Literature Review:
Using Machine Learning Methods for Sentiment Analysis of Internet Text-based Data

Aaron Modiyil Joseph
22018497

University of Hertfordshire
16 April 2024

Abstract

Sentiment analysis has witnessed significant advancements with the advent of machine learning techniques. This literature review explores the application of various machine learning models. The review critically examines these methods, highlighting their strengths, limitations, and performance on different datasets and tasks within the domain of sentiment analysis. Additionally, it discusses the challenges associated with data collection, processing, and evaluation metrics, providing valuable insights for future research directions in this field.

1 Introduction

Over the last decade, Sentiment Analysis (SA) has emerged as a significant research focus in Natural Language Processing (NLP), which involves machine learning, text analysis, and computational linguistics to identify and extract subjective information from source materials. It is a powerful tool that allows organizations to detect sentiment in social data and understand customer attitudes towards products, services, or brand reputation, forming judgments and supporting wise political decision-making. Figure 1 illustrates this by showing the rising number of peer-reviewed articles on sentiment analysis published in SCOPUS every year (Venkit et al., 2023).

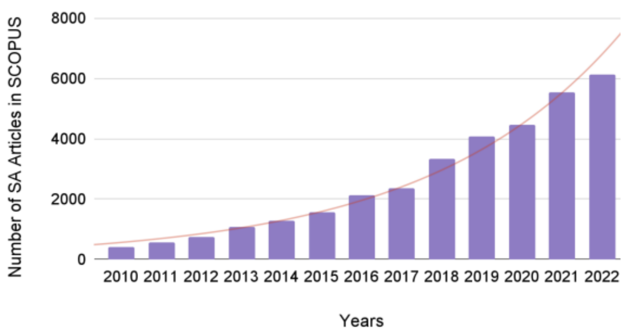

Figure 1: Number of articles published each year (from 2010 to 2022) in SCOPUS that contain the term ‘sentiment analysis’ in the title, abstract, or keywords (Venkit et al., 2023).

In recent years, machine learning (ML) techniques have been increasingly employed in SA. These techniques provide an automated approach to decipher the emotional tone behind words and offer valuable insights into the opinions, sentiments, and emotions of the individuals involved. This paper aims to explore various ML techniques used in SA and give a critical review of some of the literature published in the field.

We will delve into traditional ML methods such as Naive Bayes, Support Vector Machines, and Decision Trees, as well as more advanced techniques like Deep Learning and Neural Networks. The paper will also try to discuss SA methods, describe datasets used and the challenges involved in collecting and processing these datasets and discuss the metrics used to evaluate the ML models.

The motivation for research in this field stems from the increasing need to understand and interpret the vast amounts of unstructured data generated on the internet every day. SA offers a means of interpreting this data, allowing companies to learn more about the behaviour of their customers, enhance customer support, and make wise choices. Moreover, it is a significant area of research because it has applications in a variety of fields, including politics, healthcare, and finance.

Through this exploration, we aim to provide a comprehensive overview of how ML can be harnessed for sentiment analysis, thereby providing a valuable resource for researchers, practitioners, and enthusiasts in the field.

2 Methodology

We now chronologically explore various ML models published for SA. Wang et al. (2012) uses a Naive Bayes model on unigram features to classify tweets, messages of up to 140 characters posted on Twitter(a microblogging service), as negative, positive, neutral, or unsure and try to describe the public sentiment toward presidential candidates in the 2012 U.S. election. The features are calculated from the tokenization of the tweets, which preserves punctuation that may signify sentiment (e.g., emoticons and exclamation points) and Twitter-specific phenomena (e.g., extracting intact URLs). The sentiment model is trained on nearly 17,000 tweets labelled using crowd-sourced sentiment annotations from Amazon Mechanical Turk. The model achieves 59% accuracy surpassing the baseline of of classifying all the data as negative, the most prevalent sentiment category (56%). The choice of the model was not solely motivated by global accuracy but also took into account class-wise performance to ensure the model performs well in each sentiment category. A correlational analysis of aggregated sentiment with political events, news, and indicators such as poll and election results is then conducted to explore whether variations in Twitter sentiment and tweet volume are predictive or reflective of real-world events and public opinion.

Mohammad et al. (2013) creates two state-of-the-art Support Vector Machine (SVM) classifiers, one to detect the sentiment of messages such as tweets and SMS (message-level task) and one to detect the sentiment of a term within a message (term-level task) and classify as positive, negative, or neutral. A combination of existing, manually created sentiment lexicons(a list of words with associations to positive and negative sentiments) and new, tweet-specific, automatically generated sentiment lexicons are used to enhance sentiment analysis in tweets. The SVM classifier is trained on 9,912 annotated tweets, with 8,258 tweets in the training set and 1,654 in the development set. The model is then applied to the test sets of 3,813 unseen tweets and 2,094 SMS messages. The model is assessed with the macro-averaged F-score, which provides a balanced measure of the model's accuracy across different sentiment classes. The model obtains 69.02 on the tweet set and 68.46 on the SMS set when the baseline results of a majority classifier that always predicts the most frequent class as output is 29.19 and 19.03 on the tweet set and SMS set respectively. The submission ranked first in the SemEval-2013 competition.

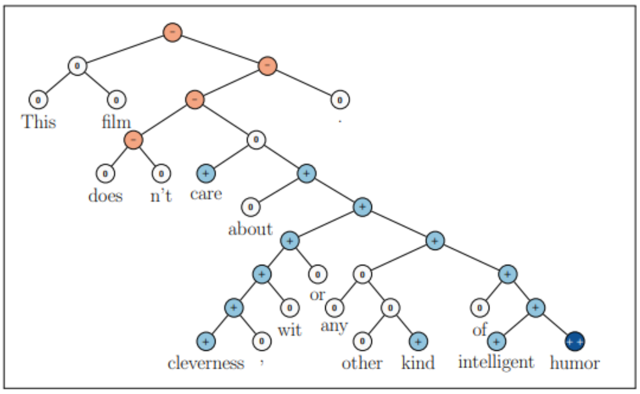

Figure 2: Example of Recursive Neural Tensor Network accurately predicting 5 sentiment classes, very negative to very positive (– –, –, 0, +, + +), at every node of a parse tree and capturing the negation and its scope in this sentence (Socher et al., 2013).

Socher et al. (2013) introduce the Recursive Neural Tensor Network (RNTN) for sentiment analysis. RNTN extends recursive neural networks by incorporating tensor-based interactions between words in a sentence to capture compositional relationships between words and phrases in a sentence more effectively. It builds a parse tree representation of a sentence and then applies tensor-based operations to learn complex interactions between words at different levels of the parse tree. The model is trained on a newly introduced Sentiment Treebank, which includes 11,855 single sentences from movie reviews from rottentomatoes.com. Each sentence is parsed and annotated with fine-grained sentiment labels by three human judges. The dataset contains 215,154 unique phrases extracted from these parse trees. The Sentiment Treebank is the first corpus with fully labelled parse trees that enable a detailed analysis of the compositional effects of sentiment in language. The granularity and size of the dataset allow for capturing complex linguistic phenomena and understanding the intricacies of sentiment analysis. The data provides a rich source for training and evaluating models of sentiment compositionality, with a focus on capturing the effects of sentiment at different levels of linguistic structure. This new dataset enables us to analyze sentiment and capture complex linguistic phenomena. Figure 2 depicts one of the many examples with a clear compositional structure. The model pushes the accuracy of single sentence positive/negative classification from 80% to 85.4%. Additionally, the RNTN accurately predicts fine-grained sentiment labels for all phrases with an accuracy of 80.7%, showing a substantial improvement over baseline methods. The model is also the only one till then capable of accurately capturing the effects of negation at various levels of the parse tree for both positive and negative phrases.

Dey et al. (2016) employed two supervised machine learning algorithms, Naïve Bayes' and K-Nearest Neighbour (K-NN), for sentiment classification of movie reviews and hotel reviews. Movie reviews were obtained from www.imdb.com, while hotel reviews were downloaded from the OpinRank Review Dataset. The dataset comprised 5000 positive and 5000 negative reviews from each source, totalling 10,000 reviews for analysis. Despite its simplicity, Naïve Bayes' outperformed K-NN in movie reviews, achieving over 80% accuracy. For hotel reviews, both classifiers yielded similar results with lower accuracies. Thus they conclude the Naïve Bayes' classifier can be used successfully to analyse movie reviews.

Baziotis et al. (2017) present two deep-learning models for sentiment analysis on Twitter data, a Message-level Sentiment Analysis (MSA) Model and a Topic-based Sentiment Analysis (TSA) Model. The authors collected a large unlabeled dataset of 330 million English Twitter messages from December 2012 to July 2016. This unlabeled dataset was used to calculate word statistics for the text preprocessing tool and to train the pre-trained word embeddings. The MSA model uses a 2-layer bidirectional Long short-term memory (BiLSTM) Recurrent neural network with an attention mechanism. The Attention layer assigns weights to the word annotations, allowing the model to focus on the most important words for sentiment expression. The TSA model Siamese BiLSTM network jointly represents the tweet and the topic.

Context-aware annotations are obtained by concatenating the tweet word annotations with the topic annotation and are used to highlight the words that express sentiment towards the given topic. MSA model ranked 1st (tie) in Subtask A (message-level sentiment classification) and TSA model ranked 2nd in Subtasks B, C, and D (topic-based sentiment classification and quantification) and ranked 11th out of 12 in Subtask E (topic-based sentiment quantification), which the authors attributed to their quantification approach in the SemEval-2017 competition. The attention mechanisms and the Siamese BiLSTM architecture with context-aware attention proved to be effective approaches for this task.

Ma et al. (2018) propose two key novel neural architectures for targeted aspect-based sentiment analysis (TABSA). The Hierarchical Attention Model consists of two attention mechanisms. The Target-Level Attention focuses on the most sentiment-relevant parts of the target expression. It computes a self-attention vector over the words in the target to generate a vector representation of the target. The Sentence-Level Attention identifies the most relevant words in the sentence concerning the target and a specific aspect. It computes an attention vector over all the words in the sentence to generate an aspect-specific sentence representation. The target and the aspect-specific sentence representation are then used for sentiment classification. The Sentic LSTM is an extension of the standard LSTM encoder that aims to better incorporate effective commonsense knowledge from the SenticNet knowledge base. The forget gate, input gate, and output gate are implemented to take the commonsense concept embeddings as additional input, allowing the model to control the flow of information from the concepts; and a knowledge output gate is introduced to output a concept-level representation that is combined with the token-level output of the LSTM. The models are evaluated on two datasets: SentiHood, which contains 5,215 sentences about locations in London, with 2,977 targets annotated for aspect-based sentiment. A SemEval-2015 subset containing 1,197 targets in the training set and 542 targets in the test set. The results show that the proposed hierarchical attention mechanism and the Sentic LSTM extension can effectively leverage the commonsense knowledge from SenticNet to better capture the sentiment and aspect information in the text.

Sun et al. (2019) Fine-tune pre-trained Bidirectional Encoder Representations from Transformers (BERT) for ABSA. The authors explore two main approaches: The BERT-single model, which directly fine-tunes BERT on the original single-sentence classification tasks, and the BERT-pair models, which transform the ABSA and TABSA tasks into sentence-pair classification problems. The authors construct auxiliary sentences from the target and aspect information by proposing QA-M: Generating a question-like auxiliary sentence without sentiment polarity information, NLI-M: a pseudo-sentence auxiliary without sentiment polarity information, QA-B: Generating a question-like auxiliary sentence with sentiment polarity information and NLI-B: Generating a pseudo-sentence auxiliary with sentiment polarity information. They use the SentiHood dataset and SemEval-2014 Task 4 dataset. On the SentiHood dataset, the BERT-pair models significantly outperform the BERT-single model and other state-of-the-art methods, achieving new state-of-the-art results on both aspect detection and sentiment classification. On the SemEval-2014 Task 4 dataset, the BERT-pair models also achieve better performance than the BERT-single model and previous state-of-the-art approaches. Among the BERT-pair models, the BERT-pair-NLI-B and BERT-pair-QA-B models generally perform the best, demonstrating the benefits of incorporating the sentiment polarity information into the auxiliary sentences.

Rietzler et al. (2019) employ a two-step machine learning approach to tackle the Aspect-Target Sentiment Classification (ATSC) task. First, they utilize self-supervised domain-specific language model fine-tuning on the BERT architecture by fine-tuning it on Yelp reviews for the restaurants domain and Amazon laptop reviews for the laptops domain, after filtering out overlapping reviews with the SemEval2014 datasets. This fine-tuning process involves two objectives - masked language modelling, where the model learns to predict randomly masked tokens, and next-sentence prediction, where the model learns to predict if a given sequence follows naturally from a preceding sequence. This domain-specific language model finetuning allows the BERT model to adapt to the target domains. In the second step, the authors take the domain-adapted BERT model and fine-tune it in a supervised manner on the downstream ATSC task, using the SemEval 2014 datasets. The input to the BERT model is formatted as a sequence pair, with the sentence and the aspect-target concatenated together. The model then predicts the sentiment polarity (positive, negative, neutral) towards the given aspect-target, using the final hidden representation of the special [CLS] token. The model effectively leverages both unsupervised domain knowledge and supervised task-specific information, leading to strong performance on the ATSC task. The detailed analysis and strong empirical results on benchmark datasets make this a compelling contribution to the field of aspect-based sentiment.

Wu and Ong (2020) propose two Context-Guided BERT (CG-BERT) model variants for TABSA and ABSA tasks. The first model, CG-BERT integrates context information into the BERT architecture by modifying the query and key matrices used for self-attention. The second model, Quasi-Attention CG-BERT (QACG-BERT), introduces a novel quasi-attention mechanism that allows the model to learn both additive and subtractive attention weights, further incorporating context information. Both CG-BERT and QACG-BERT are built upon the pre-trained BERT model and fine-tuned on the target TABSA and ABSA datasets. The results demonstrate the benefits of incorporating context-awareness into the BERT architecture, with the QACG-BERT model in particular showing significant improvements over previous approaches.

Hannak et al. (2021) propose bagged decision trees to predict aggregate sentiment on Twitter and correlate it with the weather at the time and location of the tweets to find the aggregate sentiment that follows distinct climate, temporal, and seasonal patterns. The researchers aggregated the tweets into hour-long buckets for each metropolitan area, taking the average sentiment of all tweets in a bucket as the sentiment score. They then use bagged decision trees as the machine learning algorithm, which involves constructing 1,000 decision trees, each trained on an independent bootstrap sample of the training data. The final prediction is the average of the 1,000 tree predictions. To measure the accuracy of the sentiment prediction, the authors use the Area under the Receiver Operating Characteristic (ROC) curve, which represents the probability that the predictor ranks time periods with more positive sentiment higher than time periods with more negative sentiment. They find that using all the input variables together (geography, season, time, and weather) results in a ROC area of 0.7857, indicating high predictive accuracy.

Wang et al. (2022) employed BERT to analyze the sentiment expressed in social media posts during the COVID-19 pandemic across over 100 countries in 65 languages. They introduce a comprehensive dataset of over 600 million geotagged social media posts from Twitter and Weibo (the Chinese microblogging platform) collected between January 1 and May 31, 2020. The data came from 10.56 million unique users across over 100 countries, covering approximately 74% of the world's population. To focus on general emotional states rather than reactions to the pandemic itself, the team excluded all posts containing COVID-19-related terms, after translating them into the 30 most common languages. They also employed a one-class classification approach to detect and remove Twitter bot accounts from the analysis. A simple logistic-regression classifier is trained on the first 100 principal component analysis dimensions of the Sentence-BERT social media post embeddings. The training data we used are a set of 1,600,000 tweets labelled as positive or negative52. Since representations are consistent across languages, we were able to train our sentiment classifier in English and predict sentiment in the 104 languages supported by Multilingual BERT27, which covers 65 identifiable languages on Twitter and Weibo. The model obtains a classification accuracy of 0.84 for English content and 0.75 on average in other languages

Li et al. (2023) use Sentic Graph Convolutional Network (GCN) for ABSA and investigate the effect of customer-generated content (i.e., online reviews) in predicting restaurant survival. The model is trained with the restaurant domain of SemEval series data and is utilized to calculate the sentiment polarity of aspect terms in the review text. The output of the model consists of aspect terms with sentiment polarity, which are then manually classified into aspect categories such as service and tastiness. It uses LSTM layers to acquire contextual representations and GCN layers to identify connections between contextual words in specific aspects. This approach considers the dependencies of contextual terms and aspect terms, as well as the effective information between opinion terms and aspect terms. The model can provide exact sentiment characteristics that correlate with multiple aspects, and it captures sentiment-related dependencies between contextual words and a given aspect with high accuracy. Additionally, the Sentic GCN algorithm enhances the sentiment representations by using SenticNet to adorn the graph

3 Results and Discussion

3.1 ML methods

We now have a critical discussion of the various ML methods used in SA. While earlier approaches such as Naive Bayes, K-NN, and SVMs contributed significantly to the field, they have been largely replaced by more advanced deep learning techniques.

Deep learning models, such as attention-based BiLSTM networks (Baziotis et al., 2017) and hierarchical attention models (Ma et al., 2018), have excelled at capturing contextual and compositional information, which is critical for accurate sentiment analysis. Furthermore, the introduction of the large pre-trained language model BERT has transformed the field, with fine-tuning approaches (Sun et al., 2019; Rietzler et al., 2019; Wu and Ong, 2020) yielding cutting-edge results on a variety of SA tasks, including aspect-based and targeted sentiment analysis.

However, it is important to note that traditional methods such as Naive Bayes, K-NN, and SVMs can still be useful in certain situations, especially when computational resources are limited or when dealing with specific domains or data types where simpler models perform well.

3.2 SA methods

Venkit et al. (2023) describe two categories: sentiment categorization and sentiment regression. Sentiment categorization divides the text into categories based on positive or negative sentiment, or subjective and objective tone (Wang et al., 2012; Mohammad et al., 2013; Socher et al., 2013; Dey et al 2016). However, these categories are not well defined and differ between models. There is no synchronization between the categories.

Sentiment regression, on the other hand, assigns a numerical value to a sentence and then categorizes it as positive, neutral, or negative.(Bazitios et al., 2017; Ma et al., 2018; Sun et al., 2019; Rietzler et al., 2019; Wu and Ong, 2020; Hannak et al., 2021; Wang et al., 2020; Li et al., 2023). Regression-based scales usually use scores ranging from -1 to +1 to quantify the sentence's intensity and sentiment.

The lack of standardized measures makes it difficult to compare results, establish a shared understanding of sentiment, and benchmark performance. These metrics do not measure the same quantity, even if they fall under the category of sentiment. Standardizing sentiment measures would address these issues by improving consistency, interpretation, and integration with social media platforms (Venkit et al., 2023).

3.3 Datasets

Wang et al. (2012), Mohammad et al. (2013), Baziotis et al. (2017), Hannak et al. (2021), and Wang et al. (2022) all use Twitter as their data source. Collecting large amounts of Twitter data necessitates adherence to Twitter's API policies and rate limits, which can be tedious and time-consuming. Furthermore, tweets frequently include informal language, slang, emoticons (which express sentiment), abbreviations, and noise (e.g., hashtags, URLs), making text processing and sentiment analysis difficult. Twitter data can also be multilingual, necessitating language detection and translation capabilities for accurate analysis across multiple languages. A significant portion of the data must be manually labelled for sentiment, which is a costly and labour-intensive process, frequently involving crowd-sourcing platforms such as Amazon Mechanical Turk(Wang et al., 2012).

Another commonly used dataset is review data from sources such as IMDB (movies), OpinRank (hotels), and SemEval (restaurants), as demonstrated in studies by Dey et al. (2016), Ma et al. (2018), and Li et al. (2023). To obtain review data, you may need to use web scraping or negotiate access to proprietary datasets. For aspect-based sentiment analysis, reviews must be annotated with specific aspect categories (e.g., service, food quality), which is a time-consuming task. Reviews can be written in multiple languages, necessitating language detection and processing abilities.

Furthermore, review data from different domains (e.g., movies, hotels, restaurants) may have different linguistic characteristics, necessitating domain adaptation methods. Manually labelled datasets, such as the Sentiment Treebank and SemEval competition data used by Socher et al. (2013), Mohammad et al. (2013), and Baziotis et al. (2017), face issues with annotation quality, consistency across annotators, and limited domain/language coverage.

3.4 Metrics

The accuracy metric is widely used to assess the effectiveness of SC approaches (Wang et al., 2012; Socher et al., 2013; Dey et al., 2016; Ma et al., 2018; Rietzler et al., 2019). This term refers to the proportion of correctly classified samples in the overall sample. However, other metrics, such as precision (Dey et al., 2016), recall (Dey et al., 2016; Baziotis et al., 2017), and f-measure (Mohammad et al., 2013; Baziotis et al., 2017; Ma et al., 2018; Rietzler et al., 2019), provide better insights (Kayed et al., 2023). According to Rietzler et al. (2019), the Macro-F1 metric provides better insights in imbalanced datasets with a high proportion of positive labels.

4 Future Progress

Like any other research field, SA has limitations and requires further development. First, data availability is limited; additional factors required for research are frequently unavailable or limited. For example, Rietzler et al. (2019) mention a lack of data that allows us to learn more abstract relationships for TABSA or ABSA, data from Twitter or Weibo only includes those who use these platforms to communicate (Wang et al., 2022), and Li et al. (2023) require additional factors such as restaurants' marketing tactics, managerial practices, and financial status. These attributes should be used whenever relevant data is available. Second, the credibility of the data; data available online may not express effective aspects of subjective well-being (Wang et al., 2022) or current sentiment, and may also be fraudulent or manipulated (Li et al., 2023). Subsequent work to ensure the data's credibility should be completed.

Third, SA lacks standardized metrics, resulting in various scales and categorizations. This complicates model comparisons and leads to a lack of consensus among different SA frameworks. Addressing these issues is critical to increasing the reliability and applicability of SA. Lastly, there is evidence of bias and limitations in the generalizability of SA models. Models frequently show biases against specific social groups, which can be attributed to a lack of standards. Furthermore, the use of different scales and categorizations limits the generalizability of SA models, making them less applicable in real-world scenarios, especially when they cause harmful misclassification of minority groups due to a lack of understanding of their context and language. Venkit et al. (2023) propose data statements and ethics sheets to address these limitations.

5 Conclusion

Sentiment analysis has emerged as a vital field, allowing the interpretation of massive amounts of unstructured data from a variety of sources. Machine learning techniques, particularly deep learning models such as BERT and attention-based architectures, have significantly advanced the state-of-the-art in SA tasks. However, challenges persist, including a lack of standardized metrics, a scarcity of high-quality labelled data across diverse domains and languages, and the potential of biases and limitations in generalizability. Future research should address these issues by establishing consistent evaluation metrics, curating larger and more diverse datasets, and developing more robust and inclusive models.

References

Baziotis, C., Pelekis, N. and Doulkeridis, C. (2017). *DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis*. Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pp 747–754. Available at:https://aclanthology.org/S17-2126 (Accessed: 16 April 2024).

Dey, L., Chakraborty, S., Biswas, A., Bose, B. and Tiwari, S. (2016). *Sentiment Analysis of Review Datasets Using Naïve Bayes‘ and K-NN Classifier*. CoRR. abs/1610.09982. Available at:http://arxiv.org/abs/1610.09982 (Accessed: 16 April 2024).

Hannak, A., Anderson, E., Feldman Barrett, L., Lehmann, S., Mislove, A. and Riedewald, M. (2021). *Tweetin' in the Rain: Exploring Societal-Scale Effects of Weather on Mood*. Proceedings of the International AAAI Conference on Web and Social Media. 6(1), 479-482. Available at:https://doi.org/10.1609/icwsm.v6i1.14322 (Accessed: 16 April 2024).

Kayed, M., Díaz-Redondo, R.P., Mabrouk, A. (2023). *Context-Guided BERT for Targeted Aspect-Based Sentiment Analysis*. IEEE Access. 8. Available at:https://arxiv.org/abs/2312.17253 (Accessed: 16 April 2024).

Li, H., Yu, B.X.B., Li, G. and Gao, H. (2023). *Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews*. Tourism Management. 96. Available at:https://doi.org/10.1016/j.tourman.2022.104707 (Accessed: 16 April 2024).

Ma, Y., Peng, H. and Cambria, E. (2018). *Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM*. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). Available at:https://doi.org/10.1609/aaai.v32i1.12048 (Accessed: 16 April 2024).

Mohammad, S.M., Kiritchenko, S. and Zhu, X. (2013). *NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets*. CoRR. abs/1308.6242. Available at:https://arxiv.org/abs/1308.6242 (Accessed: 16 April 2024).

Rietzler, A., Stabinger, S., Opitz, P. and Engl, S. (2019). *Adapt or Get Left Behind: Domain Adaptation through BERT Language Model Finetuning for Aspect-Target Sentiment Classification*. CoRR.abs/1908.11860. Available at:https://arxiv.org/abs/1908.11860 (Accessed: 16 April 2024).

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C.D., Ng, A., and Potts, C. (2013). *Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank*. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp 1631–1642. Available at:https://aclanthology.org/D13-1170 (Accessed: 16 April 2024).

Sun, C., Huang, L. and Qiu, X. (2019). *Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence*. CoRR abs/1903.09588. Available at:https://arxiv.org/abs/1903.09588 (Accessed: 16 April 2024).

Venkit, P.N., Srinath, M., Gautam, S., Venkatraman, S., Gupta, V., Passonneau, R.J., Wilson, S. (2023). *The Sentiment Problem: A Critical Survey towards Deconstructing Sentiment Analysis*. Available at:https://arxiv.org/abs/2310.12318 (Accessed: 16 April 2024).

Wang, H., Can, D., Kazemzadeh, A., Bar, F. and Narayanan, S. (2012). *A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle*. Proceedings of the ACL 2012 System Demonstrations, pp 115–120. Available at:https://aclanthology.org/P12-3020 (Accessed: 16 April 2024).

Wang, J., Fan, Y., Palacios, J., Chai, Y., Guetta-Jeanrenaud, N., Obradovich, N., Zhou, C. and Siqi Zheng, S. (2022). *Global evidence of expressed sentiment alterations during the COVID-19 pandemic*. Nat Hum Behav. 6, pp349–358 Available at:https://doi.org/10.1038/s41562-022-01312-y (Accessed: 16 April 2024).

Wu, Z. and Ong, D.C. (2020). *Context-Guided BERT for Targeted Aspect-Based Sentiment Analysis*. CoRR. abs/2010.07523. Available at:https://arxiv.org/abs/2010.07523 (Accessed: 16 April 2024).