



A systematic review of aspect-based sentiment analysis: domains, methods, and trends

Yan Cathy Hua¹ · Paul Denny¹ · Jörg Wicker¹ · Katerina Taskova¹

Accepted: 6 August 2024 / Published online: 17 September 2024
© The Author(s) 2024

Abstract

Aspect-based sentiment analysis (ABSA) is a fine-grained type of sentiment analysis that identifies aspects and their associated opinions from a given text. With the surge of digital opinionated text data, ABSA gained increasing popularity for its ability to mine more detailed and targeted insights. Many review papers on ABSA subtasks and solution methodologies exist, however, few focus on trends over time or systemic issues relating to research application domains, datasets, and solution approaches. To fill the gap, this paper presents a systematic literature review (SLR) of ABSA studies with a focus on trends and high-level relationships among these fundamental components. This review is one of the largest SLRs on ABSA. To our knowledge, it is also the first to systematically examine the interrelations among ABSA research and data distribution across domains, as well as trends in solution paradigms and approaches. Our sample includes 727 primary studies screened from 8550 search results without time constraints via an innovative automatic filtering process. Our quantitative analysis not only identifies trends in nearly two decades of ABSA research development but also unveils a systemic lack of dataset and domain diversity as well as domain mismatch that may hinder the development of future ABSA research. We discuss these findings and their implications and propose suggestions for future research.

Keywords ABSA · Aspect-based sentiment analysis · Systematic literature review · Natural language processing

✉ Yan Cathy Hua
yhua219@aucklanduni.ac.nz

Paul Denny
p.denny@auckland.ac.nz

Jörg Wicker
j.wicker@auckland.ac.nz

Katerina Taskova
katerina.taskova@auckland.ac.nz

¹ School of Computer Science, The University of Auckland, Auckland, New Zealand

1 Introduction

In the digital era, a vast amount of online opinionated text is generated daily through which people express views and feelings (i.e. sentiment) towards certain subjects, such as user reviews, social media posts, and open-ended survey question responses (Kumar and Gupta 2021). Understanding the sentiment of these opinionated text data is essential for gaining insights into people's preferences and behaviours and supporting decision-making across a wide variety of domains (Sharma and Shekhar 2020; Wankhade et al 2022; Tubishat et al 2021; García-Pablos et al 2018; Poria et al 2016). The analyses of opinionated text usually aim at answering questions such as "*What subjects were mentioned?*", "*What did people think of (a specific subject)?*", and "*How are the subjects and/or opinions distributed across the sample?*" (e.g. (Dragoni et al 2019; Krishnakumari and Sivasankar 2018; Fukumoto et al 2016; Zarindast et al 2021)). These objectives, along with today's enormous volume of digital opinionated text, require an automated solution for identifying, extracting and classifying the subjects and their associated opinions from the raw text. Aspect-based sentiment analysis (ABSA) is one such solution.

1.1 Review focus and research questions

This work presents a systematic literature review (SLR) of existing ABSA studies with a large-scale sample and quantitative results. We focus on trends and high-level patterns instead of methodological details that were well covered by the existing surveys mentioned above. We aim to benefit both ABSA newcomers by introducing the basics of the topic, as well as existing ABSA researchers by sharing perspectives and findings that are useful to the ABSA community and can only be obtained beyond the immediate research tasks and technicalities.

We seek to answer the following sets of research questions (RQs):

RQ1. To what extent is ABSA research and its dataset resources dominated by the commercial (especially the product and service review) domain? What proportion of ABSA research focuses on other domains and dataset resources?

RQ2. What are the most common ABSA problem formulations via subtask combinations, and what proportion of ABSA studies only focus on a specific subtask?

RQ3. What is the trend in the ABSA solution approaches over time? Are linguistic and traditional machine-learning approaches still in use?

This review makes a number of unique contributions to the ABSA research field: (1) It is one of the largest scoped SLRs on ABSA, with a main review and a Phase-2 targeted review of a combined 727 primary studies published in 2008–2024, selected from 8550 search results without time constraint. (2) To our knowledge, it is the first SLR that systematically examines the ABSA data resource distribution in relation to research application domains and methodologies; and (3) Our review methodology adopted an innovative automatic filtering process based on PDF-mining, which enhanced screening quality and reliability. Our quantitative results not only revealed trends in nearly two decades of ABSA research literature but also highlighted potential systemic issues that could limit the development of future ABSA research.

1.2 Organisation of this review

In Sect. 2 (“Background”), we introduce ABSA and highlight the motivation and uniqueness of this review. Section 3 (“Methods”) outlines our SLR procedures, and Sect. 4 (“Results”) answers the research questions with the SLR results. We then discuss the key findings and acknowledge limitations in Sects. 5 and 6 (“Discussion” and “Conclusion”).

For those interested in more details, Appendix A provides an in-depth introduction to ABSA and its subtasks. Appendix B describes the full details of our Methods, and additional figures from the Results are provided in Appendix C.

2 Background

2.1 ABSA: a fine-grained sentiment analysis

Aspect-based sentiment analysis (ABSA) is a sub-field of Sentiment Analysis (SA), which is a core task of natural language processing (NLP). SA, also known as “opinion mining” (García-Pablos et al 2018; Poria et al 2016; Liang et al 2022; López and Arco 2019; Tran et al 2020), solves the problem of identifying and classifying given text corpora’s affect or sentiment orientation (Akhtar et al 2020; Tubishat et al 2021) into polarity categories (e.g. “positive, neutral, negative”) (Brauwiers and Frasincar 2023; Hu and Liu 2004a), intensity/strength scores (e.g. from 1 to 5) (Wang et al 2017), or other categories. The “*identifying the subjects of opinions*” part of the quest relates to the granularity of SA. Traditional SA mostly focuses on document- or sentence-level sentiment and thus assumes a single subject of opinions (Nazir et al 2022b; Liu et al 2020). In recent decades, the explosion of online opinion text has attracted increasing interest in distilling more targeted insights on specific entities or their aspects within each sentence through finer-grained SA (Nazir et al 2022b; Liu et al 2020; Akhtar et al 2020; You et al 2022; Ettaleb et al 2022). This is the problem ABSA aims to solve.

2.2 ABSA and its subtasks

ABSA involves identifying the sentiments toward specific entities or their attributes, called *aspects*. These aspects can be explicitly mentioned in the text or implied from the context (“implicit aspects”), and can be grouped into aspect categories (Nazir et al 2022a; Akhtar et al 2020; Maitama et al 2020; Xu et al 2020b; Chauhan et al 2019; Akhtar et al 2018). Appendix A.1 presents a more detailed definition of ABSA, including its key components and examples.

A complete ABSA solution as described above traditionally involves a combination of subtasks, with the fundamental ones (Li et al 2022a; Huan et al 2022; Li et al 2020; Fei et al 2023b; Pathan and Prakash 2022) being Aspect (term) Extraction (AE), Opinion (term) Extraction (OE), and Aspect-Sentiment Classification (ASC), or in an aggregated form via Aspect-Category Detection (ACD) and Aspect Category Sentiment Analysis (ACSA).

The choice of subtasks in an ABSA solution reflects both the problem formulation and, to a large extent, the technologies and resources available at the time. The solutions to these fundamental ABSA subtasks evolved from pure linguistic and statistical solutions to

the dominant machine learning (ML) approaches (Maitama et al 2020; Cortis and Davis 2021; Liu et al 2020; Federici and Dragoni 2016), usually with multiple subtask models or modules orchestrated in a pipeline (Li et al 2022b; Nazir and Rao 2022). More recently, the rise of multi-task learning brought an increase in End-to-end (E2E) ABSA solutions that can better capture the inter-task relations via shared learning (Liu et al 2024), and many only involve a single model that provides the full ABSA solution via one composite task (Huan et al 2022; Li et al 2022b; Zhang et al 2022b). The most typical composite ABSA tasks include Aspect-Opinion Pair Extraction (AOPE) (Nazir and Rao 2022; Li et al 2022c; Wu et al 2021), Aspect-Polarity Co-Extraction (APCE) (Huan et al 2022; He et al 2019), Aspect-Sentiment Triplet Extraction (ASTE) (Huan et al 2022; Li et al 2022b; Du et al 2021; Fei et al 2023b), and Aspect-Sentiment Quadruplet Extraction/Prediction (ASQE/ASQP) (Zhang et al 2022a; Lim and Buntine 2014; Zhang et al 2021a, 2024a). We provide a more detailed introduction to ABSA subtasks in Appendix A.2.

2.3 The context- and domain-dependency challenges

The nature and the interconnection of its components and subtasks determine that ABSA is heavily domain- and context-dependent (Nazir et al 2022b; Chebolu et al 2023; Howard et al 2022). Domain refers to the ABSA task (training or application) topic domains, and context can be either the “global” context of the document or the “local” context from the text surrounding a target word token or word chunks. At least in English, the same word or phrase could mean different things or bear different sentiments depending on the context and topic domains. For example, “a big fan” could be an electric appliance or a person, depending on the sentence and the domain; “cold” could be positive for ice cream but negative for customer service; and “DPS” (damage per second) could be either a gaming aspect or non-aspect in other domains. Thus, the ability to incorporate relevant context is essential for ABSA solutions; and those with zero or very small context windows, such as n-gram and Markov models, are rare in ABSA literature and can only tackle a limited range of subtasks (e.g. Presannakumar and Mohamed 2021).

Moreover, although many language models (e.g. Bidirectional Encoder Representations from Transformers (BERT, Devlin et al 2019), Generative pre-trained transformers (GPT, Brown et al 2020), recurrent neural network (RNN)-based models) already incorporated local context from the input-sequence and/or general context through pre-trained embeddings, they still performed unsatisfactorily on some ABSA domains and subtasks, especially Implicit AE (IAE), AE with multi-word aspects, AE and ACD on mixed-domain corpora, and context-dependent ASC (Phan and Ogunbona 2020; You et al 2022; Liang et al 2022; Howard et al 2022). Many studies showed that ABSA task performance benefits from expanding the feature space beyond the generic and input textual context. This includes incorporating domain-specific dataset/representations and additional input features such as Part-of-Speech (POS) tags, syntactic dependency relations, lexical databases, and domain knowledge graphs or ontologies (Howard et al 2022; You et al 2022; Liang et al 2022). Nonetheless, annotated datasets and domain-specific resources are costly to produce and limited in availability, and domain adaptation, as one solution to this, has been an ongoing challenge for ABSA (Chen and Qian 2022; Zhang et al 2022b; Nazir et al 2022b; Howard et al 2022; Satyarthi and Sharma 2023).

The above highlights the critical role of domain-specific datasets and resources in ABSA solution quality, especially for supervised approaches. On the other hand, it suggests the possibility that the prevalence of dataset-reliant solutions in the field, and a skewed

ABSA dataset domain distribution, could systemically hinder ABSA solution performance and generalisability (Chen and Qian 2022; Fei et al 2023a), thus confining ABSA research and solutions close to the resource-rich domains and languages. This idea underpins this literature review's motivation and research questions.

2.4 Review rationale

This review is motivated by the following rationales:

First, the shift towards ML, especially supervised and/or DL solutions for ABSA, highlights the importance of dataset resources. In particular, annotated large benchmark datasets are crucial for the quality and development of ABSA research. Meanwhile, the finer granularity of ABSA also brings the persistent challenge of domain dependency described in Sect. 2.3. The diversity of ABSA datasets and their domains can have a direct and systematic impact on research and applications.

The early seminal works in ABSA were motivated by commercial applications and focused on product and service reviews (Liu et al 2020; Rana and Cheah 2016; Do et al 2019), such as Ganu et al (2009), Hu and Liu (2004b), and Pontiki et al (2014, 2015, 2016) that laid influential foundations with widely-used product and service review ABSA benchmark datasets (Rana and Cheah 2016; Do et al 2019). Nevertheless, the need for mining insights from opinions far exceeds this single domain. Many other areas, especially the public sector, also have an abundance of opinionated text data and can benefit from ABSA, such as helping policy-makers understand public attitudes and reactions towards events or changes (Sharma and Shekhar 2020), improving healthcare services and treatments via patient experience and concerns in clinical visits, symptoms, drug efficacy and side-effects (Cavalcanti and Prudêncio 2017; Gui and He 2021), and guiding educators in meeting teacher and learner needs and improving their experience (Wankhade et al 2022; Tubishat et al 2021; García-Pablos et al 2018; Poria et al 2016). While the more general SA research has been applied to “nearly every domain” (Nazir et al 2022b, p. 1), this does not seem to be the case for ABSA. Chebolu et al (2023) reviewed 62 public ABSA datasets released between 2004 and 2020 covering “over 25 domains” (Chebolu et al 2023, p. 1). However, 53 out of these 62 datasets were reviews of restaurants, hotels, and digital products; only five were not related to commercial products or services, and merely one was on the public sector domain (university reviews).

The above-mentioned evidence raises questions: Will this dataset domain homogeneity be found with a larger sample of primary studies? Does this domain skewness reflect the concentration of ABSA research focus or merely the lack of dataset diversity? This motivated our **RQ1** (“*To what extent is ABSA research and its dataset resources dominated by the commercial (especially the product and service review) domain? What proportion of ABSA research focuses on other domains and dataset resources?*”) Answers to these questions could inform and shape future ABSA research through individual research decisions and community resource collaboration.

Second, there are many good survey papers on ABSA, most focused on introducing the methodological details of common ABSA subtasks and solutions (e.g. Maitama et al 2020; Sabeeh and Dewang 2019; Rana and Cheah 2016; Soni and Rambola 2022; Gangawar and Rajalakshmi 2019; Zhou et al 2019) or specific approaches such as DL methods for ABSA (e.g. Liu et al 2020; Do et al 2019; Wang et al 2021a; Chen and Fnu 2022; Mughal et al 2024; Zhang et al 2022c; Satyarthi and Sharma 2023). We list these surveys in Appendix A.3 as additional resources for the reader. Nonetheless, many of these

reviews only explored each subtask and/or technique individually and often by iterating through reviewed studies, and few examined their combinations or changes over time and with quantitative evidence. For example, although the above-listed reviews (Liu et al 2020; Do et al 2019; Wang et al 2021a; Chen and Fnu 2022) reported the rise of DL approaches in ABSA similar to that of NLP as a whole, it is unclear whether ABSA research was also increasingly dominated by the attention mechanism from the Transformer architecture (Vaswani et al 2017) and pre-trained large language models since 2018 (Manning 2022), and if linguistic and traditional ML approaches were still active. In addition, most of these surveys used a smaller and selected sample that could not support conclusions on trends. As the field matures, we believe it is necessary and important to examine trends and matters outside the problem solution itself, so as to inform research decisions, identify issues, and call for necessary community awareness and actions. We thus proposed **RQ2** (“*What are the most common ABSA problem formulations via subtask combinations, and what proportion of ABSA studies only focus on a specific sub-task?*”) and **RQ3** (“*What is the trend in the ABSA solution approaches over time? Are linguistic and traditional machine-learning approaches still in use?*”).

In order to identify patterns and trends for our RQs, a sufficiently sized representative sample and systematic approach are required. We chose to conduct an SLR, as this type of review aims to answer specific research questions from all available primary research evidence following well-defined review protocols (Kitchenham and Charters 2007). Moreover, none of the existing SLRs on ABSA share the same focus and RQs as ours: Among the 192 survey/review papers obtained from four major digital database searches detailed in Sect. 3, only eight were SLRs on ABSA, within which four focused on non-English language(s) (Alyami et al 2022; Obiedat et al 2021; Hoti et al 2022; Rani and Kumar 2019), two on specific domains (software development, social media) (Cortis and Davis 2021; Lin et al 2022), one on a single subtask (Maitama et al 2020), and one mentioned ABSA subtasks as a side-note under the main topic of SA (Lighthart et al 2021).

In summary, this review aims to address gaps in the ABSA literature. The high-level nature of our research questions is best answered through a large-scale SLR to provide solid evidence. The next section presents our SLR approach and sample.

3 Methods

Following the guidance of Kitchenham and Charters (2007), we conducted this SLR with pre-planned scope, criteria, and procedures highlighted below. The complete SLR methods and process are detailed in Appendix B.

3.1 Main procedures

For the main SLR sample, we sourced the primary studies in October 2022 from four major peer-reviewed digital databases: ACM Digital Library, IEEE Xplore, Science Direct, and SpringerLink. First, we manually searched and extracted 4191 database results without publication-year constraints. Appendix B.1 provides more details of the search strategies

Table 1 Inclusion and exclusion criteria used for this Systematic Literature Review (SLR)

Inclusion criteria	Exclusion criteria
1. Published in the English language	1. The main text of the article is not in the English language
2. Has both the sentiment analysis component and entity/aspect-level granularity	2. Missing either the sentiment analysis or entity/aspect-level granularity in the research focus
3. Focuses on text data	3. Only contains search keyword in the reference section
4. Is a primary study with quantitative elements in the ABSA approach	4. Only contains less than 5 ABSA-related search keywords outside the reference section
5. Contains original solutions or ideas for ABSA task(s)	5. Is not a primary study (e.g. review articles, meta-analysis)
6. Involves experiment and results on ABSA task(s)	6. Does not provide quantitative experiment results on ABSA task(s)
	7. Has fewer than three pages
	8. Contains multimodal (i.e. non-text) input data for ABSA task(s)
	9. The research focus is not on ABSA task(s), even though ABSA models might be involved (e.g. recommender system)
	10. Focuses on transferring existing ABSA approaches between languages
	11. The ABSA tasks are integrated into a model built for other purposes, and there are no stand-alone ABSA method details and/or evaluation results

and results. Next, we applied the inclusion and exclusion criteria listed in Table 1 via automatic¹ and manual screening steps and identified 519 in-scope peer-reviewed research publications for the review. The complete screening process, including that of the automatic screening, is described in Appendix B.2. We then manually reviewed the in-scope primary studies and recorded data following a planned scheme. Lastly, we checked, cleaned, and processed the extracted data and performed quantitative analysis against our RQs.

3.2 Main SLR sample summary

Figure 1 shows the number of total reviewed vs. included studies across all publication years for the 4191 SLR search results. The search results include studies published between 1995 and 2023 ($N = 1$), although all of the pre-2008 ones (2 from the 90s, 8 from 2003–2006, 17 from 2007) were not ABSA-focused and were excluded during automatic screening. The earliest in-scope ABSA study in the sample was published in 2008, followed by a very sparse period until 2013. The numbers of extracted and in-scope publications have both grown noticeably since 2014, a likely result of the emergence of deep learning approaches, especially sequence models such as RNNs (Manning 2022; Sutskever et al 2014). We also present a breakdown of the included studies by publication year and type in Figure 9 in Appendix C.

¹ Our PDF mining for automatic review screening code is available at <https://doi.org/10.5281/zenodo.12872948>.

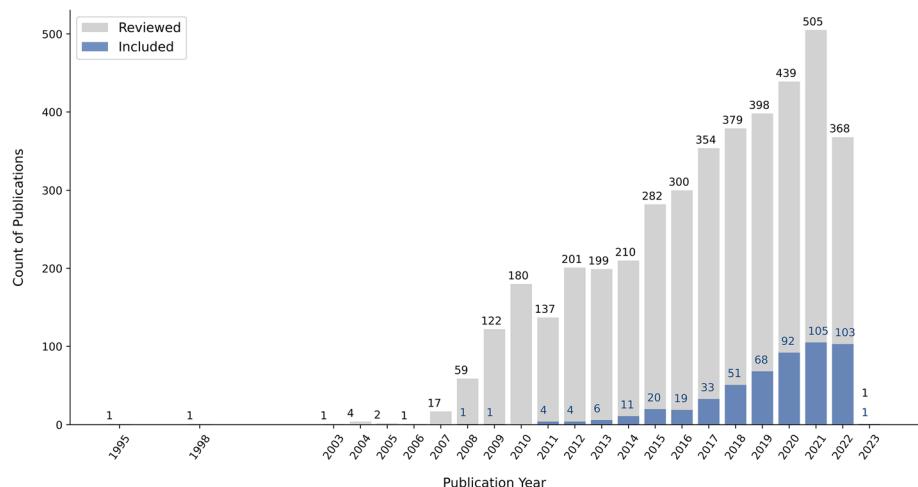


Fig. 1 Number of studies by publication year: total reviewed (N = 4191) vs. included (N = 519)

3.3 Note on “domain” mapping

In order to answer RQ1, we made the distinction between “research application domain” (“research domain” in short) and “dataset domain”, and manually examined and classified each study and its datasets into domain categories.

We considered each study’s research domain to be “non-specific” unless the study mentioned a specific application domain or use case as its motivation. For the dataset domain, we examined each dataset used by our sample, standardised its name, and recorded the domain from which it was drawn/selected based on the description provided by the author or the dataset source webpage. Datasets without a specific domain (e.g. Twitter tweets crawled without a specific domain filter) were labelled as “non-specific”.

We then manually grouped the research and dataset domains into 19 common categories used for analysis. More details and examples on domain mapping are available in Appendix B.3.

3.4 Phase-2 targeted review on in-context learning

Additionally, generative “foundation models” (Bommasani et al 2022), defined as models with billions of parameters pre-trained on enormous general-purpose data and adaptable to diverse downstream NLP tasks, have become ubiquitous after our SLR data collection (e.g. ChatGPT OpenAI 2023, released in November 2022). We use the term “foundation models” to distinguish them from the earlier pre-trained Large Language Models (LLMs) such as BERT (Devlin et al 2019), BART (Lewis et al 2020), and T5 (Raffel et al 2020), which have relatively fewer parameters and typically require fine-tuning for task adaptation (Zhang et al 2022c). These generative foundation models brought a new paradigm of “In-context Learning” (ICL) (Brown et al 2020, p. 4), where task adaptation can occur solely via conditioning the model on the text input instructions (“prompts”) with zero (“zero-shot ICL”) or few (“few-shot ICL”) examples and no model parameter changes (Brown et al

2020; Dong et al 2024). To capture and analyse this new development while balancing feasibility and currency, we conducted a Phase-2 targeted review in July 2024.

This Phase-2 targeted review focuses solely on the ICL implementations of pre-trained generative models for ABSA tasks, excluding those involving fine-tuning to draw a distinction from other non-ICL deep-learning approaches covered in the SLR. To extend the SLR sample beyond the original extraction time, we conducted a new database search² in July 2024 for studies published from 2022 onwards and removed the ones already included in the SLR sample. The new search results were screened using the SLR criteria described in Table 1 and then combined with the 519 SLR final samples. We then applied an additional filtering condition “Gen-LLM” to all the in-scope ABSA primary studies, which further selected publications with at least one occurrence of any of the following keywords outside the Reference section: “generative”, “in-context”, “in context learning”, “genai”, “bart”, “t5”, “flan-t5”, “gpt”, “chatgpt”, “llama”, and “mistral”. With the help of our automatic screening pipeline detailed in Appendix B.2, we were able to efficiently auto-screen the new search results and re-screen the previous SLR sample for the “Gen-LLM” keywords in less than one hour.

In total, the new search yielded 271 additional in-scope ABSA primary studies from 4359 search results. After applying the “Gen-LLM” filtering condition to the combined 790 in-scope ABSA primary studies, we obtained 208 Phase-2 samples for manual review, which comprised 91 studies from the new search and 117 from the earlier SLR sample, ranging from 2008 to 2024. The Phase-2 targeted review results are presented in Sect. 4.5. Unless specified otherwise, the results below only refer to those of the SLR.

4 Results

This section presents the SLR results corresponding to each of the RQs:

4.1 Results for RQ1

RQ1. To what extent is ABSA research and its dataset resources dominated by the commercial (especially the product and service review) domain? What proportion of ABSA research focuses on other domains and dataset resources?

To answer RQ1, we examined the distribution of reviewed studies by their research (application) domains, dataset domains, and the relationship between the two. From the 519 reviewed studies, we recorded 218 datasets, 19 domain categories (15 research domains and 17 dataset domains), and obtained 1179 distinct “study-dataset” pairs and 630 unique “study & dataset-domain” combinations. The key results are summarised below and presented in Table 2 and Fig. 2. We also list the datasets used by more than one reviewed study in the Appendix Table 15.

In summary, our results answer RQ1 by showing that: (1) The majority (65.32%) of the reviewed studies were not for any specific application domain and only 24.28% targeted “product/service review”. (2) The dataset resources used in the sample were mostly domain-specific (84.44%) and dominated by the “product/service review”

² This new database search followed the same procedures and criteria as the SLR, except that we aborted the SpringerLink search due to persistent database interface search result navigation issues during our data collection period.

Table 2 Number of in-scope studies per each research (application) and dataset domain category

Domain	Count of studies per research domain	% of studies per research domain (%)	Count of studies per dataset domain	% of studies per dataset domain (%)
Non-specific	339	65.32	98	15.56
Product/service review	126	24.28	447	70.95
Student feedback/education review	12	2.31	19	3.02
Politics/policy-reaction	8	1.54	5	0.79
Healthcare/medicine	7	1.35	9	1.43
Video/movie review	6	1.16	19	3.02
News	5	0.96	8	1.27
Finance	5	0.96	4	0.63
Research/academic reviews	3	0.58	3	0.48
Disease	3	0.58	3	0.48
Employer review	1	0.19	1	0.16
Nuclear energy	1	0.19	0	0.00
Natural disaster	1	0.19	1	0.16
Music review	1	0.19	1	0.16
Multiple domain	1	0.19	0	0.00
Location review	0	0.00	8	1.27
Book review	0	0.00	2	0.32
Biology	0	0.00	1	0.16
Singer review	0	0.00	1	0.16
Total	519	100.00	630	100.00

Bold indicates the highest value in each column, as a way to highlight the mismatch between research and dataset domains

datasets (70.95%). (3) Both the research effort and dataset resources were scant in the non-commercial domains, especially the main public sector areas, with fewer than 13 studies across 14 years in each of the healthcare, policy, and education domains, where about half of the used datasets were created from scratch for the study.

Beyond RQ1, (1) and (2) above also suggest a significant mismatch between the research and dataset domains as visualised in Fig. 2. Further, when filtering out datasets used by less than 10 studies, we discovered an alarming lack of dataset diversity as only 12 datasets remained, of which 10 were product/service reviews. When examining the three-way relationship among research domain, dataset domain, and dataset name, we further identified an over-representation (78.20%) of the four SemEval restaurant and laptop review benchmark datasets. This is illustrated in Fig. 4.

4.1.1 Detailed results for RQ1

For research (application) domains indicated by the stated research use case or motivation, the majority (65.32%, N = 339) of the 519 reviewed studies have a “non-specific” research

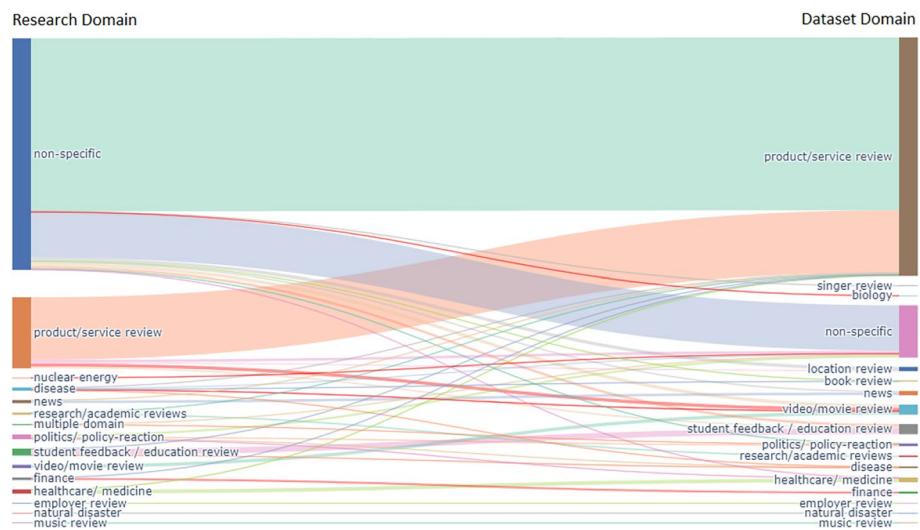


Fig. 2 Distribution of unique “study–dataset” pairs ($N = 1179$, with 519 studies and 218 datasets) by research (application) domains (left) and dataset domains (right). Note (1) The top flow visualises a mismatch between the two domains: the majority of studies without a specific research domain used datasets from the product/service review domain. (2) The disproportionately small number of samples in both domains that were neither “non-specific” nor “product/service review”

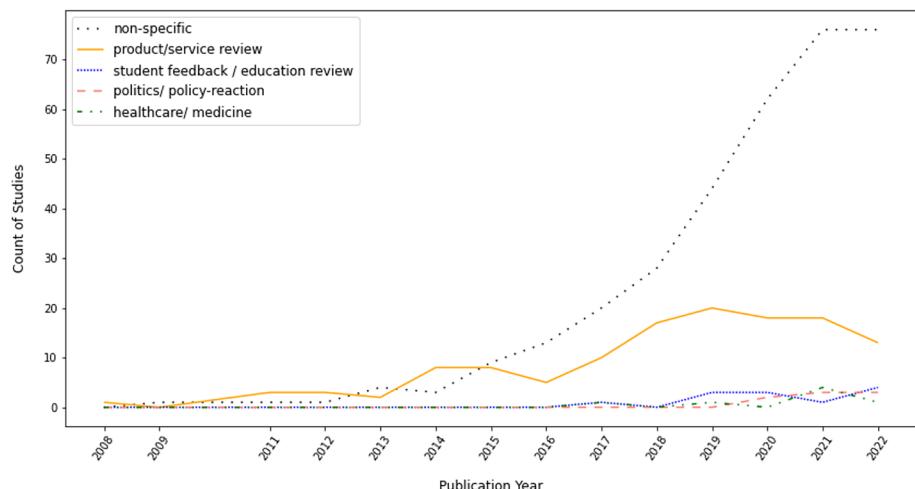


Fig. 3 Number of in-scope studies by research (application) domain and publication year ($N = 518$). This graph excludes the one 2023 study (extracted in October 2022) to avoid trend confusion

domain, followed by just a quarter (24.28%, $N = 126$) in the “product/service review” category. However, the number of studies in the rest of the research domains is magnitudes smaller in comparison, with only 12 studies (2.31%) in the third largest category “student feedback/education review” since 2008, followed by 8 in Politics/policy-reaction (1.54%), and only 7 in Healthcare/medicine (1.35%). Figure 3 revealed further insights from the

trend of research domain categories with five or more reviewed studies. Interestingly, “product/service review” has been a persistently major category over time, and has only been consistently taken over by “non-specific” since 2015. The sharp increase of domain-“non-specific” studies since 2018 could be partly driven by the rise of pre-trained language models such as BERT and the greater sequence processing power from the Transformer architecture and the attention mechanism (Manning 2022), as more researchers explore the technicalities of ABSA solutions.

As to the dataset domains, Table 2 suggests that among the 630 unique “study & dataset-domain” pairs, the majority (70.95%, N = 447) are in the “product/service review” category, followed by 15.56% (N = 98) in “Non-specific”. The third place is shared by two magnitude-smaller categories: “student feedback/ education review” (3.02%, N = 19) and “video/movie review” (3.02%, N = 19). The numbers of studies with datasets from the Healthcare/medicine (1.43%, N = 9) and Politics/policy-reaction (0.79%, N = 5) domains were again single-digit. Moreover, nearly half of the unique datasets in the public domains were created by the authors for the first time: 5/9 in Healthcare/medicine, 2/4 in Politics/policy-reaction, and 8/12 in Student feedback/ Education review.

Furthermore, to understand the dataset diversity across samples and domains, we grouped the 1179 unique “study-dataset” pairs by “research-domain, dataset-domain, dataset-name” combinations and zoomed into the 757 entries with ten or more study counts each. As shown in Table 3 and illustrated in Fig. 4, among these 757 unique combinations, 95.77% (N = 725) are in the “non-specific” research domain, of which 90.48% (N = 656) used “product/service review” datasets. Most interestingly, these 757 entries only involve 12 distinct datasets of which 10 were product and service reviews, and 78.20% (N = 592) are taken up by the four SemEval datasets from the early pioneer work (Pontiki et al 2014, 2015, 2016) mentioned in Sect. 2.4: SemEval 2014 Restaurant, SemEval 2014 Laptop (these two alone account for 50.33% of all 757 entries), SemEval 2016 Restaurant, and SemEval 2015 Restaurant. This finding echos (Xing et al 2020; Chebolu et al 2023): “The SemEval challenge datasets... are the most extensively used corpora for aspect-based sentiment analysis” (Chebolu et al 2023, p.4). Meanwhile, the top dataset used under “product/service review” research and dataset domains is the original product review dataset created by the researchers. Chebolu et al (2023) and Wikipedia (2023) provides a detailed introduction to the SemEval datasets.

It is noteworthy that among the 519 reviewed studies, 20 focused on cross-domain or domain-agnostic ABSA, and 19 of them did not have a specific research application domain. However, while all 20 studies used multiple datasets, 17 solely involved the “product/service review” domain category by using reviews of restaurants and different products, and 14 used at least one SemEval dataset. The only three studies that went beyond the “product/service review” dataset domain added in movie reviews, singer reviews, and generic tweets.

4.2 Results for RQ2

RQ2. What are the most common ABSA problem formulations via subtask combinations, and what proportion of ABSA studies only focus on a specific sub-task?

For RQ2, we examined the 13 recorded subtasks and 805 unique “study-subtask” pairs to identify the most explored ABSA subtasks and subtask combinations across the 519 reviewed studies. As shown in Fig. 5a, 32.37% (N = 168) of the studies developed

Table 3 Number of studies per each research (application) domain, dataset domain, and dataset combination for all datasets used by ten or more in-scope studies (N = 757)

Research domain	Dataset domain	Dataset	Count of studies	% of studies (%)
Non-specific	Product/service review	SemEval 2014 Restaurant	200	26.42
Non-specific	Product/service review	SemEval 2014 Laptop	181	23.91
Non-specific	Product/service review	SemEval 2016 Restaurant	110	14.53
non-specific	product/service review	SemEval 2015 Restaurant	101	13.34
Non-specific	Non-specific	Twitter (Dong et al. 2014)	53	7.00
Non-specific	Product/service review	Amazon customer review datasets (Hu and Liu 2004a)	25	3.30
Product/service review	Product/service review	Product review (original)	17	2.25
Non-specific	Non-specific	Twitter (original)	16	2.11
Non-specific	Product/service review	SemEval 2015 Laptop	16	2.11
Product/service review	Product/service review	Amazon product review (original)	15	1.98
Non-specific	Product/service review	Yelp Dataset Challenge Reviews	12	1.59
Non-specific	Product/service review	SemEval 2016 Laptop	11	1.45
Total			757	100.00

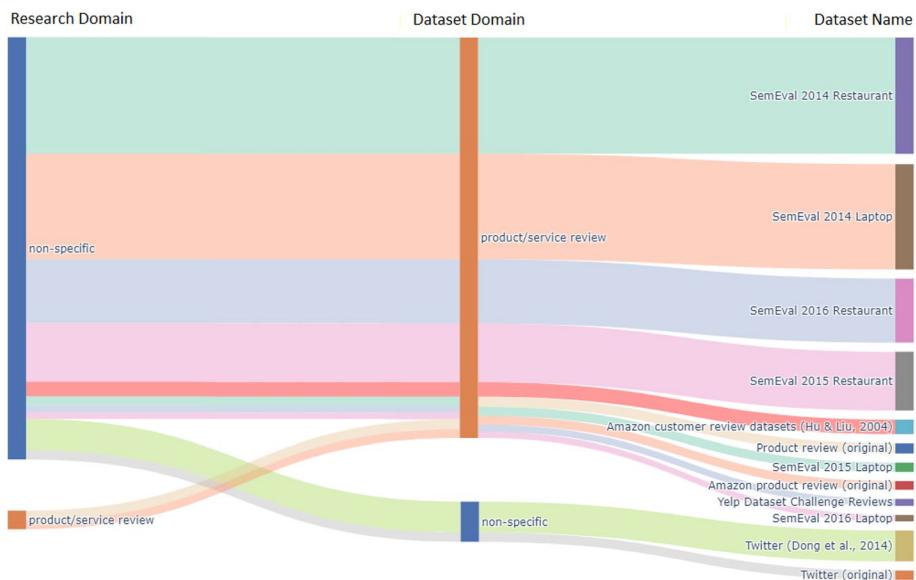
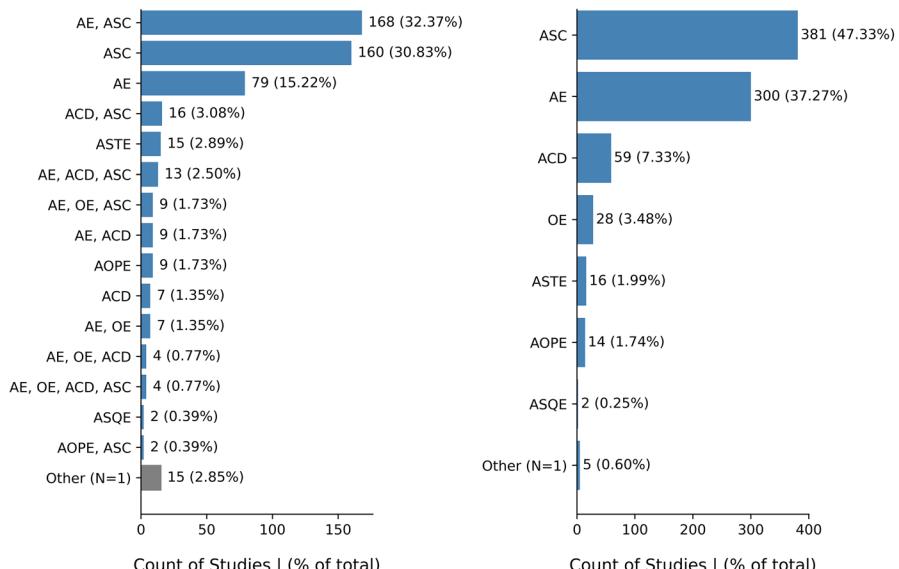


Fig. 4 Number of studies per each research (application) domain (left), dataset domain (middle), and dataset (right) combination, filtered by datasets used by 10 or more in-scope studies ($N = 757$). The three-way relationship highlights that not only did the majority of the sample studies with “non-specific” research domain use datasets from the ‘product/service review’ domain, but their datasets were also dominated by only four SemEval datasets on two types of product and service reviews



(a) Number of Studies per ABSA Subtask Combinations ($N=519$)

(b) Number of Studies per Individual ABSA Subtask ($N=805$)

Fig. 5 Number of studies by ABSA subtask

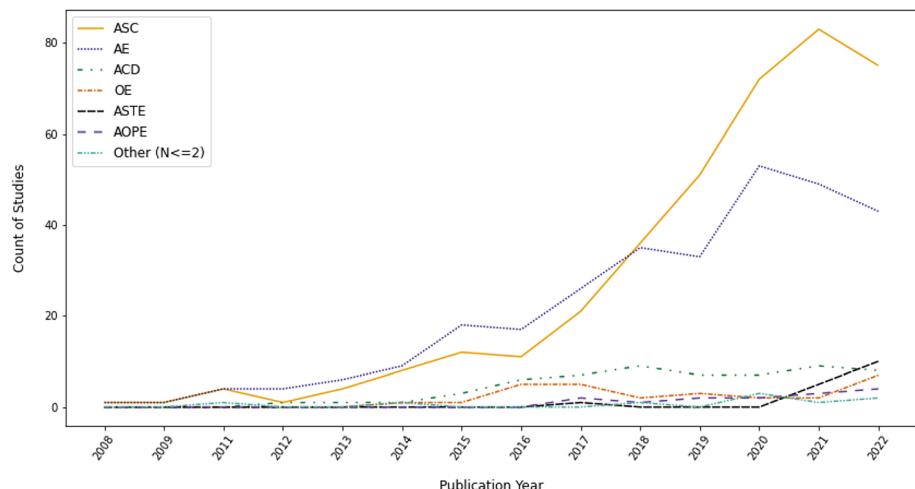


Fig. 6 Distribution of unique “Study–ABSA subtask” pairs by publication year ($N = 805$). This graph excludes the one 2023 study (extracted in October 2022) to avoid trend confusion

full-ABSA solutions through the combination of AE and ASC, and a similar proportion (30.83%, $N = 160$) focused on ASC alone, usually formulating the research problem as contextualised sentiment analysis with given aspects and the full input text. Only 15.22% ($N = 79$) of the studies solely explored the AE problem. This is consistent with the number of studies by individual subtasks shown in Fig. 5b, where ASC is the most explored sub-task, followed by AE and ACD.

Moreover, Fig. 6 reveals a small but noticeable rise in composite subtask ASTE since 2020 ($N = 1, 5$ and 10 in 2017, 2021, 2022) and a decline in ASC and AE around the same period. This could signify a problem formulation shift driven by deep-learning, especially multi-task learning methods for E2E ABSA. Our Phase-2 targeted review findings in Sect. 4.5 add more insights into this.

4.3 Results for RQ3

RQ3. What is the trend in the ABSA solution approaches over time? Are linguistic and traditional machine-learning approaches still in use?

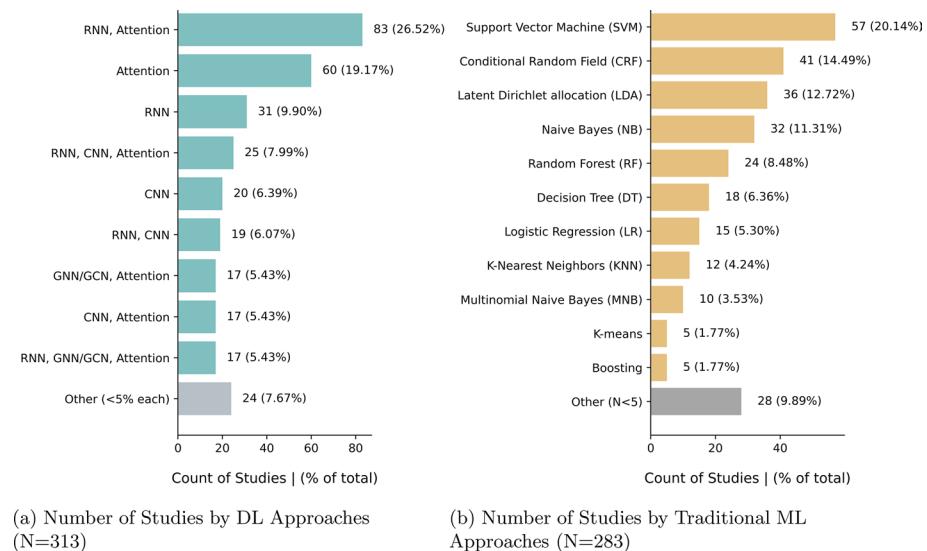
To answer RQ3, we examined the 519 in-scope studies along two dimensions, which we call “paradigm” and “approach”. We use “paradigm” to indicate whether a study employed techniques along the supervised-unsupervised dimension and other types, such as reinforcement learning. We classify non-machine-learning approaches under the “unsupervised” paradigm, as our focus is on dataset and resource dependency. By “approach”, we refer to the more specific type of techniques, such as deep learning (DL), traditional machine learning (traditional ML), linguistic rules (“rules” for short), syntactic features and relations (“syntactics” for short), lexicon lists or databases (“lexicon” for short), and ontology or knowledge-driven approaches (“ontology” for short).

Overall, the results suggest that our samples are dominated by fully- (60.89%) and partially-supervised (5.40%) ML methods that are more reliant on annotated datasets and prone to their impact. As to ABSA solution approaches, the sample shows that DL methods

Table 4 Number of studies by paradigm (N = 519)

Paradigm	Count of studies	% of studies (%)
Fully-supervised	316	60.89
Unsupervised	97	18.69
Hybrid	73	14.07
Semi-supervised	22	4.24
Weakly-supervised	6	1.16
Reinforcement learning	3	0.58
Self-supervised	2	0.39
Total	519	100.00

The unsupervised category includes non-ML approaches



(a) Number of Studies by DL Approaches (N=313)

(b) Number of Studies by Traditional ML Approaches (N=283)

Fig. 7 Number of studies using DL and traditional ML approaches

have rapidly overtaken traditional ML methods since 2017, particularly with the prevalent RNN family (55.91%) and its combination with the fast-surge attention mechanism (26.52%). Meanwhile, traditional ML and linguistic approaches have remained a small but steady force even in the most recent years. Context engineering through introducing linguistic and knowledge features to DL and traditional ML approaches was very common. More detailed results and richer findings are presented below.

4.3.1 Paradigms

Table 4 lists the number of studies per each of the main paradigms. Among the 519 reviewed studies, 66.28% (N = 344) is taken up by those using somewhat- (i.e. fully-, semi- and weakly-) supervised paradigms that have varied levels of dependency on labelled datasets, where the fully-supervised ones alone account for 60.89% (N = 316). Only

Table 5 Number of studies by paradigm and approaches (N = 519)

Paradigm	Approaches	Count of studies	% of studies (%)
Supervised	DL	149	47.15
	Syntactics, DL	53	16.77
	DL, traditional ML	23	7.28
	Other (< 5 each)	91	28.80
Total		316	100.00
Semi-supervised	DL	7	31.82
	Syntactics, Lexicon, traditional ML	3	13.64
	traditional ML	2	9.09
	Rules, Syntactics, Lexicon	2	9.09
	Other (< 5 each)	8	36.36
Total		22	100.00
Weakly-supervised	DL, traditional ML	2	33.33
	Traditional ML	1	16.67
	DL	1	16.67
	Syntactics, DL	1	16.67
	Syntactics, Ontology, DL	1	16.67
	Other (< 5 each)	0	0.00
Total		6	100.00
Self-supervised	DL	2	100.00
	Other (< 5 each)	0	0.00
Total		2	100.00
Unsupervised	Rules, Syntactics, Lexicon	24	24.74
	Rules, Syntactics, Lexicon, Ontology	15	15.46
	Rules, Syntactics	8	8.25
	Traditional ML	7	7.22
	Other (< 5 each)	43	44.33
Total		97	100.00
Hybrid	Rules, Syntactics, traditional ML	6	8.22
	Rules, Syntactics, Lexicon, traditional ML	6	8.22
	Traditional ML	5	6.85
	Syntactics, traditional ML	4	5.48
	Rules, Syntactics, Lexicon	4	5.48
	Rules, Syntactics, Lexicon, Ontology, traditional ML	4	5.48
	Other (< 5 each)	44	60.27
Total		73	100.00
Reinforcement learning	DL	1	33.33
	Syntactics, DL	1	33.33
	Other (< 5 each)	1	33.33
Total		3	100.00
TOTAL		519	100.00

Bold indicates subtotals of the corresponding rows

19.65% ($N = 102$) of the studies do not require labelled data, which are mostly unsupervised (18.69%, $N = 97$). In addition, hybrid studies are the third largest group (14.07%, $N = 73$).

We further analysed the approaches under each paradigm and focused on three for more details: deep learning (DL), traditional machine learning (ML), and Linguistic and Statistical Approaches. The results are detailed below and presented in Fig. 7 and Tables 5, 6.

4.4 Approaches

As shown in Fig. 7a and Table 5, among the 519 reviewed studies, 60.31% ($N = 313$) employed DL approaches, and 30.83% ($N = 160$) are DL-only. The DL-only approach is particularly prominent among fully-supervised (47.15%, $N = 149$) and semi-supervised (31.82%, $N = 7$) studies. Supplementing DL with syntactical features is also the second most popular approach in fully-supervised studies (16.77%, $N = 53$).

(1) *DL Approaches*

Figure 7a suggests that the 313 studies involving DL approaches are dominated by Recurrent Neural Network (RNN)-based solutions (55.91%, $N = 175$), of which nearly half used a combination of RNN and the attention mechanism (26.52%, $N = 83$), followed by attention-only (19.17%, $N = 60$) and RNN-only (9.90%, $N = 31$) models. The RNN family mainly consists of Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Gated Recurrent Unit (GRU). These neural-networks are featured by sequential processing that captures temporal dependencies of text tokens, and can thus incorporate surrounding text as context for prediction (Liu et al 2020; Satyarthi and Sharma 2023). On the other hand, the sequential nature poses challenges with parallelisation and the exploding and vanishing gradient problems associated with long sequences (Vaswani et al 2017; Liu et al 2020). Although LSTM and GRU can mitigate these issues somewhat through cell state and memory controls, efficiency and long-dependency challenges still hinder their performance (Vaswani et al 2017; Liu et al 2020; Satyarthi and Sharma 2023). The attention mechanism complements RNNs by dynamically updating weights across the input sequence based on each element's relevance to the current task, and thus guides the model to focus on the most relevant elements (Vaswani et al 2017).

In addition, convolutional and graph-neural approaches [e.g. convolutional neural networks (CNN), graph neural networks (GNN), graph convolutional networks (GCN)] also play smaller but noticeable roles in DL-based ABSA studies. While CNN was commonly used as an alternative to the sequence models such as RNNs (Liu et al 2020; J et al 2021; Zhang et al 2020), the graph-based networks (GNN, GCN) were mainly used to model the non-linear relationships such as external conceptual knowledge (e.g. Liang et al 2022) and syntactic dependency structures (e.g. Fei et al 2023a, b; Li et al 2022c) that are not well captured by the sequential networks like RNNs and the flat structure of the attention modules. As a result, they inject richer context into the overall learning process (Du et al 2021; Xu et al 2020a; Wang et al 2022).

Figure 8 depicts the trend of the main approaches across the publication years. We excluded the one study pre-published for 2023 to avoid confusing trends. It is clear that DL approaches have risen sharply and taken dominance since 2017, mainly driven by the rapid growth in RNN- and attention-based studies. This coincides with the appearance of the Transformer architecture in 2017 (Vaswani et al 2017) and the resulting pre-trained

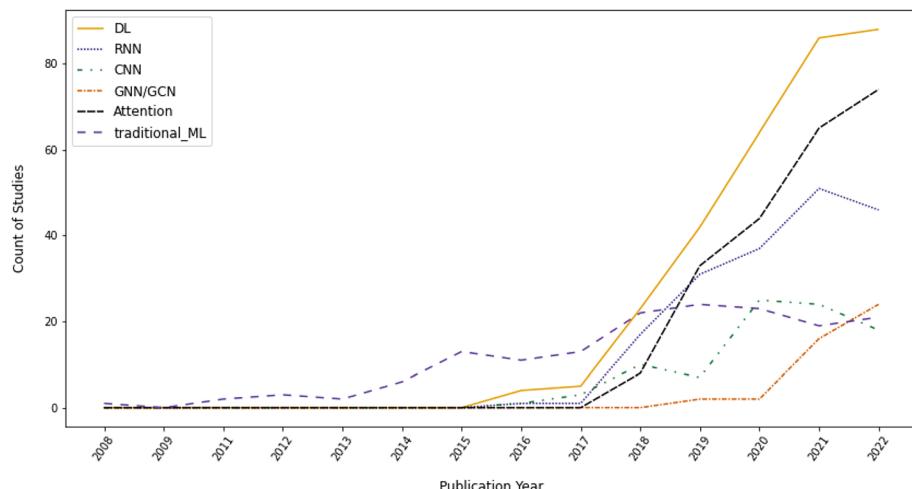


Fig. 8 Distribution of studies per method by publication year ($N = 1017$ with 519 unique studies). This graph excludes the one 2023 study (extracted in October 2022) to avoid trend confusion

models such as BERT (Devlin et al 2019) that were a popular embedding choice to be used alongside RNNs in DL and hybrid approaches (e.g., Li et al 2021; Zhang et al 2022a). GNN/GCN-based approaches remain small in number but have noticeable growth since 2020 ($N = 2, 2, 16, 24$ in each of 2019–2022, respectively), suggesting an increased effort to dynamically integrate relational context into the learning process within the DL framework.

(2) Traditional ML approaches

Interestingly, as shown in Fig. 8 traditional ML approaches remain a steady force over the decades despite the rapid rise of DL methods. Table 5 and Fig. 7b provide some insight into this: Among the 519 reviewed studies, while 60.31% employed DL approaches as mentioned in the previous sub-section, over half (54.53%, $N = 283$) also included traditional ML approaches, with the top 3 being Support Vector Machine (SVM; 20.14%, $N = 57$), Conditional Random Field (CRF; 14.49%, $N = 41$), and Latent Dirichlet allocation (LDA; 12.72%, $N = 36$). Table 5 suggests that among the major paradigms, traditional ML were often used in combination with DL approaches for fully-supervised studies (7.28%, $N = 23$), and along with linguistic rules, syntactic features, and/or lexicons and ontology in hybrid studies (27.40%, $N = 20$). Across all paradigms, traditional ML-only approaches are relatively rare (max $N = 7$).

(3) Linguistic and statistical approaches

While Table 5 illustrates the prevalence of fusing ML approaches with linguistic and statistical features or modules, there were 67 studies (12.91% out of the total 519) on pure

Table 6 Number of studies by pure linguistic or statistical approaches and publication year (N = 67)

Approach	2011–2013	2014–2016	2017–2019	2020–2022	Total	Total%
Syntactics, Rules, Lexicon	1	8	8	6	23	34.33%
Syntactics, Rules, Lexicon, Ontology	1	3	9	3	16	23.88%
Syntactics, Rules		2	4	5	11	16.42%
Statistical	2	1		1	4	5.97%
Syntactics, Rules, Lexicon, Statistical	1	1	1		3	4.48%
Syntactics, Rules, Statistical			2	1	3	4.48%
Lexicon				1	1	1.49%
Rules		1			1	1.49%
Rules, Lexicon			1		1	1.49%
Rules, Lexicon, Statistical				1	1	1.49%
Syntactics, Lexicon				1	1	1.49%
Syntactics, Lexicon, Statistical			1		1	1.49%
Syntactics, Statistical	1				1	1.49%
TOTAL	6	16	26	19	67	100.00%

linguistic or statistical approaches. As shown in Table 6, although small in number, these non-ML approaches have persisted over time. The most popular combination (34.33%, N = 23) was rules built on syntactic features (e.g. POS tags and dependency parse trees) and used along lexicon resources (e.g. domain-specific aspect lists, SentiWordNet³, MPQA⁴). A typical example is using POS tags and/or lexicon resources to narrow the scope of aspect or opinion term candidates, applying further rules based on POS tags or dependency relations for AE or OE, and/or using lexicon resources for candidate pruning, categorisation, or sentiment labelling (e.g. Asghar et al 2019; Dragoni et al 2019; Nawaz et al 2020). The second top combination (23.88%, N = 16) is the above-mentioned one plus ontology (e.g. domain-specific ontology, ConceptNet⁵, WordNet⁶) to bring in external knowledge of concepts and relations (e.g. Federici and Dragoni 2016; Marstawi et al 2017). Pure statistical methods were relatively rare (5.97%, N = 4), and mainly included frequency-based methods such as N-gram and TF-IDF, and other statistical modelling methods that were not commonly seen in the ML field.

4.5 ICL and generative approach in ABSA - Phase-2 targeted review results

ICL is a subgenre of the DL approach. However, we discuss the relevant results in this separate subsection due to the Phase-2 review's more focused sample and finer granularity. Despite the trending popularity of the ICL approach in NLP research and applications since 2022 (Dong et al 2024), our results suggest that the ABSA research community is just beginning to explore it with caution. Among the 208 ABSA studies from 2008 to 2024

³ <https://github.com/aesuli/SentiWordNet>.

⁴ <https://mpqa.cs.pitt.edu/>.

⁵ <https://conceptnet.io/>.

⁶ <https://wordnet.princeton.edu/>.

Table 7 Studies with in-context learning (ICL) approach on ABSA tasks ($N = 5$ out of 208 samples from 2008–2024)

Paper	Task	Non-ICL models	ICL models	ICL approach	Result
Zhou et al (2024)	ASQE	RoBERTa	ChatGPT	5-shot ICL	ChatGPT performed worse than almost all non-ICL methods with about 20–30% lower micro-F1
Liu et al (2024)	ASTE	BERT	ChatGPT	Zero-shot, 5-shot ICL	ChatGPT 5-shot ICL performed better than 0-shot ICL, but was still around 20% lower in F1 score than the main method
Su et al (2024)	ASC	T5	Llama2-7b-chat, Llama2-13b-chat, ChatGPT-3.5	N.A	ChatGPT 3.5 performed slightly better than the Llama-2 models across datasets, but was still up to about 20% lower than the main model in accuracy and F1
Amin et al (2024)	AE, ASC, OE	LSTM, RoBERTa (not fine-tuned)	GPT 3.5-Turbo, GPT-4	Zero-shot	On AE and ASC tasks, RoBERTa performed significantly better than the two GPT models with up to about 20% higher accuracy. GPT 3.5-Turbo performed better than GPT-4 on AE and ASC tasks, and outperformed all the other models with the OE task with up to 15% higher accuracy
Mughal et al (2024)	ASC, ACSA	LSTM, Flan-T5, DeBERTa, PaLM	GPT 3.5-Turbo, PaLM-bison	Zero-shot	PaLM showed the best overall performance on ASC and ACSA tasks in terms of accuracy and F1, closely followed by fine-tuned DeBERTa. However, PaLM had the lowest accuracy and F1 scores with the multi-aspect multi-sentiment (MAMS) dataset, with 21–48% lower accuracy and F1 than fine-tuned DeBERTa

The Non-ICL Models were trained or fine-tuned unless specified otherwise

Table 8 Studies with fine-tuned generative large language models (LLMs) on ABSA tasks (N = 18 out of 208 samples from 2008–2024)

Paper	Task	GenAI model
Hoang et al (2022)	ASQE	BART
Kang et al (2022)	OE	BART
Gong and Li (2022)	ASTE	BERT
William and Khodra (2022)	ASTE	T5
Lil et al (2023)	ASQE	BART (with BERT for upstream embedding)
Li et al (2023)	ASQE	BART
Suchady and Purwarianti (2023)	ASTE, AOPE, AE, OE, UABSA ^a	mT5 ^b
Yu et al (2023a)	AESC ^c , AOPE, ASTE	BART + GAT (Graph Attention Networks)
Yan et al (2023)	ASC, ASTE	T5
Yu et al (2023b)	AE, OE, ASQE	T5 (data generator + self-training) + BERT pair-classifier as descriminator
Lee and Kim (2023)	ASTE, ASQE, ASC	T5
Liu et al (2024)	ASTE	BERT + GPN (Global Pointer Network)
Lee et al (2024)	AE, ASC	GPT-2 fine-tuned by LoRA
Dang et al (2024)	AE, OE, ASC, ACD, AOPE, ACSA, ASTE, ACSD, ASQE	mT5, ViT5 ^b
Su et al (2024)	ASC	T5
Zhou et al (2024)	ASQE	RoBERTa
Wang et al (2024)	AE, OE, ACD, ASC, ACSA, AOPE, ASTE, ASQE, ASPE ^d , CSPE ^e	T5
Zhang et al (2024b)	ASQE	T5

^aUABSA Unified aspect-based sentiment analysis^bmT5, ViT5: multi-lingual variants of T5^cAESC aspect term extraction and sentiment classification^dASPE aspect sentiment pair extraction^eCSPE category sentiment pair extraction

containing at least one occurrence of the “Gen-LLM” keywords, only five (all published in 2024) applied ICL to both composite and traditional ABSA tasks. All of these studies were exploring the performance of foundation models via ICL against other approaches, rather than focusing on an ICL ABSA solution. Table 7 summarises the models, ABSA tasks, and key findings from these studies. Overall, four of the five studies found that zero-shot and even 5-shot ICL on foundation models (mainly GPTs) could not reach the performance of fine-tuned or fully trained DL models, especially those leveraging pre-trained LLMs to fine-tune a contextual-embedding.

In addition, we identified an emerging trend by examining the Phase-2 review non-ICL samples: Those employing fine-tuned generative LLMs mostly formulated the ABSA tasks as Sequence-to-Sequence (Seq2Seq) text generation problems, with a particular focus on composite tasks such as ASTE and ASQE. As shown in Table 8, within the 208 samples, a total of 18 studies (all from the new search) published in 2022–2024 applied pre-trained

generative LLMs with fine-tuning. The majority of these studies used models based on T5 ($N = 9$) and BART ($N = 5$) with the full Transformer (Vaswani et al 2017) encoder-decoder architecture, followed by encoder-only ($N = 3$, BERT and RoBERTa Liu et al 2019) and decoder-only ($N = 1$, GPT-2 Radford et al 2019) models. All but two of these 18 studies were on composite ABSA tasks, mainly ASTE and ASQE. Moreover, two studies (Yu et al 2023b; Zhang et al 2024b) also leveraged the generative capability of these LLMs to augment training data to enrich the fine-tuned embedding.

Compared with this Seq2Seq generation approach, the common applications of pre-trained LLMs in earlier studies from the main SLR sample often formulate the ABSA task as a classification problem (Zhang et al 2022c). These studies mostly use encoder-only LLMs for their pre-trained representations to fine-tune a contextual embedding (Zhang et al 2022c), which is then connected to other context-injection or relationship-learning modules and a classifier output layer. For instance, Zhang et al (2022a) employed pre-trained BERT with BiLSTM, a feed-forward neural network (FFNN), and CRF. Li et al (2021) used pre-trained BERT as an encoder and a decoder featuring a GRU. In contrast, the Seq2Seq generative approach can be illustrated by the signature “Generative Aspect-based Sentiment analysis (GAS)” proposed by Zhang et al (2021b), which leveraged the LLM’s pre-trained and fine-tuned encoder module for context-aware embedding and used the fine-tuned decoder module to generate text representations of the label sets (e.g., triplets) or as annotations next to the original input text (Zhang et al 2021b, 2022c).

5 Discussion

This review was motivated by the literature gap in capturing trends in ABSA research to answer higher-level questions beyond technical details, and the concern that the domain-dependent nature could predispose ABSA research to systemic hindrance from a combination of resource-reliant approaches and skewed resource domain distribution. By systematically reviewing the two waves of 727 in-scope primary studies published between 2008 and 2024, our quantitative analysis results identified trends in ABSA solution approaches, confirmed the above-mentioned concern, and provided detailed insights into the relevant issues. In this section, we examine the primary findings, share ideas for future research, and reflect on the limitations of this review.

5.1 Significant findings and trends

5.1.1 The out-of-sync research and dataset domains

Under RQ1, we examined the distributions of and relationships between our sample’s research (application) domains and dataset domains. The results showed strong skewness in both types of domains and a significant mismatch between them: While the majority (65.32%, $N = 339$) of the 519 studies did not aim for a specific research domain, a greater proportion (70.95%, $N = 447$) used datasets from the “product/service review” domain. A closer inspection of the link between the two domains revealed a clear mismatch: Among the 757 unique “research-domain, dataset-domain, dataset-name” combinations with ten or more studies: 90.48% ($N = 656$) of the studies in the “non-specific” research domain (95.77%, $N = 725$) used datasets from the “product/service review” domain. This suggests

that the lack of non-commercial-domain datasets could have forced generic technical studies to use benchmark datasets from a single popular domain. Given ABSA problem's domain-dependent nature, this could have indirectly hindered the solution development and evaluation across domains.

The results also showed that the other important and prevalent ABSA application domains such as education, medicine/healthcare, and public policy, were clearly under-researched and under-resourced. Among the reviewed samples from these three public-sector domains, about half of their datasets were created for the studies by their authors, indicating a lack of public dataset resources, hence the cost and challenge of developing ABSA research in these areas. As a likely consequence, even the most researched domain among these three had only 12 studies (2.31% out of 519) since 2008. The dataset resource scarcity in these public sector domains deserves more research community attention and support, especially given these domains' overall low research resources vs. the high cost and domain knowledge required for quality data annotation. In particular, for domains such as "Student feedback/education review" that often face strict data privacy and consent restrictions, it is crucial that the ABSA research community focus on creating ethical and open-access datasets to leverage community resources.

5.1.2 The dominance and limitations of the SemEval datasets

The results under RQ1 also revealed further issues with dataset diversity, even within the dominant "product/service review" domain. Out of the 757 unique "research-domain, dataset-domain, dataset-name" combinations with ten or more studies, 78.20% ($N = 592$) are taken up by the four popular SemEval datasets: The SemEval 2014 Restaurant and Laptop datasets alone account for 50.33% of all 757 entries, and the other two (SemEval 2015 and 2016 Restaurant).

The level of dominance of the SemEval datasets is alerting, not only because of their narrow domain range, but also for the inheritance and impact of the SemEval datasets' limitations. Several studies (e.g. Chebolu et al 2023; Xing et al 2020; Jiang et al 2019; Fei et al 2023a) suggest that these datasets fail to capture sufficient complexity and granularity of the real-world ABSA scenarios, as they primarily only include single-aspect or multi-aspect-but-same-polarity sentences, and thus mainly reflect sentence-level ABSA tasks and ignored subtasks such as multi-aspect multi-sentiment ABSA. The experiment results from Xing et al (2020), Jiang et al (2019) and Fei et al (2023a) consistently showed that all 35 ABSA models (including those that were state-of-the-art at the time) (9 in Xing et al 2020, 16 in Jiang et al 2019, 10 in Fei et al 2023a) that were trained and performed well on the SemEval 2014 ABSA datasets showed various extents of performance drop (by up to 69.73% in Xing et al 2020) when tested on same-source datasets created for more complex ABSA subtasks and robustness challenges. Given that the SemEval datasets are heavily used as both training data and "benchmark" to measure ABSA solution performance, their limitations and prevalence are likely to form a self-reinforcing loop that confines ABSA research. To break free from this dataset-performance self-reinforcing loop, it is critical that the ABSA research community be aware of this issue, and develop and adopt datasets and practices that are robustness-oriented, such as the automatic data-generation framework and the resulting Aspect Robustness Test Set (ARTS) developed by Xing et al (2020) for probing model robustness in distinguishing target and non-target aspects, and the Multi-Aspect Multi-Sentiment (MAMS) dataset created by Jiang et al (2019) to reflect

more realistic challenges and complexities in aspect-term sentiment analysis (ATSA) and aspect-category sentiment analysis (ACSA) tasks.

5.1.3 The reliance on labelled-datasets

The domain and dataset issues discussed above would not be as problematic if most ABSA studies employed methods that are dataset-agnostic. However, our results under RQ3 show the opposite. Only 19.65% ($N = 102$, with 97 being unsupervised) of the 519 reviewed studies do not require labelled data, whereas 66.28% ($N = 344$) are somewhat-supervised, and fully-supervised studies alone account for 60.89% ($N = 316$).

As demonstrated in Sect. 2.3, the domain can directly affect whether a chunk of text is considered an aspect or the relevant sentiment term, and plays a crucial role in contextual inferences such as implicit aspect extraction and multi-aspect multi-sentiment pairing. The domain knowledge reflected via ABSA labelled datasets can further shape the linguistic rules, lexicons, and knowledge graphs for non-machine-learning approaches; and define the underpinning feature space, representations, and acquired relationships and inferences for trained machine-learning models. When applying a solution built on datasets from a domain that is very remote from or much narrower than the intended application domain, it is predictable that the solution performance would be capped at subpar and even fail at more context-heavy tasks (Phan and Ogunbona 2020; You et al 2022; Liang et al 2022; Howard et al 2022; Chen and Qian 2022; Zhang et al 2022b; Nazir et al 2022b). Thus, domain transfer is crucially necessary for balancing the uneven ABSA research and resource distributions across domains. However, our finding that 17 out of the 20 reviewed cross-domain or domain-agnostic ABSA studies solely used datasets from the “product/service review” domain raised questions about these approaches’ generalisability and robustness in other domains, as well as whether such dataset choices became another reinforcement of concentrating research effort and benchmarks within this one dominant domain.

The rapid rise of deep learning (DL) in ABSA research could further add to the challenge of overcoming the negative impact of this domain mismatch and dataset limitations via the non-linear multi-layer dissemination of bias in the representation and learned relations, thus making problem-tracking and solution-targeting difficult. In reality, of the 519 reviewed studies, 60.31% ($N = 313$) employed DL approaches, and nearly half (47.15%, $N = 149$) of the fully-supervised studies and 30.83% ($N = 160$) of all reviewed studies were DL-only.

Moreover, RNN-based solutions dominate the DL approaches (55.91%, $N = 175$), mainly with the RNN-attention combination (26.52%, $N = 83$) and RNN-only (9.90%, $N = 31$) models. RNN and its variants such as LSTM, BiLSTM, and GRU are known for their limitations in capturing long-distance relations due to their sequential nature and the subsequent memory constraints (Vaswani et al 2017; Liu et al 2020). Although the addition of the attention mechanism enhances the model’s focus on more important features such as aspect terms (Vaswani et al 2017; Liu et al 2020), traditional attention weights calculation struggles with multi-word aspects or multi-aspect sentences (Liu et al 2020; Fan et al 2018). In addition, whilst 16.77% ($N = 53$) of the fully-supervised studies introduced syntactical features to their DL solutions, additional features also increased the input size. According to Prather et al (2020), sequential models, even the state-of-the-art LLMs, showed impaired performance as the input grew longer and could not always benefit from additional features.

5.1.4 The potential of generative LLMs and foundation models

Lastly, the Phase-2 targeted review highlights the ABSA community's caution toward the direct adoption of generative foundation models, with only five out of 208 recent studies testing the ICL approach and most yielding subpar results compared to other methods. However, most of these studies only tested zero-shot instructions with simple model settings. It is worth further exploring the potential of foundational models and ICL in ABSA by focusing more on instruction and example engineering, model parameter optimisation, and task re-formulation (Dong et al 2024).

On the other hand, the fine-tuning of smaller generative LLMs has seen increasing adoption through the "ABSA as Seq2Seq text generation" approach, demonstrating promising task performance. Although this generative approach can incorporate data augmentation and self-training to reduce reliance on labelled datasets, the cost of fine-tuning, the need for labelled base data, and the domain-transfer problem remain significant challenges (Zhang et al 2022c). In this context, the task adaptability and multi-domain pre-trained knowledge of foundation models could provide potential solutions.

As Zhang et al (2022c) noted, progress in applying pre-trained LLMs and foundation models to ABSA could be impeded by dataset resources constraints. To match the parameter size of these models, more diverse, complex, and larger datasets are required for effective fine-tuning or comprehensive testing. In low-resource domains where dataset resources are already limited, this requirement could further complicate the adoption of these technologies (Satyarthi and Sharma 2023).

5.2 Ideas for future research

Overall, by adopting a "systematic perspective, i.e., model, data, and training" (Fei et al 2023a, p.28) combined with a quantitative approach, we identified high-level trends unveiling the development and direction of ABSA research, and found clear evidence of large-scale issues that affect the majority of the existing ABSA research. The skewed domain distributions of resources and benchmarks could also restrict the choice of new studies. On the other hand, this evidence also highlights areas that need more attention and exploration, including: ABSA solutions and resource development for the less-studied domains (e.g. education and public health), low-resource and/or data-agnostic ABSA, domain adaptation, alternative training schemes such as adversarial (e.g. Fei et al 2023a; Chen et al 2021) and reinforcement learning (e.g. Vasanthi et al 2022; Wang et al 2021b), and more effective feature and knowledge injection. Future research could contribute to addressing these issues by focusing on ethically producing and sharing more diverse and challenging datasets in minority domains such as education and public health, improving data synthesis and augmentation techniques, exploring methods that are less data-dependent and resource-intensive, and leveraging the rapid advancements in pre-trained LLMs and foundation models.

In addition, our results also revealed emerging trends and new ideas. The relatively recent growth of end-to-end models and composite ABSA subtasks provide opportunities for further exploration and evaluation. The fact that hybrid approaches with non-machine-learning techniques and non-textual features remain steady forces in the field after nearly three decades suggests valuable characteristics that are worth re-examining under the light of new paradigms and techniques. Moreover, the small number of Phase-2 samples using

ICL and fine-tuning generative LLM approaches may suggest that we have only captured early adopters. More thorough exploration of these approaches and continued tracking of their development alongside other methods are necessary to understand how the ABSA community can leverage the resources and capabilities embedded within LLMs and foundation models.

Lastly, it is crucial that the community invest in solution robustness, especially for machine-learning approaches (Xing et al 2020; Jiang et al 2019; Fei et al 2023a). This could mean critical examination of the choice of evaluation metrics, tasks, and benchmarks, and being conscious of their limitations vs. the real-world challenges. The “State-Of-The-Art” (SOTA) performance based on certain benchmark datasets should never become the motivation and holy grail of research, especially in fields like ABSA where the real use cases are often complex and even SOTA models do not generalise far beyond the training datasets. More attention and effort should be paid to analysing the limitations and mistakes of ABSA solutions, and drawing from the ideas of other disciplines and areas to fill the gaps.

5.3 Limitations

We acknowledge the following limitations of this review: First, our sample scope is by no means exhaustive, as it only includes primary studies from four peer-reviewed digital databases and only those published in the English language. Although this can be representative of a core proportion of ABSA research, it does not generalise beyond this without assumptions. The “peer-reviewed” criteria also meant that we overlooked preprint servers such as arXiv.org that more closely track the latest development of ML and NLP research. Second, no search string is perfect. Our database search syntax and auto-screening keywords represent our best effort in capturing ABSA primary studies, but may have missed some relevant ones, especially with the artificial “total pages < 3” and “total keyword (except SA, OM) outside Reference < 5” exclusion criteria. Moreover, our search completeness might have been affected by the performance of the database search engines. This is evidenced by the significant number of extracted search results that were entirely irrelevant to the search keywords, as well as our abandonment of the 2024 SpringerLink search due to interface issues. Enhancements in digital database search capabilities could significantly improve the effectiveness and reliability of future literature review studies, particularly SLRs. Third, we may have missed datasets, paradigms, and approaches that are not clearly described in the primary studies, and our categorisation of them is also subject to the limitations of our knowledge and decisions. Future review studies could consider a more innovative approach to enhance analytical precision and efficiency, such as applying ABSA and text summarisation alongside the screening and reviewing process. Fourth, we did not compare solution performance across studies due to the review focus, sample size, and the variability in experimental settings across studies. Evaluating the effectiveness of comparable methods and the suitability of evaluation metrics would enhance our findings and offer more valuable insights.

6 Conclusion

ABSA research is riding the wave of the explosion of online digital opinionated text data and the rapid development of NLP resources and ideas. However, its context- and domain-dependent nature and the complexity and inter-relations among its subtasks pose challenges to improving ABSA solutions and applying them to a wider range of domains. In this review, we systematically examined existing ABSA literature in terms of their research application domain, dataset domain, and research methodologies. The results suggest a number of potential systemic issues in the ABSA research literature, including the predominance of the “product/service review” dataset domain among the majority of studies that did not have a specific research application domain, coupled with the prevalence of dataset-reliant methods such as supervised machine learning. We discussed the implication of these issues to ABSA research and applications, as well as their implicit effect in shaping the future of this research field through the mutual reinforcement between resources and methodologies. We suggested areas that need future research attention and proposed ideas for exploration.

Appendix A: Aspect-based sentiment analysis (ABSA)

Appendix A.1: Definition and examples

Aspect-based sentiment analysis (ABSA) is a sub-domain of fine-grained SA (Nazir et al 2022a). ABSA focuses on identifying the sentiments towards specific entities or their attributes/ features called *aspects* (Nazir et al 2022a; Akhtar et al 2020). An aspect can be explicitly expressed in the text (*explicit aspect*) or absent from the text but implied from the context (*implicit aspects*) (Maitama et al 2020; Xu et al 2020b). Moreover, the aspect-level sentiment could differ across aspects and be different from the overall sentiment of the sentence or the document (e.g. Akhtar et al 2020, 2017; Li et al 2022a). Some studies further distinguish aspect into *aspect term* and *aspect category*, with the former referring to the aspect expression in the input text (e.g. “pizza”), and the latter a latent construct that is usually a high-level category across aspect terms (e.g. “food”) that are either identified or given (Chauhan et al 2019; Akhtar et al 2018).

The following examples illustrate the ABSA terminologies:

Example 1 (From a restaurant review⁷): “*The restaurant was expensive, but the menu was great.*” This sentence has one explicit aspect “menu” (sentiment term: “great”, sentiment polarity: positive), one implicit aspect “price” (sentiment term: “expensive”, sentiment polarity: negative). Depending on the target/given categories, the aspects can be further classified into categories, such as “menu” into “general” and “price” into “price”.

Example 2 (From a laptop review⁸): “*It is extremely portable and easily connects to WIFI at the library and elsewhere.*” This sentence has two implicit aspects: “portability” (sentiment term: “portable”, sentiment polarity: positive), “connectivity” (sentiment term:

⁷ <https://alt.qcri.org/semeval2014/task4/>.

⁸ <https://alt.qcri.org/semeval2015/task12/>.

“easily”, sentiment polarity: positive). The aspects can be further classified into categories, such as both under “laptop” (as opposed to “software” or “support”).

Example 3 (Text from a course review): *“It was too difficult and had an insane amount of work, I wouldn’t recommend it to new students even though the tutorial and the lecturer were really helpful.”* The two explicit aspects in Example 3 are “tutorial” and “lecturer” (sentiment terms: “helpful”, polarities: positive). The implicit aspects are “content” (sentiment term: “too difficult”, sentiment polarity: negative), “workload” (sentiment term: “insane amount”, sentiment polarity: negative), and “course” (sentiment term: “would not recommend”, sentiment polarity: negative). An illustration of aspect categories would be assigning the aspect “lecturer” to the more general category “staff” and “tutorial” to the category “course component”.

As demonstrated above, the fine granularity makes ABSA more targetable and informative than document- or sentence-level SA. Thus, ABSA can precede downstream applications such as attribute weighting in overall review ratings (e.g. Da'u et al 2020), aspect-based opinion summarisation (e.g. Yauris and Khodra 2017; Kumar et al 2022; Almatrafi and Johri 2022), and automated personalised recommendation systems (e.g. Ma et al 2017; Nawaz et al 2020).

Compared with document- or sentence-level SA, while being the most detailed and informative, ABSA is also the most complex and challenging (Huan et al 2022). The most noticeable challenges include the number of ABSA subtasks, their interrelations and context dependencies, and the generalisability of solutions across topic domains.

Appendix A.2: ABSA Subtasks

A full ABSA solution has more subtasks than coarser-grained SA. The most fundamental ones (Li et al 2022a; Huan et al 2022; Li et al 2020; Fei et al 2023b) include:

Aspect (term) extraction/identification (AE), which has a slight variation in meaning depending on the overall ABSA approach. Some authors (e.g. (Zhang et al 2023; Luo et al 2019; Ruskanda et al 2019)) consider AE as identifying the attribute or entity that is the target of an opinion expressed in the text and sometimes call it “opinion target extraction” (Guo et al 2018). In these cases, opinion terms were often identified in order to find their target aspect terms. Others (e.g. Akhtar et al 2020; Gunes 2016; Li et al 2020; Ettaleb et al 2022; Tran et al 2020) define AE as identifying the key or all attributes of entities mentioned in the text. *Implicit-Aspect Extraction (IAE)* is often mentioned as a task by itself due to its technical challenge.

Opinion (term) Extraction/Identification (OE), which relates to identifying the “opinion terms” or the sentiment expression of a specific entity/aspect (e.g. Li et al 2022a; Wang et al 2018; Fernando et al 2019; Fei et al 2023b). In Example 1 above, an OE task would extract the sentiment terms “great” (associated with the aspect term “menu”) and “expensive” (associated with the implicit aspect “price”).

Aspect-Sentiment Classification (ASC), which refers to obtaining the sentiment polarity category (e.g. negative, neutral, positive, conflict) or sentiment score (e.g. 1 to 5 or -1 to 1 along the scale from negative to positive) associated with a given aspect or aspect category (e.g. Akhtar et al 2020; Gojali and Khodra 2016; Castellanos et al 2011). This is often done via evaluating the associated opinion term(s), and sentiment lexicon

resources such as the SentiWordNet (Baccianella et al 2010) and SenticNet (Cambria et al 2016) can be used to assign polarity scores (Gojali and Khodra 2016). Sentiment scores can be further aggregated across opinion terms for the same aspect, or across aspect terms to generate higher-level ratings, such as aspect-category ratings within or across documents (Gojali and Khodra 2016; Castellanos et al 2011).

As an extension of AE, some studies also involve *Aspect-Category Detection (ACD)* and *Aspect Category Sentiment Analysis (ACSA)* when the focus of sentiment analysis is on (often pre-defined) latent topics or concepts and requires classifying aspect terms into categories (Pathan and Prakash 2022).

Traditional full ABSA solutions often perform the subtasks in a pipeline manner (Li et al 2022b; Nazir and Rao 2022) using one or more of the linguistic (e.g. lexicons, syntactic rules, dependency relations), statistical (e.g. n-gram, Hidden Markov Model (HMM)), and machine-learning approaches (Maitama et al 2020; Cortis and Davis 2021; Federici and Dragoni 2016). For instance, for AE and OE, some studies used linguistic rules and sentiment lexicons to first identify opinion terms and then the associated aspect terms of each opinion term, or vice versa (e.g. You et al 2022; Cavalcanti and Prudêncio 2017), and then moved on to ASC or ACD using a supervised model or unsupervised clustering and/or ontology (Nawaz et al 2020; Gojali and Khodra 2016). Hybrid approaches are common given the task combinations in a pipeline.

With the rise of multi-task learning and deep learning (Chen et al 2022), an increasing number of studies explore ABSA under an *End-to-end (E2E)* framework that performs multiple fundamental ABSA subtasks in one model to better capture the inter-task relations (Liu et al 2024), and some combine them into a single composite task (Huan et al 2022; Li et al 2022b; Zhang et al 2022b). These composite tasks are most commonly formulated as a sequence- or span-based tagging problem (Huan et al 2022; Li et al 2022b; Nazir and Rao 2022). The most common composite tasks are: *Aspect-Opinion Pair Extraction (AOPE)*, which directly outputs {aspect, opinion} pairs from text input (Nazir and Rao 2022; Li et al 2022c; Wu et al 2021) such as “⟨menu, great⟩” from Example 1; *Aspect-Polarity Co-Extraction (APCE)* (Huan et al 2022; He et al 2019), which outputs {aspect, sentiment polarity} pairs such as “⟨menu, positive⟩”; *Aspect-Sentiment Triplet Extraction (ASTE)* (Huan et al 2022; Li et al 2022b; Du et al 2021; Fei et al 2023b), which outputs {aspect, opinion, sentiment category} triplets, such as “⟨menu, great, positive⟩”; and *Aspect-Sentiment Quadruplet Extraction/Prediction (ASQE/ASQP)* (Zhang et al 2022a; Lim and Buntine 2014; Zhang et al 2021a, 2024a) that outputs {aspect, opinion, aspect category, sentiment category} quadruplets, such as “⟨menu, great, general, positive⟩”.

Appendix A.3: Other ABSA reviews

As this review focuses on trends instead of detailed solutions and methodologies, we refer interested readers to existing review papers that provide comprehensive and in-depth summaries of common ABSA subtask solutions and approaches, for example:

- *Explicit and implicit AE*: Rana and Cheah (2016), Ganganwar and Rajalakshmi (2019), Soni and Rambola (2022), Maitama et al (2020).
- *Deep learning (DL) methods for ABSA*: Do et al (2019), Liu et al (2020), Wang et al (2021a), Chen and Fnu (2022), Zhang et al (2022c), Mughal et al (2024). Specifically:

Table 9 Digital databases and search details used for this systematic literature review (SLR)

Database	Search string	Search criteria	PDFs exported	Type of exported publications
ACM Digital Library	((“aspect based” OR “aspect-based”) AND (sentiment OR extraction OR extract OR mining) OR “opinion mining”)	<ul style="list-style-type: none"> No year filter Search scope = Full article Content Type = Research Article Media Format = PDF Publications = Journals OR Proceedings OR Newsletters 	1514	<ul style="list-style-type: none"> • 201 articles • 1283 conference papers • 30 newsletters
IEEE Xplore	((“aspect based” OR “aspect-based”) AND (sentiment OR extraction OR extract OR mining) OR “opinion mining”)	<ul style="list-style-type: none"> Year filter = 2004–2022 (pilot search suggested that 1995–2003 results were irrelevant) Search scope = Full article Publications = Not “Books” 	1639	<ul style="list-style-type: none"> • 165 articles • 1445 conference papers • 29 magazine pieces
Science Direct	((“aspect based” OR “aspect-based”) AND (sentiment OR extraction OR extract OR mining) OR “opinion mining”)	<ul style="list-style-type: none"> No year filter Search scope = Title, abstract or author-specified keywords Publications = Research Articles only 	497	<ul style="list-style-type: none"> • 497 articles
SpringerLink	((“aspect based” OR “aspect-based”) AND (sentiment OR extraction OR extract OR mining) OR “opinion mining”)	<ul style="list-style-type: none"> No year filter English results only Publications = Article, Conference Paper 	541	<ul style="list-style-type: none"> • 218 articles • 323 conference papers

- *DL methods for ASC*: Zhou et al (2019), Satyarthi and Sharma (2023).
 - *E2E ABSA, composite tasks, and pre-trained Large Language Models (LLMs) in ABSA*: Zhang et al (2022c) provided a comprehensive review and shared extensive reading lists and dataset resource links via <https://github.com/IsakZhang/ABSA-Survey>. Mughal et al (2024) introduced common benchmark datasets, including more challenging ones for composite ABSA tasks. They also reviewed and tested the ABSA task performance of representative RNN-based models and pre-trained LLMs.
- *Multimodal ABSA*: Zhao et al (2024).

Appendix B: Full SLR methodology

This section provides a complete, detailed description of the SLR methodology and procedures.

Appendix B.1: Research identification

To obtain the files for review, we conducted database searches between 24–25 October 2022, when we manually queried and exported a total of 4191 research papers’ PDF and BibTeX (or the equivalent) files via the web interfaces of four databases. Table 9 details the search string, search criteria, and the PDF files exported from each database.

Given the limited search parameters allowed in these digital databases, we adopted a “search broad and filter later” strategy. These database search strings were selected based on pilot trials to capture the ABSA topic name, the relatively prevalent yet unique ABSA subtask term (“extraction”), and the interchangeable use between ABSA and opinion mining; while avoiding generating false positives from the highly active, broader field of SA. The “filter later” step was carried out during the “selection of primary studies” stage introduced in the next section, which aimed at excluding cases where the keywords are only mentioned in the reference list or sparsely mentioned as a side context, and opinion mining studies that were at document or sentence levels.

Appendix B.2: Selection of primary studies

After obtaining the 4191 initial search results, we conducted a pilot manual file examination of 100 files to refine the pre-defined inclusion and exclusion criteria. We found that some search results only contained the search keywords in the reference list or Appendix, which was also reported in Prather et al (2020). In addition, there are a number of papers that only mentioned ABSA-specific keywords in their literature review or introduction sections, and the studies themselves were on coarser-grained sentiment analysis or opinion mining. Lastly, there were instances of very short research reports that provided insufficient details of the primary studies. Informed by these observations, we refined our inclusion and exclusion criteria to those in Table 1 in Sect. 3. Note that we did not include popularity criteria such as citation numbers so we can better identify novel practices and

Table 10 The automatic and manual screening processes

Inclusion/exclusion types	File count
Total extracted files	4191
<i>Auto-excluded</i>	3277
Step_1—Duplicate DOI across databases	19
Step_2—Duplicate Title across databases	11
Step_3—Survey/Not primary study papers	149
Step_4—Total keyword outside reference match = 0	26
Step_5—keyword matched only to SA and/or OM outside reference ^a	2235
Step_6—No sentiment\W+analysis in keyword outside reference	243
Step_7—Total pages < 3	7
Step_8—Total keyword (except SA, OM) outside reference < 5 ^{a,b}	587
<i>Manually excluded</i>	395
Type—Article withdrawn	1
Type—Not published in English	1
Type—Dataset unclear	1
Type—Low quality, unclear definition of ABSA	1
Type—No original method	2
Type—Duplicate of another included paper	3
Type—Sentence level SA	4
Type—Not text-data focused	6
Type—Not primary study	15
Type—Lack of details on ABSA tasks	31
Type—Review (Not primary study)	43
Type—No ABSA task experiment details/results	49
Type—Not ABSA focused; No ABSA task experiment details/results	238
Final included for review	519

^aSA, OM: the Regex keyword patterns ‘sentiment\W+analysis’, ‘opinion\W+mining’ respectively

^bThe occurrence threshold 5 was chosen based on pilot file examination, which suggested that files with target keyword occurrence below this threshold tended to be non-ABSA-focused

Bold indicates subtotals of the corresponding rows

avoid mainstream method over-dominance introduced by the citation chain (Chu and Evans 2021).

To implement the inclusion and exclusion criteria, we first applied PDF mining to automatically exclude files that meet the exclusion criteria, and then refined the selection with manual screening under the exclusion and inclusion criteria. Both of these processes are detailed below. Our PDF mining for automatic review screening code is also available at <https://doi.org/10.5281/zenodo.12872948>.

The automatic screening consists of a pipeline with two Python packages: Pandas (Team 2023) and PyMuPDF.⁹ We first used Pandas to extract into a dataframe (i.e. table) all exported papers’ file locations and key BibTex or equivalent information including title, year, page number, DOI, and ISBN. Next, we used PyMuPDF to iterate through

⁹ <https://pypi.org/project/PyMuPDF/>.

each PDF file and add to the dataframe multiple data fields: whether the file was successfully decoded¹⁰ for text extraction (if marked unsuccessful, the file was marked for manual screening), the occurrence count of each Regex keyword pattern listed below, and whether each keyword occurs after the section headings that fit into Regex patterns that represent variations of “references” and “bibliography” (referred to as “non-target sections” below). We then marked the files for exclusion by evaluating the eight criteria listed under “Auto-excluded” in Table 10 against the information recorded in the dataframe. Each of the auto-exclusion results from Steps 1–4 and 7 in Table 10 were manually checked, and those under Steps 5, 6, and 8 were spot-checked. These steps excluded 3277 out of the 4194 exported files.

Below are the regex patterns used for automatic keyword extraction and occurrence calculation:

PDF search keyword Regex list: ['absa', 'aspect\W+base\w*', 'aspect\W+extrac\w*', 'aspect\W+term\w*', 'aspect\W+level\w*', 'term\W+level', 'sentiment\W+analysis', 'opinion\W+mining']

For the 914 files filtered through the auto-exclusion process, we manually screened them individually according to the inclusion and exclusion criteria. As shown in the second half of Table 10, this final screening step refined the review scope to 519 papers.

Appendix B.3: Data extraction and synthesis

In the final step of the SLR, we manually reviewed each of the 519 in-scope publications and recorded information according to a pre-designed data extraction form. The key information recorded includes each study’s research focus, research application domain (“research domain” below), ABSA subtasks involved, name or description of all the datasets directly used, model name (for machine-learning solutions), architecture, whether a certain approach or paradigm is present in the study (e.g. supervised learning, deep learning, end-to-end framework, ontology, rule-based, syntactic-components), and the specific approach used (e.g. attention mechanism, Naïve Bayes classifier) under the deep learning and traditional machine learning categories.

After the data extraction, we performed data cleaning to identify and fix recording errors and inconsistencies, such as data entry typos and naming variations of the same dataset across studies. Then we created two mappings for the research and dataset domains described below.

For each reviewed study, its research domain was defaulted to “non-specific” unless the study mentioned a specific application domain or use case as its motivation, in which case that domain description was recorded instead.

The dataset domain was recorded and processed at the individual dataset level, as many reviewed studies used multiple datasets. We standardised the recorded dataset names, checked and verified the recorded dataset domain descriptions provided by the authors or the source web-pages, and then manually categorised each domain description into a domain category. For published/well-known datasets, we unified the recorded naming variations and checked the original datasets or their descriptions to verify the domain descriptions. For datasets created (e.g. web-crawled) by the authors of the reviewed studies, we

¹⁰ <https://pdfminersix.readthedocs.io/en/latest/faq.html>.

named them following the “[source] [domain] (original)” format, e.g. “Yelp restaurant review (original)”, or “Twitter (original)” if there was no distinct domain, and did not differentiate among the same-name variations. In all of the above cases, if a dataset was not created with a specific domain filter (e.g. general Twitter tweets), then it was classified as “non-specific”.

The recorded research and dataset domain descriptions were then manually grouped into 19 common domain categories. We tried to maintain consistency between the research and dataset domain categories. The following are two examples of possible mapping outcomes:

1. A study on a full ABSA solution without mentioning a specific application domain and using Yelp restaurant review and Amazon product review datasets would be assigned a research domain of “non-specific” and a dataset domain of “product/service review”.
2. A study mentioning “helping companies improve product design based on customer reviews” as the motivation would have a research domain of “product/service review”, and if they used a product review dataset and Twitter tweets crawled without filtering, the dataset domains would be “product/service review” and “non-specific”.

After applying the above-mentioned standardisation and mappings, we analysed the synthesised data quantitatively using the Pandas (Team 2023) library to obtain an overview of the reviewed studies and explore the answers to our RQs.

Appendix C: Additional results

See Figs. 9, 10 and Tables 11, 12, 13, 14, 15.

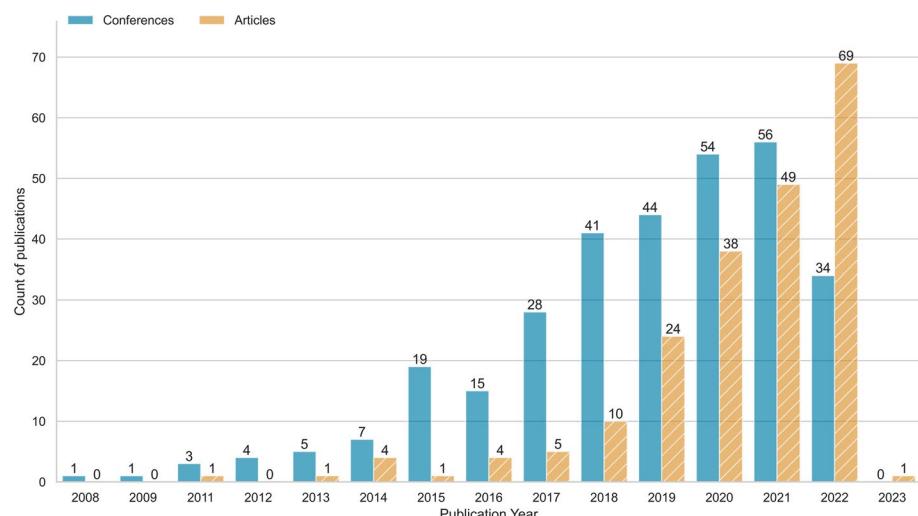


Fig. 9 Number of included studies by publication year and type (N = 519). Note Although our original search scope included journal articles, conference papers, newsletters, and magazine articles, the final 519 in-scope studies consist of only journal articles and conference papers. Conference papers noticeably outnumbered journal articles in all years until 2022, with the gap closing since 2016. We think this trend could be due to multiple factors, such as the fact that our search was conducted in late October 2022 when some conference publications were still not available; the publication lag for journal articles due to a longer processing period; and potentially a change in publication channels that is outside the scope of this review

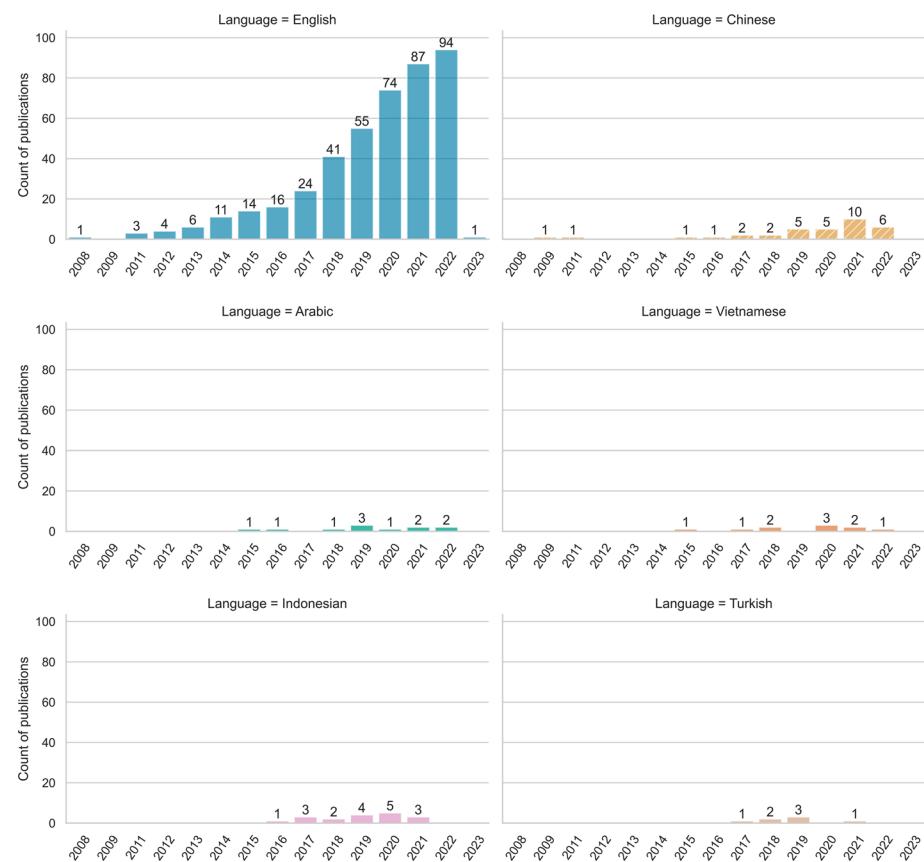


Fig. 10 Number of included studies with the top 5 dataset languages by publication year

Table 11 Number of studies by ABSA subtask combinations (N = 519)

Subtask combination	Count of studies	% of studies (%)
AE, ASC	168	32.37
ASC	160	30.83
AE	79	15.22
ACD, ASC	16	3.08
ASTE	15	2.89
AE, ACD, ASC	13	2.50
AE, OE, ASC	9	1.73
AE, ACD	9	1.73
AOPE	9	1.73
ACD	7	1.35
AE, OE	7	1.35
AE, OE, ACD	4	0.77
AE, OE, ACD, ASC	4	0.77
ASQE	2	0.39
AOPE, ASC	2	0.39
OE	1	0.19
OTE (opinion target extraction), ASC	1	0.19
Aspect-based embedding, ACD, ASC	1	0.19
ASC, OE	1	0.19
Aspect and synthetic sample discrimination	1	0.19
AE, OE, ASC, AOPE, ASTE	1	0.19
AOPE, ASC, ACD	1	0.19
AOPE, ACD	1	0.19
AE, review-level SA	1	0.19
ACD, AE, ASC	1	0.19
AE, ASC, Aspect-based sentence segmentation	1	0.19
AE, ASC, Aspect-based embedding	1	0.19
AE, ASC, ACD	1	0.19
AE, ACD, OE	1	0.19
Data augmentation	1	0.19
Total	519	100.00

^aThis table corresponds to Fig. 5a

Table 12 Number of studies by individual ABSA subtasks (N = 805)

Individual subtask	Count of studies	% of Studies (%)
ASC	381	47.33
AE	300	37.27
ACD	59	7.33
OE	28	3.48
ASTE	16	1.99
AOPE	14	1.74
ASQE	2	0.25
Aspect and synthetic sample discrimination	1	0.12
OTE (opinion target extraction)	1	0.12
Aspect-based sentence segmentation	1	0.12
Data augmentation	1	0.12
Review-level SA	1	0.12
Total	805	100

^aThis table corresponds to Fig. 5b

Table 13 Number of studies by deep learning (DL) approaches (N = 313)

DL approach	Count of studies	% of studies (%)
RNN, Attention	83	26.52
Attention	60	19.17
RNN	31	9.90
RNN, CNN, Attention	25	7.99
CNN	20	6.39
RNN, CNN	19	6.07
GNN/GCN, Attention	17	5.43
CNN, Attention	17	5.43
RNN, GNN/GCN, Attention	17	5.43
Other (< 5 % each)	24	7.67
Total	313	100.00

^aThis table corresponds to Fig. 7a

Table 14 Number of studies by traditional machine learning approaches (N = 283)

Traditional ML	Count of studies	% of studies (%)
Support vector machine (SVM)	57	20.14
Conditional random field (CRF)	41	14.49
Latent Dirichlet allocation (LDA)	36	12.72
Naïve Bayes (NB)	32	11.31
Random forest (RF)	24	8.48
Decision tree (DT)	18	6.36
Logistic regression (LR)	15	5.30
K-Nearest neighbors (KNN)	12	4.24
Multinomial Naïve Bayes (MNB)	10	3.53
K-means	5	1.77
Boosting	5	1.77
Other (N < 5)	28	9.89
Total	283	100.00

^aThis table corresponds to Fig. 7b

Table 15 Number of studies per dataset (N = 1179, with 519 studies and 218 datasets)

Datasets	Count of studies	% of studies (%)
SemEval 2014 Restaurant	211	17.90
SemEval 2014 Laptop	189	16.03
SemEval 2016 Restaurant	118	10.01
SemEval 2015 Restaurant	106	8.99
Twitter (Dong et al. 2014)	55	4.66
Amazon customer review datasets (Hu and Liu 2004a)	33	2.80
Amazon product review (original)	23	1.95
Product review (original)	22	1.87
Twitter (original)	21	1.78
SemEval 2015 Laptop	20	1.70
Yelp Dataset Challenge Reviews	16	1.36
SemEval 2016 Laptop	14	1.19
TripAdvisor Hotel review (original)	11	0.93
Hotel review (original)	10	0.85
TripAdvisor Hotel review (Wang et al. 2010)	9	0.76
TripAdvisor Restaurant review (original)	9	0.76
Movie review (original)	8	0.68
MAMS Multi-Aspect Multi-Sentiment dataset (Jiang et al. 2019)	8	0.68
SemEval 2016 Hotel	7	0.59
Twitter (Mitchell et al. 2013)	7	0.59
Amazon product review (McAuley et al. 2015)	7	0.59
VLSP 2018 Restaurant review	6	0.51
Restaurant review (Ganu et al. 2009)	6	0.51
Student feedback (original)	6	0.51
VLSP 2018 Hotel review	6	0.51
Chinese product review (Peng et al. 2018)	5	0.42
Stanford Twitter Sentiment / Sentiment140 (Go et al. 2009)	5	0.42
Yelp restaurant review (original)	5	0.42
SentiHood (Saeidi et al. 2016)	4	0.34
Game review (original)	3	0.25
Sanders Twitter Corpus (STC) (Sanders 2011)	3	0.25
TripAdvisor Tourist review (original)	3	0.25
Coursera course review (original)	3	0.25
Online drug review (original)	3	0.25
Restaurant review (original)	3	0.25
Chinese Restaurant review (original)	3	0.25
Financial Tweets and News Headlines dataset (FiQA 2018)	2	0.17
SentiRuEval-2015	2	0.17
Amazon laptop reviews (He and McAuley 2016)	2	0.17
Hindi ABSA dataset (Akhtar et al. 2016)	2	0.17
Hotel review (Kaggle)	2	0.17
Service review (Toprak etc. 2010)	2	0.17
Indonesian Product review (original)	2	0.17
Indonesian Tweets—Political topic (original)	2	0.17

Table 15 (continued)

Datasets	Count of studies	% of studies (%)
Steam game review (original)	2	0.17
Stanford Sentiment Treebank review data (Socher et al. 2013)	2	0.17
Kaggle movie review	2	0.17
Zomato restaurant review (original)	2	0.17
Turkish Product review (original)	2	0.17
SemEval 2015 Hotel	2	0.17
Weibo comments (original)	2	0.17
Amazon 50-domain Product review (Chen and Liu 2014)	2	0.17
Bangla ABSA Cricket, Restaurant dataset (Rahman and Dey 2018)	2	0.17
YouTube comments (original)	2	0.17
Beer review (McAuley et al. 2012)	2	0.17
CCF BDCI 2018 Chinese auto review dataset	2	0.17
ReLi Portuguese book reviews (Freitas et al. 2014)	2	0.17
Twitter product review (original)	2	0.17
SemEval 2016 Tweets	2	0.17
Product review dataset (Liu et al. 2015)	2	0.17
Product review dataset (Cruz et al. 2014)	2	0.17
Chinese Product review (original)	2	0.17
ICLR Open Reviews dataset	2	0.17
Amazon Electronics review data (Jo and Oh 2011)	2	0.17
Amazon product review (Wang et al. 2011)	2	0.17
SemEval 2015 (unspecified)	2	0.17
Product & service review (Kaggle)	2	0.17
Other datasets (N = 1 each)	149	12.64
Total	1179	100.00

Author contributions Y. C. H designed, conducted, and wrote this review. P. D., J. W., and K. T. guided the design of the review methodology and review protocol, and reviewed and provided feedback on the manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Conflict of interest The authors have no Conflict of interest as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akhtar MS, Ekbal A, Bhattacharyya P (2016) Aspect based sentiment analysis in Hindi: Resource creation and evaluation. In: Calzolari N, Choukri K, Declerck T, et al (eds) Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). European Language Resources Association (ELRA), Portorož, Slovenia, pp 2703–2709. <https://aclanthology.org/L16-1429>
- Akhtar MS, Gupta D, Ekbal A et al (2017) Feature selection and ensemble construction: a two-step method for aspect based sentiment analysis. *Knowl Based Syst* 125:116–135. <https://doi.org/10.1016/j.knosys.2017.03.020>. <https://linkinghub.elsevier.com/retrieve/pii/S095070511730148X>
- Akhtar MS, Ekbal A, Bhattacharyya P (2018) Aspect based sentiment analysis: category detection and sentiment classification for Hindi. In: Gelbukh A (ed) Computational linguistics and intelligent text processing, vol 9624. Lecture Notes in Computer Science. Springer, Cham, pp 246–257. https://doi.org/10.1007/978-3-319-75487-1_19
- Akhtar MS, Garg T, Ekbal A (2020) Multi-task learning for aspect term extraction and aspect sentiment classification. *Neurocomputing* 398:247–256. <https://doi.org/10.1016/j.neucom.2020.02.093>. <https://linkinghub.elsevier.com/retrieve/pii/S0925231220302897>
- Almatrafi O, Johri A (2022) Improving MOOCs using information from discussion forums: an opinion summarization and suggestion mining approach. *IEEE Access* 10:15565–15573. <https://doi.org/10.1109/ACCESS.2022.3149271>. <https://ieeexplore.ieee.org/document/9706374/>
- Alyami S, Alhothali A, Jamal A (2022) Systematic literature review of Arabic aspect-based sentiment analysis. *J King Saud Univ Comput Inf Sci* 34(9):6524–6551. <https://doi.org/10.1016/j.jksuci.2022.07.001>. <https://linkinghub.elsevier.com/retrieve/pii/S1319157822002282>
- Amin MM, Mao R, Cambria E et al (2024) A wide evaluation of chatgpt on affective computing tasks. *IEEE Trans Affect Comput* 1–9. <https://doi.org/10.1109/TAFFC.2024.3419593>. <https://ieeexplore.ieee.org/document/10572294>
- Asghar MZ, Khan A, Zahra SR et al (2019) Aspect-based opinion mining framework using heuristic patterns. *Cluster Comput* 22(S3):7181–7199. <https://doi.org/10.1007/s10586-017-1096-9>
- Baccianella S, Esuli A, Sebastiani F (2010) SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Calzolari N, Choukri K, Maegaard B et al (eds) Proceedings of the seventh international conference on language resources and evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, Malta. http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf
- Bommasani R, Hudson DA, Adeli E et al (2022) On the opportunities and risks of foundation models. [arXiv:2108.07258](https://arxiv.org/abs/2108.07258)
- Brauwers G, Frasincar F (2023) A survey on aspect-based sentiment classification. *ACM Comput Surv* 55(4):1–37. <https://doi.org/10.1145/3503044>
- Brown TB, Mann B, Ryder N et al (2020) Language models are few-shot learners. [arXiv:2005.14165](https://arxiv.org/abs/2005.14165)
- Cambria E, Poria S, Bajpai R et al (2016) SenticNet 4: a semantic resource for sentiment analysis based on conceptual primitives. In: Matsumoto Y, Prasad R (eds) Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers. The COLING 2016 Organizing Committee, Osaka, Japan, pp 2666–2677. <https://aclanthology.org/C16-1251>
- Castellanos M, Dayal U, Hsu M, et al (2011) LCI: A social channel analysis platform for live customer intelligence. In: Proceedings of the 2011 ACM SIGMOD international conference on management of data, SIGMOD'11. Association for Computing Machinery, New York, pp 1049–1058. <https://doi.org/10.1145/1989323.1989436>
- Cavalcanti D, Prudêncio R (2017) Aspect-based opinion mining in drug reviews. In: Oliveira E, Gama J, Vale Z et al (eds) Progress in artificial intelligence, vol 10423. Lecture Notes in Computer Science. Springer, Cham, pp 815–827. https://doi.org/10.1007/978-3-319-65340-2_66
- Chauhan GS, Agrawal P, Meena YK (2019) Aspect-based sentiment analysis of students' feedback to improve teaching-learning process. In: Satapathy SC, Joshi A (eds) Information and communication technology for intelligent systems, vol 107. Smart Innovation, Systems and Technologies. Springer Singapore, Singapore, pp 259–266. https://doi.org/10.1007/978-981-13-1747-7_25
- Chebolu SUS, Dernoncourt F, Lipka N et al (2023) Survey of aspect-based sentiment analysis datasets. [arXiv:2204.05232](https://arxiv.org/abs/2204.05232)
- Chen S, Fnu G (2022) Deep learning techniques for aspect based sentiment analysis. In: 2022 14th International conference on computer research and development (ICCRD), pp 69–73. <https://doi.org/10.1109/ICCRD54409.2022.9730443>. <https://ieeexplore.ieee.org/document/9730443>
- Chen Z, Liu B (2014) Topic modeling using topics from many domains, lifelong learning and big data. In: Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32. JMLR.org, ICML'14, p II–703–II–711. <https://proceedings.mlr.press/v32/chenf14.html>

- Chen Z, Qian T (2022) Retrieve-and-edit domain adaptation for end2end aspect based sentiment analysis. *IEEE/ACM Trans Audio Speech Lang Process* 30:659–672. <https://doi.org/10.1109/TASLP.2022.3146052>. <https://ieeexplore.ieee.org/document/9693267/>
- Chen M, Wu W, Zhang Y et al (2021) Combining adversarial training and relational graph attention network for aspect-based sentiment analysis with BERT. In: 2021 14th international congress on image and signal processing, BioMedical engineering and informatics (CISP-BMEI), pp 1–6. <https://doi.org/10.1109/CISP-BMEI53629.2021.9624384>. <https://ieeexplore.ieee.org/document/9624384>
- Chen F, Yang Z, Huang Y (2022) A multi-task learning framework for end-to-end aspect sentiment triplet extraction. *Neurocomputing* 479:12–21. <https://doi.org/10.1016/j.neucom.2022.01.021>. <https://linkinghub.elsevier.com/retrieve/pii/S0925231222000406>
- Chu JSG, Evans JA (2021) Slowed canonical progress in large fields of science. *Proc Natl Acad Sci* 118(41):e2021636118. <https://doi.org/10.1073/pnas.2021636118>. <https://pnas.org/doi/full/10.1073/pnas.2021636118>
- Cortis K, Davis B (2021) Over a decade of social opinion mining: a systematic review. *Artif Intell Rev* 54(7):4873–4965. <https://doi.org/10.1007/s10462-021-10030-2>
- Cruz I, Gelbukh AF, Sidorov G (2014) Implicit aspect indicator extraction for aspect based opinion mining. *Int J Comput Linguistics Appl* 5(2):135–152. <https://www.semanticscholar.org/paper/Implicit-Aspect-Indicator-Extraction-for-Aspect-Cruz-Gelbukh/8768fc3374b27c0ac023f5bf60da9ab50714b37e>
- Dang TV, Hao D, Nguyen N (2024) Vi-AbSQA: multi-task prompt instruction tuning model for Vietnamese aspect-based sentiment quadruple analysis. *ACM Trans Asian Low-Resour Lang Inf Process*. <https://doi.org/10.1145/3676886>, just Accepted
- Da'u A, Salim N, Rabiu I et al (2020) Weighted aspect-based opinion mining using deep learning for recommender system. *Expert Syst Appl* 140:112871. <https://doi.org/10.1016/j.eswa.2019.112871>. <https://linkinghub.elsevier.com/retrieve/pii/S0957417419305810>
- Devlin J, Chang M, Lee K et al (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: Burstein J, Doran C, Solorio T (eds) Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2–7, 2019, volume 1 (long and short papers). Association for Computational Linguistics, pp 4171–4186. <https://doi.org/10.18653/V1/N19-1423>
- Do HH, Prasad P, Maag A et al (2019) Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Syst Appl* 118:272–299. <https://doi.org/10.1016/j.eswa.2018.10.003>. <https://www.sciencedirect.com/science/article/pii/S0957417418306456>
- Dong L, Wei F, Tan C, et al (2014) Adaptive recursive neural network for target-dependent Twitter sentiment classification. In: Toutanova K, Wu H (eds) Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Association for Computational Linguistics, Baltimore, Maryland, pp 49–54. <https://doi.org/10.3115/v1/P14-2009>. <https://aclanthology.org/P14-2009>
- Dong Q, Li L, Dai D et al (2024) A survey on in-context learning. [arXiv:2301.00234](https://arxiv.org/abs/2301.00234)
- Dragoni M, Federici M, Rexha A (2019) An unsupervised aspect extraction strategy for monitoring real-time reviews stream. *Inf Process Manag* 56(3):1103–1118. <https://doi.org/10.1016/j.ipm.2018.04.010>. <https://linkinghub.elsevier.com/retrieve/pii/S0306457317305174>
- Du C, Wang J, Sun H et al (2021) Syntax-type-aware graph convolutional networks for natural language understanding. *Appl Soft Comput* 102:107080. <https://doi.org/10.1016/j.asoc.2021.107080>. <https://linkinghub.elsevier.com/retrieve/pii/S156849462100003X>
- Ettaleb M, Barhoumi A, Camelin N et al (2022) Evaluation of weakly-supervised methods for aspect extraction. *Proc Comput Sci* 207:2688–2697. <https://doi.org/10.1016/j.procs.2022.09.327>. <https://linkinghub.elsevier.com/retrieve/pii/S1877050922012169>
- Fan F, Feng Y, Zhao D (2018) Multi-grained attention network for aspect-level sentiment classification. In: Conference on empirical methods in natural language processing. <https://api.semanticscholar.org/CorpusID:53080156>
- Federici M, Dragoni M (2016) A knowledge-based approach for aspect-based opinion mining. In: Sack H, Dietze S, Tordai A et al (eds) Semantic web challenges, vol 641. Communications in Computer and Information Science. Springer, Cham, pp 141–152. https://doi.org/10.1007/978-3-319-46565-4_11
- Fei H, Chua TS, Li C et al (2023) On the robustness of aspect-based sentiment analysis: rethinking model, data, and training. *ACM Trans Inf Syst* 41(2):1–32. <https://doi.org/10.1145/3564281>. <https://dl.acm.org/doi/10.1145/3564281>

- Fei H, Ren Y, Zhang Y et al (2023) Nonautoregressive encoder-decoder neural framework for end-to-end aspect-based sentiment triplet extraction. *IEEE Trans Neural Netw Learn Syst* 34(9):5544–5556. <https://doi.org/10.1109/TNNLS.2021.3129483>. <https://ieeexplore.ieee.org/document/9634849/>
- Fernando J, Khodra ML, Septiandri AA (2019) Aspect and opinion terms extraction using double embeddings and attention mechanism for Indonesian hotel reviews. In: 2019 International conference of advanced informatics: concepts, theory and applications (ICAICTA). IEEE, Yogyakarta, pp 1–6. <https://doi.org/10.1109/ICACITA.2019.8904124>. <https://ieeexplore.ieee.org/document/8904124/>
- FiQA (2018) Financial opinion mining and question answering. <https://sites.google.com/view/fiqa/home>
- Freitas C, Motta E, Miliidi 'u RL, et al (2014) Sparkling vampire...lol! annotating opinions in a book review corpus. New language technologies and linguistic research: a two-way Road pp 128–146. https://www.researchgate.net/publication/271836545_Sparkling_Vampire_lol_Annotating_Opinions_in_a_Book_Review_Corpus
- Fukumoto F, Sugiyama H, Suzuki Y et al (2016) Exploiting guest preferences with aspect-based sentiment analysis for hotel recommendation. In: Fred A, Dietz JL, Aveiro D et al (eds) Knowledge discovery, knowledge engineering and knowledge management, vol 631. Communications in Computer and Information Science. Springer, Cham, pp 31–46. https://doi.org/10.1007/978-3-319-52758-1_3
- Ganganwar V, Rajalakshmi R (2019) Implicit aspect extraction for sentiment analysis: a survey of recent approaches. *Proc Comput Sci* 165:485–491. <https://doi.org/10.1016/j.procs.2020.01.010>. <https://www.sciencedirect.com/science/article/pii/S1877050920300181>, 2nd International Conference on Recent Trends in Advanced Computing ICRTAC-DISRUP - TIV INNOVATION, 2019 November 11–12, 2019
- Ganu G, Elhadad N, Marian A (2009) Beyond the stars: improving rating predictions using review text content. In: International workshop on the web and databases. <https://api.semanticscholar.org/CorpusID:18345070>
- García-Pablos A, Cuadros M, Rigau G (2018) W2vlda: Almost unsupervised system for aspect based sentiment analysis. *Expert Syst Appl* 91:127–137. <https://doi.org/10.1016/j.eswa.2017.08.049>. <https://linkinghub.elsevier.com/retrieve/pii/S0957417417305961>
- Go A (2009) Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1–12. <https://api.semanticscholar.org/CorpusID:18635269>
- Gojali S, Khodra ML (2016) Aspect based sentiment analysis for review rating prediction. In: 2016 International conference on advanced informatics: concepts, theory and application (ICAICTA). IEEE, Penang, pp 1–6. <https://doi.org/10.1109/ICACITA.2016.7803110>. <http://ieeexplore.ieee.org/document/7803110/>
- Gong Z, Li B (2022) Emotional text generation with hard constraints. In: 2022 4th International conference on frontiers technology of information and computer (ICFTIC), pp 68–73. <https://doi.org/10.1109/ICFTIC57696.2022.10075091>. <https://ieeexplore.ieee.org/document/10075091>
- Gui L, He Y (2021) Understanding patient reviews with minimum supervision. *Artif Intell Med* 120:102160. <https://doi.org/10.1016/j.artmed.2021.102160>. <https://www.sciencedirect.com/science/article/pii/S0933365721001536>
- Gunes O (2016) Aspect term and opinion target extraction from web product reviews using semi-Markov conditional random fields with word embeddings as features. In: Proceedings of the 6th international conference on web intelligence, mining and semantics. ACM, Nimes, pp 1–5. <https://doi.org/10.1145/2912845.2936809>
- Guo L, Jiang S, Du W et al (2018) Recurrent neural CRF for aspect term extraction with dependency transmission. In: Zhang M, Ng V, Zhao D et al (eds) Natural language processing and Chinese computing, vol 11108. Lecture Notes in Computer Science. Springer, Cham, pp 378–390. https://doi.org/10.1007/978-3-319-99495-6_32
- He R, McAuley J (2016) Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In: Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, WWW '16, p 507–517. <https://doi.org/10.1145/2872427.2883037>
- He R, Lee WS, Ng HT et al (2019) An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In: Proceedings of the 57th annual meeting of the association for computational linguistics. Association for Computational Linguistics, Florence, pp 504–515. <https://doi.org/10.18653/v1/P19-1048>. <https://www.aclweb.org/anthology/P19-1048>
- Hoang CD, Dinh QV, Tran NH (2022) Aspect-category-opinion-sentiment extraction using generative transformer model. In: 2022 RIVF international conference on computing and communication technologies (RIVF), pp 1–6. <https://doi.org/10.1109/RIVF55975.2022.10013820>. <https://ieeexplore.ieee.org/document/10013820>
- Hoti MH, Ajdari J, Hamiti M et al (2022) Text mining, clustering and sentiment analysis: a systematic literature review. In: 2022 11th Mediterranean conference on embedded computing (MECO).

- IEEE, Budva, Montenegro, pp 1–6. <https://doi.org/10.1109/MECO55406.2022.9797203>. <https://ieeexplore.ieee.org/document/9797203/>
- Howard P, Ma A, Lal V et al (2022) Cross-domain aspect extraction using transformers augmented with knowledge graphs. In: Proceedings of the 31st ACM international conference on information & knowledge management. ACM, Atlanta, pp 780–790. <https://doi.org/10.1145/3511808.3557275>
- Huan H, He Z, Xie Y et al (2022) A multi-task dual-encoder framework for aspect sentiment triplet extraction. IEEE Access 10:103187–103199. <https://doi.org/10.1109/ACCESS.2022.3210180>. <https://ieeexplore.ieee.org/document/9903619/>
- Hu M, Liu B (2004a) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, Seattle, pp 168–177. <https://doi.org/10.1145/1014052.1014073>
- Hu M, Liu B (2004b) Mining opinion features in customer reviews. In: AAAI conference on artificial intelligence. <https://api.semanticscholar.org/CorpusID:5724860>
- J AK, Trueman TE, Cambria E (2021) A convolutional stacked bidirectional LSTM with a multiplicative attention mechanism for aspect category and sentiment detection. Cogn Comput 13(6):1423–1432. <https://doi.org/10.1007/s12559-021-09948-0>
- Jiang Q, Chen L, Xu R et al (2019) A challenge dataset and effective models for aspect-based sentiment analysis. In: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, pp 6279–6284. <https://doi.org/10.18653/v1/D19-1654>. <https://www.aclweb.org/anthology/D19-1654>
- Jo Y, Oh AH (2011) Aspect and sentiment unification model for online review analysis. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. Association for Computing Machinery, New York, NY, USA, WSDM '11, p 815–824. <https://doi.org/10.1145/1935826.1935932>
- Kang T, Kim S, Yun H et al (2022) Gated relational encoder-decoder model for target-oriented opinion word extraction. IEEE Access 10:130507–130517. <https://doi.org/10.1109/ACCESS.2022.3228835>. <https://ieeexplore.ieee.org/document/9982601>
- Kitchenham BA, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering. https://www.elsevier.com/_data/promis_misc/525444systemicreviewsguide.pdf. Backup Publisher: Keele University and Durham University Joint Report
- Krishnakumari K, Sivasankar E (2018) Scalable aspect-based summarization in the hadoop environment. In: Aggarwal VB, Bhatnagar V, Mishra DK (eds) Big data analytics, vol 654. Advances in Intelligent Systems and Computing. Springer Singapore, Singapore, pp 439–449. https://doi.org/10.1007/978-981-10-6620-7_42
- Kumar A, Gupta D (2021) Sentiment analysis as a restricted NLP problem:. In: Pinarbasi F, Taskiran MN (eds) Advances in business information systems and analytics. IGI Global, pp 65–96. <https://doi.org/10.4018/978-1-7998-4240-8.ch004>. <http://services.igi-global.com/resolveddoi/resolve.aspx?doi=10.4018/978-1-7998-4240-8.ch004>
- Kumar A, Seth S, Gupta S et al (2022) Sentic computing for aspect-based opinion summarization using multi-head attention with feature pooled pointer generator network. Cogn Comput 14(1):130–148. <https://doi.org/10.1007/s12559-021-09835-8>
- Lee SK, Kim JH (2023) Sener: Sentiment element named entity recognition for aspect-based sentiment analysis. In: ICASSP 2023—2023 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 1–5. <https://doi.org/10.1109/ICASSP49357.2023.10095101>. <https://ieeexplore.ieee.org/document/10095101>
- Lee C, Lee H, Kim K et al (2024) An efficient fine-tuning of generative language model for aspect-based sentiment analysis. In: 2024 IEEE international conference on consumer electronics (ICCE), pp 1–4. <https://doi.org/10.1109/ICCE59016.2024.10444216>. <https://ieeexplore.ieee.org/document/10444216>
- Lewis M, Liu Y, Goyal N et al (2020) BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Jurafsky D, Chai J, Schluter N et al (eds) Proceedings of the 58th annual meeting of the association for computational linguistics. Association for Computational Linguistics, Online, pp 7871–7880. <https://doi.org/10.18653/v1/2020.acl-main.703>. <https://aclanthology.org/2020.acl-main.703>
- Li X, Wang B, Li L et al (2020) Deep2s: Improving aspect extraction in opinion mining with deep semantic representation. IEEE Access 8:104026–104038. <https://doi.org/10.1109/ACCESS.2020.2999673>. <https://ieeexplore.ieee.org/document/9107147/>
- Li Z, Li L, Zhou A et al (2021) JTSG: A joint term-sentiment generator for aspect-based sentiment analysis. Neurocomputing 459:1–9. <https://doi.org/10.1016/j.neucom.2021.06.045>. <https://www.sciencedirect.com/science/article/pii/S0925231221009693>

- Li J, Zhao Y, Jin Z et al (2022a) SK2: Integrating implicit sentiment knowledge and explicit syntax knowledge for aspect-based sentiment analysis. In: Proceedings of the 31st ACM international conference on information & knowledge management. ACM, Atlanta, pp 1114–1123. <https://doi.org/10.1145/3511808.3557452>
- Li Y, Lin Y, Lin Y et al (2022) A span-sharing joint extraction framework for harvesting aspect sentiment triplets. *Knowl Based Syst* 242:108366. <https://doi.org/10.1016/j.knosys.2022.108366>. <https://linkinghub.elsevier.com/retrieve/pii/S0950705122001381>
- Li Y, Wang C, Lin Y et al (2022) Span-based relational graph transformer network for aspect-opinion pair extraction. *Knowl Inf Syst* 64(5):1305–1322. <https://doi.org/10.1007/s10115-022-01675-8>
- Li S, Zhang Y, Lan Y et al (2023) From implicit to explicit: a simple generative method for aspect-category-opinion-sentiment quadruple extraction. In: 2023 international joint conference on neural networks (IJCNN), pp 1–8. <https://doi.org/10.1109/IJCNN5450.2023.10191098>. <https://ieeexplore.ieee.org/document/10191098>
- Liang B, Su H, Gui L et al (2022) Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowl Based Syst* 235:107643. <https://doi.org/10.1016/j.knosys.2021.107643>. <https://linkinghub.elsevier.com/retrieve/pii/S0950705121009059>
- Lighthart A, Catal C, Tekinerdogan B (2021) Systematic reviews in sentiment analysis: a tertiary study. *Artif Intell Rev* 54(7):4997–5053. <https://doi.org/10.1007/s10462-021-09973-3>
- Lil Z, Yang Z, Li X et al (2023) Two-stage aspect sentiment quadruple prediction based on MRC and text generation. In: 2023 IEEE International conference on systems, man, and cybernetics (SMC), pp 2118–2125. <https://doi.org/10.1109/SMC53992.2023.10394369>. <https://ieeexplore.ieee.org.ezproxy.auckland.ac.nz/document/10394369>
- Lim KW, Buntine W (2014) Twitter opinion topic model: extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. In: Proceedings of the 23rd ACM international conference on conference on information and knowledge management. ACM, Shanghai, pp 1319–1328. <https://doi.org/10.1145/2661829.2662005>
- Lin B, Cassee N, Serebrenik A et al (2022) Opinion mining for software development: a systematic literature review. *ACM Trans Softw Eng Methodol* 31(3):1–41. <https://doi.org/10.1145/3490388>
- Liu Q, Gao Z, Liu B, et al (2015) Automated rule selection for aspect extraction in opinion mining. In: Proceedings of the 24th International Conference on Artificial Intelligence. AAAI Press, IJCAI' 15, p 1291–1297. <https://doi.org/10.5555/2832415.2832429>
- Liu Y, Ott M, Goyal N et al (2019) Roberta: a robustly optimized BERT pretraining approach. [arXiv:1907.11692](https://arxiv.org/abs/1907.11692)
- Liu H, Chatterjee I, Zhou M et al (2020) Aspect-based sentiment analysis: a survey of deep learning methods. *IEEE Trans Comput Soc Syst* 7(6):1358–1375. <https://doi.org/10.1109/TCSS.2020.3033302>. <https://ieeexplore.ieee.org/document/9260162/>
- Liu J, Chen T, Guo H et al (2024) Exploiting duality in aspect sentiment triplet extraction with sequential prompting. *IEEE Trans Knowl Data Eng* 1–12. <https://doi.org/10.1109/TKDE.2024.3391381>. <https://ieeexplore.ieee.org/document/10505831>
- López D, Arco L (2019) Multi-domain aspect extraction based on deep and lifelong learning. In: Nyström I, Hernández Heredia Y, Milián Núñez V (eds) Progress in pattern recognition, image analysis, computer vision, and applications, vol 11896. Lecture Notes in Computer Science. Springer, Cham, pp 556–565. https://doi.org/10.1007/978-3-030-33904-3_52
- Luo H, Li T, Liu B et al (2019) Improving aspect term extraction with bidirectional dependency tree representation. *IEEE/ACM Trans Audio Speech Lang Process* 27(7):1201–1212. <https://doi.org/10.1109/TASLP.2019.2913094>. <https://ieeexplore.ieee.org/document/8698340/>
- Ma Y, Chen G, Wei Q (2017) Finding users preferences from large-scale online reviews for personalized recommendation. *Electron Commer Res* 17(1):3–29. <https://doi.org/10.1007/s10660-016-9240-9>
- Maitama JZ, Idris N, Abdi A et al (2020) A systematic review on implicit and explicit aspect extraction in sentiment analysis. *IEEE Access* 8:194166–194191. <https://doi.org/10.1109/ACCESS.2020.3031217>. <https://ieeexplore.ieee.org/document/9234464/>
- Manning CD (2022) Human language understanding & reasoning. *Daedalus* 151(2):127–138. https://doi.org/10.1162/daed_a_01905. <https://direct.mit.edu/daed/article/151/2/127/110621/Human-Language-Understanding-and-RReasoning>
- Marstawi A, Sharef NM, Aris TNM, et al (2017) Ontology-based aspect extraction for an improved sentiment analysis in summarization of product reviews. In: Proceedings of the 8th international conference on computer modeling and simulation, ICCMS'17. Association for Computing Machinery, New York, pp 100–104. <https://doi.org/10.1145/3036331.3036362>

- McAuley J, Leskovec J, Jurafsky D (2012) Learning attitudes and attributes from multi-aspect reviews. In: Proceedings of the 2012 IEEE 12th International Conference on Data Mining. IEEE Computer Society, USA, ICDM '12, p 1020–1025. <https://doi.org/10.1109/ICDM.2012.110>
- McAuley J, Targett C, Shi Q, Van Den Hengel A (2015) Image-based recommendations on styles and substitutes. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval Association for Computing Machinery, New York, NY, USA, SIGIR'15, p 43–52. <https://doi.org/10.1145/2766462.2767755>
- Mitchell M, Aguilar J, Wilson T, et al (2013) Open domain targeted sentiment. In: Yarowsky D, Baldwin T, Korhonen A, et al (eds) Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Seattle, Washington, USA, pp 1643–1654. <https://aclanthology.org/D13-1171>
- Mughal N, Mujtaba G, Shaikh S et al (2024) Comparative analysis of deep natural networks and large language models for aspect-based sentiment analysis. IEEE Access 12:60943–60959. <https://doi.org/10.1109/ACCESS.2024.3386969>. <https://ieeexplore.ieee.org/document/10504711>
- Nawaz A, Awan AA, Ali T et al (2020) Product's behaviour recommendations using free text: an aspect based sentiment analysis approach. Clust Comput 23(2):1267–1279. <https://doi.org/10.1007/s10586-019-02995-1>
- Nazir A, Rao Y (2022) IAOTP: An interactive end-to-end solution for aspect-opinion term pairs extraction. In: Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval. ACM, Madrid, pp 1588–1598. <https://doi.org/10.1145/3477495.3532085>
- Nazir A, Rao Y, Wu L et al (2022) IAF-LG: an interactive attention fusion network with local and global perspective for aspect-based sentiment analysis. IEEE Trans Affect Comput 13(4):1730–1742. <https://doi.org/10.1109/TAFFC.2022.3208216>. <https://ieeexplore.ieee.org/document/9896931/>
- Nazir A, Rao Y, Wu L et al (2022) Issues and challenges of aspect-based sentiment analysis: a comprehensive survey. IEEE Trans Affect Comput 13(2):845–863. <https://doi.org/10.1109/TAFFC.2020.2970399>. <https://ieeexplore.ieee.org/document/8976252/>
- Obiedat R, Al-Darras D, Alzaghouli E et al (2021) Arabic aspect-based sentiment analysis: a systematic literature review. IEEE Access 9:152628–152645. <https://doi.org/10.1109/ACCESS.2021.3127140>. <https://ieeexplore.ieee.org/document/9611271/>
- OpenAI (2023) Chatgpt (mar 14 version) [large language model]. <https://chat.openai.com/chat>
- Pathan AF, Prakash C (2022) Cross-domain aspect detection and categorization using machine learning for aspect-based opinion mining. Int J Inf Manag Data Insights 2(2):100099. <https://doi.org/10.1016/j.jjimei.2022.100099>. <https://www.sciencedirect.com/science/article/pii/S2667096822000428>
- Peng H, Ma Y, Li Y, et al (2018) Learning multi-grained aspect target sequence for chinese sentiment analysis. Knowledge-Based Syst 148:167–176. <https://doi.org/10.1016/j.knosys.2018.02.034>. <https://www.sciencedirect.com/science/article/pii/S0950705118300972>
- Phan MH, Ogunbona PO (2020) Modelling context and syntactical features for aspect-based sentiment analysis. In: Proceedings of the 58th annual meeting of the association for computational linguistics. Association for Computational Linguistics, Online, pp 3211–3220. <https://doi.org/10.18653/v1/2020.acl-main.293>. <https://www.aclweb.org/anthology/2020.acl-main.293>
- Pontiki M, Galanis D, Pavlopoulos J et al (2014) SemEval-2014 task 4: aspect based sentiment analysis. In: Nakov P, Zesch T (eds) Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). Association for Computational Linguistics, Dublin, pp 27–35. <https://doi.org/10.3115/v1/S14-2004>. <https://aclanthology.org/S14-2004>
- Pontiki M, Galanis D, Papageorgiou H et al (2015) SemEval-2015 task 12: aspect based sentiment analysis. In: Nakov P, Zesch T, Cer D et al (eds) Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015). Association for Computational Linguistics, Denver, pp 486–495. <https://doi.org/10.18653/v1/S15-2082>. <https://aclanthology.org/S15-2082>
- Pontiki M, Galanis D, Papageorgiou H et al (2016) SemEval-2016 task 5: aspect based sentiment analysis. In: Bethard S, Carpuat M, Cer D et al (eds) Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016). Association for Computational Linguistics, San Diego, pp 19–30. <https://doi.org/10.18653/v1/S16-1002>. <https://aclanthology.org/S16-1002>
- Poria S, Chaturvedi I, Cambria E et al (2016) Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. In: 2016 International joint conference on neural networks (IJCNN). IEEE, Vancouver, pp 4465–4473. <https://doi.org/10.1109/IJCNN.2016.7727784>. <https://ieeexplore.ieee.org/document/7727784/>
- Prather J, Becker BA, Craig M et al (2020) What do we think we think we are doing?: Metacognition and self-regulation in programming. In: Proceedings of the 2020 ACM conference on international

- computing education research. ACM, Virtual Event New Zealand, pp 2–13. <https://doi.org/10.1145/3372782.3406263>
- Presannakumar K, Mohamed A (2021) An enhanced method for review mining using n-gram approaches. In: Raj JS, Ilyasu AM, Bestak R et al (eds) Innovative data communication technologies and application, vol 59. Lecture Notes on Data Engineering and Communications Technologies. Springer Singapore, Singapore, pp 615–626. https://doi.org/10.1007/978-981-15-9651-3_51
- Radford A, Wu J, Child R et al (2019) Language models are unsupervised multitask learners. <https://api.semanticscholar.org/CorpusID:160025533>
- Raffel C, Shazeer N, Roberts A et al (2020) Exploring the limits of transfer learning with a unified text-to-text transformer. *J Mach Learn Res* 21(1). <https://dl.acm.org/doi/abs/10.5555/3455716.3455856>
- Rahman MA, Kumar Dey E (2018) Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation. *Data* 3(2). <https://doi.org/10.3390/data3020015>. <https://www.mdpi.com/2306-5729/3/2/15>
- Rana TA, Cheah YN (2016) Aspect extraction in sentiment analysis: comparative analysis and survey. *Artif Intell Rev* 46:459–483. <https://api.semanticscholar.org/CorpusID:24401592>
- Rani S, Kumar P (2019) A journey of Indian languages over sentiment analysis: a systematic review. *Artif Intell Rev* 52(2):1415–1462. <https://doi.org/10.1007/s10462-018-9670-y>
- Ruskanda FZ, Widiantoro DH, Purwarianti A (2019) Sequential covering rule learning for language rule-based aspect extraction. In: 2019 International conference on advanced computer science and information systems (ICACSIS). IEEE, Bali, pp 229–234. <https://doi.org/10.1109/ICACSIS47736.2019.8979743>. <https://ieeexplore.ieee.org/document/8979743>
- Sabeeh A, Dewang RK (2019) Comparison, classification and survey of aspect based sentiment analysis. In: Luhach AK, Singh D, Hsiung PA et al (eds) Advanced informatics for computing research. Springer Singapore, Singapore, pp 612–629. https://doi.org/10.1007/978-981-13-3140-4_55
- Saeidi M, Bouchard G, Liakata M, et al (2016) SentiHood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods. In: Matsumoto Y, Prasad R (eds) Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. The COLING 2016 Organizing Committee, Osaka, Japan, pp 1546–1556. <https://aclanthology.org/C16-1146>
- Sanders NJ (2011) Sanders-twitter sentiment corpus. Sanders Analytics LLC
- Satyarthi S, Sharma S (2023) Identification of effective deep learning approaches for classifying sentiments at aspect level in different domain. In: 2023 IEEE International conference on paradigm shift in information technologies with innovative applications in global scenario (ICPSITIAGS), pp 496–508. <https://doi.org/10.1109/ICPSITIAGS59213.2023.10527695>. <https://ieeexplore-ieee-org.ezproxy.auckland.ac.nz/document/10527695>
- Sharma A, Shekhar H (2020) Intelligent learning based opinion mining model for governmental decision making. *Proc Comput Sci* 173:216–224. <https://doi.org/10.1016/j.procs.2020.06.026>. <https://linkinghub.elsevier.com/retrieve/pii/S1877050920315301>
- Socher R, Perelygin A, Wu J, et al (2013) Recursive deep models for semantic compositionality over a sentiment treebank. In: Yarowsky D, Baldwin T, Korhonen A, et al (eds) Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Seattle, Washington, USA, pp 1631–1642. <https://aclanthology.org/D13-1170>
- Soni PK, Rambola R (2022) A survey on implicit aspect detection for sentiment analysis: terminology, issues, and scope. *IEEE Access* 10:63932–63957. <https://doi.org/10.1109/ACCESS.2022.3183205>. <https://ieeexplore.ieee.org/document/9796523>
- Suchrady RZ, Purwarianti A (2023) Indo LEGO-ABSA: a multitask generative aspect based sentiment analysis for Indonesian language. In: 2023 International conference on electrical engineering and informatics (ICEEI), pp 1–6. <https://doi.org/10.1109/ICEEI59426.2023.10346852>. <https://ieeexplore-ieee-org.ezproxy.auckland.ac.nz/document/10346852>
- Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In: Proceedings of the 27th international conference on neural information processing systems, NIPS'14, vol 2. MIT Press, Montreal, pp 3104–3112
- Su H, Wang X, Li J et al (2024) Enhanced implicit sentiment understanding with prototype learning and demonstration for aspect-based sentiment analysis. *IEEE Trans Comput Soc Syst* 1–16. <https://doi.org/10.1109/TCSS.2024.3368171>. <https://ieeexplore-ieee-org.ezproxy.auckland.ac.nz/document/10584152>
- Team TPD (2023) pandas-dev/pandas: Pandas. <https://doi.org/10.5281/ZENODO.3509134>. <https://zenodo.org/record/3509134>
- Toprak C, Jakob N, Gurevych I (2010) Sentence and expression level annotation of opinions in user-generated discourse. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, USA, ACL '10, p 575–584. <https://doi.org/10.5555/1858681.1858740>

- Tran TU, Hoang HTT, Huynh HX (2020) Bidirectional independently long short-term memory and conditional random field integrated model for aspect extraction in sentiment analysis. In: Satapathy SC, Bhateja V, Nguyen BL et al (eds) Frontiers in intelligent computing: theory and applications, vol 1014. Advances in Intelligent Systems and Computing. Springer Singapore, Singapore, pp 131–140. https://doi.org/10.1007/978-981-13-9920-6_14
- Tubishat M, Idris N, Abushariah M (2021) Explicit aspects extraction in sentiment analysis using optimal rules combination. *Futur Gener Comput Syst* 114:448–480. <https://doi.org/10.1016/j.future.2020.08.019>. <https://linkinghub.elsevier.com/retrieve/pii/S0167739X1933081X>
- Vasantha A, Kumar H, Karanraj R (2022) An RL approach for ABSA using transformers. In: 2022 6th International conference on trends in electronics and informatics (ICOEI), pp 354–361. <https://doi.org/10.1109/ICOEI53556.2022.9776915>. <https://ieeexplore.ieee.org/document/9776915>
- Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need. In: Proceedings of the 31st international conference on neural information processing systems, NIPS'17. Curran Associates Inc., Red Hook, pp 6000–6010. <https://doi.org/10.5555/3295222.3295349>
- Wang H, Lu Y, Zhai C (2010) Latent aspect rating analysis on review text data: a rating regression approach. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, KDD '10, p 783–792. <https://doi.org/10.1145/1835804.1835903>
- Wang H, Lu Y, Zhai C (2011) Latent aspect rating analysis without aspect keyword supervision. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, KDD '11, p 618–626. <https://doi.org/10.1145/2020408.2020505>
- Wang Y, Huang Y, Wang M (2017) Aspect-based rating prediction on reviews using sentiment strength analysis. In: Benferhat S, Tabia K, Ali M (eds) Advances in artificial intelligence: from theory to practice, vol 10351. Lecture Notes in Computer Science. Springer, Cham, pp 439–447. https://doi.org/10.1007/978-3-319-60045-1_45
- Wang W, Pan SJ, Dahlmeier D (2018) Memory networks for fine-grained opinion mining. *Artif Intell* 265:1–17. <https://doi.org/10.1016/j.artint.2018.09.002>. <https://linkinghub.elsevier.com/retrieve/pii/S000437021830599X>
- Wang J, Xu B, Zu Y (2021a) Deep learning for aspect-based sentiment analysis. In: 2021 International conference on machine learning and intelligent systems engineering (MLISE), pp 267–271. <https://doi.org/10.1109/MLISE54096.2021.00056>. <https://ieeexplore.ieee.org/document/9611705>
- Wang L, Zong B, Liu Y et al (2021b) Aspect-based sentiment classification via reinforcement learning. In: 2021 IEEE international conference on data mining (ICDM), pp 1391–1396. <https://doi.org/10.1109/ICDM51629.2021.00177>. <https://ieeexplore.ieee.org/document/9679112>
- Wang X, Liu P, Zhu Z et al (2022) Interactive double graph convolutional networks for aspect-based sentiment analysis. In: 2022 International joint conference on neural networks (IJCNN). IEEE, Padua, Italy, pp 1–7. <https://doi.org/10.1109/IJCNN55064.2022.9892934>. <https://ieeexplore.ieee.org/document/9892934/>
- Wang Z, Xia R, Yu J (2024) Unified ABSA via annotation-decoupled multi-task instruction tuning. *IEEE Trans Knowl Data Eng* 1–13. <https://doi.org/10.1109/TKDE.2024.3392836>. <https://ieeexplore.ieee.org/document/10507027>
- Wankhade M, Rao ACS, Kulkarni C (2022) A survey on sentiment analysis methods, applications, and challenges. *Artif Intel Rev* 55(7):5731–5780. <https://doi.org/10.1007/s10462-022-10144-1>
- Wikipedia (2023) SemEval. <https://en.wikipedia.org/wiki/SemEval>
- William, Khodra ML (2022) Generative opinion triplet extraction using pretrained language model. In: 2022 9th International conference on advanced informatics: concepts, theory and applications (ICAICTA), pp 1–6. <https://doi.org/10.1109/ICAICTA56449.2022.9933004>. <https://ieeexplore.ieee.org/document/9933004>
- Wu S, Fei H, Ren Y et al (2021) High-order pair-wise aspect and opinion terms extraction with edge-enhanced syntactic graph convolution. *IEEE/ACM Trans Audio Speech Lang Process* 29:2396–2406. <https://doi.org/10.1109/TASLP.2021.3095672>. <https://ieeexplore.ieee.org/document/9478183/>
- Xing X, Jin Z, Jin D et al (2020) Tasty burgers, soggy fries: probing aspect robustness in aspect-based sentiment analysis. In: Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP). Association for Computational Linguistics, Online, pp 3594–3605. <https://doi.org/10.18653/v1/2020.emnlp-main.292>. <https://www.aclweb.org/anthology/2020.emnlp-main.292>
- Xu K, Zhao H, Liu T (2020) Aspect-specific heterogeneous graph convolutional network for aspect-based sentiment classification. *IEEE Access* 8:139346–139355. <https://doi.org/10.1109/ACCESS.2020.3012637>. <https://ieeexplore.ieee.org/document/9152016/>

- Xu Q, Zhu L, Dai T et al (2020) Non-negative matrix factorization for implicit aspect identification. *J Ambient Intell Humaniz Comput* 11(7):2683–2699. <https://doi.org/10.1007/s12652-019-01328-9>
- Yan K, Tang L, Wu M et al (2023) Aspect-based sentiment analysis method using text generation. In: Proceedings of the 2023 7th international conference on big data and internet of things, BDIOIT'23. Association for Computing Machinery, New York, pp 156–161. <https://doi.org/10.1145/3617695.3617709>
- Yauris K, Khodra ML (2017) Aspect-based summarization for game review using double propagation. In: 2017 International conference on advanced informatics, concepts, theory, and applications (ICAICTA). IEEE, Denpasar, pp 1–6. <https://doi.org/10.1109/ICAICTA.2017.8090997>. <http://ieeexplore.ieee.org/document/8090997/>
- You L, Han F, Peng J et al (2022) ASK-RoBERTa: a pretraining model for aspect-based sentiment classification via sentiment knowledge mining. *Knowl Based Syst* 253:109511. <https://doi.org/10.1016/j.knosys.2022.109511>. <https://linkinghub.elsevier.com/retrieve/pii/S0950705122007584>
- Yu C, Wu T, Li J et al (2023a) Syngen: A syntactic plug-and-play module for generative aspect-based sentiment analysis. In: ICASSP 2023–2023 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 1–5. <https://doi.org/10.1109/ICASSP49357.2023.10094591>. <https://ieeexplore.ieee.org/document/10094591>
- Yu Y, Zhao M, Zhou S (2023b) Boosting aspect sentiment quad prediction by data augmentation and self-training. In: 2023 International joint conference on neural networks (IJCNN), pp 1–8. <https://doi.org/10.1109/IJCNN54540.2023.10191634>. <https://ieeexplore.ieee.org/document/10191634>
- Zarindast A, Sharma A, Wood J (2021) Application of text mining in smart lighting literature—an analysis of existing literature and a research agenda. *Int J Inf Manag Data Insights* 1(2):100032. <https://doi.org/10.1016/j.jjimei.2021.100032>. <https://linkinghub.elsevier.com/retrieve/pii/S2667096821000252>
- Zhang Y, Xu B, Zhao T (2020) Convolutional multi-head self-attention on memory for aspect sentiment classification. *IEEE/CAA J Automatica Sinica* 7(4):1038–1044. <https://doi.org/10.1109/JAS.2020.1003243>. <https://ieeexplore.ieee.org/document/9128078/>
- Zhang W, Deng Y, Li X et al (2021a) Aspect sentiment quad prediction as paraphrase generation. In: Moens MF, Huang X, Specia L et al (eds) Proceedings of the 2021 conference on empirical methods in natural language processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, pp 9209–9219. <https://doi.org/10.18653/v1/2021.emnlp-main.726>. <https://aclanthology.org/2021.emnlp-main.726>
- Zhang W, Li X, Deng Y et al (2021b) Towards generative aspect-based sentiment analysis. In: Zong C, Xia F, Li W et al (eds) Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 2: short papers). Association for Computational Linguistics, Online, pp 504–510. <https://doi.org/10.18653/v1/2021.acl-short.64>. <https://aclanthology.org/2021.acl-short.64>
- Zhang H, Chen Z, Chen B et al (2022) Complete quadruple extraction using a two-stage neural model for aspect-based sentiment analysis. *Neurocomputing* 492:452–463. <https://doi.org/10.1016/j.neucom.2022.04.027>. <https://www.sciencedirect.com/science/article/pii/S0925231222003939>
- Zhang W, Li X, Deng Y et al (2022b) A survey on aspect-based sentiment analysis: tasks, methods, and challenges. [arXiv:2203.01054](https://arxiv.org/abs/2203.01054)
- Zhang W, Li X, Deng Y et al (2022) A survey on aspect-based sentiment analysis: tasks, methods, and challenges. *IEEE Trans on Knowl and Data Eng* 35(11):11019–11038. <https://doi.org/10.1109/TKDE.2022.3230975>
- Zhang X, Xu J, Cai Y et al (2023) Detecting dependency-related sentiment features for aspect-level sentiment classification. *IEEE Trans Affect Comput* 14(1):196–210. <https://doi.org/10.1109/TAFFC.2021.3063259>. <https://ieeexplore.ieee.org/document/9368987/>
- Zhang W, Zhang X, Cui S, et al (2024a) Adaptive data augmentation for aspect sentiment quad prediction. In: ICASSP 2024–2024 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 11176–11180. <https://doi.org/10.1109/ICASSP48485.2024.10447700>
- Zhang W, Zhang X, Cui S et al (2024b) Adaptive data augmentation for aspect sentiment quad prediction. In: ICASSP 2024–2024 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 11176–11180. <https://doi.org/10.1109/ICASSP48485.2024.10447700>. <https://ieeexplore.ieee.org.ezproxy.auckland.ac.nz/document/10447700>
- Zhao H, Yang M, Bai X et al (2024) A survey on multimodal aspect-based sentiment analysis. *IEEE Access* 12:12039–12052. <https://doi.org/10.1109/ACCESS.2024.3354844>. <https://ieeexplore.ieee.org/document/10401113>
- Zhou J, Huang JX, Chen Q et al (2019) Deep learning for aspect-level sentiment classification: survey, vision, and challenges. *IEEE Access* 7:78454–78483. <https://doi.org/10.1109/ACCESS.2019.2920075>. <https://ieeexplore.ieee.org/document/8726353>

Zhou C, Wu Z, Song D et al (2024) Span-pair interaction and tagging for dialogue-level aspect-based sentiment quadruple analysis. In: Proceedings of the ACM on web conference 2024, WWW'24. Association for Computing Machinery, New York, pp 3995–4005. <https://doi.org/10.1145/3589334.3645355>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.