A comprehensive survey of contemporary Arabic sentiment analysis: Methods, Challenges, and Future Directions

Zhiqiang Shi
University of Edinburgh
shizhiqiang126@163.com

Ruchit Agrawal
University of Birmingham
r.r.agrawal@bham.ac.uk

Abstract

Sentiment Analysis, a popular subtask of Natural Language Processing, employs computational methods to extract sentiment, opinions, and other subjective aspects from linguistic data. Given its crucial role in understanding human sentiment, research in sentiment analysis has witnessed significant growth in the recent years. However, the majority of approaches are aimed at the English language, and research towards Arabic sentiment analysis remains relatively unexplored. This paper presents a comprehensive and contemporary survey of Arabic Sentiment Analysis, identifies the challenges and limitations of existing literature in this field and presents avenues for future research. We present a systematic review of Arabic sentiment analysis methods, focusing specifically on research utilizing deep learning. We then situate Arabic Sentiment Analysis within the broader context, highlighting research gaps in Arabic sentiment analysis as compared to general sentiment analysis. Finally, we outline the main challenges and promising future directions for research in Arabic sentiment analysis.

1 Introduction

Sentiment Analysis (SA), also referred to as opinion mining, leverages computational models to extract individuals' sentiments and opinions from data (Liu, 2015). This field has garnered significant attention from both academic and industrial sectors, as evidenced by the multitude of studies conducted to comprehend human sentiment (Liu, 2015), (Zeng and Li, 2022), (Varathan et al., 2017). However, although Arabic is a widely popular language spoken by over 372 million people across the globe, the volume of research dedicated to Arabic Sentiment Analysis (ASA) remains considerably lower compared to high-resourced languages such as English and French. This study presents a systematic review of existing literature on Arabic sentiment analysis, with a particular focus on approaches that employ deep learning methodologies.

Several prior surveys on Arabic sentiment analysis (ASA) exist (Almurqren et al., 2024), (Al-Ayyoub et al., 2019), (Al-Twairesh et al., 2014), (Biltawi et al., 2016), (Abdelhameed and Muñoz-Hern'andez, 2017), (Abo et al., 2019), (Rahma et al., 2023). However, the majority of these do not cover contemporary deep learning based methods. Additionally, these do not present a detailed analysis of the gaps, challenges and future directions for ASA. This paper fills this gap and presents a comprehensive survey of contemporary methods for Arabic sentiment analysis, systematically organizing recent literature in the field and highlighting the key contributions and limitations of current SA methods. We also situate these approaches within the broader framework of general sentiment analysis and approaches for high-resource languages, facilitating an understanding of the developments as well as the gaps in ASA. The major contributions of this study are summarised below:

- We present a systematic survey of contemporary research conducted in Arabic sentiment analysis, with a focus on deep learning methodologies. We present an analysis of the key contributions and limitations of state-of-the-art ASA methods across various dimensions such as modality (uni-modal, multi-modal), granularity (coarse-grained, fine-grained) and context (sentence-level, document-level, aspect-level).

- We situate Arabic sentiment analysis within the broader framework of general sentiment analysis, identifying research gaps in Arabic sentiment analysis, and highlighting areas where advancements are needed to bridge the gap with high-resource languages.

- We outline the key challenges in developing robust sentiment analysis models for the Arabic language, and present promising directions to guide future research in this field.

2 The Evolution of Arabic Sentiment Analysis

In this section, we describe the evolution of Arabic sentiment analysis, from lexicon based methods to deep learning based methods. To highlight the importance of these traditional methods, we conduct a case study using Arabic lexicons, in which we highlight how lexicons can be utilised to improve deep learning based methods.

2.1 Lexicon Based Methods

Lexicon based methods utilise a pre-defined lexicon to determine the sentiment of the given text. The words in the lexicon are annotated with polarity or sentiment scores. The overall sentiment of the text are calculated by summing up all the words' sentiment score. Given their crucial role in lexicon based methods, we briefly mention some widely used Arabic lexicons.

Arabic Senti-Lexicon (Abdullah;, 2018): Arabic Senti-Lexcion contains 3880 synsets that are annotated with part of speech, polarity scores and inflected forms.

ArsenL (Arabic Sentiment Lexicon) (Badaro et al., 2014): ArsenL are constructed from multiple resources, including English WordNet (EWN), Arabic WordNet (AWN), English SentiWordNet (ESWN), and SAMA (Standard Arabic Morphological Analyzer). It contains 157969 synsets and has positve, negative and neutral sentiment scores.

2.2 Machine learning based methods

Lexicon based methods are simple and fast, but they heavily rely on the lexicons and the sentiment scores of the lexicons lack context. Machine learning based methods can help to overcome these limitations by learning the sentiment features from data rather than pre-defining them by humans. However, for traditional machine learning methods, feature engineering is required as a precursor to the ML algorithms.

Some widely used feature engineering methods include bag-of-words (Qader et al., 2019), TF-IDF (Sammut and Webb, 2010) and word embedding (Almeida and Xexéo, 2019). After the features have been extracted, machine learning methods such as naive Bayes (Al-Horaibi and Khan, 2016), support vector machines (Duwairi et al., 2016) and random forests based methods (Altawaier and Tiun, 2016) can be used for sentiment analysis. In the following sections, we will systematically review deep learning based approaches.

2.3 The importance of traditional methods: a case study on sentiment lexicons

It is important to note that even with the rise of deep learning based methods for sentiment analysis, sentiment lexicons like ArSenL(Badaro et al., 2014) can still be valuable as they provide a foundational understanding of sentiment that can enhance model performance, especially in low-resource scenarios or when dealing with domain-specific language that may not be well-represented in training data. Some use cases of these lexicons include data preprocessing where the irrelevant terms are filtered out based on the sentiment lexicon, sentiment weighting (Zhang et al., 2011) where the lexicon is employed to help the model weight sentiment-relevant features more effectively. In the following paragraphs, we will use some examples to illustrate how sentiment lexicons can be utilised to improve the sentiment analysis performance.

Feature Augmentation: Lexicons can be utilised to augment the features. In (Heikal et al., 2018), a sentiment lexicon is integrated to augment the features for deep learning based modes. (Xiang et al., 2021) explores part-of-speech-focused lexical substitution for data augmentation to enhance sentiment analysis performance. (Li and Zou, 2024) discusses how lexicon-based approaches assign sentiment polarities and scores to keywords, which can be used for feature augmentation in hybrid models.

Interpretability: Arabic sentiment lexicons enhance interpretability in sentiment analysis by providing a clear framework for understanding how specific words and phrases contribute to sentiment assessments. By combining lexicons with advanced methods like attention-based LSTM and explainable AI techniques, such as LIME (Ribeiro et al., 2016), researchers can further clarify which features are most significant in determining sentiment, thereby enhancing the transparency of their findings (Abdelwahab et al., 2022). These lexicons facilitate the identification of the sentiment polarity of individual terms, making it easier to trace the reasoning behind a sentiment classification (van der Veen and Bleich, 2025). Moreover, the integration of lexicon-based approaches with machine learning techniques can improve the interpretability of complex models, as researchers can analyze how lexicon entries influence the overall sentiment predictions (Ambreen et al., 2024).

Dataset	Modality	Granularity	Context	Dialect
LABR (Aly and Atiya, 2013)	text	SC	document-level	MSA and various other dialects
ASTD (Nabil et al., 2015)	text	SC	document-level	MSA and Egyptian Arabic
ArSentD-LEV (Baly et al., 2019)	text	SC, ABSA, MAST	document-level	Levantine dialect
ArSarcasm (Abu Farha and Magdy, 2020)	text	SC, MAST	document-level	various dialects
ArSarcasm-v2 (Abu Farha et al., 2021)	text	SC, MAST	document-level	various dialects
Arsen-20 (Fang and Xu, 2024)	text	SC	document-level	various dialects
Arabic multimodal dataset (Haouhat et al., 2023)	text, audio, video	SC	document-level (video segments)	various dialects

Table 1: Datasets for Arabic Sentiment Analysis, organised according to modality, granularity and context.
SC: Sentiment Classification
MAST: Multifacted Analysis of Subjective Text
ABSA: Aspect based Sentiment Analysis.

Methods	Contributions	Limitations	Accuracy
(Dahou et al., 2016)	Develops Arabic word embeddings and employs CNN as a classifier	Task-specific method, static word embeddings	LABR (89.6%), ASTD (79.07%)
(Medhaffar et al., 2017)	Annotates a Tunisian dialect corpus and evaluates models on different dialects	Only experiments with traditional machine learning methods	-
(Baly et al., 2017)	Performs a characterization study that analyses tweets from different Arab regions, and compares ML-based vs. deep-learning methods for Arabic SA	Does not experiment on different dialects and topics	ASTD (SVM 51.7%, RNTN 58.5%)
(Guellil et al., 2018)	Automatically constructs an Algerian dialect corpus	Evaluation is carried out only using traditional ML methods	-
(Attia et al., 2018)	Proposes a language independent, multi-class model for SA using CNNs	Evaluation for Arabic is only carried out on ASTD (Nabil et al., 2015), unclear whether the model generalises well to other datasets	ASTD (67.93%)
(Alyafeai et al., 2021)	Compares different tokenizers for different Arabic classification tasks	Does not evaluate on complex architecture like attention-based models -	
(Atabuzzaman et al., 2023)	Proposes an explainable sentiment classification framework for Arabic	Does not conduct experimentation on Transformer-based models	LABR (88.0%)

Table 2: Task-specific methods for Arabic sentiment classification.
LABR: Large Scale Arabic Book Reviews Dataset
ASTD: Arabic Sentiment Tweets Dataset

3 Situating ASA methods vis-à-vis general SA approaches

We situate the research in Arabic Sentiment Analysis (ASA) and juxtapose it with general trends in sentiment analysis (SA) in this section. We present an overview of several sentiment analysis tasks, and for each task we highlight the advancements in general sentiment analysis research, followed by a focus on Arabic-specific sentiment analysis. While not exhaustive, the selected approaches illustrate key differences and current trends between general and Arabic sentiment analysis. Table 1 provides an overview of datasets for ASA organised according to the modality, granularity and context involved. We refer to these datasets in the subsequent subsections.

3.1 Sentiment Classification

Sentiment classification involves assigning a sentiment label (positive, neutral, negative) or a sentiment rating (e.g., one to five) to a given input, which can be text or data from other modalities. As one of the earliest sentiment analysis tasks (Liu, 2015), sentiment classification has attracted significant research interest.

3.1.1 General Sentiment Classification

The development of general sentiment classification reflects the ongoing paradigm shifts within the field of natural language processing (Liu et al., 2021). Early works in sentiment classification primarily relied on task-specific models, employing either traditional machine learning methods like Support Vector Machines (SVMs) or deep learning-based approaches. These models were trained on labeled data and limited to solving specific tasks.

However, the emergence of pre-trained language models such as BERT (Devlin et al., 2019) has revolutionized the field. These models, typically based on components of the Transformer architecture, are pre-trained on massive amounts of unlabeled data and subsequently fine-tuned for specific tasks, including sentiment classification. Large language models like GPT-3 (Brown et al., 2020) further push the boundaries of model size, demonstrating the ability to acquire various emergent capabilities such as in-context learning when scaled sufficiently large (Wei et al., 2022). A systematic analysis of large language models' effectiveness in tackling various sentiment analysis tasks, including sentiment classification, is provided by (Zhang et al., 2023).

3.1.2 Arabic Sentiment Classification

The development of Arabic sentiment classification follows a similar trajectory to that of general sentiment classification. Early research predominantly focused on task-specific models trained on sentiment classification datasets for Arabic text. (Dahou et al., 2016) constructed Arabic word embeddings and subsequently employed a Convolutional Neural Network (CNN) as a classifier. (Attia et al., 2018) proposed a language-independent framework for text classification, evaluating its performance on Arabic sentiment classification tasks as well. Table 2 provides detailed descriptions of various task-specific methods for Arabic sentiment classification, along with their contributions and limitations. As highlighted in the table, the biggest limitation of such methods is that they are task-specific and do not generalise well to other tasks or dialects.

With the remarkable success of pre-trained language models based on bidirectional transformers, such as BERT (Devlin et al., 2019), on diverse natural language understanding tasks, numerous studies have explored their utilization for Arabic sentiment classification. (ElJundi et al., 2019) developed hULMonA, a pre-trained language model specifically for Arabic, and fine-tuned it for Arabic sentiment analysis. AraBERT (Antoun et al., 2020) builds upon this work by pre-training the model entirely on an Arabic corpus and evaluating its performance on various tasks. (Abdul-Mageed et al., 2021) introduced ARBERT and MARBERT, language models pre-trained on dialectal Arabic. Table 3 offers detailed descriptions of different pre-trained language model-based methods for Arabic sentiment classification.

3.2 Multifaceted Analysis of Subjective Text (MAST)

Multifaceted analysis of subjective text (MAST) represents an extension of sentiment classification that delves deeper into task granularity. It shifts the focus towards more specialized tasks, such as irony detection (Zeng and Li, 2022) and comparative opinion mining (Varathan et al., 2017).

3.2.1 General MAST

The development trajectory of general MAST mirrors that of general sentiment classification, as discussed in Section 3.1.1. Due to the focus on specialized tasks within MAST, it encompasses a multitude of sub-tasks. While these sub-tasks have been extensively explored in the field, a detailed de-

Methods	Contributions	Limitations	Accuracy
hULMonA (ElJundi et al., 2019)	Develops a pre-trained LM for Arabic and fine-tunes it for SA	Does not employ an Arabic specific tokenizer and only evaluates on the SA task	ASTD (86.5%), ArSenTD-LEV (52.4%)
AraBERT (Antoun et al., 2020)	Pre-trains an Arabic LM AraBERT and evaluates performance on different tasks	Does not systematically evaluate the model on different dialects	LABR (86.7%), ASTD (92.6%), ArSenTD-Lev (59.4%)
(Alkaoud and Syed, 2020)	Proposes tokenization methods for static and contextual word embeddings and improves their performance	Does not study generalisation ability of the proposed method	LABR (89.87%)
(Abdul-Mageed et al., 2021)	Introduces ARBERT and MARBERT, pre-trains models on dialectal Arabic, introduces ARLUE benchmark	The models have a high memory requirement, thereby impeding computational efficiency	LABR (92.51%), ASTD (95.24%), ArSenTD-Lev (61.38%)
(Alyafeai and Ahmad, 2021)	Employs distillation and quantization to train compact Arabic language models	The effects of hyperparameter tuning are not analysed	LABR (87.5%), ASTD (86.2%)
(El Mekki et al., 2021)	Introduces an unsupervised domain adaptation method for Arabic cross-domain and cross-dialect SA	Does not study the effect of domain adaptation from high-resource languages to Arabic	-
(Abu Kwaik et al., 2022)	Compares feature-based, deep learning and pre-trained LM based methods on dialectal Arabic SA	Lacks an error analysis and a comparison of feature-based vs pre-trained LMs in different situations	-
(Refai et al., 2022)	Proposes a data augmentation method for Arabic text classification using Transformer based models	Does not study the generalisation ability of their method	-

Table 3: Pre-trained language model (LM) based methods for Arabic sentiment classification.
hULMonA: The First Universal Language Model in Arabic
LABR: Large Scale Arabic Book Reviews Dataset
ASTD: Arabic Sentiment Tweets Dataset
ArSenTD-Lev: Arabic Sentiment Twitter Dataset for the Levantine dialect

scription falls outside the scope of this survey. We encourage readers to refer to comprehensive surveys on specific sub-tasks, such as those by (Zeng and Li, 2022) and another work referenced here (Varathan et al., 2017).

3.2.2 Arabic MAST

Compared to general MAST, research on Arabic MAST remains less developed. This section will solely introduce research on Arabic sarcasm detection, as it has garnered a relatively larger body of work following the release of the ArSarcasm (Abu Farha and Magdy, 2020) and ArSarcasm-v2 (Abu Farha et al., 2021) datasets, alongside a shared task organized by WANLP (Abu Farha et al., 2021).

While various methods have been employed, including traditional machine learning approaches, task-specific deep learning methods, and pre-trained language model-based methods, the latter category combined with various optimizations has emerged as the most effective approach. (Hengle et al., 2021) propose a hybrid model that leverages contextual representations from AraBERT (Antoun et al., 2020) alongside static word vectors.

Additionally, recent research explores various machine learning techniques such as down-sampling and augmentation (Israeli et al., 2021) for this task. (Faraj et al., 2021) employ an ensemble approach, combining different pre-trained language models with a hard voting technique. (Talafha et al., 2021) propose framing the problem as a regression task, predicting the level of sarcasm. Table 4 provides detailed descriptions of methods for Arabic sarcasm detection, along with their contributions and limitations.

3.3 Aspect-Based Sentiment Analysis (ABSA)

Aspect-based sentiment analysis (ABSA) extends sentiment analysis by introducing a finer-grained level of task granularity. Unlike sentiment classification, where the output is typically a binary or multi-class label, ABSA delves deeper, focusing on aspects within the sentiment analysis process.

3.3.1 General ABSA

Similar to MAST, general ABSA encompasses various sub-tasks, ranging from simpler single ABSA tasks like aspect term extraction to more complex compound ABSA tasks such as aspect sentiment triplet extraction (Zhang et al., 2022). (Zhang et al., 2022) provide a comprehensive survey on general ABSA, we recommend referring to their work for

further details on general trends in this direction.

3.3.2 Arabic ABSA

Research on Arabic ABSA lags behind that of general ABSA. The majority of existing works in Arabic ABSA primarily address aspect sentiment classification, which essentially translates to sentiment classification applied at the aspect level. Additionally, many studies rely solely on feature-based approaches and traditional machine learning methods. This section will focus exclusively on deep learning-based methods for Arabic ABSA.

The development of Arabic ABSA parallels that of Arabic sentiment classification, as discussed in Section 3.1.2. A growing number of studies are employing deep learning and pre-trained language model-based methods. (Al-Smadi et al., 2017) and (Alshammari and Almansour, 2020) compare traditional machine learning and deep learning methods for Arabic ABSA. (Al-Dabet et al., 2021) propose different network architectures tailored to specific Arabic ABSA tasks. (Abdelgwad et al., 2021) develops a BERT-based model for Arabic ABSA. Table 5 provides detailed descriptions of methods for Arabic ABSA.

3.4 Comparison within Arabic Sentiment Analysis Methods

Although task-specific methods are dedicated to ASA tasks, the most effective methods are those that combine pre-trained language models and various optimisation techniques. hULMonA (ElJundi et al., 2019) first demonstrate the effectiveness of pre-trained language models by developing a pre-trained LM for Arabic and fine-tuning it for ASA, which significantly improves the performance. Later pre-trained Arabic LMs such as AraBERT (Antoun et al., 2020) and ARBERT (Abdul-Mageed et al., 2021) further push the boundary of ASA by using Arabic specific tokenisation and pre-training models on dialectal Arabic.

Another research direction extends the existing methods using various approaches such as domain adaptation (El Mekki et al., 2021) and data augmentation (Refai et al., 2022). However, the models based on pre-trained LMs are not computationally efficient, and involve a significant computational overhead, whereas approaches such as (Alyafeai and Ahmad, 2021) maintain the balance between performance and efficiency by distillation and quantisation.

Methods	Contributions	Limitations	Accuracy
(Hengle et al., 2021)	Proposes a hybrid model which combines contextual representations from AraBERT (Antoun et al., 2020) and static word vectors	Hybrid model, computational efficiency is impeded owing to use of both contextual and static embeddings.	74.1%
(Israeli et al., 2021)	Employs pre-trained Transformer based models and various machine learning techniques such as down-sampling and augmentation	Does not explain the effects of these techniques	76.7%
(Talafha et al., 2021)	Annotates an Arabic sarcasm detection dataset, trains a regression model and predicts the level of sarcasm	While the model can output the level of sarcasm, it is reliant on a binary classification dataset	-
(Khondaker et al., 2022)	Applies contrastive learning to Arabic social meaning tasks	Does not study the generalisation ability of their method	-
(Faraj et al., 2021)	Ensembles different pre-trained language models and employs hard voting technique	The method is not efficient as it needs various pre-trained language models.	78.3%
(El Mahdaouy et al., 2021)	Proposes an end-to-end multi-task model for Arabic sentiment analysis and sarcasm detection	Does not present experimentation on other tasks	76.8%
(Kaseb and Farouk, 2022)	Proposes SAIDS that uses its prediction of sarcasm and dialect as known information to predict the sentiment	Does not study the generalisation ability of their method	-

Table 4: Methods for Arabic sarcasm detection. The accuracy is evaluated on ArSarcam-v2 dataset.

Methods	Contributions	Limitations
(Ashi et al., 2019)	Employs pre-trained word embeddings for Arabic ABSA	Only uses traditional machine learning methods as classifier
(Al-Smadi et al., 2017)	Compares RNN and SVM for Arabic ABSA	The dataset is relatively small, does not use other deep learning models such as LSTM
(Alshammari and Almansour, 2020)	Compares CNN and traditional machine learning methods for Arabic ABSA	Task-specific method, does not compare different deep learning methods
(Al-Dabet et al., 2021)	Proposes different network architectures for various Arabic ABSA tasks	Task-specific method
(Abdelgwad et al., 2021)	Develops a BERT based model for Arabic ABSA	Does not present experimentation on other tasks

Table 5: Methods for Arabic aspect based sentiment analysis (ABSA).

3.5 Gaps and challenges in Arabic Sentiment Analysis

3.5.1 Gaps

This section outlines key research gaps between Arabic sentiment analysis and general sentiment analysis across three dimensions:

- **Modality:** Multimodality has recently garnered significant interest within general sentiment analysis, with a surge in research activity (Lai et al., 2023). However, investigations into multi-modal Arabic sentiment analysis remain limited. Most datasets for Arabic sentiment analysis are restricted to the text modality.

- **Granularity:** The majority of research in Arabic sentiment analysis focuses solely on Arabic sentiment classification. As evidenced in the previous sections, even studies exploring Arabic MAST and Arabic ABSA often target simpler tasks. Consequently, Arabic sentiment analysis lags behind general sentiment analysis in terms of MAST and ABSA tasks.

- **Context:** While datasets for general sentiment analysis encompass various levels ranging from document level to aspect level, most datasets for Arabic sentiment analysis remain at the document level. Even some recently released datasets lack annotations at sentence level and aspect level.

3.5.2 Challenges

The Arabic language is characterized by its high morphological complexity, which entails intricate word formation processes that may obscure meaning (Habash, 2010). Additionally, the high degree of ambiguity and polysemy inherent in Arabic lexicon complicates semantic interpretation. The presence of negation and the extensive range of dialects further exacerbate these challenges, as they introduce variations that must be meticulously accounted for in linguistic models (El-Beltagy and Ali, 2013).

Data scarcity and cultural contextualization present additional challenges for Arabic. There is a scarcity of large, labeled datasets for many dialects, making it difficult to train robust models. Moreover, sentiment expression can vary significantly based on cultural nuances, requiring models to understand context beyond mere text.

4 Recent trends in Arabic Sentiment Analysis

Several research efforts are ongoing to develop robust Arabic-specific methods and overcome the challenges presented in Section 3.5. We organise and present these efforts below:

4.1 Addressing Dialectal Variations

The issue of dialectal variation has received significant attention in both task-specific and pre-trained language model-based approaches. (Baly et al., 2017) conducted a characterization study analyzing tweets from different Arab regions, highlighting the importance of addressing the dialectal problems in Arabic SA. The efforts that tackle this challenge are presented below, grouped into the broad approach employed:

4.1.1 Dataset Creation, Domain Adaptation and Data Augmentation

(Medhaffar et al., 2017) addressed this challenge by annotating a corpus specifically for the Tunisian dialect and evaluating their models on data from various dialects. A similar approach to this method was the one proposed by (Guellil et al., 2018). They presented a method for automatically constructing an Algerian dialect corpus. (El Mekki et al., 2021) introduced an unsupervised domain adaptation method for cross-domain and cross-dialect sentiment analysis in Arabic. (Refai et al., 2022) proposed a data augmentation method specifically designed for Arabic classification tasks using transformer-based models.

4.1.2 Increasing use of Deep learning

Researchers are increasingly using deep learning models, particularly transformer based models, to effectively capture the nuances of different Arabic dialects. For example, ARBERT and MARBERT (Abdul-Mageed et al., 2021) were specifically pre-trained on dialectal Arabic to address these dialectal variations.

4.1.3 Transfer Learning and Multilingual Models

Transfer learning approaches are being used to leverage knowledge from models trained on larger datasets in other languages or MSA, facilitating better performance on dialect data with limited resources. Multilingual transformer models like mBERT are also applied for handling multiple Arabic dialects (Devlin et al., 2019).

4.2 Arabic-specific Tokenization

Recent research has also explored the importance of developing tokenization methods specifically for the Arabic language. (Alyafeai et al., 2021) compared the performance of different tokenizers for various Arabic classification tasks. (Alkaoud and Syed, 2020) proposed tokenization strategies specifically tailored for both static and contextual Arabic word embeddings, demonstrating significant performance improvements. The efforts in this direction can be grouped into the following trends:

Morphological Analysis Implementation of advanced morphological analysis tools to accurately identify roots, prefixes, and suffixes, ensuring proper tokenization of complex words. Noteworthy contributions in this area include the MADAMIRA tool, which provides robust morphological analysis and disambiguation for modern written Arabic, showcasing significant improvement in processing complex Arabic morphological structures (Pasha et al., 2014).

Dialect-Specific Tokenizers: Development of tokenization models tailored to specific Arabic dialects to handle regional vocabulary and expressions effectively. The CALIMA-Star project exemplifies efforts to create comprehensive morphological lexicons specific to different Arabic dialects, allowing more precise tokenization and analysis for dialectal data (Taji et al., 2018).

Contextual Tokenization: Use of context-aware tokenization methods to understand the meaning of words in context, assisting in disambiguating similar words. Contextual models like AraBERT and its advancements in tokenization strategies demonstrate the power of context-aware embeddings to capture nuanced language variations in sentiment analysis (Antoun et al., 2020).

Incorporating Diacritics: Desining tokenizers that handle diacritics appropriately, either by retaining them for analysis or by normalizing words without diacritics while preserving meaning. Research by (Alqahtani et al., 2020) highlights the role of diacritics in enhancing sentiment analysis, emphasizing the necessity for tokenizers that efficiently process diacritized text data without losing critical semantic information.

While these trends have demonstrated improved performance for ASA, significant research efforts need to be directed in order to bridge the gap between ASA and general SA for high-resource languages.

5 Future Directions

To conclude, we present promising research directions to foster the development of robust models for Arabic sentiment analysis.

Creation of richer datasets: Future efforts should prioritize the development of datasets that encompass richer annotations across the following dimensions:

- **Multimodality:** Datasets that integrate various modalities (text, audio, video) to capture richer sentiment information.

- **Fine-grained tasks:** Datasets designed for exploring more intricate sentiment analysis tasks beyond sentiment classification.

- **Multi-context annotations:** Datasets with annotations at finer levels (sentence level, aspect level) to facilitate in-depth analysis.

Multimodal Sentiment Analysis: While limited research has been conducted in multimodal ASA, leveraging information from multiple modalities holds significant potential for accurate sentiment analysis. Future research should explore effective techniques for incorporating multimodal data and develop robust models for this task.

Large Language Models (LLMs) for ASA: Recent advancements in LLMs have yielded remarkable performance on various tasks. Arabic LLMs like AceGPT-LL (Huang et al., 2023) have also emerged. However, a systematic analysis of LLMs for sentiment analysis, particularly in the context of Arabic, is lacking. Future research should investigate how to best utilize LLMs for Arabic sentiment analysis.

Interpretable Sentiment Analysis: Existing Arabic sentiment analysis methods primarily provide final sentiment labels without explanations for their outputs. Recent work on improving the interpretability of question answering by examining model reasoning (Huang and Chang, 2023) suggests a promising approach that can be adapted to sentiment analysis. For example, models could be designed to output reasoning steps leading to their final sentiment polarity predictions.

Exploration of Fine-Grained Tasks: General sentiment analysis research has shifted towards increasingly fine-grained tasks. However, most Arabic sentiment analysis studies remain focused on sentiment classification at the document level. A systematic exploration of other fine-grained tasks, particularly those within MAST and ABSA, would be beneficial for advancing the field.

6 Limitations

This survey acknowledges some limitations. Firstly, it primarily focuses on works utilizing deep learning methods. As highlighted in (Abu Kwaik et al., 2022), feature-based methods can outperform pretrained language model-based methods in some instances. Future surveys may benefit from including an exploration of feature-based approaches. Additionally, while this work compares Arabic sentiment analysis with general sentiment analysis, it would also be valuable to situate Arabic sentiment analysis within the broader context of Arabic classification tasks and Arabic natural language processing tasks in general.

References

Mohammed M. Abdelgwad, Taysir Hassan A. Soliman, and Ahmed I. Taloba. 2021. Arabic aspect sentiment polarity classification using BERT. *Journal of Big Data*, 9:1–15.

Huda Jamal Abdelhameed and Susana Muñoz-Hern'andez. 2017. Emotion and opinion retrieval from social media in Arabic language: Survey. In *2017 Joint International Conference on Information and Communication Technologies for Education and Training and International Conference on Computing in Arabic (ICCA-TICET)*, pages 1–8.

Youmna Abdelwahab, Mohamed Hamed Kholief, and Ahmed Ahmed Hesham Sedky. 2022. Justifying Arabic text sentiment analysis using Explainable AI (XAI): LASIK surgeries case study. *Inf.*, 13:536.

Muhammad Abdul-Mageed, AbdelRahim Elmadany, and El Moatez Billah Nagoudi. 2021. ARBERT & MARBERT: Deep bidirectional transformers for Arabic. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7088–7105, Online. Association for Computational Linguistics.

Tareq Al-Moslmi;Mohammed Albared;Adel Al-Shabi;Nazlia Omar;Salwani Abdullah;. 2018. Arabic senti-lexicon: Constructing publicly available language resources for arabic sentiment analysis. *Journal of Information Science*, 44(3):345–362.

Mohamed Elhag Mohamed Abo, Ram Gopal Raj, and Atika Qazi. 2019. A review on Arabic sentiment analysis: State-of-the-art, taxonomy and open research challenges. *IEEE Access*, 7:162008–162024.

Ibrahim Abu Farha and Walid Magdy. 2020. From Arabic sentiment analysis to sarcasm detection: The ArSarcasm dataset. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 32–39, Marseille, France. European Language Resource Association.

Ibrahim Abu Farha, Wajdi Zaghouani, and Walid Magdy. 2021. Overview of the WANLP 2021 shared task on sarcasm and sentiment detection in Arabic. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 296–305, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Kathrein Abu Kwaik, Stergios Chatzikyriakidis, and Simon Dobnik. 2022. Pre-trained models or feature engineering: The case of dialectal Arabic. In *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection*, pages 41–50, Marseille, France. European Language Resources Association.

Mahmoud Al-Ayyoub, Abed Allah Khamaiseh, Yaser Jararweh, and Mohammed N. Al-Kabi. 2019. A comprehensive survey of Arabic sentiment analysis. *Information Processing and Management*, 56(2):320–342. Advance Arabic Natural Language Processing (ANLP) and its Applications.

Saja Al-Dabet, Sara Tedmori, and Mohammad Al-Smadi. 2021. Enhancing Arabic aspect-based sentiment analysis using deep learning models. *Comput. Speech Lang.*, 69:101224.

Lamia Al-Horaibi and Muhammad Badruddin Khan. 2016. Sentiment analysis of Arabic tweets using text mining techniques. In *International Workshop on Pattern Recognition, 2016*.

Mohammad Al-Smadi, Omar Qawasmeh, Mahmoud Al-Ayyoub, Yaser Jararweh, and Brij Bhooshan Gupta. 2017. Deep Recurrent Neural Network vs. Support Vector Machine for aspect-based sentiment analysis of Arabic hotels' reviews. *J. Comput. Sci.*, 27:386–393.

Nora Al-Twairesh, Hend Al-Khalifa, and AbdulMalik Al-Salman. 2014. Subjectivity and sentiment analysis of Arabic: Trends and challenges. In *2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)*, pages 148–155.

Mohamed Alkaoud and Mairaj Syed. 2020. On the importance of tokenization in Arabic embedding models. In *Proceedings of the Fifth Arabic Natural Language Processing Workshop*, pages 119–129, Barcelona, Spain (Online). Association for Computational Linguistics.

Felipe Almeida and Geraldo Bonorino Xexéo. 2019. Word embeddings: A survey. *ArXiv*, abs/1901.09069.

Latifah Almurqren, Ryan Hodgson, and A Ioana Cristea. 2024. Arabic text sentiment analysis: Reinforcing human-performed surveys with wider topic analysis. *ArXiv*, abs/2403.01921.

Sawsan Alqahtani, Ajay Mishra, and Mona Diab. 2020. A multitask learning approach for diacritic restoration. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8238–8247, Online. Association for Computational Linguistics.

Norah Fahad Alshammari and Amal Almansour. 2020. Aspect-based sentiment analysis for Arabic content in social media. *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pages 1–6.

Merfat.M. Altawaier and Sabrina Tiun. 2016. Comparison of machine learning approaches on arabic twitter sentiment analysis. *International Journal on Advanced Science, Engineering and Information Technology*, 6:1067–1073.

Mohamed Aly and Amir Atiya. 2013. LABR: A large scale Arabic book reviews dataset. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 494–498, Sofia, Bulgaria. Association for Computational Linguistics.

Zaid Alyafeai and Irfan Ahmad. 2021. Arabic compact language modelling for resource limited devices. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 53–59, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Zaid Alyafeai, Maged S. Al-Shaibani, Mustafa Ghaleb, and Irfan Ahmad. 2021. Evaluating various tokenizers for Arabic text classification. *Neural Processing Letters*, 55:2911–2933.

Shela Ambreen, Muhammad Iqbal, Muhammad Zubair Asghar, Tehseen Mazhar, Umar Farooq Khattak, Muhammad Amir Khan, and Habib Hamam. 2024. Predicting customer sentiment: The fusion of deep learning and a fuzzy system for sentiment analysis of arabic text. *Soc. Netw. Anal. Min.*, 14:206.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. AraBERT: Transformer-based model for Arabic language understanding. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 9–15, Marseille, France. European Language Resource Association.

Mohammed Ashi, Muazzam Siddiqui, and Farrukh Nadeem. 2019. *Pre-trained Word Embeddings for Arabic Aspect-Based Sentiment Analysis of Airline Tweets*, pages 241–251. Springer.

Md. Atabuzzaman, Md Shajalal, Maksuda Bilkis Baby, and Alexander Boden. 2023. Arabic sentiment analysis with noisy deep explainable model. In *International Conference on Natural Language Processing and Information Retrieval*.

Mohammed Attia, Younes Samih, Ali Elkahky, and Laura Kallmeyer. 2018. Multilingual multi-class sentiment classification using convolutional neural networks. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Gilbert Badaro, Ramy Baly, Hazem Hajj, Nizar Habash, and Wassim El-Hajj. 2014. A large scale Arabic sentiment lexicon for Arabic opinion mining. In *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*, pages 165–173, Doha, Qatar. Association for Computational Linguistics.

Ramy Baly, Gilbert Badaro, Georges El-Khoury, Rawan Moukalled, Rita Aoun, Hazem Hajj, Wassim El-Hajj, Nizar Habash, and Khaled Shaban. 2017. A characterization study of Arabic Twitter data with a benchmarking for state-of-the-art opinion mining models. In *Proceedings of the Third Arabic Natural Language Processing Workshop*, pages 110–118, Valencia, Spain. Association for Computational Linguistics.

Ramy Baly, Alaa Khaddaj, Hazem M. Hajj, Wassim El-Hajj, and Khaled Bashir Shaban. 2019. ArSentD-LEV: A multi-topic corpus for target-based sentiment analysis in Arabic Levantine tweets. *CoRR*, abs/1906.01830.

Mariam Biltawi, Wael Etaiwi, Sara Tedmori, Amjad Hudaib, and Arafat Awajan. 2016. Sentiment classification techniques for Arabic language: A survey. In *2016 7th International Conference on Information and Communication Systems (ICICS)*, pages 339–346.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeff Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *ArXiv*, abs/2005.14165.

Abdelghani Dahou, Shengwu Xiong, Junwei Zhou, Mohamed Houcine Haddoud, and Pengfei Duan. 2016. Word embeddings and convolutional neural network for Arabic sentiment classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2418–2427, Osaka, Japan. The COLING 2016 Organizing Committee.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Rehab M. Duwairi, Mosab Alfaqeh, Mohammad Wardat, and Areen Alrabadi. 2016. Sentiment analysis for Arabizi text. In *2016 7th International Conference on Information and Communication Systems (ICICS)*, pages 127–132.

Samhaa R. El-Beltagy and Ahmed Ali. 2013. Open issues in the sentiment analysis of Arabic social media: a case study. In *2013 IEEE International Conference on Innovations in Information Technology (IIT)*, pages 215–220. IEEE.

Abdelkader El Mahdaouy, Abdellah El Mekki, Kabil Essefar, Nabil El Mamoun, Ismail Berrada, and Ahmed Khoumsi. 2021. Deep multi-task model for sarcasm detection and sentiment analysis in Arabic language. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 334–339, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Abdellah El Mekki, Abdelkader El Mahdaouy, Ismail Berrada, and Ahmed Khoumsi. 2021. Domain adaptation for Arabic cross-domain and cross-dialect sentiment analysis from contextualized word embedding. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2824–2837, Online. Association for Computational Linguistics.

Obeida ElJundi, Wissam Antoun, Nour El Droubi, Hazem Hajj, Wassim El-Hajj, and Khaled Shaban. 2019. hULMonA: The universal language model in Arabic. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pages 68–77, Florence, Italy. Association for Computational Linguistics.

Yang Fang and Cheng Xu. 2024. ArSen-20: A new benchmark for Arabic sentiment detection. In *5th Workshop on African Natural Language Processing*.

Dalya Faraj, Dalya Faraj, and Malak Abdullah. 2021. SarcasmDet at sarcasm detection task 2021 in Arabic using AraBERT pretrained model. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 345–350, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Imane Guellil, Ahsan Adeel, Faical Azouaou, and Amir Hussain. 2018. SentiALG: Automated corpus annotation for Algerian sentiment analysis. *ArXiv*, abs/1808.05079.

Nizar Habash. 2010. *Introduction to Arabic Natural Language Processing*. Morgan & Claypool Publishers.

Abdelhamid Haouhat, Slimane Bellaouar, Attia Nehar, and Hadda Cherroun. 2023. Towards Arabic multimodal dataset for sentiment analysis. *2023 Fourth International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, pages 126–133.

Maha Heikal, Marwan Torki, and Nagwa M. El-Makky. 2018. Sentiment analysis of Arabic tweets using deep learning. In *International Conference on Arabic Computational Linguistics*.

Amey Hengle, Atharva Kshirsagar, Shaily Desai, and Manisha Marathe. 2021. Combining context-free and contextualized representations for Arabic sarcasm detection and sentiment identification. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 357–363, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Huang Huang, Fei Yu, Jianqing Zhu, Xuening Sun, Hao Cheng, Dingjie Song, Zhihong Chen, Abdulmohsen Alharthi, Bang An, Ziche Liu, Zhiyi Zhang, Junying Chen, Jianquan Li, Benyou Wang, Lian Zhang, Ruoyu Sun, Xiang Wan, Haizhou Li, and Jinchao Xu. 2023. AceGPT, localizing large language models in Arabic. *ArXiv*, abs/2309.12053.

Jie Huang and Kevin Chen-Chuan Chang. 2023. Towards reasoning in large language models: A survey. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1049–1065, Toronto, Canada. Association for Computational Linguistics.

Abraham Israeli, Yotam Nahum, Shai Fine, and Kfir Bar. 2021. The IDC system for sentiment classification and sarcasm detection in Arabic. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 370–375, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Abdelrahman Kaseb and Mona Farouk. 2022. SAIDS: A novel approach for sentiment analysis informed of dialect and sarcasm. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 22–30, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Md Tawkat Islam Khondaker, El Moatez Billah Nagoudi, AbdelRahim Elmadany, Muhammad Abdul-Mageed, and Laks Lakshmanan, V.S. 2022. A benchmark study of contrastive learning for Arabic social meaning. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 63–75, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Songning Lai, Hao Xu, Xifeng Hu, Zhaoxia Ren, and Zhi Liu. 2023. Multimodal sentiment analysis: A survey. *ArXiv*, abs/2305.07611.

Zhenyu Li and Zongfeng Zou. 2024. Punctuation and lexicon aid representation: A hybrid model for short text sentiment analysis on social media platform. *Journal of King Saud University - Computer and Information Sciences*, 36(3):102010.

Bing Liu. 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pretrain, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55:1 – 35.

Salima Medhaffar, Fethi Bougares, Yannick Estève, and Lamia Hadrich-Belguith. 2017. Sentiment analysis of Tunisian dialects: Linguistic ressources and experiments. In *Proceedings of the Third Arabic Natural Language Processing Workshop*, pages 55–61, Valencia, Spain. Association for Computational Linguistics.

Mahmoud Nabil, Mohamed Aly, and Amir Atiya. 2015. ASTD: Arabic sentiment tweets dataset. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2515–2519, Lisbon, Portugal. Association for Computational Linguistics.

Arfath Pasha et al. 2014. Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of arabic. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)*.

Wisam A. Qader, Musa M. Ameen, and Bilal I. Ahmed. 2019. An overview of Bag of Words: Importance, implementation, applications, and challenges. In *2019 International Engineering Conference (IEC)*, pages 200–204.

Alaa Rahma, Shahira Shaaban Azab, and Ammar Mohammed. 2023. A comprehensive survey on Arabic sarcasm detection: Approaches, challenges and future trends. *IEEE Access*, 11:18261–18280.

Dania Refai, Saleh Abu-Soud, and Mohammad J. Abdel-Rahman. 2022. Data augmentation using transformers and similarity measures for improving Arabic text classification. *IEEE Access*, 11:132516–132531.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Claude Sammut and Geoffrey I. Webb, editors. 2010. TF–IDF, pages 986–987. Springer US, Boston, MA.

Dima Taji, Salam Khalifa, Ossama Obeid, Fadhl Eryani, and Nizar Habash. 2018. An Arabic morphological analyzer and generator with copious features. In *Proceedings of the Fifteenth Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 140–150, Brussels, Belgium. Association for Computational Linguistics.

Bashar Talafha, Muhy Eddin Za'ter, Samer Suleiman, Mahmoud Al-Ayyoub, and Mohammed N. Al-Kabi. 2021. Sarcasm detection and quantification in Arabic tweets. *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 1121–1125.

A. Maurits van der Veen and Erik Bleich. 2025. The advantages of lexicon-based sentiment analysis in an age of machine learning. *PLOS ONE*, 20(1):1–19.

Kasturi Dewi Varathan, Anastasia Giachanou, and Fabio Crestani. 2017. Comparative opinion mining: A review. *Journal of the Association for Information Science and Technology*, 68(4):811–829.

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed Huai hsin Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. 2022. Emergent abilities of large language models. *ArXiv*, abs/2206.07682.

Rong Xiang, Emmanuele Chersoni, Qin Lu, Chu-Ren Huang, Wenjie Li, and Yunfei Long. 2021. Lexical data augmentation for sentiment analysis. *Journal of the Association for Information Science and Technology*, 72:1432 – 1447.

Qingcheng Zeng and An-Ran Li. 2022. A survey in automatic irony processing: Linguistic, cognitive, and multi-X perspectives. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 824–836, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Lei Zhang, Riddhiman Ghosh, Mohamed Dekhil, Meichun Hsu, and B. Liu. 2011. Combining lexicon-based and learning-based methods for twitter sentiment analysis. In *HP Labs Technical Reports*.

Wenxuan Zhang, Yue Deng, Bing-Quan Liu, Sinno Jialin Pan, and Lidong Bing. 2023. Sentiment analysis in the era of large language models: A reality check. *ArXiv*, abs/2305.15005.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2022. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 35:11019–11038.