ENHANCING THE UNDERSTANDING OF E-COMMERCE REVIEWS THROUGH ASPECT EXTRACTION TECHNIQUES: A BERT-BASED APPROACH

Davoodi, Laleh

Published in:
Proceedings of 36th Bled eConference

DOI:
10.18690/um.fov.6.2023

Published: 01/06/2023

Document Version
Final published version

ENHANCING THE UNDERSTANDING OF E-COMMERCE REVIEWS THROUGH ASPECT EXTRACTION TECHNIQUES: A BERT-BASED APPROACH

LALEH DAVOODI

Åbo Akademi University, Faculty of Social Sciences, Business and Economics, and Law, Turku, Finland
laleh.davoodi@abo.fi

The growth of online customer reviews on e-commerce platforms has led to an overwhelming volume and variety of data, making manual analysis impractical for both consumers and managers. Consequently, machine learning techniques, such as Aspect-Based Sentiment Analysis (ABSA), have gained prominence for their ability to determine sentiment at the aspect level. This study aims to fine-tune natural language processing models for aspect extraction in e-commerce customer reviews. We manually annotated 2781 online user review sentences in English and employed different extensions of the BERT model to identify implicit and explicit aspects. This approach diverges from prior studies, as our dataset comprises real user reviews from five prominent e-commerce platforms. The findings demonstrate the models' effectiveness in extracting aspects from diverse e-commerce user reviews, yielding a deeper understanding of user-generated content and customer satisfaction trends, and providing valuable insights for managerial decision-making. This study contributes to the ABSA literature and offers practical implications for e-commerce platforms aiming to improve their products and services based on customer feedback.

Keywords:
customer reviews, e-commerce, aspect extraction, BERT, machine learning

University of Maribor Press

DOI https://doi.org/10.18690/um.fov.4.2023.15
ISBN 978-961-286-751-5

1 Introduction

Currently, customers heavily rely on review content posted by other users on various e-commerce sites prior to making a purchase decision. However, with the ever-increasing volume and variety of online content, it has become challenging for customers to manually sift through the vast amounts of information available (Ansari et al., 2020). Moreover, managerial decision-making requires a constant flow of up-to-date information to provide insight into the trends and dynamics of customer satisfaction and manual classification of reviews has become increasingly difficult. Therefore, the popularity of machine learning techniques, such as opinion mining and aspect term detection, has increased in recent years due to the rise in online data volume and diversity. Aspect-Based Sentiment Analysis (ABSA) is a subfield of Sentiment Analysis (SA) that primarily focuses on determining the sentiment of products or services at the aspect level. To gain a deeper understanding of user-generated content, an essential task in ABSA is to extract both implicit and explicit aspects from various online reviews.

This paper aims to fine-tune the Bidirectional Encoder Representations from Transformers (BERT) and two of its extensions for aspect extraction in user-generated content in the form of customer reviews of e-commerce platforms. The research question we set forth to answer is the following:

How can aspect extraction techniques from manually annotated customer reviews be used to enhance the understanding of customer opinions in e-commerce?

This was accomplished by manually annotating 2781 online user review sentences in English. Unlike the majority of literature that assesses aspect extraction tasks using only the Laptops and Restaurants datasets from the SemEval 2014, 2015, and 2016 ABSA aspect extraction context (Xu et al., 2019, Pereget al., 2019; Chauhan et al., 2022; Venugopalan et al, 2022), this study employed BERT to categorize implicit and explicit aspects of our manually annotated dataset into 14 distinct groups. The data used for this analysis was obtained from real user reviews of five prominent e-commerce platforms that are available online. The present study demonstrates the effectiveness of using BERT for aspect extraction from a diverse range of user reviews in the e-commerce domain. By utilizing the manually annotated dataset and categorizing implicit and explicit aspects into 14 distinct groups, this study has provided a deeper understanding of user-generated content and customer satisfaction trends offering novel academic insights. Furthermore, the findings of this study can inform managerial decision-making and help e-commerce platforms to improve their products and services to meet the needs and expectations of their customers.

The rest of the paper is structured as follows. Section 2 presents a literature review on academic contributions focusing on aspect detection in customer reviews. The methodology and data preparation are discussed in Section 3, with the main results presented and discussed in Section 4. Finally, some conclusions, limitations and future research directions are provided in Section 5.

2 Aspect detection and extraction

Sentiment analysis (SA), also referred to as opinion mining, is the process that involves identifying, recognizing, and categorizing users' emotions or opinions on various services such as movies, products, events, or any attributes as positive, negative, or neutral. The data used in the analysis can be gathered from diverse sources such as review websites, forum discussions, blogs, Twitter, etc. SA is a powerful tool as it provides valuable information about people's preferences and can help companies gain a clear perspective regarding their product or service features (Mehta et al., 2020). By analyzing customer sentiments, companies can identify areas of improvement, determine customer satisfaction levels, and gauge market trends. SA has been widely applied in various industries, including hospitality, healthcare, and e-commerce, among others.

Previous studies have generally categorized SA into document-level, sentence-level, and aspect-level SA. These levels aim to classify whether a whole document, a sentence (subjective or objective), and an aspect express a sentiment (Nazir et al., 2020). Target-based sentiment analysis (TBSA) or Aspect-based Sentiment Analysis (ABSA) is a sub-task of SA that provides a better understanding of the problem of SA at a fine-grained level (Liu 2012; Pontiki 2014) because it focuses directly on sentiments rather than on language structure. ABSA analyzes specific aspects or entities of a product, service, or topic and determines the sentiment associated with them. This technique provides a more in-depth and nuanced understanding of public opinion, making it useful for companies to develop marketing strategies, improve product features, and enhance customer satisfaction. ABSA includes subtasks such as aspect/target term extraction (ATE), opinion term extraction (OTE), aspect/target term sentiment classification (ATC), and others (Peng et al, 2020).

The objective of ATE is to identify and extract terms that represent aspects of a given sentence. For instance, in the sentence "Excellent customer service and delivery", the aspect terms are “customer service” and “delivery”. ATE typically involves two sub-tasks: (1) extracting all aspect terms (such as "delivery") from the text, and (2) grouping aspect terms with similar meanings into categories, where each category represents a single aspect (e.g., “delivery”, “shipping”, and “track number” will be clustered into the shipping aspect). It should be noted that ABSA categorizes aspects into two types: explicit and implicit aspects. Explicit aspects are directly mentioned in a text. In contrast, implicit aspects are not explicitly indicated by any specific word or term (Alqaryouti et al., 2020). For instance, in the sentence “If you need something tomorrow go somewhere else”, “shipping” is an implicit aspect.

Prior research on aspect extraction can be classified into four approaches: rule-based (Poria et al., 2014; Liu et al., 2015), supervised (Maitama et al., 2021; Poria et al., 2016), unsupervised (Chauhan et al., 2020; Luo et al., 2019), and semi-supervised (Ansari et al., 2020; Anand et al., 2016). According to He et al., (2017), rule-based methods typically do not group extracted aspect terms into categories, while supervised learning requires annotated data and may face domain adaptation issues. Unsupervised methods are utilized to avoid reliance on labeled data. In addition to the methods such as statistical analysis, topic modeling, and dependency parsing used in prior unsupervised aspect extraction studies, supervised aspect extraction techniques such as Conditional Random Field (CRF) and long short-term memory have also been employed (Maitama et al, 2021).

In recent years, deep learning has become one of the most effective approaches for natural language processing tasks, due to its supervised training process with large amounts of training data. However, acquiring a large amount of supervised data can be a difficult and time-consuming process, particularly for NLP tasks in low-resource languages. In these cases, transfer learning can offer a solution by allowing a model to be pre-trained on a large amount of unsupervised data before being fine-tuned for a specific task under supervised conditions. BERT is one of the latest popular algorithms that employ this transfer learning approach (Yanuar et al., 2021).

3 Methodology

This section will present the NLP-based aspect extraction and classification methodology. Our main goal is to construct a model that is able to detect the aspects of customer feedback automatically. As presented in Figure 1, we have applied five steps to achieve our research objective. In this section, we discuss data collection, annotation, and preprocessing, and present the machine-learning models used for aspect extraction and classification.

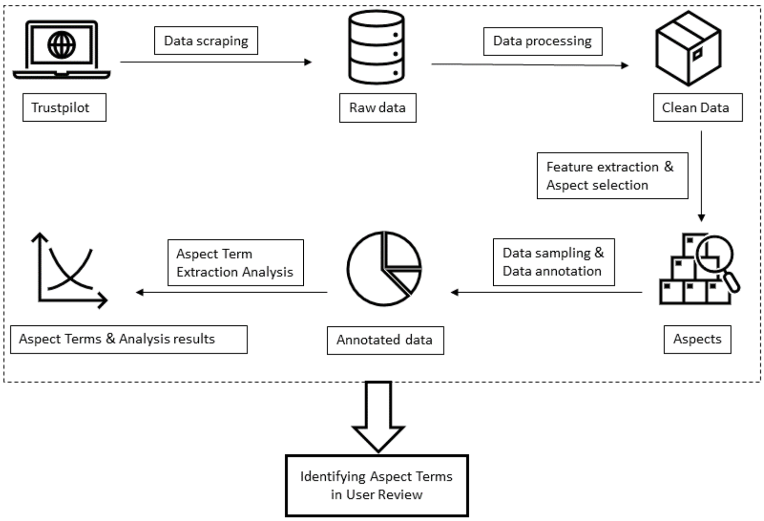

Figure 1: The stages of the research process

3.1 Data collection and processing

In order to study the user-generated content in e-commerce and evaluate their potential insights with ABSA, we collected 12,000 English reviews from Trustpilot, one of the biggest platforms in hosting online reviews, between 2013 and 2021. Next, we randomly selected 3,500 messages for aspect extraction and manual data annotation, excluding messages that did not meet certain criteria, such as not being written in English or not discussing any of the considered aspects. The final dataset consists of 2,782 data points consisting of the reviews from five online stores, namely Zalando, Wish, Sheinside, Boozt, and Nelly. Reviewers from 73 different countries contributed to the reviews, with the UK (1166 reviews) and the US (778 reviews) having the highest number of reviews among all 73 countries.

Prior to extracting our probable candidate aspects, we executed a series of seven text-cleaning steps: converting the text to lowercase, eliminating any non-English text, discarding all non-alphabetic characters, stripping out HTML tags and URLs, expanding contractions in the reviews, eliminating stop words, and performing lemmatization. Following text processing, the total count of distinct words was determined to be 4580. In preparation for manual data annotation, it was necessary to specify the aspects relevant to the domain. The final set of aspects, as presented in Table 1, is extracted from the literature and analysis of the reviews. Some further details can be found in Davoodi et al. (2022).

Table 1: Final aspects

1	Shipping	The quality of the delivery can refer to various factors, such as the cost and timeline.
2	Trust	The customers' overall sentiment towards the store.
3	Item quality	The level of quality of the products.
4	Customer service	The quality of customers' direct interactions with the store's representatives, such as their helpfulness, professionalism, and effectiveness
5	Pricing	The prices offered by the store, as well as the availability of discounts and promotional campaigns.
6	Product features	The quality of the product images and size guides provided on the website and implicit aspects such as described or expected
7	Refund process	The speed and quality of the refund process, as well as the effectiveness and efficiency of the store's handling of refund-related issues.
8	Return process	The speed, convenience, and cost associated with the return process for customers.
9	App experience	The experience that a user has while interacting with the website of the store.
10	Delivered product status	The state of the products that have been delivered, such as if they are broken or have an odor
11	Information	The presence and accuracy of the information, including the quality of advertisements and whether they may be deceptive.
12	Packaging	The visual appeal and quality of the product’s packaging.
13	Payment	The level of quality in the financial transaction process, including accuracy, security, efficiency, and different payment options.
14	Product availability	The range of products or brands that are being offered.

3.2 Data annotation

Prior to annotating the data, we utilize Python libraries such as contextualSpellCheck and Caribe to conduct spell and grammar checks on the reviews, thereby converting them into standard English text. Subsequently, we established a set of rules to convert implicit aspects into explicit ones while endeavoring to maintain the original text as closely as possible.

Following the text-cleaning process, two annotators manually annotated the data separately. The sentences were divided into individual words, and labels were assigned to the nouns or noun phrases only if they conveyed a sentiment related to at least one of the aspects, subsequently disregarding aspects with a neutral sentiment. Previous studies have revealed that aspects are often represented by nouns or noun phrases, while opinions are typically conveyed through adjectives or adjective phrases (Samah et al., 2014). Research on sentiment analysis has also demonstrated that certain tag combinations are instrumental in identifying aspects and opinions. In contrast to these prior investigations, the present study relied heavily on sentence parsing by considering a broader range of sentence components as potential aspects and/or opinions (Hu et al., 2004; Hu et al., 2004; Subrahmanian et al., 2008; Samha et al., 2014). It is important to note that the aspect terms were extracted solely based on their positive or negative sentiment in the given sentence without considering the preceding or subsequent sentence, i.e., the review as a whole.

The ATE task is typically approached as a sequence labeling problem, where each input word is assigned one of three labels: "B" for the beginning of an aspect term, "I" for the continuation of an aspect term, or "O" for non-aspect terms. The "I" label is necessary because some aspect terms can consist of multiple words, and the system must identify and extract all of them as aspects (Karimi et al., 2020). However, in this study, we provided 29 labels as follows: *O, App_experience_B, App_experience_I, Trust_B, Trust_I, Customer_service_B, Customer_service_I, Delivered_product_status_B, Delivered_product_status_I, Information_B, Information_I, Item_quality_B, Item_quality_I, Packaging_B, Packaging_I, Payment_B, Payment_I, Pricing_B, Pricing_I, Product_availability_B, 'Product_availability_I, Product_features_B, Product_features_I, Refund_process_B, Refund_process_I, Return_process_B, Return_process_I, Shipping_B, and Shipping_I.*

In the given example, "*Love this site fantastic saving for quality stuff.*", the text is first transformed into a set of words, and then the annotators assign corresponding labels to each word. The label sequence for this example is as follows: ['O', 'O', 'Trust_B', 'O', 'Pricing_B', 'O', 'Item_quality_B', 'Item_quality_I', 'O']. Once the individual annotations were completed, disagreements were identified and discussed to arrive at a final agreement between the annotators.

3.3 Machine learning models for aspect extraction

To evaluate our annotated model, we selected three recent transformer-based machine-learning models: BERT, RoBERTa, and BERT_Review.

BERT (Devlin et al., 2018) is a language model that can evaluate the context of a word from both the left and right sides simultaneously, unlike traditional language models that process sentences from one direction. This is achieved using the masked language modeling (MLM) technique, which randomly masks a word in a sentence and replaces it with a [MASK] token. The model then attempts to predict the masked word based on the context from both sides of the masked word. In addition to MLM, BERT also includes a next-sentence prediction (NSP) task, which involves predicting whether a given sentence follows another sentence or not. This helps BERT to capture the relationship between sentences and improve its ability to perform tasks such as question answering and text classification. Overall, BERT's bidirectional feature and MLM technique make it a powerful language model that can provide more contextual features from a sentence compared to other models such as ELMO.

RoBERTa (Liu et al., 2019) and BERT_Review (Xu et al., 2019) are variants of the BERT model, with the former showing promising results in various natural language processing (NLP) tasks. These variants are designed to enhance BERT's performance by incorporating different modifications. RoBERTa, for example, utilizes a transformer architecture like BERT but is trained differently and for a longer duration. Additionally, RoBERTa employs the entire sentence as input and eliminates the following sentence prediction objective.

BERT_Review is designed to address Review Reading Comprehension (RRC). For this purpose, Xu et al. (2019) created a dataset, which is based on a well-known benchmark used for aspect-based sentiment analysis. However, due to the limited training examples available for RRC and aspect-based sentiment analysis in their data, they subsequently explored a post-training method using BERT to improve the fine-tuning process for RRC. Additionally, they applied the proposed post-training approach to other review-based tasks, such as aspect extraction and aspect sentiment classification in aspect-based sentiment analysis, to demonstrate its universality. The results indicated that the proposed post-training approach is highly efficient (Xu et al., 2019).

3.4 Model building

To evaluate our annotated model, we selected three recent transformer-based machine-learning models: BERT, RoBERTa, and BERT_Review. Given the limited size of our dataset, we opted to employ cross-validation as a means of achieving a more reliable estimation of our models' performance. Specifically, we partitioned our manually annotated dataset into distinct training and validation sets by means of a five-fold cross-validation methodology, involving ten steps of iteration per fold. Subsequently, to obtain an aggregate summary of model performance, we calculated the mean value across all folds for each performance metric. To transform the input text to numeric features, we utilized a transformer tokenizer for all models. We utilized loss and accuracy as the performance metric for all models, with the main model serving as the baseline. For training BERT, RoBERTa, and BERT_Review our network model includes the main model with 12 layers and 768 hidden dimensions as well as the initial learning rate of 1e-4. Moreover, the training batch size was defined as 8. The programming language used for model building and data analysis is Python 3.10.11. We evaluated the models' performance using two commonly used measures: accuracy and F1 score.

4 Results

In this section, we will present the outcomes of our experiments and demonstrate the efficacy of different machine-learning models in the domain of aspect extraction and classification. We will also compare our findings to prior academic research in this area. Additionally, we will discuss several noteworthy observations that are relevant to managing expectations regarding performance.

4.1 Performance analysis

As detailed in the preceding section, a total of three distinct models were built and trained. The outcomes of these models are presented in Table 2. The table displays the mean loss, accuracy, and F1 value of the validation set after performing 5-fold cross-validation. Additionally, the execution time required to build the models is reported.

Table 2: Aspect extraction performance

Method	Loss. validation set	Accuracy-validation set	F1-validation set	Execution time (in seconds)
BERT_Review	**0.104**	**0.972**	**0.841**	**3,077**
BERT	0.113	0.969	0.829	3,075
RoBERTa	0.109	0.969	0.828	2,905

After conducting 5-fold cross-validation for three transformer-based models, we calculated the mean validation accuracy for each model. To test for statistical significance between the models, we performed a one-way ANOVA with post-hoc Tukey HSD tests. The results showed that BERT_Review had a significantly higher mean accuracy compared to the other two models. However, there was no significant difference in mean accuracy between BERT and RoBERTa. As demonstrated in Table 2, the BERT_Review model displayed marginally superior performance when compared to two other models, achieving an F1 score of 84%. Given the fact that the study involved 29 distinct categories, this level of performance can be considered promising. After running the cross-validation, to analyze the performance of the best model we partitioned the dataset into the fixed train (70%), validation (15%), and test (15%) sets. We used the test set for the error analysis. The analysis results are presented in Table 3.

Table 3: Aspect extraction performance of the best model

Method	Accuracy-validation	F1-validation	Accuracy-test	F1-test	Number of misclassifications (test)
BERT_Review	0.975	0.86	0.972	0.85	169

In regard to the specific misclassifications generated by the BERT_Review model, a portion of the errors can be attributed to the identification of neutral aspects in the messages. For instance, the following review was misclassified by the model as a return process. However, there should not be any aspect marked in this message, as the sentiment expressed towards the return process in this message is neutral: *"I ordered around 300 dollars worth of clothes for my kids and I need to return 100 dollars worth of clothes."*

Secondly, there are certain aspects in the dataset that occur less frequently, such as payment. Consequently, the model did not have sufficient samples to accurately identify these aspects.

Thirdly, in the majority of the misclassification cases, the model was able to extract the correct aspects. However, there were slight differences in the position of the true and predicted aspects, which we believe can be easily addressed by increasing the training sample. For instance, in the following example: *"No more purchase from Wish, no support after delivery of a product."* The true aspects are *Company, Customer Service, and Shipping*, and the model was able to correctly identify them. However, for the *Shipping* aspect, the model should have identified *"delivery of a product"* as the *shipping* aspect but instead marked each "*delivery*" and "*product*" word separately as *shipping*.

Finally, consider the following example: *"Order arrived really fast and was well wrapped."* In this case, the aspects being referred to are *shipping* and *packaging*. However, the model used in the study only predicted the aspect of *shipping*, whereas the true label encompasses both *shipping* and *packaging*. As a result, both predicted aspects are technically true, but the issue of accurately identifying multiple aspects needs to be addressed further in future research. Moreover, it is important to acknowledge that human annotation errors may occur during the manual annotation process, and such errors have the potential to negatively impact the performance of the machine learning models trained in this research.

4.2 Discussion

The input for classification models and data annotation, which involves extracting relevant aspect terms from reviews, is crucial in natural language processing. Previous research in this area has predominantly utilized datasets from SemEval 2014, SemEval 2015, and SemEval 2016 (Wang et al, 2021; Dai et al., 2019) for their experimental studies. However, our study focused on e-commerce businesses that do not have physical shops when collecting data, aiming to identify and understand the most prevalent sources of satisfaction and dissatisfaction in this context. For this purpose, data were collected from five distinguished and widely used e-commerce platforms that cater to a diverse range of customers. The comprehensiveness and broadness of this data make it applicable to research endeavors aimed at analyzing smaller online retailers and comparable enterprises.

While automated annotations are less time-consuming and costly than manual annotations, they are generally less accurate. In this study, we opted for manual annotation to gain a comprehensive understanding of the review content and its relevance to the target companies. By presenting a manually annotated dataset that includes extracted aspect terms with negative and positive sentiments, companies can ensure that they concentrate on the correct elements by obtaining precise aspect detection results. Additionally, the combination of sentiment classification of the extracted aspects can enable automated and accurate identification of the sources of dissatisfaction.

From our review of the literature, we have found that BERT and RoBERTa are among the most frequently employed models for aspect extraction in user reviews (Chauhan et al., 2020; Tian et al., 2020; Yanuar et al., 2020; Lopes et al., 2021) and the F1 score is achievable in the range of 0.738-0.85 depending on the domain and the language of the reviews. Our experimental results revealed that the BERT_Review model outperformed the other two models under consideration. Among all categories, the Trust_B aspect attained the highest F1 score of 0.92, while Payment_I had the lowest score of 0.4. This finding indicates that the inclusion of infrequent aspects may negatively impact the model's performance. One potential solution to address this issue is to automatically generate labeled data for infrequent aspects.

5 Conclusions

This study involved the development of a novel manually annotated dataset for the aspect extraction task, which was evaluated by using three state-of-the-art transformer-based models. The evaluation results demonstrate that the models performed well, but further improvements could be achieved by adding more samples to the dataset, particularly for infrequent aspects such as payment and packaging. A possible direction for future research could involve the use of this manually annotated dataset to generate automated labeled data, thereby enhancing the performance of the models. For example, this could entail fine-tuning Large Language Models like GPT-2 (Radford et al., 2019) and GPT-3 (Brown et al., 2022), as explored by Veyseh et al. (2022) and Ding et al. (2022). Moreover, this approach could help to address the key challenges associated with aspect extraction tasks, which is the scarcity of training data that limits the training of large neural networks. Large Language Models have proven remarkable few-shot capabilities in various NLP tasks (Min et al., 2021). Given this, a logical use case would be utilizing it for data annotation purposes.

References

Alqaryouti, O., Siyam, N., Abdel Monem, A., & Shaalan, K. (2020). Aspect-based sentiment analysis using smart government review data. Applied Computing and Informatics. Vol. ahead-of-print, 20 pages.

Anand, D., & Naorem, D. (2016). Semi-supervised aspect-based sentiment analysis for movies using review filtering. Procedia Computer Science, 84, 86-93.

Ansari, G., Saxena, C., Ahmad, T., & Doja, M. N. (2020). Aspect term extraction using graph-based semi-supervised learning. Procedia Computer Science, 167, 2080-2090.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

Chauhan, G. S., Meena, Y. K., Gopalani, D., & Nahta, R. (2020). A two-step hybrid unsupervised model with attention mechanism for aspect extraction. Expert systems with applications, 161, 113673.

Chauhan, G. S., Meena, Y. K., Gopalani, D., & Nahta, R. (2022). A mixed unsupervised method for aspect extraction using BERT. Multimedia Tools and Applications, 81, 31881-31906.

Dai, H., & Song, Y. (2019). Neural aspect and opinion term extraction with mined rules as weak supervision. arXiv preprint arXiv:1907.03750.

Davoodi, L., & Mezei, J. (2022). A comparative study of machine learning models for sentiment analysis: Customer reviews of e-commerce platforms. In Proceedings of the 35th Bled eConference Digital Restructuring and Human {Re} (pp. 217-230).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Ding, B., Qin, C., Liu, L., Bing, L., Joty, S., & Li, B. (2022). Is GPT-3 a Good Data Annotator? arXiv preprint arXiv:2212.10450.

He, R., Lee, W. S., Ng, H. T., & Dahlmeier, D. (2017, July). An unsupervised neural attention model for aspect extraction. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (pp. 388-397).

Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177).

Hu, M., & Liu, B. (2004, July). Mining opinion features in customer reviews. In Proceedings of the 19th national conference on American Association for Artificial Intelligence (pp. 755-760).

Im, J., Song, T., Lee, Y., & Kim, J. (2019). Confirmatory aspect-based opinion mining processes. arXiv preprint arXiv:1907.12850.

Karimi, A., Rossi, L., & Prati, A. (2021, January). Adversarial training for aspect-based sentiment analysis with bert. In Proceedings of 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 8797-8803).

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5, 1-167.

Liu, Q., Gao, Z., Liu, B., & Zhang, Y. (2015, June). Automated rule selection for aspect extraction in opinion mining. In the Proceedings of the Twenty-Fourth international joint conference on artificial intelligence (pp. 1291-1297).

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Luo, L., Ao, X., Song, Y., Li, J., Yang, X., He, Q., & Yu, D. (2019, August). Unsupervised Neural Aspect Extraction with Sememes. In Proceedings of The Twenty-Eighth International Joint Conference on Artificial Intelligence {IJCAI-19} (pp. 5123-5129).

Lopes, É., Correa, U., & Freitas, L. (2021, April). Exploring Bert for aspect extraction in Portuguese language. In Proceedings of The International FLAIRS Conference (pp. 1-4).

Maitama, J. Z., Idris, N., Abdi, A., & Bimba, A. T. (2021, May). Aspect extraction in sentiment analysis based on emotional affect using supervised approach. In Proceedings of the 4th International Conference on Artificial Intelligence and Big Data (ICAIBD) (pp. 372-376).

Mehta, P., & Pandya, S. (2020). A review on sentiment analysis methodologies, practices, and applications. International Journal of Scientific and Technology Research, 9, 601-609.

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., ... & Roth, D. (2021). Recent advances in natural language processing via large pre-trained language models: A survey. arXiv preprint arXiv:2111.01243.

Nazir, A., Rao, Y., Wu, L., & Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. IEEE Transactions on Affective Computing, 13, 845-863.

Peng, H., Xu, L., Bing, L., Huang, F., Lu, W., & Si, L. (2020, April). Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 8600-8607).

Pereg, O., Korat, D., Wasserblat, M., Mamou, J., & Dagan, I. (2019). ABSApp: a portable weakly-supervised aspect-based sentiment extraction system. arXiv preprint arXiv:1909.05608.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., ... & Eryiğit, G. (2016). Semeval-2016 task 5: Aspect-based sentiment analysis. In Proceedings of the ProWorkshop on Semantic Evaluation (SemEval-2016) (pp. 19-30).

Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. Knowledge-Based Systems, 108, 42-49.

Poria, S., Cambria, E., Ku, L. W., Gui, C., & Gelbukh, A. (2014, August). A rule-based approach to aspect extraction from product reviews. In Proceedings of the second workshop on natural language processing for social media (SocialNLP) (pp. 28-37).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multi task learners. OpenAI blog, 1, 9.

Samha, A. K., Li, Y., & Zhang, J. (2014). Aspect-based opinion extraction from customer reviews. arXiv preprint arXiv:1404.1982.

Subrahmanian, V. S., & Reforgiato, D. (2008). AVA: Adjective-verb-adverb combinations for sentiment analysis. IEEE Intelligent Systems, 23, 43-50.

Tian, H., & White, M. (2020, July). A pipeline of aspect detection and sentiment analysis for E-commerce customer reviews. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval {SIGIR} (pp. 1-9).

Venugopalan, M., & Gupta, D. (2022). An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis. Knowledge-based systems, 246, 108668.

Veyseh, A. P. B., Dernoncourt, F., Min, B., & Nguyen, T. H. (2022, July). Generating Complement Data for Aspect Term Extraction with GPT-2. In Proceedings of the Third Workshop on Deep Learning for Low-Resource Natural Language Processing (pp. 203-213).

Wang, Q., Wen, Z., Zhao, Q., Yang, M., & Xu, R. (2021, November). Progressive self-training with discriminator for aspect term extraction. In Proceedings of the 2021 conference on empirical methods in natural language processing (pp. 257-268).

Xu, H., Liu, B., Shu, L., & Yu, P. S. (2019). BERT post-training for review reading comprehension and aspect-based sentiment analysis. arXiv preprint arXiv:1904.02232.

Yanuar, M. R., & Shiramatsu, S. (2020, February). Aspect extraction for tourist spot review in Indonesian language using BERT. In Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 298-302).