# Adversarial Weakly Supervised Domain Adaptation for Few Shot Sentiment Analysis

Seyyed Ehsan Taher
Faculty of Computer Science and Engineering
Shahid Beheshti University
Tehran, Iran
e.taher@mail.sbu.ac.ir

Mehrnoush Shamsfard
Faculty of Computer Science and Engineering
Shahid Beheshti University
Tehran, Iran
m-shams@sbu.ac.ir

***Abstract*— The ability of deep neural networks to generate state-of-the-art results on many NLP problems has been apparent to everyone for some years now. However, when there is not enough labeled data or the test dataset has domain shift, these networks face many challenges and results are getting worse.**

**In this article, we present a method for adapting the domain from formal to colloquial (in sentiment classification). Our method uses two approaches, adversarial training and weak supervision, and only needs a few shots of labeled data.**

**In the first stage, we label a crawled dataset (containing colloquial and formal sentences) with weakly supervised sentiment tags using a sentiment vocabulary network. Then we fine-tune a pre-trained model with adversarial training on this weak dataset to generate domain-independent representations. In the last stage, we train the above fine-tuned neural network with just 50 samples of data (formal domain) and test it on colloquial.**

**We achieved 15% better F1 compared to the current SOTA model (Pars BERT) with the same data.**

***Keywords*— *adversarial training; weak Supervision; domain adaptation; few shot learning; persian sentiment analysis; low resource languages***

## I. Introduction

The analysis of sentiment is one of the topics of natural language processing, which deals with the study of sentiments, emotions, and attitudes of people about a product, a service, an organization, etc.

Due to the growing amount of text content on the web (e.g. blogs, social media, and e-commerce websites) the importance of automatic classification of text sentiment, using artificial intelligence, is more than ever. Hence, sentiment analysis is followed as a hot research field among NLP researchers to achieve automated sentiment analysis with higher precision and higher quality models [1].

Today, one of the most common approaches for solving NLP problems (especially in language understanding) is the transfer of knowledge from a pre-trained model (which is trained with lots of data on a general task like language modeling) to the target problem. This pre-trained model acts as a feature extractor and generates vector representations that will fine-tune with the labeled dataset. Under the fine-tuning process, the knowledge will be transferred from the pre-trained model to the target problem [2]–[11].

However, one of the main challenges of this approach is that the fine-tuned model does not work properly in a new domain, which has not been observed in fine-tuning. There is no labeled data for many problems, and in many domains, we just have a few shots of labeled data (lack of Y). Also, there are zero or a few shots of unlabeled data in some problems and domains (lack of X).

In such low resource settings, researchers present a group of solutions, which are known as domain adaptation. Domain adaptation is a subset of transfer learning and multi-task learning, in which a model that has good performance at the source domain is used to achieve acceptable performance in the target domain.

Adversarial training is one of the ways for regularization in deep neural networks that lead to the more general models[12]–[14]. In [14] adversarial training has been used as an efficient method for domain adaptation. The general idea of this approach is to generate representations that are domain-independent. One of the advantages of this method is that it does not need a parallel domain dataset and it is enough to have unlabeled data in each domain separately. We used this method in the representation phase to represent each sentence regardless of its domain.

A weakly-supervised dataset refers to a dataset that is labeled with an inaccurate or noisy or irrelevant labeling source. If such data is treated as supervised data and used in training a machine learning model, this model is called a weakly supervised model. Using weak supervision greatly reduces the cost of labeling datasets and applying such datasets can help us build a robust and general model.

In this paper, we first weakly label a dataset including both the formal and the colloquial domains crawled from the web. Then, we train a model to generate domain-independent representations under an adversarial training process. Finally, we adapt the domain of the model from formal domain to the colloquial domain using a few shots of human-labeled data in the formal domain (some examples are given in Table I.).

Our contributions are:

- Creating a formal/colloquial classifier model
- Presenting a two phases training procedure for domain adaptation in low resource setting (for sentiment classification):
    1. Using the adversarial training alongside the use of weakly supervised data for generating representation vectors that are domain-independent
    2. Adapting the domain from formal to colloquial using only a few shots of human-labeled data in the formal domain (50 sentences)

In this article, we firstly review the related works in the field of domain adaptation and adversarial methods. Then, in the proposed method section, we explain the data resources, formal/colloquial classifier, weakly supervised labeling method, proposed architecture, and training Procedure, respectively. then finally we present the results of our method and compare it to other methods.

Table I. Some example sentences from sentipers with domain and sentiment label

| Sentence | Domain | Polarity Class |
|---|---|---|
| برای اپلی‌ها انتخاب مناسبیه.<br>It is a suitable choice for Apple owners. | colloquial | positive |
| زود لق میشه.<br>It will loosen soon. | colloquial | negative |
| PS3 واقعا فوق العاده اس.<br>The PS3 is really fantastic. | colloquial | positive |
| واقعا بی نظیره، من اگر بازم بخواهم دوباره هارد اکسترنال بگیرم انتخابم همین مارک و مدل هست، امنیتش که خیلی برام مهم بود یکی دیگر از مزیتاشه. در کل الان ۶ ماهه که گرفتم و ازش راضی هستم ممنون از سایت خوب دیجی کالا<br>Really unique. If I want to buy an external hard drive again, my choice will be the same brand and model. Its security, which was very important to me, is another advantage. In total, I have taken it for 6 months now and I am satisfied with it. Thanks to Digikala good site. | colloquial | positive |
| هماهنگی خوب سخت‌افزار و نرم‌افزار در این مدل، باعث گردیده که در هنگام کار با آن، سرعت عمل انجام دستورات را احساس کنید.<br>The good coordination of hardware and software in this model, makes you feel the speed of .execution of commands when working with it | formal | positive |
| این قلم نه از نظر ظاهری و نه از لحاظ کارایی، شباهت زیادی به قلم Noteهای سامسونگ ندارد و بیشتر می‌توان آن‌را یک قلم کمکی نامید که در کار با گوشی به کاربر کمک می‌نماید اما قابلیت‌های فوق‌العاده‌ای ندارد.<br>This pen is not very similar to Samsung Note pens in terms of appearance or performance, and it can be called an auxiliary pen that helps the user to work with the cell phone, but it does not have extraordinary capabilities. | formal | negative |
| صفحه نمایش و بلندگوها: صفحات نمایشی که بر روی این لپ‌تاپ‌ها قرار دارد، واقعا خوب است.<br>Screens and speakers: The screens on these laptops are really good. | formal | positive |
| به همین دلیل، کاربران Sleek book زمان بیشتری را صرف تمیز کردن نوت بوک خود سپری خواهند کرد.<br>Because of this, Sleek book users will spend more time cleaning their notebooks. | formal | negative |

## II. Related Work

In this section, we briefly outline the most important approaches for domain adaptation and adversarial training.

In natural language processing, the concept of domain is widely used in many contexts, and there is no common ground between researchers. However, the domain generally refers to the different types of text in a corpus. These types may be related to the topic (politics, social, economic, etc), writing style (formal, colloquial, etc), genre (fiction, news, scientific report, etc), or any other classification [15].

If we have a solution to task T in domain A and want to use the same solution for the same task in domain B, we can use domain adaptation techniques. Common methods for domain adaptation in natural language processing fall into three general categories:

1. Model-centric: In this category, the focus is on feature space, the generation of input representations, loss function, adversarial training, model architecture, and parameters [15]–[18].

2. Data-centric: In this category, the focus is on data, semi-automatic labeling methods and weak supervision, data selection, active learning, and data augmentation [19]–[21].

3. Hybrid: In this category, a combination of the above 2 methods is used[22]–[26].

There is also another categorization for domain adaptation methods according to whether or not it needs the human-labeled data. In recent years, several methods have been introduced for domain adaptation without the need for labeled data, which generally use adversarial training[14], [15].

In [14] adversarial training has been used to adapt the domain in image classification. This paper introduces a gradient reversal layer (GRL) that adversarially trains a classifier in such a way that it is not able to discriminate the input domain from the target domain. The GRL layer is a parameterless layer that in the feed-forward phase gives input to output with no changes, but in the back-propagation phase multiplies the gradient of the next layer by -λ and gives it to the previous layer.

## III. The Proposed Method

### A. Data and Resources

The following data resources have been used in this paper:

1. Dataset of formal and colloquial sentences (crawled from Digikala) include 20,000 sentences taken from expert review texts and 20,000 sentences taken from user comments at the bottom of products. We hypothesized that the first (expert reviews) is in the formal domain and the second (user comments) is in the colloquial domain. We tested this hypothesis by examining a hundred sentences from each domain, and 100% of them had the correct domain tag. The sentences of this dataset do not contain any sentiment tags.

2. Domain-labeled Sentipers dataset (Polarity Classification): Sentipers [27], [28] dataset includes the opinions of users and experts of Digikala which are human-labeled (an integer between -2 to +2). First, we converted the sentiment score of sentences into binary polarity. So that the positive scores became a positive class and the negative scores became a negative class (we also deleted the sentences with zero scores). Then, using the formal/colloquial classifier model (further details are provided), we divided the Sentipers sentences into formal and colloquial domains and extracted the following modified dataset from the Sentipers:
   - Training data:
     - 50 formal sentences (including 25 positives and 25 negatives).
   - Test data:
     - 501 colloquial sentences (including 404 positives and 97 negatives)
     - 962 formal sentences (including 804 positives and 158 negatives)

   Some of the sentences in this dataset are randomly listed in Table I.

3. HesNegar [29] Persian sentiment vocabulary network: This network includes 100062 unique words with the weight of each word in the positive and negative classes. Some words of HesNegar are shown in Table II.

Table II. Some example words from HesNegar with positive and negative weights

| Word | Average weight of positive class | Average weight of negative class |
|---|---|---|
| خوب (Good) | 0.46 | 0.04 |
| بد (Bad) | 0.11 | 0.55 |
| عشق (Love) | 0.19 | 0.11 |
| معیوب (Damaged) | 0.14 | 0.31 |
| سالم (Healthy) | 0.33 | 0.07 |

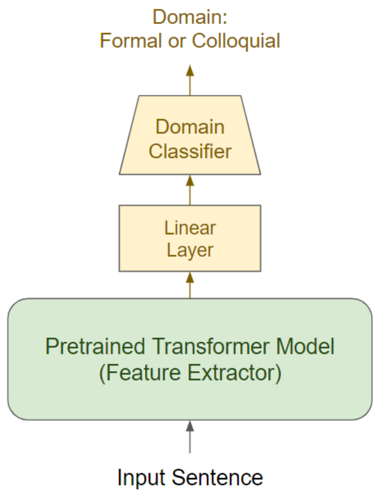


Fig. 1. The schema of formal/colloquial classifier

### B. Formal/Colloquial Classifier

The schema of the model is given in Fig. 1. We initialize the feature extractor model with ParsBERT weights [11]. We also randomly initialize the weights of the linear layer and the output classifier. The crawled dataset was split into two sets of training including 30000 training data (15000 formal and 15000 colloquial) and 10000 testing data (5000 formal and 5000 colloquial). Other hyperparameters and details of the training are given in Table III. Results are shown in Table IV.

Table III. Details of formal/colloquial classifier training

| | |
|---|---|
| Learning rate | 10e-6 |
| Epochs | 10 |
| GPU Details | Tesla T4 (15109MiB) |
| GPU Numbers | 1 |

Table IV. Results of formal/colloquial classifier test

| Data set | Num sentences | F1 |
|---|---|---|
| Train | 30000 | 99 |
| Test | 10000 | 97 |

### C. Weakly Supervised Labeled Dataset

For weak labeling the dataset with HesNegar, we sum the average of the positive and negative weight of the sentence words individually. If the average weight of positive is more, we take the sentence positive and otherwise we take the sentence negative.

The pseudocode of the above weak labeling method is given in Fig. 2.

Using this weak supervision labeling method, we labeled the test sets extracted from the Sentipers. The results are given in Table V. As we would expect from a weakly supervised dataset, its accuracy is far from ideal human performance.

**Weakly Supervised Sentence Polarity Labeling Pseudo Code:**

```
Sentence_positivity = 0
Sentence_negativity = 0
For each Word in Sentence:
        Sentence_positivity = Sentence_positivity +
                positive weight of Word in Hes Negar
        Sentence_negativity = Sentence_negativity +
                negative weight of Word in Hes Negar
Sentence_positivity = Sentence_positivity / (length of Sentence)
Sentence_negativity = Sentence_negativity / (length of Sentence)

If Sentence_positivity > Sentence_negativity:
        Polarity of Sentence is Positive
Else:
        Polarity of Sentence is Negative
```

Fig. 2. Pseudo Code of Weakly Supervised Polarity Labeling

Table V. Results of weakly supervised labeling method test

| Test Set Domain | Num of Sentences | F1 | Accuracy |
|---|---|---|---|
| colloquial | 501 | 59% | 70٪ |
| Formal | 962 | 57% | 70٪ |

### *D. The Proposed Architecture and Training Procedure*

The schema of the proposed model is given in Fig. 3 and the details of architecture and training procedure are shown in Table VI.

Our proposed training procedure involves the following steps:

- Initial preparation. In this step, we initialize the feature extractor model with ParsBERT pre-trained weights [11]. We randomly initialize the other weights as well.

- Step one. In the first step, we use the weakly supervised dataset. We run a multitasking process in which the feature extractor model (ParsBERT) and the right and left heads of the model are trained simultaneously. The right head tends to predict the domain of the sentence (while the GRL layer is adversarially blocking the domain prediction) and the left head tends to predict the sentiment class of the sentence.
  We first set the value of λ to zero and gradually increase it to 1 according to (1) (taken from [15]).

  t: Percentage of training progress

$$\lambda = \frac{2}{1 + exp\,(-t)} - 1 \tag{1}$$

  The loss function of this step was designed according to (2)

  F: output of the left head (probability distribution on sentiment classes)
  G: output of the right head (probability distribution on domain classes)
  $y_i$: sentiment label of sentence (binary)
  $d_i$: domain label of sentence (binary)

$$Loss_{Sentiment\ Cls} = -\frac{1}{N}\sum_{i=1}^{N} y_i \,.\log\left(F_{y_i}\right) + (1 - y_i).\log\left(1 - F_{1-y_i}\right)$$

$$Loss_{Domain\ Cls} = -\frac{1}{N}\sum_{i=1}^{N} d_i \,.\, log\left(G_{d_i}\right) + (1 - d_i)\,.\, log\left(1 - G_{1-d_i}\right)$$

$$Loss_{Feature\ Extractor} = L_{Sentiment\ Cls} - \lambda\, L_{Domain\ Cls}$$

(2)

- Step two. In this step, we use the modified training data set extracted from the Sentipers (including 50 formal training data) and only train the feature extractor model and the left head. (In this step, we throw away the right head)
  The loss function of this step was designed according to (3)

$$Loss_{Feature\ Extractor} = Loss_{Sentiment\ Cls} \tag{3}$$

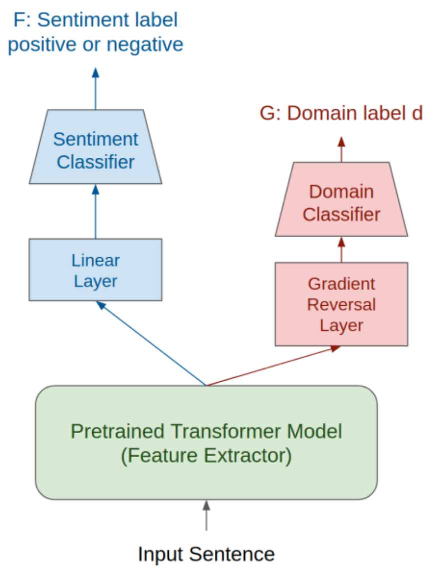


Fig. 3. The schema of the proposed model

Table VI. Details of proposed architecture and training procedure

| | |
|---|---|
| Number of parameters | 178 M |
| Learning rate | Pre-trained encoder: 10e-6<br>Sentiment & domain heads: 10e-5 |
| Epochs | Step one: 3<br>Step two: 1 |
| GPU Details | Tesla T4 (15109MiB) |
| GPU Numbers | 1 |

### *E. Results*

The result of our proposed method and other compared methods are shown in Table VII. Also, the confusion matrix for each domain is given in Fig. 4 and Fig. 5.

- In the first and second row, we just train respectively mBERT and ParsBERT pre-trained models without any addition.
- In the first three rows model trains in a low resource setting (with just 50 human-labeled formal sentences)

- In the fourth row, ParsBERT trains with all of the sentipers formal data (2884 sentences) so there are no low resource conditions for formal domain in this row.
- In the fifth row, ParsBERT trains with all of sentipers data (1501 colloquial sentences and 2884 formal sentences) so there are no low resource conditions for formal and colloquial domain in this row.
- The test data was the same for all rows.

Table VII. Results

| # | method | F1 colloquial | F1 formal | Accuracy colloquial | Accuracy formal |
|---|---|---|---|---|---|
| 1 | Multi lingual BERT | 52 | 71 | 53 | 78 |
| 2 | Pars BERT | 54 | 77 | 56 | 85 |
| 3 | **Our Method** | **69** | **80** | **77** | **87** |
| *4* | *Full data (Formal)* | *76* | *84* | *84* | *91* |
| *5* | *Full data* | *83* | *86* | *88* | *92* |

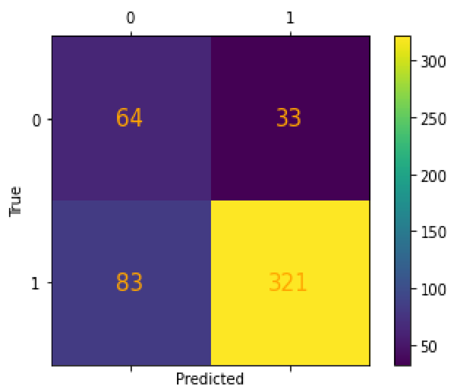


Fig. 4. Confusion matrix for colloquial domain

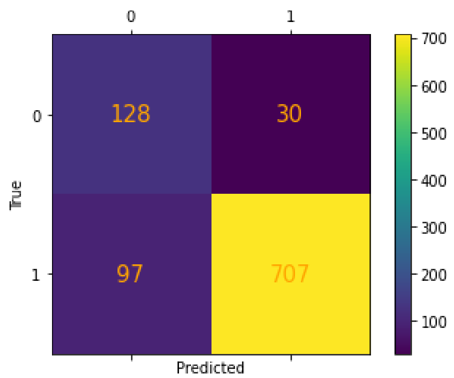


Fig. 5. Confusion matrix for formal domain

## IV. Conclusions and Future Work

In this article, we introduced a method that reduces the gap between the presence and absence of large amounts of data by taking advantage of adversarial training and weak supervision.

According to the results in Table VII, the F1 of the ParsBERT model is 54% (on colloquial) when we have only 50 formal data. The same model (ParsBERT), when trained with all of sentipers data (including 4385 formal and colloquial sentences), reaches an F1 of 83% (on colloquial), indicating a 29% gap between the presence and absence of abundant data.

However, our method with only 50 formal data has raised the F1 to 69%, which represents a 15% increase, halving the distance to the model trained with all of sentipers data.

Future works that will help to improve the above method will include the use of auxiliary tasks in the form of a multi-task learning architecture such as classification of POS, NER, or other language processing tasks. The use of cross-lingual embedding methods is another method that can help reduce the dependency of the feature extractor to the domain and generate better domain-independent representations of the sentence.

## References

[1] B. Liu and L. Zhang, “A survey of opinion mining and sentiment analysis,” *Mining Text Data*, pp. 415–463, Aug. 2013, doi: 10.1007/978-1-4614-3223-4_13.

[2] J. Howard and S. Ruder, “ULM-FiT: Universal Language Model Fine-tuning for Text Classification,” in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia, Jul. 2018, pp. 328–339, [Online]. Available: https://www.aclweb.org/anthology/P18-1031.

[3] J. Devlin, M.-W. Chang, K. Lee, and 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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186, Accessed: Jun. 30, 2019. [Online]. Available: https://aclweb.org/anthology/papers/N/N19/N19-1423/.

[4] M. Peters *et al.*, “ELMO: Deep Contextualized Word Representations,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, Louisiana, Jun. 2018, pp. 2227–2237, doi: 10.18653/v1/N18-1202.

[5] M. Peters *et al.*, “Deep Contextualized Word Representations,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, Louisiana, Jun. 2018, pp. 2227–2237, doi: 10.18653/v1/N18-1202.

[6] X. Liu, P. He, W. Chen, and J. Gao, “Multi-Task Deep Neural Networks for Natural Language Understanding,” *arXiv:1901.11504 [cs]*, Jan. 2019, Accessed: Jul. 31, 2019. [Online]. Available: http://arxiv.org/abs/1901.11504.

[7] C. Raffel *et al.*, “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer,” *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020.

[8] T. Brown *et al.*, “Language Models are Few-Shot Learners,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.

[9] K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, “ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators,” *ICLR*, 2020.

[10] E. Taher, S. A. Hoseini, and M. Shamsfard, “Beheshti-NER: Persian named entity recognition Using BERT,” in *Proceedings of The First International Workshop on NLP Solutions for Under Resourced Languages (NSURL 2019) co-located with ICNLSP 2019 - Short Papers*, Trento, Italy, 2019, pp. 37–42, Accessed: Mar. 19, 2021. [Online]. Available: https://www.aclweb.org/anthology/2019.nsurl-1.6.

[11] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, “ParsBERT: Transformer-based Model for Persian Language Understanding,” *arXiv:2005.12515 [cs]*, May 2020, Accessed: Mar. 19, 2021. [Online]. Available: http://arxiv.org/abs/2005.12515.

[12] I. Goodfellow *et al.*, “Generative Adversarial Nets,” in *Advances in Neural Information Processing Systems*, 2014, vol. 27, [Online]. Available: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

[13] T. Miyato, A. M. Dai, and I. Goodfellow, “Adversarial Training Methods for Semi-Supervised Text Classification,” *arXiv:1605.07725 [cs, stat]*, May 2017, Accessed: Mar. 19, 2021. [Online]. Available: http://arxiv.org/abs/1605.07725.

[14] Y. Ganin *et al.*, “Domain-Adversarial Training of Neural Networks,” in *Domain Adaptation in Computer Vision Applications*, G. Csurka, Ed. Cham: Springer International Publishing, 2017, pp. 189–209.

[15] Y. Ganin and V. Lempitsky, “Unsupervised Domain Adaptation by Backpropagation,” in *International Conference on Machine Learning*,

Jun. 2015, pp. 1180–1189, Accessed: Mar. 19, 2021. [Online]. Available: http://proceedings.mlr.press/v37/ganin15.html.

[16] G. Vernikos, K. Margatina, A. Chronopoulou, and I. Androutsopoulos, "Domain Adversarial Fine-Tuning as an Effective Regularizer," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, Online, Nov. 2020, pp. 3103–3112, doi: 10.18653/v1/2020.findings-emnlp.278.

[17] D. Ghosal, D. Hazarika, A. Roy, N. Majumder, R. Mihalcea, and S. Poria, "KinGDOM: Knowledge-Guided DOMain Adaptation for Sentiment Analysis," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, Jul. 2020, pp. 3198–3210, doi: 10.18653/v1/2020.acl-main.292.

[18] A. Naik and C. Rose, "Towards Open Domain Event Trigger Identification using Adversarial Domain Adaptation," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, Jul. 2020, pp. 7618–7624, doi: 10.18653/v1/2020.acl-main.681.

[19] S. Ruder and B. Plank, "Strong Baselines for Neural Semi-Supervised Learning under Domain Shift," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia, Jul. 2018, pp. 1044–1054, doi: 10.18653/v1/P18-1096.

[20] X. Cui and D. Bollegala, "Self-Adaptation for Unsupervised Domain Adaptation," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, Varna, Bulgaria, Sep. 2019, pp. 213–222, doi: 10.26615/978-954-452-056-4_025.

[21] S. Gururangan *et al.*, "Don't Stop Pretraining: Adapt Language Models to Domains and Tasks," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, Jul. 2020, pp. 8342–8360, doi: 10.18653/v1/2020.acl-main.740.

[22] C. Jia, X. Liang, and Y. Zhang, "Cross-Domain NER using Cross-Domain Language Modeling," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, Jul. 2019, pp. 2464–2474, doi: 10.18653/v1/P19-1236.

[23] P. Zhao and B. Zhou, "Discriminative Adversarial Domain Adaptation Based on Circle Loss," in *2020 International Conference on Communications, Information System and Computer Engineering (CISCE)*, Jul. 2020, pp. 30–33, doi: 10.1109/CISCE50729.2020.00012.

[24] K. Lim, J. Y. Lee, J. Carbonell, and T. Poibeau, "Semi-Supervised Learning on Meta Structure: Multi-Task Tagging and Parsing in Low-Resource Scenarios," *AAAI*, vol. 34, no. 05, Art. no. 05, Apr. 2020, doi: 10.1609/aaai.v34i05.6351.

[25] E. Ben-David, C. Rabinovitz, and R. Reichart, "PERL: Pivot-based Domain Adaptation for Pre-trained Deep Contextualized Embedding Models," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 504–521, 2020, doi: 10.1162/tacl_a_00328.

[26] H. Guo, R. Pasunuru, and M. Bansal, "Multi-Source Domain Adaptation for Text Classification via DistanceNet-Bandits," *AAAI*, 2020, doi: 10.1609/AAAI.V34I05.6288.

[27] P. Hosseini, A. Ramaki, H. Maleki, M. Anvari, and S. A. Mirroshandel, "SentiPers: A Sentiment Analysis Corpus for Persian," *ArXiv*, 2018.

[28] J. P. R. Sharami, P. A. Sarabestani, and S. A. Mirroshandel, "DeepSentiPers: Novel Deep Learning Models Trained Over Proposed Augmented Persian Sentiment Corpus," *ArXiv*, 2020.

[29] E. Askariyan, M. Kahani, and S. Sharifi, "HesNegar: persian sentiment vocabulary network," *Signal and Data Processing*, vol. 15, no. 1, pp. 71–86, Jun. 2018, doi: 10.29252/jsdp.15.1.71.