

Article

Few-Shot Methods for Aspect-Level Sentiment Analysis

Aleksander Wawer 

Institute of Computer Science, Polish Academy of Sciences, Jana Kazimierza 5, 01-248 Warsaw, Poland; axw@ipipan.waw.pl

Abstract: In this paper, we explore the approaches to the problem of cross-domain few-shot classification of sentiment aspects. By cross-domain few-shot, we mean a setting where the model is trained on large data in one domain (for example, hotel reviews) and is intended to perform on another (for example, restaurant reviews) with only a few labelled examples in the target domain. We start with pre-trained monolingual language models. Using the Polish language dataset AspectEmo, we compare model training using standard gradient-based learning to a zero-shot approach and two dedicated few-shot methods: ProtoNet and NNShot. We find both dedicated methods much superior to both gradient learning and zero-shot setup, with a small advantage held by NNShot. Overall, we find few-shot to be a compelling alternative, achieving a surprising amount of performance compared to gradient training on full-size data.

Keywords: few-shot; cross-domain; aspect-based sentiment analysis

1. Introduction

1.1. Cross-Domain Sentiment Analysis

Sentiment analysis is a process of computationally identifying and categorizing opinions expressed in texts to determine their polarities. Despite a certain set of words universally used to express polar attitudes (usually polar adjectives, e.g., ‘good’ or ‘bad’), sentiment analysis techniques involve dealing with a highly diverse vocabulary, often very domain-specific. In natural language processing, domains refer to different areas or text types where language is used, such as news articles, social media, legal documents, or hotel reviews.

The recent success of computational sentiment analysis can be attributed to several factors, such as using deep transformer neural networks [1] and taking advantage of the vast amounts of labelled data for model training. However, in most scenarios, the labelled data are limited to one or just a few domains and do not cover all possible applications.

Cross-domain text sentiment analysis refers to using sentiment-labelled data in the source domain to train a sentiment classifier and use it to classify sentiment in the target domain data.

Cross-domain sentiment analysis poses a difficult problem for natural language processing techniques due to the following reasons:

- Domain-dependent expressions. Each domain has its own set of domain-specific words and phrases. A word with a positive sentiment in one domain might be neutral or negative in another. This difference in vocabulary makes it hard for models trained on one domain to accurately interpret sentiment in another [2]. This challenge is also called sparsity [3].
- Feature distribution. The distribution of words and phrases vary across domains. This poses a problem for models trained in one domain to generalize well to another [4,5]. In [3], this challenge is called feature divergence.
- Context variation. A phrase that indicates positive sentiment in one domain might be interpreted differently in another due to changes in context. For instance, “easy”



Citation: Wawer, A. Few-Shot Methods for Aspect-Level Sentiment Analysis. *Information* **2024**, *15*, 664. <https://doi.org/10.3390/info15110664>

Academic Editor: Diego Reforgiato Recupero

Received: 3 September 2024

Revised: 8 October 2024

Accepted: 14 October 2024

Published: 22 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

could be positive in one domain but negative in another, called polarity divergence in [3].

Sentiment classification across domains is solved using multiple approaches. Examples of older techniques before 2017 include Spectral Feature Alignment, Structured Correspondence Learning, and the Joint Sentiment–Topic Model. A comprehensive review in [3] summarizes as many as 28 cross-domain sentiment analysis papers. Among newer methods, one can mention contrastive learning [6] and attempts to use large language models (LLMs) via prompts [7].

1.2. Aspect-Based Sentiment Analysis (ABSA)

Sentiment analysis can be performed at the document, sentence, or aspect levels. In document-level classification, sentiment is extracted from the entire opinionated text; the goal is to compute its overall polarity. In sentence-level analysis, classification is performed to label the overall dominating polarity of a sentence. In some cases, mainly when the input sentence includes phrases of opposing polarities, the decision as to which sentiment dominates the sentence is not easy and is determined by the compositional semantics of phrases, using rules driven by the syntactic structure [8].

Problems associated with sentence-level analysis have been addressed by introducing a more detailed level of sentiment analysis, namely aspect-based sentiment analysis (ABSA). It refers to determining the opinions or sentiments expressed on different features or aspects of entities. An aspect (also called feature) is an attribute or component of an entity, as for example, a phone's battery or a bank's customer service [9].

The papers directly addressing the combination of ABSA and a cross-domain scenario are quite limited in number. Notably, in [10], the authors propose a technique called CSC-PLDAT (Cross-domain Aspect-based Sentiment Classification with Prompt Learning and Domain Adversarial Training) based on a hybrid prompt comprising transferable and task-specific parts.

1.3. Few-Shot

In the broadest sense, few-shot learning involves showing a model a few samples so that it can be used to process unseen input, for example, in a classification task. This differs from supervised training on large labelled datasets and does not involve gradient updates.

Few-shot techniques are often applied using prompts when working with generative large language models. Specifically, a prompt may include a few examples for a model to learn from few-shot demonstrations specified purely via text interaction with the model. It was demonstrated that scaling up language models significantly improves task-agnostic few-shot performance [11].

Few-shot learning in generative, prompt-based scenarios requires dedicated prompts. Methods of effective prompt-tuning for few-shot aspect-based sentiment analysis were described in [12], where the authors fine-tuned a T5 model with instructional prompts, covering all the ABSA sub-tasks.

In this paper, we use the few-shot approach in a cross-domain setting. In this setup, we have a small sample of labelled examples in the target domain and a model trained on the source domain data. Retraining the model using gradient techniques to accommodate the small sample data is often not feasible, hence the need for a machine learning framework that enables the pre-trained model to generalize over new types of data that the pre-trained model has not seen during training. Utilizing this sample data efficiently to boost the model performance in the target domain is the goal of few-shot learning techniques.

In our work, we will use a few-shot framework based on support and query sets, and backbone encoders as described in [13]. The concepts can be defined as follows:

- Support set: A labelled sample of novel data type, which a pre-trained model will use to generalize from to classify examples in the query set.
- Query set: This consists of the unlabelled samples for evaluation; it may contain both the new and old types of data on which the model needs to generalize using previously acquired knowledge and information gained from the support set.

Figure 1 illustrates the operation in few-shot mode. It can be described as follows. In the first step, based on a large set of labelled data in the source domain, a backbone model is trained to calculate object embeddings (here, vectors reflecting represented texts). In the second step, this model is used to calculate the embedding vectors of the support and query sets. The support set comprises a small number of labelled examples for each class or task, while the query set consists of unlabeled examples used for evaluation. The query set is used to predict labels. The predictions are made by a similarity function which compares the input query embeddings with the support set embeddings. The label of each item in the query set comes from (is copied from) the most similar item in the support set. See Figure 2 for an example of this process.

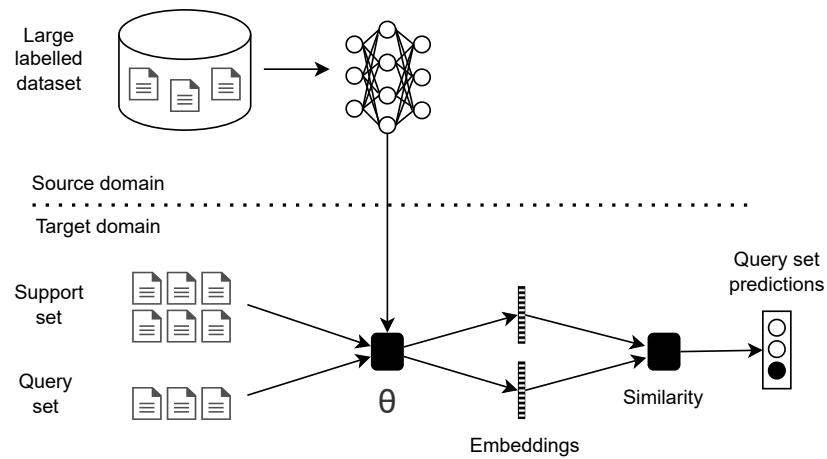


Figure 1. Cross-domain few-shot learning diagram.

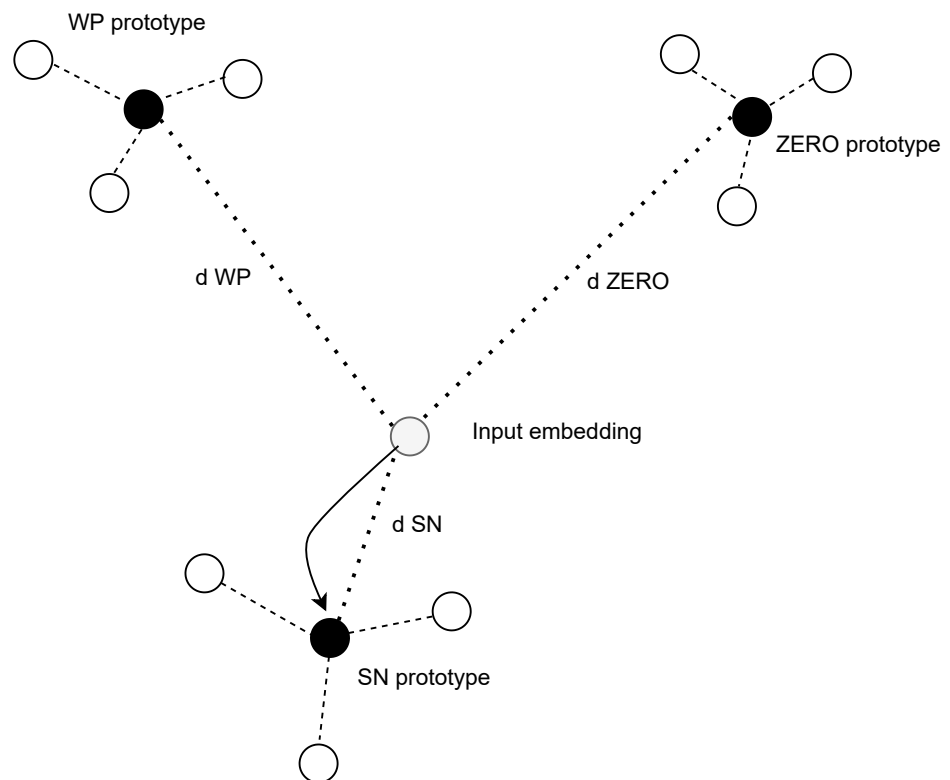


Figure 2. Prototypical networks.

1.4. Motivation

Recent advances in few-shot machine learning have demonstrated impressive capabilities, especially with the largest language models. However, these advancements typically rely on massive models with billions of parameters, which require extensive computational resources and multiple GPUs. Specifically, few-shot capabilities were demonstrated using GPT-3, an autoregressive language model with 175 billion parameters [11].

This paper focuses on developing methods that enable smaller models (of hundreds of millions of parameters) to perform few-shot learning effectively. The benefits include a reduced reliance on costly infrastructure, more energy-efficient computation, and possible usage in scenarios of low latency and on-device processing. In short, this research has the potential to bridge the gap between cutting-edge AI capabilities and the need for practical, accessible, and efficient solutions when applied to the problems of cross-domain aspect-level sentiment analysis.

This paper aims to answer the following research questions:

- What is the cross-domain performance of large language models fine-tuned on aspect-level sentiment analysis task?
- What is the performance of further fine-tuning such models using traditional gradient-based learning on small labelled samples in the target domain?
- What is the performance of dedicated few-shot methods, when applied to models fine-tuned on aspect-level sentiment analysis data from other domains, using small labelled samples in the target domain?

2. Data

2.1. AspectEmo

As the large labelled dataset in the source domain, we used AspectEmo 1.0 Corpus [14]. It is an extended version of the publicly available PolEmo 2.0 corpus of Polish customer reviews. It was used to train and test sentiment analysis models. The AspectEmo corpus consists of four subcorpora, each containing online customer reviews.

- The *school* subcorpus consists of 1000 documents (94,642 tokens, approximately 95 tokens per text). Texts come from a discussion forum and are opinions on university courses and lecturers.
- The *medicine* subcorpus consists of 3510 documents (478,505 tokens, approximately 136 tokens per text). The texts were typed by patients on a website intended to help patients find a good doctor.
- The *hotels* subcorpus consists of 4200 documents (578,259 tokens, approximately 137 tokens per text). Texts originate from the English version of tripadvisor.com.
- The *products* subcorpus consists of 1000 documents (135,217 tokens, approximately 135 tokens per text). The texts come from a price comparison website.

All tokens in the corpus are annotated at the aspect level with sentiment categories. Table 1 contains their description and their frequencies, as reported in [14].

Table 1. Token class labels and their frequencies, as reported in [14], TABLE III.

Label	Description	# of Tokens
SN	strong negative	8046
WN	weak negative	933
ZERO	neutral	6685
WP	weak positive	855
SP	strong positive	9109
AMB	ambiguous	744

The ambiguous label (AMB) refers to aspect tokens whose polarity cannot be easily determined due to unclear wording or sentence structure.

The dataset is publicly available <https://clarin-pl.eu/dspace/handle/11321/849> (accessed on 13 October 2024) and contains 1465 texts.

2.2. Few-Shot Cross-Domain Samples

AspectEmo corpus is limited to the four types of reviews listed above. To explore out-of-domain performance and few-shot methods, we prepared small samples of reviews annotated using the same guidelines as applied to the main AspectEmo corpus. The dataset consists of text snippets from the following domains (the number is given in parentheses):

- Fitness clubs (51);
- Movies (68);
- Restaurants (65).

These are fragments of texts downloaded from appropriate thematic websites of the Polish-language Internet. They are available at the same GitHub repository as the source code of experiments.

3. Methods

The basic steps of this work can be described as follows:

1. Training the base model in the source domain on the AspectEmo dataset (AspectEmo-HerBERT).
2. Testing on the target domain using various methods, including domain adaptation and few-shot learning solutions.

3.1. AspectEmo-HerBERT

Due to the best results achieved in [14], the HerBERT model [15] was selected for experiments. HerBERT is a BERT-based Language Model trained on Polish corpora using Masked Language Modelling (MLM) and Sentence Structural Objective (SSO) with the dynamic masking of whole words.

A variant of the HerBERT model adapted to the AspectEmo dataset is named AspectEmo-HerBERT. The number of training epochs was set to 20, and the hyperparameter values were consistent with those given in [14]. The testing part of the AspectEmo collection was used to select the best model (as validation data).

Based on the best results on less frequent classes (as measured on the validation set, here, the 'test' part of the AspectEmo), a model snapshot from epoch 6 was selected for further experiments as the best performing.

The results are partially linked to the class frequencies reported in Table 1. Both strong classes, SN and SP, are the most abundant in the data, and the models have less difficulty in recognizing them. Both weak classes were more problematic, while the lowest performance was observed for the AMB label, which was not only inherently tricky for humans but also low in frequency.

Both weak classes were more problematic, while the lowest performance was observed for the AMB label, which was not only inherently difficult for humans, but also low in frequency.

Our obtained results were generally comparable but worse between 2% and 10%, depending on label class, than the results reported in [14]. In the case of the F1 score, the comparison of our results to scores contained in TABLE VI of [14] is presented in Table 2. Provided that the hyperparameters are the same, it is difficult to explain this discrepancy. It is possible that the replication issue was caused by different versions of drivers, deep learning libraries or different versions of the dataset. Hopefully, this difference does not affect conclusions related to the few-shot results reported below.

Table 2. Class-level results (F1) of our AspectEmo-HerBERT model and the model reported in [14], TABLE VI.

	Our Model	Reported in [14]
AMB	11.29	15.65
SN	59.64	64.87
SP	72.44	77.42
WP	26.09	28.24
WN	25.37	28.85
ZERO	46.15	56.85

3.2. Gradient-Based Supervised Learning

The first and most obvious strategy for adapting the AspectEmo-HerBERT model to the target domain is to use regular gradient-based weight optimization. Initially, we attempted to use the same learning rate of 5×10^{-5} and a batch size of 16 as when training AspectEmo-HerBERT on the AspectEmo dataset. However, we observed that a larger learning rate of 1×10^{-4} is required for the training process to be effective.

In other words, this approach is based on regular training on the source domain data for 20 epochs and then continuing training for 3 epochs on the target domain data.

3.3. ProtoNet

The Prototypical Network [13] implementation in our experiments was inspired by [16]. We used the last hidden layer of the AspectEmo-HerBERT model for token representation. Following the Section 1.3, we divide the set in the target domain into two subsets:

- Support set—Based on this set, we count the representations of the embedding vector for each class (e.g., WP prototype, ZERO prototype, etc.) as the average of the embeddings of tokens belonging to a given class in the set S . This is an analogy to the training set.
- Query set—The evaluation of tokens in this set is carried out in a manner similar to that of nearest neighbours in relation to prototypes: by counting the distance of each embedding from the query set to the prototype embeddings of classes calculated on the support set.

An illustration of the process is shown in Figure 2. For a given input embedding, the classification decision is made based on selecting the minimum distance d to each class prototype. In this example case, the nearest distance was the d SN, towards the SN prototype.

3.4. NNShot

This approach is a modification of the ProtoNet proposed in [17] in that instead of computing averaged class-level prototypes, the nearest neighbour classification method is used for tokens. At inference, given a test example $x = \{x_i\}_T^1$ and a K-shot entity support set $S = \{(x_n, y_n)\}_{N=1}^{n=1}$ consisting of N text snippets, NNShot uses the AspectEmo-HerBERT encoder $f_\theta(x) = \hat{x}$ to obtain embedding representations of tokens. NNShot computes a similarity score between a token x in the test set and all tokens in the support set $\{x'\}$. Label y^* of a token x is assigned as a label of the support set token selected as follows:

$$y^* = \arg \min_{c \in \{1, \dots, C\}} d(\hat{x}, \hat{x}'_c) \quad (1)$$

where S_c is the support set token. Squared Euclidean distance is used for computing similarities between tokens in the nearest neighbour classification $d(\hat{x}, \hat{x}') = \|\hat{x} - \hat{x}'\|_2^2$.

In other words, NNShot computes the similarity between each token x in the query set and all tokens in the support set. It assigns the token x the label c , corresponding to the

most similar token in the support set. L2-normalization is performed on features before computing these distances [17].

4. Results

The evaluation results are presented at the level of token types (for example, AMB, SN, etc.) as described in Section 2.1.

4.1. Zero-Shot Setting

In this section, we present the performance of the AspectEmo-HerBERT in a zero-shot manner without any form of adaptation to the target domain. In other words, we test how the model trained on AspectEmo domains such as school, medicine, hotels and product reviews performs on fitness clubs, movies, and restaurant reviews. Table 3 contains the F1 score computed on each target domain using a model trained on the main AspectEmo corpus. The results are negative and show that token class recognition works only occasionally on WP, WN, and ZERO tokens.

Table 3. Results (F1) scores in a zero-shot setting.

	Fitness	Movies	Restaurants
AMB	0	0	0
SN	0	0	0
SP	0	0	0
WP	0.02	0.03	0.03
WN	0.01	0.03	0.02
ZERO	0.02	0.02	0.03
OVERALL	0.04	0.04	0.03

4.2. Gradient-Based Supervised Learning

Table 4 contains the results of an averaged F1 score from three experiments. In each of them, we randomly sampled 20 instances as a training set and 10 instances as evaluation data, and trained the model using the learning rate of 1×10^{-4} . Due to insufficient data size, we could not use a separate validation split to select the best model and only report the results obtained at the 10th epoch. In some cases, it was possible to recognize SN, SP, and WP labels. However, this procedure did not succeed in recognizing WN, AMB, and ZERO labels.

Table 4. Results (F1) of gradient-based supervised model fine-tuning using the target domain sample.

	Fitness	Movies	Restaurants
AMB	0	0	0
SN	0.21	0.33	0
SP	0.18	0.32	0.26
WP	0.24	0	0.02
WN	0	0	0
ZERO	0	0	0
OVERALL	0.18	0.20	0.08

4.3. ProtoNet

Table 5 contains the average F1 value from as many as twenty 10-shot experiments (Query set = 10 and Support set = 10), as well as the standard error of the mean, which measures how far the sample mean (average) of the data is likely to be from the true population mean. Here, we note that all of the token classes are much more often recognized correctly. Surprisingly, this also includes the most difficult AMB class.

Table 5. Results of ProtoNet: F1 score \pm standard error of the mean.

	Fitness	Movies	Restaurants
AMB	0.24 \pm 0.05	0.04 \pm 0.00	0.26 \pm 0.06
SN	0.46 \pm 0.1	0.28 \pm 0.06	0.31 \pm 0.07
SP	0.47 \pm 0.1	0.36 \pm 0.08	0.46 \pm 0.1
WP	0.05 \pm 0.01	0.16 \pm 0.03	0.09 \pm 0.02
WN	0.16 \pm 0.03	0.12 \pm 0.03	0.13 \pm 0.03
ZERO	0.22 \pm 0.04	0.17 \pm 0.04	0.13 \pm 0.03
OVERALL	0.34 \pm 0.08	0.24 \pm 0.05	0.31 \pm 0.07

4.4. NNShot

Table 6 contains the mean F1 value from twenty 10-shot experiments (Query set = 10 and Support set = 10), as well as the standard error of the mean.

The improvement over the gradient and zero-shot methods is similar to that obtained with ProtoNet, but the overall results are noticeably better.

Table 6. Results of NNShot: F1 score \pm standard error of the mean.

	Fitness	Movies	Restaurants
AMB	0.26 \pm 0.06	0.04 \pm 0.01	0.17 \pm 0.04
SN	0.43 \pm 0.1	0.36 \pm 0.08	0.32 \pm 0.07
SP	0.49 \pm 0.1	0.42 \pm 0.09	0.48 \pm 0.1
WP	0.35 \pm 0.08	0.19 \pm 0.04	0.24 \pm 0.05
WN	0.37 \pm 0.08	0.22 \pm 0.05	0.31 \pm 0.07
ZERO	0.27 \pm 0.06	0.26 \pm 0.06	0.18 \pm 0.04
OVERALL	0.4 \pm 0.08	0.30 \pm 0.07	0.33 \pm 0.07

5. Discussion

This paper uses few-shot techniques to explore approaches to cross-domain, aspect-level (ABSA) sentiment analysis.

We started by training a model on the AspectEmo dataset. We fine-tuned the HerBERT (named AspectEmo-HerBERT) model, reaching an overall F1 of 0.59 on the exact domains it was trained on. We achieved comparable yet slightly worse results than those reported in [14]. The difference can be explained by a different implementation of some parts of the model or the training process or by a more extensive tuning of hyperparameters. However, it does not significantly affect the results of the few-shot methods tested in our paper.

We started the cross-domain experiments by observing that the model trained on AspectEmo (domains such as school, medicine, hotels and products) does not generalize to target domains of fitness clubs, movies and restaurants. As indicated in Table 3, the overall F1 averaged over target domains in this scenario was as low as 0.04.

We then fine-tuned the AspectEmo-HerBERT model using a small number of observations (on the level of tens) from the target domains. As illustrated in Table 4, the overall F1 average over target domains improved significantly, with the F1 score of 0.08 to 0.20 depending on the target domain. While much better than the zero-shot scenario, the results were unsatisfactory, as some token classes were not recognized. Unsurprisingly, small datasets can lead to highly variable or unstable gradients, especially in complex models with many parameters. In addition, with very small training samples, regularization methods may not be as effective, possibly resulting in model overfitting.

We subsequently tested two few-shot methods. The overall results are much better, as demonstrated in Tables 5 and 6. ProtoNet achieved an overall average F1 score of 0.29 (0.24 to 0.34, depending on the domain) and NNShot as much as 0.34 (0.30 to 0.4). Both few-shot methods greatly improve the quality of predictions, with the NNShot approach having a

slight advantage, possibly due to the larger amount of information it takes into account. In NNShot, all known tokens are used in the labelling process, not just the averaged prototype of the class, as in ProtoNet.

Few-shot methods tested in this paper can significantly improve classification quality in a cross-domain scenario, requiring no more than a few dozen labelled training data points. This interesting property allows for automatic sentiment analysis without training datasets containing thousands of examples in the target domain but with well-performing models trained on data other than the target.

Apart from the backbone model used to compute the embeddings, few-shot methods do not introduce any parameters. They rely on the similarity between the embedding vectors and selecting the most similar candidate from the support set. This property ensures reducing the need for parameter selection. However, the challenge now lies in selecting the right support set.

6. Conclusions

In this paper, we address the relatively rarely undertaken problem of cross-domain sentiment analysis at the aspect level, focusing on data in the Polish language.

To begin with, we demonstrate that the recognition of aspect-based sentiment in an out-of-domain manner is a challenging problem for BERT-based models. We also show that gradient learning does not perform satisfactorily for small amounts of data in the target domain.

However, the results reported in this paper are optimistic: a small sample of tens of labelled text snippets in the target domain can help achieve two-thirds of the F1 score achieved in an in-domain scenario when using a dedicated few-shot algorithm.

Future work can include verifying how the few-shot methods described in this paper compare to the few-shot approach of generative large language models such as GPT and using prompting- instead of embedding-based similarity metrics.

Funding: Financed by the Institute of Computer Science, Polish Academy of Sciences.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Source code as a Python notebook, version 1.0, and data samples used to test the few-shot approach can be accessed at GitHub at <https://github.com/alexwz/few-shot-aspect-level> (accessed on 13 October 2024).

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Wang, A.; Singh, A.; Michael, J.; Hill, F.; Levy, O.; Bowman, S. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Brussels, Belgium, 2–4 November 2018; Linzen, T., Chrupała, G., Alishahi, A., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 353–355. [\[CrossRef\]](#)
2. Save, A.; Shekokar, N. Analysis of cross domain sentiment techniques. In Proceedings of the 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Mysuru, India, 15–16 December 2017; pp. 1–9. [\[CrossRef\]](#)
3. Al-Moslmi, T.; Omar, N.; Abdullah, S.; Albared, M. Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review. *IEEE Access* **2017**, *5*, 16173–16192. [\[CrossRef\]](#)
4. Pan, S.J.; Ni, X.; Sun, J.T.; Yang, Q.; Chen, Z. Cross-domain sentiment classification via spectral feature alignment. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 751–760.
5. Blitzer, J.; Dredze, M.; Pereira, F. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, 25–27 June 2007; pp. 440–447.

6. Luo, Y.; Guo, F.; Liu, Z.; Zhang, Y. Mere Contrastive Learning for Cross-Domain Sentiment Analysis. In Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea, 12–17 October 2022; Calzolari, N., Huang, C.R., Kim, H., Pustejovsky, J., Wanner, L., Choi, K.S., Ryu, P.M., Chen, H.H., Donatelli, L., Ji, H., et al., Eds.; International Committee on Computational Linguistics: New York, NY, USA, 2022; pp. 7099–7111.
7. Wu, H.; Shi, X. Adversarial Soft Prompt Tuning for Cross-Domain Sentiment Analysis. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; Muresan, S., Nakov, P., Villavicencio, A., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2022; pp. 2438–2447. [[CrossRef](#)]
8. Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A.; Potts, C. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; Yarowsky, D., Baldwin, T., Korhonen, A., Livescu, K., Bethard, S., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 1631–1642.
9. Hu, M.; Liu, B. Mining and summarizing customer reviews. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Seattle, WA, USA, 22–25 August 2004.
10. Yuan, S.; Li, M.; Du, Y.; Xie, Y. Cross-domain aspect-based sentiment classification with hybrid prompt. *Expert Syst. Appl.* **2024**, *255*, 124680. [[CrossRef](#)]
11. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. In *Proceedings of the Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 1877–1901.
12. Varia, S.; Wang, S.; Halder, K.; Vacareanu, R.; Ballesteros, M.; Benajiba, Y.; Anna John, N.; Anubhai, R.; Muresan, S.; Roth, D. Instruction Tuning for Few-Shot Aspect-Based Sentiment Analysis. In Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, Toronto, ON, Canada, 14 July 2023; pp. 19–27. [[CrossRef](#)]
13. Snell, J.; Swersky, K.; Zemel, R. Prototypical Networks for Few-shot Learning. In *Proceedings of the Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
14. Kocoń, J.; Radom, J.; Kaczmarz-Wawryk, E.; Wabnic, K.; Zajączkowska, A.; Zaśko-Zielińska, M. Aspectemo: Multi-domain corpus of consumer reviews for aspect-based sentiment analysis. In Proceedings of the 2021 International Conference on Data Mining Workshops (ICDMW), Auckland, New Zealand, 7–10 December 2021; pp. 166–173.
15. Mroczkowski, R.; Rybak, P.; Wróblewska, A.; Gawlik, I. HerBERT: Efficiently Pretrained Transformer-based Language Model for Polish. In Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing, Kyiv, Ukraine, 20 April 2021; pp. 1–10.
16. Ding, N.; Xu, G.; Chen, Y.; Wang, X.; Han, X.; Xie, P.; Zheng, H.; Liu, Z. Few-NERD: A Few-shot Named Entity Recognition Dataset. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Online, 1–6 August 2021; pp. 3198–3213. [[CrossRef](#)]
17. Yang, Y.; Katiyar, A. Simple and Effective Few-Shot Named Entity Recognition with Structured Nearest Neighbor Learning. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; pp. 6365–6375. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.