Incorporating Dynamic Semantics into Pre-Trained Language Model for Aspect-based Sentiment Analysis

Kai Zhang¹, Kun Zhang², Mengdi Zhang³, Hongke Zhao⁴, Qi Liu⁴,* Wei Wu³, Enhong Chen¹

¹ Anhui Province Key Lab of Big Data Analysis and Application, School of Data Science, USTC
² School of Computer Science and Information Engineering, Hefei University of Technology
³ Meituan; ⁴ College of Management and Economics, Tianjin University
Aspect-based Sentiment Analysis?

• E-commerce platforms (e.g., Amazon, Alibaba) allow users to post their reviews towards products. The reviews may contain the finer-level opinions of users.

• Value of the ABSA:
  • User-generated content analysis
  • Help for making a purchase decision
  • Assess the impact of marketing campaigns, success of product launches

---

The color of the PC looks good, but the battery life is too short.
Outline

1. Background
   a. Formal definition & previous methods
   b. Some existing problems
2. Our method: Dynamic Re-weighting BERT
3. Experiments
4. Conclusion
Outline

1. Background
   a. Formal definition & previous methods
   b. Some existing problems
2. Our method: Sentiment-aware Interactive Fusion Network
3. Experiments
4. Conclusion
Background – **formal definition**

To begin with, we first define the Aspect-based Sentiment Analysis. It predicts the sentiment tendency that a user expressed at the whole reviews based on the given aspect term.

predicts the sentiment label $y_{<a,i>}$ of given aspect $a$ in the review sentence $i$.

General schema: $y_{<a,i>} = f(\text{aspect}, \text{reviews})$

- CNN / RNN / LSTM
- Attention methods
- Graph network
- Pre-trained models

Aspect & Context relations
- Aspect-based syntax info
- ...

---

5
Background – previous methods

**CNN, RNN, Attention based methods**

- **TD-LSTM** [Tang D et al. COLING’2016]
  - Contains LSTM\(_L\) and LSTM\(_R\)
  - Utilizing Bi-LSTM to extract aspect-aware features along with the context words

- **Attention model: ATAE-LSTM** [Wang Y et al. EMNLP’2016]
  - Concatenate with each word as input
  - Concatenate with each hidden state
  - Attention mechanism to learn important part of a sentence
Background – previous methods

Pre-trained Models with Graph for ABSA

✓ R-GAT [Wang et al. ACL’2020]
  • Not all words contribute the same point
  • The purpose is to better model neighbor words

✓ T-GCN [Tian et al. NAACL’21]
  • Type-aware graph convolutional networks to comprehensively learn from dependency parsing relations

✓ BERT-PT [Xu et al. NAACL’19], BERT-SPC [Song et al. 2019], AEEN-BERT [Song et al. 2019]
Background – shortcomings

Although these works achieved significant performance improvement, they still suffer from **two intrinsic issues:**

1. largely ignoring the **dynamic semantic understanding** during sentiment mining
2. neglect to interpret the sentiment mining process intuitively

✓ directly mining the semantic information may lead to **sub-optimal sentiment prediction** because, according to neuroscience studies, the essential words during semantic comprehension are dynamically changing with the reading process and should be repeatedly considered!
Outline

1. Background
   a. Formal definition & previous methods
   b. Some existing problems
2. Our method: Dynamic Re-weighting BERT
3. Experiments
4. Conclusion
Our solution – **Dynamic Re-weighting BERT**

**Motivation:** exploring the dynamic semantic in **PLMs**:

1. The pre-trained models do not effectively boost the performance in ABSA as we expected (i.e., they prefer to mine the sentiment of the *whole sentence*)
2. It is vital to model dynamic semantics such as Human Semantic Comprehension process
Our solution – **Dynamic Re-weighting BERT**

Overall architecture of **DR-BERT**:

1. A BERT Encoder (BERT embedding Layers)
2. Dynamic Re-weighting Adapter (Re-weighting Attention + GRU)
3. Sentiment Prediction (MLP + Softmax)
Our solution – Dynamic Re-weighting BERT

Overall architecture of DR-BERT:

1. A BERT Encoder:
   - $m = \{m_i \mid i = 1,2,\ldots,l_s\}$
   - $f = \{f_i \mid i = 1,2,\ldots,l_s\}$
   - $m = \text{MultiHead}(sW_Q^m, sW^K_m, sW^V_m)$
   - $f = \max(0, mW_1 + b_1)W_2 + b_2$

2. Dynamic Re-weighting Adapter:
   - $s = [s_1, s_2,\ldots, s_l_s]$
   - $M = W_s s + (W_a h_{t-1} + W_o a) \otimes w$
   - $m = \omega^T \tanh(M)$
   - $z_t = \sigma(W_z \cdot [h_{t-1}, a_t])$
   - $r_t = \sigma(W_r \cdot [h_{t-1}, a_t])$
   - $\tilde{h}_t = \tanh(W \cdot [r_t \ast h_{t-1}, a_t])$
   - $h_t = (1 - z_t) \ast h_{t-1} + z_t \ast \tilde{h}_t$

3. Sentiment Prediction:
   - $R_t = \text{Relu}(W_t R_{t-1} + b_t)$
   - $\hat{y} = \text{softmax}(W_o R_h + b_o)$
Outline

1. Background
   a. Formal definition & previous methods
   b. Some existing problems
2. Our method: Dynamic Re-weighting BERT
3. Experiments
4. Conclusion
Experiments

Dataset

✓ Benchmark Amazon & Twitter dataset:

  • Laptop
  • Restaurant
  • Twitter

Baseline methods

① Attention-based methods: ATAE-LSTM, IAN, AOA, TNet ...

② Pre-trained methods: BERT, BERT-PT, BERT-SPC, RGAT-BERT, T-GCN ...

Accuracy and F1-score as the evaluation metric!
## Experiments

### Overall performance

1. **BERT-based** methods beat most of the **attention-based** methods (e.g., IAN, TNet)
2. The **task-specific BERT models** perform better than the **non-specific models** (T-GCN & RGAT-BERT) > AEN-BERT > BERT-PT > BERT

| Category   | Methods                        | Datasets          | Laptop (Accuracy, F1-score) | Restaurant (Accuracy, F1-score) | Twitter (Accuracy, F1-score) |
|------------|--------------------------------|-------------------|-----------------------------|---------------------------------|------------------------------|
| Attention  | ATAE-LSTM (Wang et al., 2016)  |                   | 68.57, 64.52                | 76.58, 67.39                    | 67.27, 66.43                 |
|            | IAN (Ma et al., 2017)          |                   | 70.84, 65.73                | 76.88, 68.36                    | 68.74, 67.61                 |
|            | MemNet (Tang et al., 2016)     |                   | 72.32, 67.03                | 78.12, 68.99                    | 70.19, 68.22                 |
|            | AOA (Huang et al., 2018)       |                   | 74.56, 68.77                | 79.42, 70.43                    | 71.68, 69.25                 |
|            | MGNNet (Fan et al., 2018)      |                   | 75.37, 71.26                | 81.28, 72.07                    | 72.54, 70.78                 |
|            | TNet (Li et al., 2018)         |                   | 76.54, 71.75                | 80.69, 71.27                    | 74.93, 73.60                 |
| Pre-trained| BERT (Devlin et al., 2019)     |                   | 77.29, 73.36                | 82.40, 73.17                    | 73.42, 72.17                 |
|            | BERT-PT (Xu et al., 2019a)     |                   | 78.07, 75.08                | 84.95, 76.96                    | -                            |
|            | BERT-SPC (Song et al., 2019)   |                   | 78.99, 75.03                | 84.46, 76.98                    | 74.13, 72.73                 |
|            | AEN-BERT (Song et al., 2019)   |                   | 79.93, 76.31                | 83.12, 73.76                    | 74.71, 73.13                 |
|            | RGAT-BERT (Wang et al., 2020)  |                   | 78.21, 74.07                | 86.60, 81.35                    | 76.15, 74.88                 |
|            | T-GCN (Tian et al., 2021)      |                   | 80.88, 77.03                | 86.16, 79.95                    | 76.45, 75.25                 |
| **Ours.**  | DR-BERT                        |                   | **81.45, 78.16**            | **87.72, 82.31**                | **77.24, 76.10**             |

**DR-BERT** outperforms the most advanced baseline (i.e., T-GCN or RGAT-BERT) no matter in terms of Accuracy or F1-score.
Experiments

| Model Variants                  | Laptop          |
|---------------------------------|-----------------|
|                                 | Accuracy  | F1-score |
| BERT-Base                       | 77.29     | 73.36    |
| (1): + MLP                      | 77.94     | 74.42    |
| (2): + DRA                      | 80.66     | 77.13    |
| (3): + DRA on top 3 layers      | 78.64     | 75.16    |
| (4): + DRA on top 6 layers      | 79.17     | 75.93    |
| (5): + DRA on top 9 layers      | 80.22     | 76.49    |
| (6): DR-BERT                    | **81.45** | **78.16** |

**Ablations on the Proposed Components**

- without utilizing adapters and MLPs, DR-BERT degenerates into the BERT model, which gains the worst performance
- we can easily conclude that DRA plays a more crucial role in the final sentiment prediction than MLPs
- the DRA is efficient in encoding the aspect-aware semantics over the whole sentence

**Ablations on the Scale of Adapter**

- the performance of DR-BERT first becomes better with the increasing of re-weighting length and achieving the best result at 7
- after that, as the length continues to increase, the performance continues to decline
- Our proposed model can achieve better performance with only 4 or 5 times of re-weighting at most test sets
## Experiments

| Sentence | Prediction |
|----------|------------|
| It could be a perfect laptop if it would have faster system memory and its radeon would have DDR5 instead of DDR3. | Negative |

Case study

1. The chosen sentence has three different aspects with their sentiment polarity, i.e., “System memory”-negative, “DDR5”-positive and “DDR3”-negative
2. The model tends to associate “DDR5” with the context words {“would”, “have”, “instead”} to predict the correct sentiment “positive”
## Experiments

### Case Examples

The label in brackets represents ground truth.

| Aspects: “system memory” (Neg.), “DDR5” (Pos.), “DDR3” (Neg.) | Sentence: It could be a perfect laptop if it would have faster system memory and its radeon would have DDR5 instead of DDR3. | BERT-base | RGAT-BERT | DR-BERT |
| --- | --- | --- | --- | --- |
| Pos/Neg/Neg | Neg/Pos/Pos | Neg/Pos/Neg |
| X | ✓ | ✓ | ✓ |

| Aspects: “Supplied software” (Neu.), “software” (Pos.), “Windows” (Neg.) | Sentence: Supplied software: The software that comes with this machine is greatly welcomed compared to what Windows comes with. | BERT-base | RGAT-BERT | DR-BERT |
| --- | --- | --- | --- | --- |
| Pos/Pos/Pos | Pos/Pos/Neg | Pos/Pos/Neg |
| X | ✓ | ✓ | ✓ |

| Aspects: “waiter” (Neg.), “served” (Neg.), “specials” (Pos.) | Sentence: First, the waiter who served us neglected to fill us in on the specials, which I would have chosen had I known about them. | BERT-base | RGAT-BERT | DR-BERT |
| --- | --- | --- | --- | --- |
| Neg/Neg/Neg | Neg/Neg/Neg | Neg/Neg/Pos |
| ✓ | ✓ | ✓ |

## Error Analysis

1. The vanilla BERT often makes the wrong classification since it tends to learn the overall sentiment polarity of the sentences instead of the aspect-aware semantic.

2. RGAT-BERT can alleviate the problem to a certain extent via few dependency relations.

3. Our DR-BERT model, succeeding in predicting most sentiment labels by considering the dynamic changing of the aspect-aware semantic.
Outline

1. Background
   a. Formal definition & previous methods
   b. Some existing problems
2. Our method: Dynamic Re-weighting BERT
3. Experiments
4. Conclusion
Conclusion

1. We highlight the aspect-aware semantic learning in the ABSA task, and inspired by human cognitive processes, we focus on modeling the dynamic semantics.

2. We propose a novel Dynamic Re-weighting BERT (DR-BERT) model with a new Dynamic Re-weighting Adapter (DRA) to enhance aspect-aware semantic features.

3. We conduct extensive experiments on three datasets and the results demonstrate the effectiveness and interpretability of our proposed method.
References

[1] Xiangsheng Li, Jiaxin Mao, Chao Wang, Yiqun Liu, Min Zhang, and Shaoping Ma. 2019. Teach machine how to read: reading behavior inspired relevance estimation. In Proceedings of SIGIR’2019, pages 795–804.

[2] Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. In Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI’2017), pages 4068–4074.

[3] Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018. Transformation networks for target-oriented sentiment classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL’2018), pages 946–956.

[4] Yuanhe Tian, Guimin Chen, and Yan Song. 2021. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In Proceedings of NAACL’2021, pages 2910–2922.

[5] Hu Xu, Bing Liu, Lei Shu, and S Yu Philip. 2019a. Bert post-training for review reading comprehension and aspect-based sentiment analysis. In Proceedings of NAACL’2019, pages 2324–2335.

[6] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based lstm for aspect-level sentiment classification. In Proceedings of EMNLP’2016, pages 606–615.

[7] Kai Wang, Weizhou Shen, Yunyi Yang and Xiaojun Quan. 2020. Relational graph attention network for aspect-based sentiment analysis. In Proceedings of ACL’2020, pages 3229–3238.
Thanks!