Towards Target-dependent Sentiment Classification in News Articles

Felix Hamborg[1,2][0000−0003−2444−8056], Karsten Donnay[2,3][0000−0002−9080−6539], and Bela Gipp[2,4][0000−0001−6522−3019]

[1] Dept. of Computer Science, University of Konstanz, Germany
felix.hamborg@uni-konstanz.de
[2] Heidelberg Academy of Sciences and Humanities, Germany
[3] Dept. of Political Science, University of Zurich, Switzerland
[4] Data and Knowledge Engineering, University of Wuppertal, Germany

Abstract. Extensive research on target-dependent sentiment classification (TSC) has led to strong classification performances in domains where authors tend to explicitly express sentiment about specific entities or topics, such as in reviews or on social media. We investigate TSC in news articles, a much less researched domain, despite the importance of news as an essential information source in individual and societal decision making. This article introduces NewsTSC, a manually annotated dataset to explore TSC on news articles. Investigating characteristics of sentiment in news and contrasting them to popular TSC domains, we find that sentiment in the news is expressed less explicitly, is more dependent on context and readership, and requires a greater degree of interpretation. In an extensive evaluation, we find that the current state-of-the-art in TSC performs worse on news articles than on other domains (average recall $AvgRec = 69.8$ on NewsTSC compared to $AvgRev = [75.6, 82.2]$ on established TSC datasets). Reasons include incorrectly resolved relation of target and sentiment-bearing phrases and off-context dependence. As a major improvement over previous news TSC, we find that BERT's natural language understanding capabilities capture the less explicit sentiment used in news articles.

Keywords: sentiment classification · stance detection · news bias · media bias.

1 Introduction

Target-dependent sentiment classification (TSC) is a sub-task of sentiment analysis that aims to identify the sentiment of a text, usually on sentence-level, towards a given target, such as named entities (NEs) or other semantic concepts [19]. Aspect-based sentiment classification (ABSC) [29], a closely related task, defines such targets as aspects of a given topic, e.g., "service" and "food" may be aspects of the topic "restaurant." Previous research on TSC and ABSC (due to their technical similarity we will refer to both as TSC) has focused mostly

on domains in which authors tend to express their opinions explicitly, such as reviews, surveys, and social media [7,26,24,29].

In this paper, we investigate TSC in the domain of news articles – a much less researched domain that is of critical relevance, especially in times of "fake news," echo chambers, and news owner centralization [15]. How persons and other entities are portrayed in articles on political topics is, e.g., very relevant for individual and societal opinion formation [3,17,14].

The main contributions of this paper are: (1) We introduce *NewsTSC*, a manually annotated dataset for the exploration of TSC in political news articles. (2) We discuss similarities and differences between political news and established TSC domains. (3) We perform an extensive evaluation of state-of-the-art TSC approaches on NewsTSC. To improve classification performance, we also fine-tune a BERT language model [6] on a large news dataset, thereby establishing the current state-of-the-art in TSC on political news.

We provide the dataset including code book, code to reproduce our experiments, and the fine-tuned BERT at: `https://github.com/fhamborg/newstsc`

2   Related Work

Most TSC-related research uses three annotated datasets: *Restaurant* and *Laptop*, containing reviews on restaurants and laptops [26], and *Twitter*, consisting of tweets [7]. Each example in these datasets consists of a *target*, *context* (often a single sentence or tweet), and the target's *sentiment* within its context.

The advent of word embeddings and deep learning including neural language models, such as BERT [6], has led to a performance leap in many natural language processing (NLP) disciplines including TSC, where, e.g., macro F1 gained from $F1_m = 63.3$ [21] to $F1_m = 75.8$ on the Twitter set [38]. Whereas traditional TSC research focused on careful feature engineering and dictionary creation (cf. [21]), researchers now focus on designing neural architectures suited to catch the relation between target and context [39,32,38]. By fine-tuning the underlying language model for the particular classification domain, performance can be improved further [28].

Text in news articles differs from reviews and social media in that news authors typically do not express sentiment towards a target explicitly (exceptions include opinion pieces and columns). Instead, they implicitly or indirectly express sentiment because language in news is expected to be neutral and journalists to be objective [1,10,15]. For example, news texts express sentiment by describing actions performed by a target, or by including and highlighting information in favor or against a target (or omitting and downplaying such information, respectively) [31]. Adding to the difficulty of news TSC, different readers may assess an article's sentiment towards a target differently [1], depending on their own political or ideological views (we discuss real-world examples in Section 3.3). Previous news TSC approaches mostly employ manually created [1] or semi-automatically extended [10] sentiment dictionaries. To our knowledge, there exist one dataset for evaluation of news TSC methods [34], which – perhaps due to

its small size ($N = 1274$) – has not been used or tested in recent TSC literature. Another dataset contains quotes extracted from news articles, since quotes more likely contain explicit sentiment ($N = 1592$) [1].

To our knowledge, no suitable datasets for news TSC exist nor have news TSC approaches been proposed that exploit recent advances in NLP.

3   Dataset

We describe how we create the news TSC dataset, including the collection of articles and the annotation procedure. Afterward, we discuss the characteristics of the dataset.

3.1   Data Collection and Example Extraction

We create a base set of articles of high diversity in topics covered and writing styles, e.g., whether emotional or factual words are used (cf. [8]). Using a news extractor [16], we collect news articles from the Common Crawl news crawl (CCNC, also known as CC-NEWS), consisting of over 250M articles until August 2019 [23]. To ensure diversity in writing styles, we select 14 US news outlets,[5] which are mostly major outlets that represent the political spectrum from left to right, based on selections by [4,13,2]. We cannot simply select the whole corpus, because CCNC lacks articles for some outlets and time frames. By selecting articles published between August 2017 and July 2019, we minimize such gaps while covering a time frame of two years, which is sufficiently large to include many diverse news topics. To facilitate the balanced contribution of each outlet and time-range we perform binning: we create 336 bins, one for each outlet and month, and randomly draw 10 articles reporting on politics for each bin, resulting in 3360 articles in total.[6] During binning, we remove any article duplicates by text equivalence.

To create examples for annotation, we select all mentions of NEs recognized as PERSON, NROP, or ORG for each article [37].[7] We discard NE mentions in sentences shorter than 50 characters. For each NE mention, we create an example by using the mention as the target and its surrounding sentence as its context. We remove any example duplicates. Afterward, to ensure diversity in writing styles and topics, we use the outlet-month binning described previously and randomly draw examples from each bin.

---

[5] BBC, Breitbart, Chicago Tribune, CNN, LA Daily News, Fox News, HuffPost, LA Times, NBC, NY Times, Reuters, USA Today, Washington Post, and Wall Street Journal.

[6] To classify whether an article reports on politics, we use a DistilBERT-based [30] classifier with a single dense layer and softmax trained on the HuffPost [22] and BBC datasets [12]. During the subsequent manual annotation, coders discard remaining, non-political articles.

[7] For this task, we use spaCy v2.1.

Different means may be used to address expected class imbalance, e.g., for the Twitter set, only examples that contained at least one word from a sentiment dictionary were annotated [25,24]. While doing so yields high frequencies of classes that are infrequent in real-world distribution, it also causes dataset shift and selection bias [27]. Thus, we instead investigate the effectiveness of different means to address class imbalance during training and evaluation (see Section 4).

3.2   Annotation

We set up an annotation process following best practices from the TSC literature [26,24,29,34]. For each example, we asked three coders to read the context, in which we visually highlighted the target and assess the target's sentiment. Examples were shown in random order to each coder. Coders could choose from *positive*, *neutral*, and *negative* polarity, whereby they were allowed to choose positive and negative polarity at the same time. Coders were asked to *reject* an example, e.g., if it was not political or a meaningless text fragment. Before, coders read a code book that included instructions on how to code and already annotated examples. Five coders, students, aged between 24 and 32, participated in the process.

In total, 3288 examples were annotated, from which we discard 125 (3.8%) that were rejected by at least one coder, resulting in 3163 non-rejected examples. From these, we discard 3.3% that lacked a majority class, i.e., examples where each coder assigned a different sentiment class, and 1.8% that were annotated as positive and negative sentiment at the same time, to allow for better comparison with previous TSC datasets and methods (see Section 2). Lastly, we split the remaining 3002 examples into 2301 training and 701 test examples. Table 1 shows class frequencies of the sets.

We use the full set of 3163 non-rejected examples to illustrate the degree of agreement between coders: 3.3% lack a majority class, for 62.7%, two coders assigned the same sentiment, and for 33.9% all coders agreed. On average, the accuracy of individual coders is $acc_h = 72.9\%$. We calculate two intercoder reliability (ICR) measures. For completeness, Cohen's Kappa is $\kappa = 25.1$, but it is unreliable in our case due to Kappa's sensitivity to class imbalance [5]. The mean pairwise observed agreement over all coders is 72.5.

	negative	neutral	positive	total
training	530	1600	171	2301
test	167	487	47	701
total	697	2087	218	3002

Table 1. Class frequencies of NewsTSC sets.

3.3   Characteristics of Sentiment in News Articles

In a manual, qualitative analysis of NewsTSC, we find two key differences of news compared to established domains: first, we confirm that news contains mostly implicit and indirect sentiment (see Section 2). Second, determining the sentiment in news articles typically requires a greater degree of interpretation (cf. [34]). The second difference is caused by multiple factors, particularly the implicitness of sentiment (mentioned as the first difference) and that sentiment in news articles is more often dependent on non-local, i.e., off-sentence, context. In the following, we discuss annotated examples (part of the dataset and discarded examples) to understand the characteristics of target-dependent sentiment in news texts.

We find that in news articles, a key means to express targeted sentiment is to describe actions performed by the target. This is in contrast, e.g., to product reviews where more often a target's feature, e.g., "high resolution", or the mention of the target itself, e.g., "the camera is awesome," express sentiment. For example, in "The Trump administration has worked tirelessly to impede a transition to a green economy with actions ranging from opening the long-protected Arctic National Wildlife Refuge to drilling, [...]." the target (underlined) was assigned negative sentiment due to its actions.

We find sentiment in $\approx 3\%$ of the examples to be strongly reader-dependent (cf. [1]).[8] In the previous example, the perceived sentiment may, in part, depend on the reader's own ideological or political stance, e.g., readers focusing on economic growth could perceive the described action positively whereas those concerned with environmental issues would perceive it negatively.

In some examples, targeted sentiment expressions can be interpreted differently due to ambiguity. As a consequence, we mostly find such examples in the discarded examples and thus they are not contained in NewsTSC. While this can be true for any domain (cf. "polarity ambiguity" in [26]), we think it is especially characteristic for news articles, which are lengthier than tweets and reviews, giving authors more ways to refer to non-local statements and to embed their arguments in larger argumentative structures. For instance, in "And it is true that even when using similar tactics, President Trump and President Obama have expressed very different attitudes toward immigration and espoused different goals." the target was assigned neutral sentiment. However, when considering this sentence in the context of its article [36], the target's sentiment may be shifted (slightly) negatively.

From a practical perspective, considering more context than only the current sentence seems to be an effective means to determine otherwise ambiguous sentiment expressions. By considering a broader context, e.g., the current sentence and previous sentences, annotators can get a more comprehensive understanding of the author's intention and the sentiment the author may have wanted to communicate. The greater degree of interpretation required to determine non-explicit

---

[8] We drew a random sample of 300 examples and concluded in a two-person discussion that the sentiment in 8 examples could be perceived differently.

sentiment expressions may naturally lead to a higher degree of subjectivity. Due to our majority-based consolidation method (see Section 3.2), examples with non-explicit or apparently ambiguous sentiment expressions are not contained in NewsTSC.

4   Experiments and Discussion

We evaluate three TSC methods that define the state-of-the-art on the established TSC datasets Laptop, Restaurant, and Twitter: AEN-BERT [32], BERT-SPC [6], and LCF-BERT [38]. Additionally, we test the methods using a domain-adapted language model, which we created by fine-tuning BERT (base, uncased) for 3 epochs on 10M English sentences sampled from CCNC (cf. [28]). For all methods, we test hyperparameter ranges suggested by their respective authors.[9] Additionally, we investigate the effects of two common measures to address class imbalance: weighted cross-entropy loss (using inverse class frequencies as weights) and oversampling of the training set. Of the training set, we use 2001 examples for training and 300 for validation.

We use average recall ($AvgRec$) as our primary measure, which was also chosen as the primary measure in the TSC task of the latest SemEval series, due to its robustness against class imbalance [29]. We also measure accuracy (acc), macro F1 ($F1_m$), and average F1 on positive and negative classes ($F1_{pn}$) to allow comparison to previous works [24].

Table 2 shows that LCF-BERT performs best ($AvgRec = 67.3$ using BERT and 69.8 using our news-adapted language model).[10] Class-weighted cross-entropy loss helps best to address class imbalance ($AvgRec = 69.8$ compared to 67.2 using oversampling and 64.6 without any measure).

Performance in news articles is significantly lower than in established domains, where the top model (LCF-BERT) yields in our experiments $AvgRev = 78.0$ (Laptop), 82.2 (Restaurant), and 75.6 (Twitter). For Laptop and Restaurant, we used domain-adapted language models [28]. News TSC accuracy $acc = 66.0$ is lower than single-human-level $acc_h = 72.9$ (see Section 3.3).

We carry out a manual error analysis (up to 30 randomly sampled examples for each true class). We find *target misassociation* as the most common error cause: in 40%, sentences express the predicted sentiment towards a different target. In 30%, we cannot find any apparent cause. The remaining cases contain various potential causes, including usage of euphemisms or sayings (12% of examples with negative sentiment). Infrequently, we find that sentiment is expressed by rare words or figurative speech, or is reader-dependent (the latter

---

[9] Epochs $\in \{3, 4\}$; batch size $\in \{16, 32\}$; learning rate $\in \{2e-5, 3e-5, 5e-5\}$; label smoothing regularization (LSR) [35]: $\epsilon \in \{0, 0.2\}$; dropout rate: 0.1; $\mathcal{L}_2$ regularization: $\lambda = 10^{-5}$. We use Adam optimization [20], Xaviar uniform initialization [9], and cross-entropy loss [11]. Where multiple values for a hyperparameter are given, we test all their combinations in an exhaustive search.

[10] Each row in Table 2 shows the results of the hyperparameters that performed best on the validation set.

LM	Method	AvgRec	acc	F1$_m$	F1$_{pn}$
base	AEN-BERT	59.7	62.9	55.0	47.3
	BERT-SPC	62.1	62.1	53.3	44.9
	LCF-BERT	**67.3**	61.3	54.4	46.5
news	AEN-BERT	59.8	62.9	54.5	46.2
	BERT-SPC	66.7	63.5	55.0	45.8
	LCF-BERT	**69.8**	66.0	58.8	51.4

Table 2. Experiment results. *LM* refers to the language model used, where *base* is BERT (base, uncased) and *news* is our fine-tuned BERT model.

in 2%, approximately matching the 3% of reader-dependent examples reported in Section 3.3).

Previous news TSC approaches, mostly dictionary-based, could not reliably classify implicit or indirect sentiment expressions (see Section 2). In contrast, our experiments indicate that BERT's language understanding suffices to interpret implicitly expressed sentiment correctly (cf. [6,1,10]). NewsTSC does not contain instances in which the broader context defines sentiment, since human coders could not classify them correctly in the first place. Our experiments therefore cannot elucidate this particular characteristic discussed in Section 3.3.

5   Conclusion and Future Work

We explore how target-dependent sentiment classification (TSC) can be applied to political news articles. Our main contributions are as follows: first, we introduce NewsTSC, a dataset to explore target-dependent sentiment classification (TSC) in political news articles, consisting of over 3000 manually annotated sentences.

Second, in a qualitative analysis, we find notable differences concerning how authors express sentiment towards targets as compared to other well-researched domains of TSC, such as product review or posts on social media. In these domains, authors tend to explicitly express their opinions. In contrast, in news articles, we find dominant use of implicit or indirect sentiment expressions, e.g., by describing actions, which were performed by a given target, and their consequences. Thus, sentiment expressions may be more ambiguous, and determining their polarity requires a greater degree of interpretation.

Third, in a quantitative evaluation, we find that state-of-the-art TSC methods perform lower on the news domain (average recall $AvgRec = 69.8$ using our news-adapted BERT model, $AvgRec = 67.3$ without) than on popular TSC domains ($AvgRec = [75.6, 82.2]$).

We identify multiple future research directions for news TSC. While NewsTSC contains clear sentiment expressions, it lacks other sentiment types that occur in real-world news coverage. For example, sentences that express sentiment more implicitly or ambiguously. To create a labeled TSC dataset that better

reflects real-world news coverage, we suggest to adjust annotation instructions to raise annotators' awareness of these sentiment types and clearly define how they should be labeled. Technically, apparently ambiguous sentiment expressions might be easier to label when considering a broader context, e.g., not only the current sentence but also previous sentences. Considering more context might also help to improve a classifier's performance.

We envision to integrate TSC methods into a system that identifies slanted news coverage [18,33]. For example, given a set of articles reporting on the same topic, a system could identify articles that similarly frame the actors involved in the event. To do so, the system would analyze frequently mentioned persons' polarities in each article. Then, it would group articles that similarly portray these persons.

References

1. Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., Pouliquen, B., Belyaeva, J.: Sentiment analysis in the news. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, Malta (2010)
2. Baum, M.A., Groeling, T.: New Media and the Polarization of American Political Discourse. Political Communication **25**(4), 345–365 (2008). https://doi.org/10.1080/10584600802426965
3. Bernhardt, D., Krasa, S., Polborn, M.: Political polarization and the electoral effects of media bias. Journal of Public Economics **92**(5), 1092–1104 (2008)
4. Budak, C., Goel, S., Rao, J.M.: Fair and balanced? Quantifying media bias through crowdsourced content analysis. Public Opinion Quarterly **80**(S1), 250–271 (2016). https://doi.org/10.1093/poq/nfw007
5. Cicchetti, D.V., Feinstein, A.R.: High agreement but low kappa: II. Resolving the paradoxes. Journal of Clinical Epidemiology **43**(6), 551–558 (1 1990). https://doi.org/10.1016/0895-4356(90)90159-M, `https://linkinghub.elsevier.com/retrieve/pii/089543569090159M`
6. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North. pp. 4171–4186. Association for Computational Linguistics, Stroudsburg, PA, USA (2019). https://doi.org/10.18653/v1/N19-1423, `http://aclweb.org/anthology/N19-1423`
7. Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., Xu, K.: Adaptive Recursive Neural Network for target-dependent Twitter sentiment classification. In: 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014. pp. 49–54. Baltimore, MD, USA (2014). https://doi.org/10.3115/v1/p14-2009
8. Gebhard, L., Hamborg, F.: The POLUSA Dataset: 0.9M Political News Articles Balanced by Time and Outlet Popularity. In: Proceedings of the ACM/IEEE

Joint Conference on Digital Libraries in 2020. pp. 467–468. ACM, New York, NY, USA (8 2020). https://doi.org/10.1145/3383583.3398567, `https://dl.acm.org/doi/10.1145/3383583.3398567`

9. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Journal of Machine Learning Research. pp. 249–256 (2010)
10. Godbole, N., Srinivasaiah, M., Skiena, S.: Large-Scale Sentiment Analysis for News and Blogs. In: Proceedings of the International Conference on Weblogs and Social Media (ICWSM). vol. 7, pp. 219–222. Boulder, CO, USA (2007)
11. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016), `http://www.deeplearningbook.org`
12. Greene, D., Cunningham, P.: Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering. In: Proceedings of the 23rd international conference on Machine learning. pp. 377–384. ACM Press (2006)
13. Groseclose, T., Milyo, J.: A measure of media bias. The Quarterly Journal of Economics **120**(4), 1191–1237 (2005). https://doi.org/10.1162/003355305775097542, `http://dx.doi.org/10.1162/003355305775097542`
14. Hamborg, F.: Media Bias, the Social Sciences, and NLP: Automating Frame Analyses to Identify Bias by Word Choice and Labeling. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. pp. 79–87. Association for Computational Linguistics, Stroudsburg, PA, USA (2020). https://doi.org/10.18653/v1/2020.acl-srw.12, `https://www.aclweb.org/anthology/2020.acl-srw.12`
15. Hamborg, F., Donnay, K., Gipp, B.: Automated identification of media bias in news articles: an interdisciplinary literature review. International Journal on Digital Libraries **20**(4), 391–415 (12 2019). https://doi.org/10.1007/s00799-018-0261-y, `http://link.springer.com/10.1007/s00799-018-0261-y`
16. Hamborg, F., Meuschke, N., Breitinger, C., Gipp, B.: news-please: A Generic News Crawler and Extractor. In: Proceedings of the 15th International Symposium of Information Science. pp. 218–223. Verlag Werner Hülsbusch (2017)
17. Hamborg, F., Zhukova, A., Donnay, K., Gipp, B.: Newsalyze: Enabling News Consumers to Understand Media Bias. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020. pp. 455–456. ACM, New York, NY, USA (8 2020). https://doi.org/10.1145/3383583.3398561, `https://dl.acm.org/doi/10.1145/3383583.3398561`
18. Hamborg, F., Zhukova, A., Gipp, B.: Automated Identification of Media Bias by Word Choice and Labeling in News Articles. In: 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL). pp. 196–205. IEEE, Champaign, IL, USA (6 2019). https://doi.org/10.1109/JCDL.2019.00036, `https://ieeexplore.ieee.org/document/8791197/`
19. Jiang, L., Yu, M., Zhou, M., Liu, X., Zhao, T.: Target-dependent Twitter sentiment classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. pp. 151–160. Portland, OR, USA (2011)
20. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. arXiv preprint arXiv: 1412.6980 (12 2014), `http://arxiv.org/abs/1412.6980`
21. Kiritchenko, S., Zhu, X., Cherry, C., Mohammad, S.: NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). pp. 437–442. Association for Computational Linguistics, Dublin, Ireland (2014). https://doi.org/10.3115/v1/s14-2076
22. Misra, R.: News Category Dataset (2018), `https://www.kaggle.com/rmisra/news-category-dataset`

23. Nagel, S.: Common Crawl: News Crawl (2016), `https://web.archive.org/web/20191118111519/https://commoncrawl.org/2016/10/news-dataset-available/`

24. Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F., Stoyanov, V.: SemEval-2016 Task 4: Sentiment Analysis in Twitter. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). pp. 1–18. Association for Computational Linguistics, San Diego, CA, USA (2016). https://doi.org/10.18653/v1/S16-1001

25. Nakov, P., Rosenthal, S., Kozareva, Z., Stoyanov, V., Ritter, A., Wilson, T.: SemEval-2013 Task 2: Sentiment Analysis in Twitter. In: Second Joint Conference on Lexical and Computational Semantics (SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). pp. 312–320. Association for Computational Linguistics, Atlanta, GA, USA (2013)

26. Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., Androutsopoulos, I.: SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). pp. 486–495. Association for Computational Linguistics, Stroudsburg, PA, USA (2015). https://doi.org/10.18653/v1/S15-2082, `http://aclweb.org/anthology/S15-2082`

27. Quionero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N.D.: Dataset Shift in Machine Learning. The MIT Press (2009)

28. Rietzler, A., Stabinger, S., Opitz, P., Engl, S.: Adapt or Get Left Behind: Domain Adaptation through BERT Language Model Finetuning for Aspect-Target Sentiment Classification. arXiv preprint arXiv:1908.11860 (8 2019), `http://arxiv.org/abs/1908.11860`

29. Rosenthal, S., Farra, N., Nakov, P.: SemEval-2017 Task 4: Sentiment Analysis in Twitter. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). pp. 502–518. Association for Computational Linguistics, Vancouver, Canada (2017). https://doi.org/10.18653/v1/s17-2088

30. Sanh, V., Debut, L., Chaumond, J., Wolf, T.: DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv: 1910.01108 (10 2019), `http://arxiv.org/abs/1910.01108`

31. Schreier, M.: Qualitative content analysis in practice. SAGE Publications (2012)

32. Song, Y., Wang, J., Jiang, T., Liu, Z., Rao, Y.: Targeted Sentiment Classification with Attentional Encoder Network. In: Artificial Neural Networks and Machine Learning - ICANN 2019: Text and Time Series. pp. 93–103. Springer International Publishing, Cham, US (2019). https://doi.org/10.1007/978-3-030-30490-4_9, `http://link.springer.com/10.1007/978-3-030-30490-4_9`

33. Spinde, T., Hamborg, F., Donnay, K., Becerra, A., Gipp, B.: Enabling News Consumers to View and Understand Biased News Coverage: A Study on the Perception and Visualization of Media Bias. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020. pp. 389–392. ACM, New York, NY, USA (8 2020). https://doi.org/10.1145/3383583.3398619, `https://dl.acm.org/doi/10.1145/3383583.3398619`

34. Steinberger, R., Hegele, S., Tanev, H., Della Rocca, L.: Large-scale news entity sentiment analysis. In: RANLP 2017 - Recent Advances in Natural Language Processing Meet Deep Learning. pp. 707–715. Incoma Ltd. Shoumen, Bulgaria (11 2017). https://doi.org/10.26615/978-954-452-049-6_091, `http://www.acl-bg.org/proceedings/2017/RANLP2017/pdf/RANLP091.pdf`

35. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE confer-

ence on computer vision and pattern recognition. pp. 2818–2826. IEEE (2016). https://doi.org/10.1109/CVPR.2016.308, `http://arxiv.org/abs/1512.00567`

36. Taub, A.: How Liberals Got Lost on the Story of Missing Children at the Border (2018), `https://web.archive.org/web/20191120151037/https://www.nytimes.com/2018/05/31/upshot/liberals-immigration-children-border-misinformation.html`

37. Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., Xue, N., Taylor, A., Kaufman, J., Franchini, M., El-Bachouti, M., Belvin, R., Houston, A.: OntoNotes 5.0 (2013), `https://web.archive.org/web/20190705173013/https://catalog.ldc.upenn.edu/LDC2013T19`

38. Zeng, B., Yang, H., Xu, R., Zhou, W., Han, X.: LCF: A Local Context Focus Mechanism for Aspect-Based Sentiment Classification. Applied Sciences **9**(16), 1–22 (8 2019). https://doi.org/10.3390/app9163389, `https://www.mdpi.com/2076-3417/9/16/3389`

39. Zhaoa, P., Houb, L., Wua, O.: Modeling Sentiment Dependencies with Graph Convolutional Networks for Aspect-level Sentiment Classification. arXiv preprint arXiv:1906.04501 (6 2019), `http://arxiv.org/abs/1906.04501`

Listing 1.1. Use the following `BibTeX` code to cite this article

```
@InProceedings{Hamborg2021,
  author    = {Hamborg, Felix and Donnay, Karsten and Gipp,
      Bela},
  title     = {Towards Target-dependent Sentiment
      Classification in News Articles},
  booktitle = {Proceedings of the iConference 2021},
  year      = {2021},
  month     = {Mar.},
  location  = {Beijing, China (Virtual Event)},
  topic     = {newsanalysis}
}
```