PGCD: A POSITION-GUIDED CONTRIBUTIVE DISTRIBUTION UNIT FOR ASPECT BASED SENTIMENT ANALYSIS

Zijian Zhang†, Chenxin Zhang, Qin Liu, Hongming Zhu, Jiangfeng Li†

Tongji University, Shanghai, China
{zijian96, lijf}@tongji.edu.cn

ABSTRACT

Aspect based sentiment analysis (ABSA) aims to explore the sentiment polarity towards sentence by giving aspect words. Previously models or methods typically gain an aspect-independent feature by sentence representation generation only depending on text level data. In this paper, we propose a novel Position-Guided Contributive Distribution unit, named PGCD unit, which obtains a position-dependent regulation from historical aspect-given texts. When the position of aspect words is determined, that regulation would be the contributive distribution of other positional words, which is useful to language understanding. By doing so, PGCD unit can enhance the generation of aspect-based sentence representation, and thus conducts more accurately. This paper attempts to evaluate PGCD unit on both text level and text-audio level ABSA tasks. Experimental results conclude the effectiveness of our proposed unit, which significantly outperforms the state-of-the-art models under the same settings (accuracy increase 2.38%).

***Index Terms*—** Position-Guided Contributive Contribution Aspect-based Sentiment Analysis

1. INTRODUCTION

Aspect Based Sentiment Analysis (ABSA) is one of the most important and challenging tasks in human language processing[1] . It aims to judge correct sentiment polarity of targets or aspects given sentence, which is different from traditional sentiment analysis tasks. Some human language experts show that most of sentiment analysis errors are caused by ignoring the aspect information in the original text, other than the choice of downstream model (like RNN, BERT, etc.). Thus, generate aspect-related feature for classification model is the kernel of ABSA even language technology. The Position-Guided Contributive Distribution (PGCD) proposed by this paper is a novel unit to enhance sentence feature whose aspect terms are given. The significant application of PGCD unit is based on the assumption that *aspect terms in the specific location are affected by other regular positional words*[1].

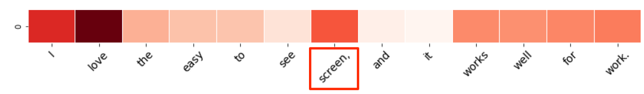

Fig. 1. Example of Positional-Guided Contributive Distribution

Traditional methods include both Non-neural methods and neural models. The former usually use low-dimensional and dense vectors to implicitly represent the syntactic or semantic features, which usually has acceptable explanation but poor performance[3] . In recent years, the latter achieve much success in many NLP tasks, especially ABSA task. On the one hand, some adjust the structure of the model to gain better predictions, such as Convolutional Neural Network (CNN)[4] and Recursive Neural Networks (RNN)[5] . However, the above model can only understand data directionally and achieve less significant compared to the computer vision (CV) field. With the development of computing power and training skills, Pre-trained Language Models (PLMs)[6] advanced the state of the art. On the other hand, feature-based method is used to obtain sentence features with more abundant aspect-based information, the most influential mechanism is the Attention. Both feature-distance-based and historical-records-based attention can only capture low-sensitive of distance aware, which is disaster in NLP tasks[2]. Although positional embedding is proposed to relief this problem, it can only mark which positions are different rather than find the essential feature difference[7] .

In this work, we propose a novel PGCD unit (code can be seen here[3]) to cultivate aspect-based information as well as improve the efficiency of several models. The major difference between PGCD unit and other feature-based methods is that PGCD displays the position of aspect terms and counts the contribution of other places from the historical data based on Shapley Value[8] . Figure 1 shows an example of PGCD output, the contribution of different locational words indicated by color depth (varied value). Specifically, we split sentence into 3 parts: left context, aspect term and right context. Then the position of aspect terms can only be decided by position. Convert above three sub-sentences into vectors by pre-trained word embedding. Afterwards, consider

[1] "I love the easy to see *screen*, and it works well for work." is an example in SemEval2014 datasets. The *screen* is the given aspect term, and the underlined part is called regulation, which can assist in judging sentiment polarity. Quantifying this regulation as a priori knowledge, which is the Contribution Distribution proposed in this paper.

[2] *Low-sensitive distance aware* generally refers to the words order through sentence can not be recognized and modeled in NLP field.

[3] https://github.com/96-Zachary/PGCD-for-ABSA

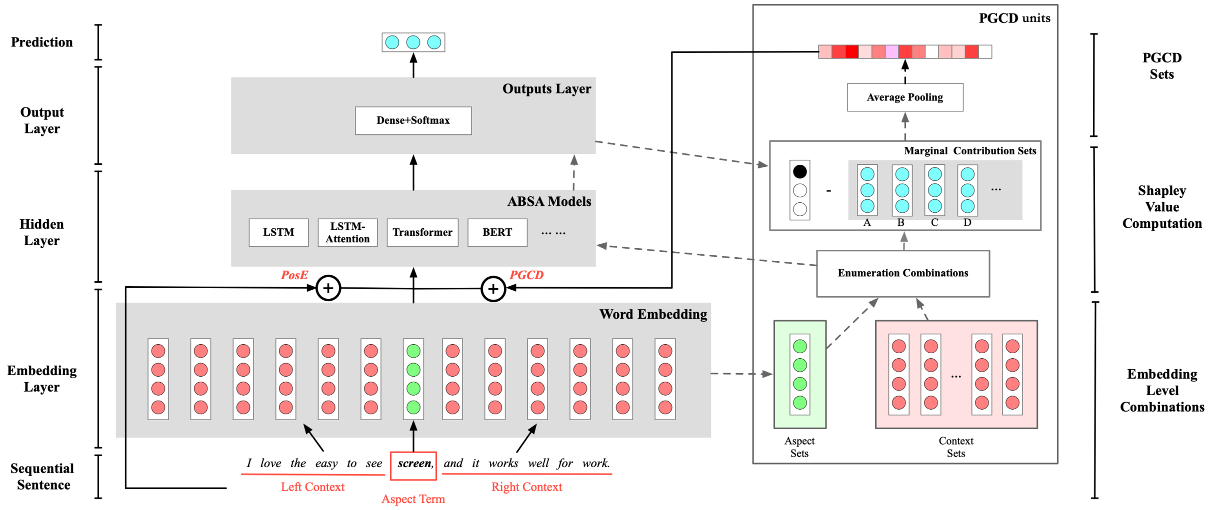

Fig. 2. Structure and Usage of proposed unit —— Positional-Guided Contributive Distribution (PGCD) units

contextual combinations containing aspect term as candidate sets, the marginal contribution of context parts in each combination is calculated. Considering all combinations, the average pooling of marginal contribution will be taken as the output result of PGCD, which can reflect the contribution of other positions to the final predictions. The contributions of our paper are as follows:

- We proposed a PGCD unit to enrich and extend sentence feature. This unit can generate sentence representation which is useful to language understanding.
- This paper restructures the SemEval dataset, adding audio modality data so that the combination of proposed unit and popular models can solve the text-audio level ABSA task.
- We combine the PGCD unit with some popular models evaluated on SemEval datasets with or without audio. PGCD unit advances the performance of state-of-the-art models. Details of the code can be seen here[3].

The structure of paper is as follows: Section 2 describes our proposed unit in detail. Section 3 shows our experimental settings and compares experiment results with baselines. Finally, we conclude this paper in section 4.

2. MODEL DESCRIPTION

In this section we describe our proposed unit and the usage in any models. Both model architecture and the PGCD unit procedure are depicted in Figure 2.

2.1. Problem Statement

This paper deals with ABSA, which explores the sentiment polarity of text that aspect terms are mentioned. We solve it as a feature-enhancement classification task.

Formally, we specifically define ABSA task and method procedure. Given an original sentence S with L words, denoted by $S = \{w_i | i = 1, \ldots, t, \ldots L\}$. w_t is the aspect term and t refers to position of w_t. In solution of such task, convert all words into signal feature $\boldsymbol{F} \in \mathbb{R}^{L\times k}$ by embeddings, where k is the embedding dimension. Then, for sentence with aspect term, the objective of our task is to assign sentiment polarity $P \in \mathbb{R}^3$, where $\mathbb{R}^3 = \{\mathrm{POS}, \mathrm{NEU}, \mathrm{NEG}\}$. The tags in P indicate categories: positive, neural and negative, respectively.

In conclusion, how to gain aspect-concerned feature $\boldsymbol{F}$ for the classifier is key point of ABSA. This paper deals with it enhanced by token/positional/PGCD embeddings.

2.2. Model Structure

The whole ABSA model starts from a pre-trained embedding space $\boldsymbol{E} \in \mathbb{R}^{V\times k}$, where V is the size of vocabulary. For each word w in the sentence, we sign feature vector $\boldsymbol{E}_w \in \mathbb{R}^k$ after embedding, We also have another two additional embeddings, $\boldsymbol{F}_2$ is positional embedding to mark the position of words in sentence and $\boldsymbol{F}_3^t$ is PGCD outputs to indicate contributive distribution. The value of $\boldsymbol{F}_3^t$ varies according to the position t of the aspect term. Then the final input representation can be computed as $\boldsymbol{F} = \boldsymbol{F}_1 \oplus \boldsymbol{F}_2 \oplus \boldsymbol{F}_3^t$.

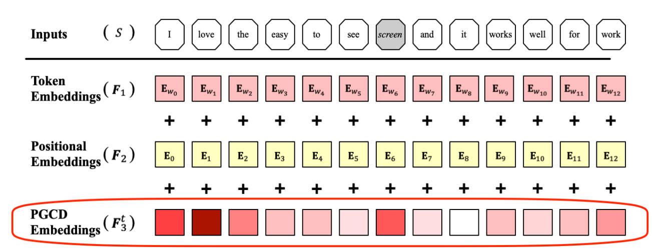

Fig. 3. Input Representation

The ABSA classification model can be simplified to a parameter matrix $\boldsymbol{M} \in \mathbb{R}^{k\times h}$, where h is hidden dimension. In this way, the middle representation H of the inputs can be obtained, which is calculated by:

$$H(\boldsymbol{F}) = tanh(\boldsymbol{F}^T \cdot \boldsymbol{M} + b) \quad (1)$$

Then, the final probability prediction vector can be gained by the output layer, that is, the sentiment prediction of the whole statement toward aspect term is obtained.

$$P(F) = softmax_{row}\left(Dense\big(H(F)\big)\right) \quad (2)$$

2.3. Position-Guided Contributive Distribution unit

From the description of 2.1, the calculation of F_3^t is the key to the whole ABSA model. When aspect term is given, the emotion of whole sentence towards this specific location term is affected by other regular positional words. We employ Position-Guided Contributive Distribution unit to learn this regulation and to attend the aspect term and its surrounding contexts simultaneously. The procedure of PGCD unit is shown on the right of Figure 2.

Firstly, enumeration combinations generated by aspect sets A and context sets S, where $|A| + |S| = L$ and $w_t \in A$. Symbol $|\cdot|$ means the size of given set. The whole sentence sets $N = A \cup S$ and the size of combinations is $\Gamma = 2^{|S|}$-1.

Then, use the method of Shapley value computation, each marginal contribution of enumeration combinations can be calculated by formula (1) and (2). Shapley value for special aspect words w_t or position t is defined as:

$$\phi_t(N, w_t) = \frac{(|N| - |S^*|)|S^*|!}{|N|!}[P(S^* \cup w_t) - P(S^*)] \quad (3)$$

where S^* is a subset of context sets S and w_t is an element or aspect words of A and N. Briefly, the Shapley value of given position word w_t is the weighted sum of the contribution of w_t in each subset of S. The contribution of w_t with respect to set S^* is computed by $P(S^* \cup w_t) - P(S^*)$.

The Shapley value is considered as a uniquely fair way for distributing the true label $P(N)$ into $(\phi_1(N), \ldots, \phi_L(N))$ for L different positional words, since it satisfies the following characteristics[9] :

- **Efficiency:** $\sum_{t \leq L} \phi_t(N, w_t) = P(N)$
- **Symmetry:** If for two different positional words w_i and w_j, $\forall S^* \subseteq N - \{i, j\}$, $P(S^* \cup w_i) = P(S^* \cup w_j)$, then $\phi_i(N, w_i) = \phi_j(N, w_j)$.
- **Dummy:** If $\forall S^* \subseteq N - \{t\}$, $P(S^* \cup w_t) = P(S^*)$, then $P(w_t) = 0$.
- **Additivity:** For any pair of combinations <N, w_i> and <N, w_j>: $\phi(N, w_i + w_j) = \phi_i(N, w_i) + \phi_j(N, w_j)$, where $\forall S^*$, $\phi(S^*, w_i + w_j) = \phi(S^*, w_i) + \phi(S^*, w_j)$.

Since predictive modeling could be viewed as a conditional game, Shapley value have been considered as a metric for contributive distribution and model explanation. In order to gain the PGCD vector towards special position, an average pooling should be used. Then the whole PGCD unit outputs can be formulated as:

$$\boldsymbol{F}_3^t = softmax_{row}\left(\frac{1}{\Gamma}\sum_{\substack{w_t \in A \\ S^* \subseteq S}} \phi_t(S^*, w_t)\right) \quad (4)$$

where $\boldsymbol{F}_3^t \in \mathbb{R}^L$ represents a contributive distribution (PGCD embeddings) for input sentence.

3. EXPERIMENTS

3.1. Dataset Description

For text level tasks, experiments are conducted on three ABSA benchmark datasets. Two review datasets of laptop, restaurants from SemEval[10] and a twitter social dataset introduced by [11] . These datasets are adopted by most of the proposed models and are the most popular datasets of text-level ABSA task. A large number of experiments have been carried out on these datasets for comparison. All aspects of the above datasets are labeled in three categories of sentiment polarities: positive, neutral, and negative.

For multimodal (text and audio modality) ABSA task, we can not find a suitable public dataset to evaluate proposed PGCD unit. Hence, this paper adds audio information to the above text level SemEval dataset. Each data segment contains aspect term, sentence and polarity, which records as text and audio types respectively. The dataset is published in the code link mentioned above, which named SemEval-Audio.

3.2. Our Models and Comparison Models

PGCD unit is a novel method proposed in paper which can gain a contributive distribution for language understanding and can be applied to any previous models. Thus, this paper compares the baseline models with or without PGCD unit to illustrate the effectiveness of the proposed method. All baselines and their introduction are as follows:

Non-Transformer based baselines:

- **Feature-based SVM**[12] , which is extensive feature-based support vector machine.
- **TD-LSTM**[13] extends LSTM model by using two LSTM networks to model the left context and the right context with target respectively.
- **ATAE-LSTM**[14] appends the target embeddings with each word embeddings, use LSTM with attention to get the final representation for classification.
- **AOA**[15] models aspects and sentences in a joint way and explicitly captures the interaction for classification.

Transformer-based model:

- **TNet**[16] which is a target-based transformer model to gain bi-directional feature for classification.
- **BERT-SPC**[17] feeds sequence "[CLS] + context + [SEP] + target + [SEP]" into the basic BERT model for sentence pair classification task.

For multimodal ABSA task, we improve the performance of multimodal fusion by applying PGCD unit only at text level. The baselines are as follows:

- **MDRE**[18] has used RNNs to encode both acoustic and textual data followed by a DNN for classification.
- **MCNN-LSTM**[19] has used a RNN and a DCNN to encode both modalities followed by fusion and an SVM for classification.
- **MHA2**[20] has used two Bidirectional Recurrent Encoders (BRE) for both modalities followed by a multimodality attention mechanism.

3.3. Experimental Settings

We download a *Glove*[21] as pre-trained embedding space with dimension $k = 300$ and *bert-base-uncased*[22] as pre-trained model parameters with dimension $k = 768$ and layer number $L = 12$. For training process, we use Adam optimizer to optimize model parameters with batch size of 16 and maximum epochs of 20. We set learning rate equals to 10^{-3} for Non-Transformer models and 2×10^{-5} for Transformer-based models. In loss function (Cross-Entropy Loss), L2 regularization is set to 10^{-5}.

3.4. Results

We follow previous works and use Accuracy and F1 score to evaluate model performance. For text-level and multimodal-level ABSA tasks, comparison results are reported in Table 1 and Table 2 respectively. Statistics demonstrate that the models use PGCD unit (signed as bold font in each block) has a significant improvement. But the improvement of the former is obviously higher than that of the latter. This is consistent with our cognition, because PGCD unit is a method that make text-level language understanding better. When multimodal features are fused, the contribution distribution learned by PGCD will be diluted.

For text-level ABSA task, the experiment results are shown in Table 1. There is 2.38% average accuracy improvement after adding PGCD unit, 2.1%, 1.16% and 3.39% improvement for *laptop*, *restaurant* and *twitter* respectively. It can be found that the performance of the dataset with higher overall accuracy is less improved, which is also in line with the general thoughts. For multimodal-level ABSA task (results shown in Table 2), 1.25% average accuracy improvement is less than text only dataset obviously. However, the improvement of accuracy and F1 score still shows the effectiveness of PGCD unit.

At the same time, taking LSTM model as an example, we explore when to update PGCD unit is the best time during the training process. This paper experiments: setting PGCD unit in epoch =1, ≤5, =5, ≤10, =10, ≤20, =20. PGCD unit is used during training process. When epoch is equal to an exact number (like "="), it means that PGCD is only calculated once. When epoch is a range (like "≤"), it means that PGCD should be continuously updated for many times. The results are shown in Table 3. It can be seen that the best effect show up when epoch = 5 or epoch = 10. From the description of section 2, the updating of PGCD unit depends on the training process of the models. Therefore, the above experimental results show that: (1) PGCD unit should be added or updated after the model has a rough performance; (2) the results of updating PGCD repeatedly not only waste computing resources, but also can't achieve good results; (3) updating PGCD later makes the model too late to adapt to the addition of PGCD embeddings, which makes the model ineffective.

Models	*Laptop*		*Restaurant*		*Twitter*	
	Acc	F1	Acc	F1	Acc	F1
SVM	69.52	0.687	77.26	0.762	62.17	0.612
TD-LSTM	66.81	0.594	76.32	0726	66.27	0.641
ATAE-LSTM	68.70	0.601	**78.29**	0.771	67.89	0.671
PGCD-LSTM	**70.47**	**0.681**	78.12	**0.776**	**70.27**	**0.689**
AOA	74.91	0.716	78.61	0.741	69.28	0.675
PGCD-AOA	**75.05**	**0.717**	**79.81**	**0.801**	**72.17**	**0.701**
TNet	76.54	0.715	80.27	0.796	71.06	0.691
PGCD-TNet	**79.14**	**0.768**	**81.72**	**0.810**	**75.23**	**0.745**
BERT-SPC	79.47	0.765	82.17	0.804	74.71	0.715
PGCD-BERT	**81.49**	**0.795**	**84.25**	**0.836**	**77.19**	**0.741**

Table 1. Experiment results for text-level ABSA. Accuracy and F1 is reported, which whole results experimented by this paper. The results using PGCD unit are shown in bold.

Models	*Laptop*		*Restaurant*		*Twitter*	
	Acc	F1	Acc	F1	Acc	F1
MDRE	67.48	0.662	**73.96**	0.717	62.77	0.607
PGCD-MDRE	**67.74**	**0.670**	73.42	**0.720**	**64.39**	**0.619**
MCNN-LSTM	65.61	0.634	**72.41**	**0701**	61.52	0.601
PGCD-M/L	**67.55**	**0.709**	72.38	**0.701**	**65.27**	**0.621**
MHA-2	69.01	0.681	72.01	0.729	66.72	0.645
PGCD-MHA2	**71.29**	**0.732**	**74.56**	**0.742**	**67.29**	**0.651**

Table 2. Experiment results for multimodal-level ABSA task. Accuracy and F1 is reported and the results using PGCD unit are shown in bold.

PGCD strategy	Avg-Acc	Avg-F1
Epoch = 1	-0.56%	-0.0017
Epoch ≤ 5	1.92%	0.0021
Epoch = 5	**2.38%**	0.0034
Epoch ≤ 10	2.01%	0.0027
Epoch = 10	2.34%	**0.0036**
Epoch ≤ 20	1.64%	0.0023
Epoch = 20	-1.02%	-0.0004

Table 3. Experiments results for updating PGCD strategy.

4. CONCLUSION

In this paper, we propose a novel unit called Position-Guided Contributive Distribution (PGCD) to capture the regulation between aspect term (special give position) and its surrounding contexts. This unit can address the "aspect with low-sensitive sentiment" issue in the ABSA task. We conduct extensive experiments on three publicly available datasets and a new restructured multimodal dataset. Experimental results demonstrate that our PGCD unit can advance the extant ABSA models on all 3 + "1" datasets, improve the input embedding feature and make the result more exactly. In addition, our PGCD units can update any time during the training process flexibly. The final results show that PGCD unit should be updated only once in the middle and early stage of training, which can not only save computing resources and time, but also achieve satisfying results.

5. REFERENCE

[1] Hongliang Dai and Yangqiu Song. "Neural aspect and opinion term extraction with mined rules as weak supervision," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 5268–5277.

[2] Duy Tin Vo and Yue Zhang, "Target-dependent twitter sentiment classification with rich automatic features," in *IJCAI*, 2015, pp. 201-214.

[3] Qiu, Xipeng, et al. "Pre-trained Models for Natural Language Processing: A Survey," *arXiv preprint arXiv: 2003.08271*, 2020.

[4] Wei Xue and Tao Li, "Aspect based sentiment analysis with gated convolutional networks," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 2018, pp. 2514-2523.

[5] Meishan Zhang, Yue Zhang, and Duy Tin Vo, "Gated neural networks for targeted sentiment analysis," in *AAAI*, 2016, pp. 3087–3093.

[6] Jacob Devlin, Ming-Wei Chang and Kenton Lee, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810. 04805*, 2018.

[7] Ashish Vaswani, Noam Shazeer and Niki Parmar, "Attention is all you need," in *NIPS,* 2017.

[8] Sisi Ma, Roshan Tourani, "Predictive and Causal Implications of using Shapley Value for Model Interpretation," in *KDD*, 2020, pp. 167-180.

[9] Lloyd String and David Bros, "Shapley. A value for n-person games," in *Contributions to the Theory of Games*, 1953, pp. 307-317.

[10] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Yanyan Zhao, Bing Qin, Orphe ́e De Clercq, et al., "Semeval-2014 task 4: Aspect based sentiment analysis," in *Proceedings of the 10th international workshop on semantic evaluation*, 2014, pp. 27–35.

[11] Li Dong, Furu Wei, Chuanqi Tan and Ke Xu, "Adaptive recursive neural network for target-dependent twitter sentiment classification," in *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics*, 2014, vol. 2, pp. 49–54.

[12] Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad, "Nrc-canada-2014: Detecting aspects and sentiment in customer reviews," in *Proceedings of WSE*, 2014, pp. 437–442.

[13] Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu, "Effective lstms for target-dependent sentiment classification," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Paper*, 2016, pp. 3298–3307.

[14] Yequan Wang, Minlie Huang, et al., "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing,* 2016, pp. 606–615.

[15] Huang, Binxuan, et al. "Aspect Level Sentiment Classifcation with Attention-over-Attention Neural Networks," *arXiv preprint arXiv:1804.06536*, 2018.

[16] Li, Xin, et al. "Transformation Networks for Target-Oriented Sentiment Classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 2018, pp. 2611-2620.

[17] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[18] S. Yoon, S. Byun, and K. Jung, "Multimodal speech emotion recognition using audio and text," in *Proceedings of SLT*, 2018, pp. 112-118.

[19] J. Cho, R. Pappagari, P. Kulkarni, J. Villalba and N. Dehak, "Deep neural networks for emotion recognition combining audio and transcripts," in *Proceedings of Interspeech,* 2018, pp. 247–251, 2018.

[20] S. Yoon, S. Byun, S. Dey, and K. Jung, "Speech emotion recognition using multi-hop attention mechanism," in *ICASSP*, IEEE 2019, pp. 2822–2826.

[21] Jeffrey Pennington, Richard Socher, and Christopher D.Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 2014, pp. 1532-1543.

[22] Rietzler, Alexander, et al. "Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification," *arXiv preprint arXiv:1908.11860*, 2019.