Syntax-Guided Pre-training and Self-training for Domain Adaptation in Aspect Term Extraction

Anguo Dong[1*], Xu Yang[1*], Cuiyun Gao[1†], Ye Ding[3], Qing Liao[1,2]
[1]*School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China*
[2]*Pengcheng Laboratory, Shenzhen, China*
[3]*School of Cyberspace Security, Dongguan University of Technology, Dongguan, China*
{20S151091, xuyang97}@stu.hit.edu.cn, dingye@dgut.edu.cn, {gaocuiyun, liaoqing}@hit.edu.cn

2025 11th IEEE International Conference on Privacy Computing and Data Security (PCDS) | 978-1-6654-7746-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/PCDS65695.2025.00013

***Abstract*—Aspect Term Extraction (ATE) aims to extract opinionated aspect terms from review texts, and it has been widely studied in both academia and industry. As a token-level and domain-specific task, the annotation cost is extremely high. Domain adaptation is a popular solution to alleviate the issue by transferring common knowledge across domains. However, most previous studies still rely on manually labeled pivot words or external knowledge to establish cross-domain associations. In this work, we propose a novel Syntax-guided Aspect Term Extraction model, named SATE, which enables automatic domain adaptation without relying on external knowledge sources. SATE utilizes syntactic structure similarity as a proxy-pivot feature to automatically construct cross-domain associations. Besides, although pre-trained language models have significantly improved the performance of this task, fine-tuning a pre-trained model on the source domain often leads to a drastic performance drop on the target domain due to domain discrepancy. To address this issue, we propose a variant of the masked language model based on the syntactic structure similarity between domains to learn a domain-invariant representation. Additionally, to further facilitate adaptation, we construct syntax-based pseudo instances for training. Experiments show that SATE achieves at least a 3.5% improvement in Micro-F1 over state-of-the-art baselines across three benchmark datasets on average.**

***Index Terms*—domain adaptation, aspect term extraction, pre-training.**

I. Introduction

Aspect Term Extraction (ATE) is a crucial sub-task in aspect-based sentiment analysis [1], [2], which aims to extract all the aspect terms present in sentences. For instance, given a review sentence "*The keyboard and mouse are both pretty decent.*", ATE aims to extract the aspect terms "*keyboard*" and "*mouse*". The release of the SemEval datasets [3]–[5] and the rise of deep learning techniques have greatly advanced ATE research. The Multi-Aspect Multi-Sentiment (MAMS) dataset [6] further increases the difficulty of the task, as each sentence contains at least two aspects with different sentiment polarities. Recent studies commonly formulate ATE as a sequence tagging or token-level classification task. Researchers focus on developing various neural sequence taggers. Although these models [7]–[11] achieve satisfactory performance, they heavily rely on in-domain labeled data. The scarcity of labeled data, due to the high cost of annotation, remains a major challenge for ATE.

Domain adaptation is a popular solution to address the issue of data scarcity, aiming to generalize a model trained on labeled data from a source domain to an unlabeled target domain. Unsupervised domain adaptation specifically refers to the scenario where labeled data are only available in the source domain, while the target domain contains no labels during training. In this paper, we focus on unsupervised domain adaptation, which is a more practical setting.

Some recent domain adaptation methods [12]–[15] aim to align the source and target domains by learning domain-invariant feature representations. Structure Correspondence Learning (SCL) [16] is one of the core techniques used in learning domain-invariant feature representations for text classification tasks. It splits the feature space into pivot and non-pivot features. Pivot features are those that meet the following two criteria: (a) they appear frequently in both domains; and (b) they are highly correlated to the task labels. Non-pivot features are those that do not satisfy at least one of these criteria. Although pivot-based models [17]–[19] have shown promising results in sentence-level classification tasks, applying them effectively to token-level tasks such as ATE remains infeasible. This is because aspect terms and opinion words vary significantly across domains, making it difficult to define explicit pivot features. Some methods [20]–[23] use aspect-opinion relations as pivot features, based on the observation that aspect terms and opinion words often co-occur. However, these methods require either external domain knowledge or manually labeled opinion words, and their effectiveness heavily depends on the similarity of opinion expressions between domains. To address this, Lekhtman et al. [24] proposed a Category-based MLM pre-training approach that uses aspect categories as proxy-pivot features to facilitate domain adaptation for ATE. However, this method relies heavily on domain-specific aspect category information.

Meanwhile, self-training could be another effective solution for domain adaptation, which can directly learn concepts from the target domain in a fully automatic manner without any human intervention [20]. Previous methods [25]–[28] apply self-training by using a source-domain model to label the target-domain data, followed by selecting a set of high-confidence pseudo-labeled instances to further train the model. The quality of the pseudo-labeled data is a critical factor in determining the effectiveness of these methods. However, due to domain

[*] Both authors contributed equally to this research.
[†] Corresponding author.

978-1-6654-7746-8/25/$31.00 ©2025 IEEE
DOI 10.1109/PCDS65695.2025.00013

discrepancies, the pseudo labels may suffer from poor quality, which can lead to significant performance degradation.

To overcome the aforementioned limitations, we propose a novel Syntax-guided Aspect Term Extraction model for domain adaptation in ATE, named SATE. We use syntactic structure similarity as a proxy-pivot feature and combine pre-training and self-training by the proxy-pivot feature. Specifically, we first encode the syntactic structures of source-domain texts and aggregate their aspect terms to obtain an average syntactic structure representation. Then, we compute the syntactic similarity between tokens and the average aspect term. Based on syntactic structure similarities, we propose three modules to transfer aspect terms across domains: (1) Syntax-based Masked Language Model (SMLM): a variant of BERT [29], where masked tokens are selected based on their syntactic similarity to aspect terms, rather than being randomly chosen. This module is trained on large-scale unlabeled corpora from both source and target domains. (2) Syntax-based Self-Training (SST): constructs a pseudo-labeled training set for the target domain based on the syntactic structure similarity and source-domain labeled data. (3) Syntax-based Loss (SL): a weighted cross-entropy loss function that assigns greater importance to tokens that are more likely to be aspect terms based on syntactic structure during classifier training. Experimental results on three benchmark datasets show that SATE improves cross-domain ATE performance by at least 3.5% in Micro-F1 score on average compared to state-of-the-art baselines. The main contributions of this paper are summarized as follows:

- We propose a novel syntax-guided domain adaptation model for aspect term extraction, which leverages syntactic structure similarity to construct cross-domain associations without relying on external knowledge.
- To the best of our knowledge, we are the first to address cross-domain aspect term extraction using self-training. We also present a variant of the BERT MLM pre-training model based on the syntactic structure.
- Extensive experiments on three benchmark datasets validate the effectiveness of the proposed model, showing consistent improvements over state-of-the-art baselines in cross-domain aspect term extraction.

II. Methodology

A. Problem Statement

We formulate cross-domain ATE as a sequence tagging task. The input is a sequence of tokens $x = \{x_1, x_2, ..., x_n\}$, and the output is a sequence of labels $y = \{y_1, y_2, ..., y_n\}$, where each $y_i \in \{B, I, O\}$ denotes the beginning of, inside of, and outside of an aspect term. In the cross-domain setting, labeled data are only available in the source domain. Given a set of labeled data from source domain $D^S = \{(x_i^S, y_i^S)\}_{i=1}^{N_S}$ and a set of unlabeled data from target domain $D^U = \{(x_i^U)\}_{i=1}^{N_U}$, our goal is to predict token labels for unseen target test data in target domain: $y_i^T = f(x_i^T), D^T = \{(x_i^T)\}_{i=1}^{N_T}$.

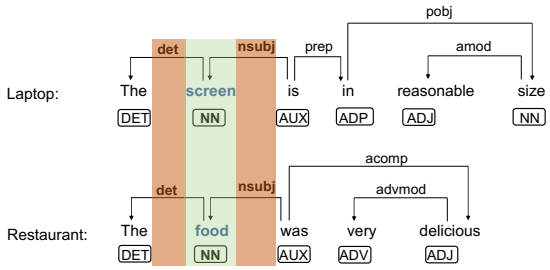

Fig. 1. Dependency tree of reviews from laptop and restaurant domains. The brown blocks indicate that aspect terms from the two domains have the same dependency relation, and the green block indicates that aspect terms from the two domains have the same POS tag.

B. Syntactic Structure Similarity

Although aspect terms may vary across domains, their syntactic roles are generally consistent [30]. We propose to capture aspect terms in the target domain based on their syntactic structure similarity to aspect terms in the source domain. Part-of-speech (POS) tags and syntactic dependency relations are employed to measure the syntactic structure similarity between tokens. As illustrated in Fig. 1, the aspect terms "*waiter*" and "*screen*" share the same POS tag "*NN*" and the same set of dependency relations "{*det*, *nsubj*}", indicating high syntactic similarity. To quantify this similarity, for each word x_i, we use a one-hot vector $b_i^{pos} \in R^{N_{pos}}$ and a multi-hot vector $b_i^{dep} \in R^{N_{dep}}$ to represent its POS tag and syntactic dependency relations, respectively, where N_{pos}/N_{dep} indicate the size of POS tags/syntactic dependency relation set.

To compute b_i^{dep}, we merge all dependency relations that involve the token x_i. To obtain the average syntactic structure representation of aspect terms $\bar{a} = (b^{pos}, b^{dep})$, we aggregate all aspect terms in the source domain: $b^{pos} = \sum_i^A \frac{b_i^{pos}}{N_A}$ and $b^{dep} = \sum_i^A \frac{b_i^{dep}}{N_A}$, where A is the set of aspect terms, N_A is the size of A. The Syntactic structure similarity between a token x_i and $\bar{a}$ is defined as:

$$S_{sim}(x_i, \bar{a}) = \cos(b_i^{pos}, b^{pos}) \times cos(b_i^{dep}, b^{dep}) \quad (1)$$

where $\cos(,)$ is the cosine similarity. We treat the syntactic structure similarity as a proxy-pivot feature to build the transferring modules.

C. Syntax-based Masked Language Model

We present a variant of the BERT MLM pre-training model, referred to as the Syntax-based Masked Language Model (SMLM), as illustrated in Fig. 2. For the standard MLM pre-training model, all input tokens have the same probability of being chosen for prediction. In contrast, SMLM leverages the syntactic structure similarity between tokens and the average aspect term from the source domain to prioritize tokens that are more likely to be aspect terms for prediction.

To capture domain-invariant representations, SMLM is pre-trained on large-scale unlabeled corpora from both the source and target domains. Specifically, we choose the top $\alpha\%$ of the

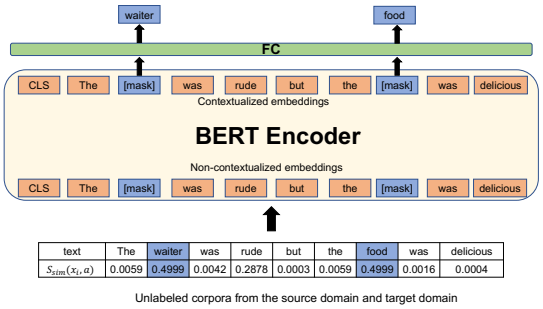

Fig. 2. Overall architecture of Syntax-based Mask Language Model (SMLM).

input tokens based on the similarity $S_{sim}(x_i, \bar{a})$ in Eq. (1), where α is the masking threshold. Following the standard MLM strategy, each selected token has an 80% probability of being replaced with the `[MASK]` token, a 10% probability of being substituted with a random token from the vocabulary, and a 10% probability of remaining unchanged.

Our token masking strategy ensures that the pre-training language model focuses on words that are likely to be aspect words in one of the domains. For example, given the input sentence "*The screen is in reasonable size, I really liked it.*" from the laptop domain, SMLM would choose the word "*screen*", which has a high syntactic structure similarity score with the average aspect term. In contrast, the original MLM randomly selects tokens for masking. By predicting these aspect-term-likely tokens, SMLM can learn task-adaptive and domain-invariant representations.

D. Syntax-based Self-training

Besides learning domain-invariant representation, self-training is another promising solution for domain adaptation. The core idea of self-training is to construct a set of high-quality pseudo-training instances for the target domain. Different from previous self-training methods [25]–[28], we present a novel module called syntax-based self-training (SST). When constructing pseudo training instances, we select a pseudo aspect term set $\mathcal{A}$ from the target domain which are similar to the aspect terms of the source domain in terms of syntactic structure. We then replace the aspect terms in the source domain with the pseudo aspect terms in set $\mathcal{A}$ to create pseudo training instances. The detailed process is described in Algorithm 1.

Specifically, words or phrases that present higher similarities to the average aspect term representation $\bar{a}$ are selected as candidates for composing the pseudo aspect term set $\mathcal{A}$. All the selected terms are ranked according to the term frequencies, and the most frequent ones compose the set $\mathcal{A}$.

The each aspect term in the source domain is randomly replaced by one pseudo aspect term in $\mathcal{A}$ to create a pseudo training set for the target domain. For each instance, we create one pseudo instance.

Algorithm 1 Building Syntax-based Pseudo Training Dataset

Input: The set of labeled source domain: D^S; The set of unlabeled target domain: D^U; The syntax-similarity threshold: σ; The size of $\mathcal{A}$: β; The average syntactic structure representation of aspect terms: $\bar{a} = (b^{pos}, b^{dep})$

Output: The pseudo training set for target domain: D^{st}

initiate $D^{st} = \{\}$; $\mathcal{A} = \{\}$
for instance X in D^U **do**
  for x_i in X **do**
    $S_{sim}(x_i, \bar{a}) = \cos(b_i^{pos}, b^{pos}) * \cos(b_i^{dep}, b^{dep})$
    if $S_{sim}(x_i, \bar{a}) > \sigma$ **then**
      $\mathcal{A} = \mathcal{A} + x_i$
    end if
  end for
end for
$\mathcal{A} = mostCommon(\mathcal{A}, \beta)$
for instance (X, Y) in D^S **do**
  $A = \{a_1, a_2, ..., a_n\}$ is the set of aspect terms
  for a in A **do**
    replace a by random($\mathcal{A}$)
  end for
  $D^{st} = D^{st} + \{(X, Y)\}$
end for
return D^{st};

E. Syntax-based Loss

ATE is a token-level classification task, each token is of different importance for classifier training. We propose to concentrate on those tokens that are more likely to be aspect terms in syntax. We modify the cross entropy loss with syntactic structure similarity, named as syntax-based loss (SL). The weighted cross entropy loss is defined as follows:

$$L = \sum_{D^S + D^{ST}} \sum_{i}^{T} S_{sim}(x_i, \bar{a}) * l(y_i, \widehat{y}_i) \qquad (2)$$

III. EXPERIMENTS

A. Data & Experiment Setup

Datasets: We adopt unlabeled corpora from the Amazon laptop reviews[1] and the Yelp restaurant reviews[2] to perform SMLM pre-training model. The labeled data from the laptop domain are taken from SemEval-2014 ABSA [3]. Following the setting in [24], for the labeled data from the restaurant domain, we combine the SemEval 2014, 2015, 2016 ABSA [3]–[5] restaurant datasets and remove the duplicated instances. To construct a more challenging evaluation setup beyond the SemEval datasets, we consider the MAMS dataset [6], which contains sentences with at least two aspects of different sentiment polarities. Detailed statistics are shown in Table I.

Settings & Implementation Details: We conduct experiments on four source-target transfer pairs such as L→R using the three domains in Table I. Following the experimental setup

[1] http://jmcauley.ucsd.edu/data/amazon/links.html
[2] https://www.yelp.com/dataset/challenge

TABLE I
STATISTICS OF THE BENCHMARK DATASETS.

Dataset	Domain	Total	Training	Testing
L	Laptop	3845	3045	800
R	Restaurant	6035	3877	2158
M	MAMS	5297	4297	1000

in [24], we remove R→M and M→R, as R and M are similar. For each transfer pair, the training data consists of labeled training data in the source domain and pseudo training data for the target domain. Meanwhile, we use the labeled test set from the source domain as the validation set, and the test set from the target domain as the evaluation set. We use Spacy[3] for dependency parsing. There are 51 classes of POS tags and 47 classes of dependency relations in total in the three datasets.

To implement the SMLM pre-training model, we use the BERT-Base-Uncased model of the Hugging-Face Transformers package [31]. We fine-tune all BERT layers and mask full words instead of sub-words to reduce the influence of the tokenizer. The masking threshold is set to $\alpha = 15$, and pre-training is conducted for 2 epochs using the AdamW optimizer with a learning rate of 3e-5, an epsilon of 1e-8, and a batch size of 16. After the pre-training phase, the fully connected layer used in the SMLM objective is discarded. To construct the pseudo training data for the target domain, we set the syntax similarity threshold to $\sigma = 0.45$ and limit the size of the pseudo aspect term set to $\beta = 300$.

To facilitate training on our ATE task, we add an additional MLP layer on top of the pre-trained language model and train it using the AdamW optimizer with a learning rate of 2e-5, an epsilon of 1e-8, and a batch size of 8.

Evaluation Metric: Following previous studies [24], [30], [32], we evaluate the models using Micro-F1, where only exact matches are considered correct. All experiments are repeated nine times and the average results are reported.

B. Baselines

We compare our model with several state-of-the-art models, as follows:

- **BERT-Cross**: BERT-Cross post-trains BERT on a combination of Yelp and Amazon corpus.
- **BERT-Base-UDA and BERT-Cross-UDA [32]**: UDA is a Transformer-based neural network that performs fine-tuning with syntactic-driven auxiliary tasks and a modified attention mechanism. BERT-Base-UDA and BERT-Cross-UDA are designed differently in the initialization. BERT-Base-UDA is initialized by the BERT-Base model while BERT-Cross-UDA is initialized by BERT-Cross.
- **Combridge [30]**: A bridge-based convolution neural network that incorporates syntactic structures and cross-semantic information, combined with an adversarial component.

[3]https://spacy.io

TABLE II
MAIN RESULTS FOR CROSS-DOMAIN ATE ON FOUR SOURCE-TARGET PAIRS.

Methods	M→L	L→M	R→L	L→R	AVG.
BERT-Cross	35.30	29.82	45.89	39.32	37.48
BERT-Base-UDA	36.52	39.59	48.32	49.52	43.50
BERT-Cross-UDA	41.29	45.62	53.51	56.12	49.14
Combridge	39.98	47.42	53.32	66.20	51.73
DILBERT	43.72	58.96	56.07	61.04	54.95
SATE	**48.11**	**65.38**	**56.26**	**69.26**	**58.45**

TABLE III
MAIN RESULTS FOR CROSS-DOMAIN ABSA ON FOUR SOURCE-TARGET PAIRS.

Methods	M→L	L→M	R→L	L→R	AVG.
RNSCN	23.74	24.63	26.63	35.65	27.66
AD-SAL	28.49	26.87	34.13	43.04	33.13
BERT-Cross	34.60	25.75	39.72	45.4	36.37
BERT-B-DANN	-	-	30.41	41.63	-
BERT-E-DANN	-	-	38.83	47.39	-
BERT-Base-UDA	27.19	27.25	33.68	45.46	33.40
BERT-Cross-UDA	32.47	33.03	43.95	49.52	39.74
SATE	**46.65**	**44.75**	**54.55**	**63.11**	**52.27**

- **DILBERT [24]**: A Transformer-based model that performs category-based masked language modeling and a category proxy prediction task. It incorporates aspect category information from an external domain and achieves the state-of-the-art performance in cross-domain aspect term extraction.
- **RNSCN [22]**: A recursive neural structural correspondence network that incorporates syntactic structures.
- **AD-SAL [33]**: A recursive neural network which can automatically capture aspect-opinion latent relations to achieve token-level adversarial adaptation.
- **BERT-Base-DANN and BERT-Cross-DANN [34]**: DANN is a domain-adversarial training neural network. Similarly, BERT-Base-DANN and BERT-Cross-DANN are initialized by BERT-Base and BERT-Cross respectively.

C. Results

Table II shows the comparison results of cross-domain ATE for all the methods. As can be seen, the proposed SATE achieves the best performance in terms of the exact-match F1 metric. For example, the SATE model significantly improves the average performance of DILBERT from 54.95% to 58.45%. Specifically, on the more challenging transfer pairs L→M and L→R, SATE improves by 6.42% and 8.22% from DILBERT, respectively.

Table III shows the comparison results for cross-domain ABSA for all the methods. As can be seen, the proposed SATE achieves the best performance in terms of the exact-match F1 metric. Specifically, BERT-Cross-UDA performs poorly on more challenging transfer pairs such as M→L and L→M (i.e., column 1∼2), only achieving 32.47% and 33.03%, respec-

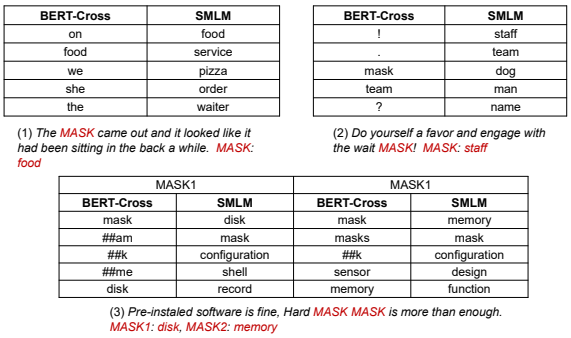

Fig. 3. Top five predictions of BERT-Cross and SMLM.

tively. In contrast, our model exhibits promising transferability on the more challenging transfer pairs, achieving 46.65% and 44.74%, respectively.

IV. ANALYSIS

To compare the ability of learning task-adaptive representation, we pre-train the SMLM model and the BERT-Cross model on unlabeled corpora from the Amazon laptop reviews and the Yelp restaurant reviews, and test them on sentences from these domains. Fig. 3 shows the comparison of the top five predictions of BERT-Cross and SMLM. Text (1) and (2) are from unlabeled Yelp restaurant reviews, and (3) is from unlabeled Amazon laptop reviews. It is obvious that the prediction of SMLM is more similar to the mask token than BERT-Cross. Our token masking strategy facilitates the BERT model to learn task-adaptive representations for aspect term extraction.

V. CONCLUSION

In this paper, we propose a novel domain adaptation method for aspect term extraction. We enhance the transferring ability by incorporating the syntactic structure similarity into the pre-training model and the self-training model. Extensive experiments on three benchmark datasets demonstrate the superiority of our approach over existing methods in cross-domain aspect extraction.

ACKNOWLEDGMENT

This work was supported in part by the National Key Research and Development Program of China under Grant No. 2023YFB3107000, the Major Key Project of PCL under Grant No. PCL2024A05, and the National Key Research and Development Program of China under Grant No. 2020YFB2104003.

REFERENCES

[1] T. T. Thet, J.-C. Na, and C. S. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," *J. Inf. Sci.*, vol. 36, no. 6, pp. 823–848, 2010.

[2] L. Bing, "Sentiment analysis and opinion mining (synthesis lectures on human language technologies)," *University of Illinois: Chicago, IL, USA*, 2012.

[3] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation*, Aug. 2014, pp. 27–35.

[4] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "Semeval-2015 task 12: Aspect based sentiment analysis," in *Proceedings of the 9th international workshop on semantic evaluation*, 2015, pp. 486–495.

[5] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," in *International workshop on semantic evaluation*, 2016, pp. 19–30.

[6] Q. Jiang, L. Chen, R. Xu, X. Ao, and M. Yang, "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*, 2019, pp. 6280–6285.

[7] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Recursive neural conditional random fields for aspect-based sentiment analysis," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016.

[8] P. Liu, S. Joty, and H. Meng, "Fine-grained opinion mining with recurrent neural networks and word embeddings," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 1433–1443.

[9] X. Li and W. Lam, "Deep multi-task learning for aspect term extraction with memory interaction," in *Proceedings of the 2017 conference on empirical methods in natural language processing*, 2017, pp. 2886–2892.

[10] H. Xu, B. Liu, L. Shu, and S. Y. Philip, "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*, 2019, pp. 2324–2335.

[11] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Recursive neural conditional random fields for aspect-based sentiment analysis," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, J. Su, X. Carreras, and K. Duh, Eds., 2016, pp. 616–626.

[12] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, "Cross-domain sentiment classification via spectral feature alignment," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 751–760.

[13] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proceedings of the 28th International Conference on Machine Learning*, 2011.

[14] D. Bollegala, D. Weir, and J. Carroll, "Cross-domain sentiment classification using a sentiment sensitive thesaurus," *IEEE Trans. Knowl. Data. Eng.*, vol. 25, no. 8, pp. 1719–1731, 2012.

[15] F. Zhuang, X. Cheng, P. Luo, S. J. Pan, and Q. He, "Supervised representation learning: Transfer learning with deep autoencoders," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015, pp. 4119–4125.

[16] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification," in *Proceedings of the 45th annual meeting of the association of computational linguistics*, 2007, pp. 440–447.

[17] Y. Ziser and R. Reichart, "Deep pivot-based modeling for cross-language cross-domain transfer with minimal guidance," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 238–249.

[18] T. Miller, "Simplified neural unsupervised domain adaptation," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, pp. 414–419.

[19] E. Ben-David, C. Rabinovitz, and R. Reichart, "Perl: Pivot-based domain adaptation for pre-trained deep contextualized embedding models," *Trans. Assoc. Comput. Linguist.*, vol. 8, pp. 504–521, 2020.

[20] Y. Liu and Y. Zhang, "Unsupervised domain adaptation for joint segmentation and pos-tagging," in *Proceedings of COLING 2012*, 2012, pp. 745–754.

[21] Y. Ding, J. Yu, and J. Jiang, "Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017, pp. 3436–3442.

[22] W. Wang and S. J. Pan, "Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction," in *Proceed-*

ings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018, pp. 2171–2181.

[23] W. Wang and S. J. Pa, "Transferable interactive memory network for domain adaptation in fine-grained opinion extraction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, pp. 7192–7199.

[24] E. Lekhtman, Y. Ziser, and R. Reichart, "Dilbert: Customized pre-training for domain adaptation with category shift, with an application to aspect extraction," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 219–230.

[25] N. Inoue, R. Furuta, T. Yamasaki, and K. Aizawa, "Cross-domain weakly-supervised object detection through progressive domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5001–5009.

[26] Y. Zou, Z. Yu, X. Liu, B. Kumar, and J. Wang, "Confidence regularized self-training," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5982–5991.

[27] K. Saito, D. Kim, S. Sclaroff, and K. Saenko, "Universal domain adaptation through self supervision," in *Advances in Neural Information Processing Systems*, 2020.

[28] P. Jiang, D. Long, Y. Sun, M. Zhang, G. Xu, and P. Xie, "A fine-grained domain adaption model for joint word segmentation and pos tagging," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 3587–3598.

[29] J. D. M.-W. C. Kenton and L. K. Toutanova, "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*, 2019, pp. 4171–4186.

[30] Z. Chen and T. Qian, "Bridge-based active domain adaptation for aspect term extraction," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 2021, pp. 317–327.

[31] T. Wolf, J. Chaumond, L. Debut, V. Sanh, C. Delangue, A. Moi, P. Cistac, M. Funtowicz, J. Davison, S. Shleifer *et al.*, "Transformers: State-of-the-art natural language processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2020, pp. 38–45.

[32] C. Gong, J. Yu, and R. Xia, "Unified feature and instance based domain adaptation for end-to-end aspect-based sentiment analysis," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 2020, pp. 7035–7045.

[33] Z. Li, X. Li, Y. Wei, L. Bing, Y. Zhang, and Q. Yang, "Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, 2019, pp. 4589–4599.

[34] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. S. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, pp. 59:1–59:35, 2016.