainly contains four components: 1) *Data augmentation*, 2) *Feature extractor*, 3) *Sentiment classification*, 4) *Contrastive learning*.

**Data Augmentation:**
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. We first train a BERT-based ABSA model $\mathcal{M}$ with the training data $\mathcal{D}$, and the training accuracy is close to 100% for each dataset. Then, we deploy two ways to construct augmentation data for each training instance.1) We replace the aspect with a special token "$[MASK]$" and derive a masked synthetic instance $s_i^M$. 2)We randomly select another aspect to replace the aspect and derive an aspect-based synthetic instance $s_i^A$. Then $s_i^M$ and $s_i^A$ are fed as testing instances into $\mathcal{M}$ to acquire the prediction results. If both the results are the same as the ground-truth label, we supply a sentiment pattern label for the instance with "*aspect-invariant*", or else with "*aspect-dependent*".

**Feature Extractor:**
For each instance $s_i$, we adopt pre-trained BERT as the Feature Extractor to acquire a $d_m$-dimensional hidden vector $\boldsymbol{h}_i \in \mathbb{R}^{d_m}$:

$$\boldsymbol{h}_i = \mathrm{BERT}([CLS]s_i[SEP]a_i[SEP]) \tag{1}$$

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

**Sentiment Classification:**
We feed the hidden vectors of the mini-batch $\mathcal{B}$ into a classifier with a softmax function to produce the predicted sentiment distribution:

$$\boldsymbol{p}_i = \mathrm{softmax}(\boldsymbol{W}\boldsymbol{h}_i + \boldsymbol{b}) \tag{2}$$

Based on the predicted sentiment probability, we employ a cross-entropy loss between predicted and ground-truth distribution $\boldsymbol{y}_i$ to train the classifier:

$$\mathcal{L}^{cla} = -\sum_{i=1}^{N_b}\sum_{j=1}^{d_p} y_i^j \log p_i^j \tag{3}$$

**Contrastive Learning:**
For an *anchor* $\boldsymbol{h}_i$, we refer to 1) $\boldsymbol{h}_i, \boldsymbol{h}_j \in \mathcal{B}$ with the same sentiment polarity as a *positive polarity* pair, 2) $\boldsymbol{h}_i, \boldsymbol{h}_j \in \mathcal{B}$ with the same sentiment pattern as a *positive pattern* pair. While the samples $\{\boldsymbol{h}_k \in \mathcal{B}, k \neq i\}$ are treated as *negative* instances towards this *anchor*. The contrastive losses $\mathcal{L}^P$ and $\mathcal{L}^A$ are as follow:

$$\mathcal{L}^P = \frac{1}{N_b}\sum_{\boldsymbol{h}_i \in \mathcal{B}} \ell^p(\boldsymbol{h}_i), \qquad \mathcal{L}^A = \frac{1}{N_b}\sum_{\boldsymbol{h}_i \in \mathcal{B}} \ell^a(\boldsymbol{h}_i) \tag{4}$$

$$\ell^p(\boldsymbol{h}_i) = -\log \frac{\sum_{j=1}^{N_b} \mathbb{1}_{[j\neq i]}\mathbb{1}_{[y_i=y_j]}\exp(\mathrm{sim}(\boldsymbol{h}_i,\boldsymbol{h}_j)/\tau)}{\sum_{k=1}^{N_b}\mathbb{1}_{[k\neq i]}\exp(\mathrm{sim}(\boldsymbol{h}_i,\boldsymbol{h}_k)/\tau)} \tag{5}$$

$$\ell^a(\boldsymbol{h}_i) = -\log \frac{\sum_{j=1}^{N_b} \mathbb{1}_{[j\neq i]}\mathbb{1}_{[z_i=z_j]}\exp(\mathrm{sim}(\boldsymbol{h}_i,\boldsymbol{h}_j)/\tau)}{\sum_{k=1}^{N_b}\mathbb{1}_{[k\neq i]}\exp(\mathrm{sim}(\boldsymbol{h}_i,\boldsymbol{h}_k)/\tau)} \tag{6}$$

**Learning Objective:**
We train the proposed framework by jointly minimizing the sum of the aforementioned three losses $\mathcal{L}^{cla}$, $\mathcal{L}^P$ and $\mathcal{L}^A$:

$$\mathcal{L} = \mathcal{L}^{cla} + \mathcal{L}^P + \mathcal{L}^A + \lambda||\Theta||^2 \tag{7}$$

$\Theta$ denotes all trainable parameters of the framework, $\lambda$ represents the coefficient of $L_2$-regularization.

## Experimental Results & Conclusion

**Main Experimental Results:**

| | Model | REST14 | LAP14 | REST15 | REST16 | MAMS |
|---|---|---|---|---|---|---|
| Attention | ATAE-LSTM | 78.60♮ | 68.88♮ | 78.48♮ | 83.77♮ | 77.05† |
| | MGAN | 81.25♮ | 75.39♮ | 79.36♮ | 87.06♮ | 75.98 |
| | CapsNet | 80.79⋆ | - | - | - | 79.78⋆ |
| Graph | ASGCN | 80.77⋆ | 75.55⋆ | 79.89⋆ | 88.99⋆ | 76.27 |
| | R-GAT | 83.30⋆ | 77.42⋆ | 80.57 | 88.79 | 77.50 |
| | BiGCN | 81.97⋆ | 74.59⋆ | 81.16⋆ | 88.96⋆ | 80.07 |
| | InterGC | 82.23⋆ | 77.86⋆ | 81.76⋆ | 89.77⋆ | 79.25 |
| BERT | BERT | 84.11♮ | 77.59♮ | 83.48♮ | 90.10♮ | 82.22† |
| | CapsNet-BERT | 85.93⋆ | - | - | - | 83.39⋆ |
| | R-GAT+BERT | 86.60⋆ | 78.21⋆ | 83.27 | 89.91 | 82.07 |
| | BERT-KVMN | 85.98⋆ | 79.78⋆ | 84.14⋆ | 90.52⋆ | - |
| Ours | BERT-SCon | **87.62** | **82.94** | **85.42** | **92.53** | **85.78** |

**Experimental Results of Model Generalizability:**

| Model | REST14 | LAP14 | REST15 | REST16 | MAMS |
|---|---|---|---|---|---|
| TD-LSTM | 78.00 | 71.83 | 76.39 | 82.16 | 74.60 |
| TD-LSTM-SCon (**ours**) | **79.75** | **72.63** | **78.04** | **85.71** | **76.18** |
| MGAN | 81.25 | 75.39 | 79.36 | 87.06 | 75.98 |
| MGAN-SCon (**ours**) | **82.96** | **77.85** | **81.75** | **89.42** | **78.33** |
| InterGCN | 82.23 | 77.86 | 81.76 | 89.77 | 79.25 |
| InterGCN-SCon (**ours**) | **83.25** | **79.33** | **83.05** | **91.07** | **81.22** |

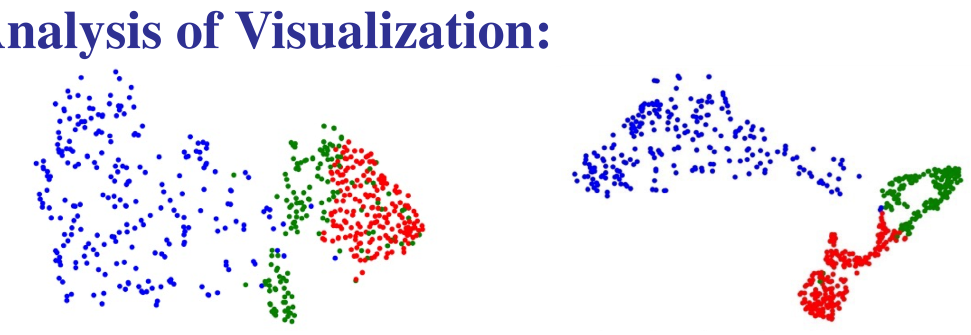
**Experimental Results of Ablation Study:**

| Model | REST14 | LAP14 | REST15 | REST16 | MAMS |
|---|---|---|---|---|---|
| BERT-SCon | **87.62** | **82.74** | **85.42** | **92.53** | **84.78** |
| BERT-Multi | 83.96 | 77.62 | 82.81 | 90.57 | 82.53 |
| w/o $\mathcal{L}^P$ | 85.39 | 79.10 | 84.10 | 91.17 | 83.61 |
| w/o $\mathcal{L}^A$ | 85.07 | 79.22 | 83.93 | 91.05 | 83.45 |

**Experimental Results of Unknown Aspect-based Data:**

| Model | REST14 | LAP14 | REST15 | REST16 | MAMS |
|---|---|---|---|---|---|
| TD-LSTM | 73.25 | 60.57 | 74.79 | 80.70 | 61.08 |
| MGAN | 75.28 | 61.34 | 76.32 | 81.21 | 60.23 |
| InterGCN | 75.99 | 67.78 | 79.70 | 88.09 | 65.71 |
| BERT | 77.98 | 71.05 | 81.33 | 88.31 | 68.66 |
| BERT-SCon | **82.26** | **78.98** | **84.09** | **92.40** | **77.82** |

**Experimental Analysis of Visualization:**



(a) BERT (b) BERT-SCon

**Conclusion:**
We propose to deploy a novel contrastive learning framework **BERT-SCon**, in which the aspect-invariant and aspect-dependent sentiment representations can be discriminated to improve the sentiment learning for the testing instances.