



Arabic sentiment analysis: studies, resources, and tools

Imane Guellil¹ · Faical Azouaou¹ · Marcelo Mendoza²

Received: 27 December 2018 / Revised: 15 September 2019 / Accepted: 17 September 2019 / Published online: 27 September 2019
© Springer-Verlag GmbH Austria, part of Springer Nature 2019

Abstract

To determine whether a document or a sentence expresses a positive or negative sentiment, three main approaches are commonly used: the lexicon-based approach, corpus-based approach, and a hybrid approach. The study of sentiment analysis in English has the highest number of sentiment analysis studies, while research is more limited for other languages, including Arabic and its dialects. Lexicon based approaches need annotated sentiment lexicons (containing the valence and intensity of its terms and expressions). Corpus-based sentiment analysis requires annotated sentences. One of the significant problems related to the treatment of Arabic and its dialects is the lack of these resources. We present in this survey the most recent resources and advances that have been done for Arabic sentiment analysis. This survey presents recent work (where the majority of these works are between 2015 and 2019). These works are classified by category (survey work or contribution work). For contribution work, we focus on the construction of sentiment lexicon and corpus. We also describe emergent trends related to Arabic sentiment analysis, principally associated with the use of deep learning techniques.

Keywords Arabic sentiment analysis · Sentiment lexicon · Corpus lexicon · Machine learning · Deep learning · Word embeddings

1 Introduction

Nowadays, if someone desires to buy a product or visit a new restaurant or a hotel, one is not limited to asking one's friends and family for getting their opinions. This is principal because there are many user reviews and discussions about the products, restaurants, and hotels in public forums on the Web. For an organization, it is not necessary to conduct surveys or focus groups to gather public opinions because such information is abundant and publicly available. Recently, the richness of social media in term of opinions

have helped reshape businesses, and affect public sentiments and emotions, which have profoundly impacted on social and political systems. Such postings have also lead to some political changes such as those that happened in some Arabian countries in 2011. Hence, it became a necessity to collect and study opinions and sentiments (Zhang et al. 2018).

The field studying people opinions, emotions, and sentiments is named opinion mining and sentiment analysis. Opinion mining and sentiment analysis are defined as an interdisciplinary domain among natural language processing (NLP), artificial intelligence (AI), and text mining (Liu 2012; Guellil and Boukhalfa 2015). To determine whether a document or a sentence expresses a positive or negative sentiment, three main approaches are commonly used: the lexicon-based approach (Taboada et al. 2011), corpus-based approach (Maas et al. 2011) and a hybrid approach (Khan et al. 2015). These approaches are principally based on annotated data. Corpus-based sentiment analysis requires an annotated sentiment corpus. The lexicon-based approach needs an annotated sentiment lexicon (containing the valence and intensity of its terms and expressions) (Guellil et al. 2018b). However, the majority of the work has been focused on English. Other languages such as Spanish, Chinese, or even Arabic (which is considered the 4th most used

✉ Imane Guellil
i_guellil@esi.dz
<http://www.esi.dz>

Faical Azouaou
<http://www.esi.dz>

Marcelo Mendoza
<http://www.inf.utfsm.cl/~mmendoza>

¹ Laboratoire des Méthodes de Conception des Systèmes,
Ecole nationale Supérieure d'Informatique, BP 68M,
16309 Oued-Smar, Alger, Algeria

² Department of Informatics, Universidad Técnica Federico
Santa María, Av. Vicña Mackenna 3939, Santiago, Chile

language in social media) (Alayba et al. 2018b) has been less studied. Regarding the work focusing on the Arabic language, the major drawback is related to the lack of annotated data. The characteristics of Arabic and its dialects (orthographic, syntactic, and semantic) make sentiment analysis tasks more challenging. Mainly due to this reason, research works that have been conducted on Arabic and its dialect are limited compared to researches that have done on other languages as in English. To provide the reader with a better insight into this survey, we firstly outline its motivations, and its major's contribution.

1.1 Motivation of the proposed survey

Users' opinions on a product, a company, or a political personality are crucial for business managers and company directors. The emergence of the internet and social media makes available large amounts of data containing significant numbers of opinions, sentiments, and emotions, thus engendering the interest of extracting and analyzing them. Hence, research in terms of sentiment analysis has attracted more attention during the last decade. This attention began to increase for the Arabic language. A simple research on *Google Scholar* with the keyword *Sentiment analysis* returns approximately 1,920,000 results where a research with the keyword *Arabic Sentiment analysis* returns approximately 82,800 and a research with the keyword *Arabic Sentiment analysis* returns approximately 1,010,000. From these statistics, it could be seen that even if the interest regarding Arabic Sentiment analysis is relatively important, it is negligible compared to the interest in English sentiment Analysis. The lack of works focusing on Arabic Sentiment Analysis is mainly due to two reasons: (1) The lack of resources dedicated to Arabic. (2) The complexity of this language and its dialects. To provide the research community with a synthesis of the constructed resources (focusing on the construction techniques, sentiment analysis methods and new tendencies related to word embedding and deep learning) dedicated to Arabic sentiment Analysis, we present this survey.

Other survey papers were presented in the research literature in order to classify the works that have been conducted on Arabic sentiment analysis such as Korayem et al. (2012), Harrag (2014), Alhumoud et al. (2015), Assiri et al. (2015), Kaseb and Ahmed (2016), Biltawi et al. (2016), El-Masri et al. (2017). These surveys are presented in more detail in Sect. 4. However, the majority of them are a little bit old where almost all of them are before 2015. Moreover, the proposed state of the art paper focused only on the work related to Arabic sentiment analysis by neglecting the work that has been done on Arabizi sentiment analysis. Also, all these surveys classified the studied work by the type of approach that the authors used (i.e., lexicon-based approach, corpus-based approach, and hybrid approach). Despite that

this classification gives a good overview of the related work, it will not provide details about the constructed resources. Moreover, All the sentiment analysis approaches require data (i.e., sentiment annotated lexicon and sentiment annotated corpus). The presentation and the synthesis of the techniques used for constructing these resources, highlighting the publicly available resources by showing how they improved the research results represent a firm insight/overview for the reader in the field of sentiment analysis.

1.2 Contribution of the proposed survey

The primary purpose of this survey is to present a paper focusing on the most recent works on Arabic sentiment analysis. The main contributions of this survey are summarised in the following points:

- The majority of the work presented in this article covers the period 2015–2019.
- This survey is concentrating on all works done on Arabic sentiment analysis, those concentrating on Arabic script, as well as those concentrating on Arabizi script.
- We consider the resources (annotated lexicons and corpus) as the classification pivot. Hence the presented work are grouped into those presenting sentiment lexicon (and using them) from those proposing sentiment corpus and using them).
- We integrate the presented lexicon and corpus into summarising table including the name of the constructed resources and the work proposing it, the year of its construction, its size, the related work using this resource and finally the link to get it (if available).
- Highlighting the work using the constructed resources and the link of these resources could give the reader an idea about the effect of the resource with meaningful information about the availability of Arabic sentiment analysis resources.
- We are not focusing on the constructed resources only; we also present the presented methods for validating these resources as well as new tendencies related to sentiment analysis. In this part, we present the emerging models for features extraction, approach, and new models for sentiment classification.

This survey is organized as follows. We first present an overview of Sentiment analysis and Arabic language processing in both Sects. 2 and 3. Afterwards, we present the different surveys that have been conducted on Arabic sentiment analysis in Sect. 4. The different resources (lexicons and corpora), which are constructed are presented in Sect. 5. The methods used in order to validate the constructed resources are presented in Sect. 6. Then, we present new trends related to Arabic sentiment analysis in Sect. 7. Analysis and discussion

are included in Sect. 8. We conclude in Sect. 9 by providing remarks and outlining research perspectives.

2 Sentiment analysis: an overview

Sentiment analysis (SA), also called opinion mining, is the field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions toward entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents an important and active research area in computer science (Liu 2012).

Several types of research (Bhuta et al. 2014; El-Masri et al. 2017; Zia et al. 2018) presented the general process of sentiment analysis consisting of four steps: data extraction, text pre-processing, data analysis, and identification of useful knowledge. Three kinds of data are used in the research literature; namely, *existing corpus* used in previous research work, *manual corpus*, manually extracted from social media or *automatic corpus* obtained using APIs provided by different social media. The pre-processing step may include multiple tasks such as splitting the text into sentences or word lemmatization (transformation of verbs to their infinitive forms), a treatment that identifies different grammatical parts (e.g., POS¹ tags) of the words in sentences (Guellil and Boukhalfa 2015). Most of research works use the term "*sentiment classification*" rather than "*sentiment analysis*" to indicate the third step of sentiment analysis process (Singh et al. 2013; Liang and Dai 2013; Akaichi 2013; Antai 2014; Colace et al. 2013). According to Colace et al. (2013), Karamibekr and Ghorbani (2012), sentiment classification is the main task of sentiment analysis. Classification can be done using three automatic learning approaches: *supervised*, *unsupervised* or *hybrid*. Supervised approaches (also called the corpus-based method) typically use labeled corpus to train the sentiment classifier (Wan 2012). Unsupervised approaches (also called lexicon-based method) are based on sentiment lexicons like dictionaries (Medhat et al. 2014). Hybrid approaches combine the two precedents approaches (Bahrainian and Dengel 2013). The knowledge generated through the sentiment analysis process can be related to the population studied their posts or comments to determine the general sentiment associated with a given product, company, institution, or celebrity.

Researchers have mainly studied sentiment analysis at three levels: document level, sentence level, and aspect level. In document-level tasks, the goal is related to the classification of a personal document (e.g., a product review) as expressing an overall positive or negative opinion. It

considers the whole document as the primary information unit, and it assumes that the text is known to be subjective and contains opinions about a single entity (e.g., a particular phone). Sentiment analysis at sentence level classifies individual sentences. However, each sentence cannot be assumed to be subjective. Traditionally, the first task consists of classifying a sentence as subjective or objective, which is called subjectivity classification. Then the resulting subjective sentences are classified as expressing positive or negative opinions. Sentence level sentiment classification can also be formulated as a three-class classification problem, that is, to classify a sentence as neutral, positive, or negative.

Aspect-based sentiment analysis is a fine-grained task. It consists of extracting and summarising people's opinions expressed on aspects of named entities. These entities are often called targets. A product review aims to summarise positive and negative opinions on different aspects of the product, respectively. The general sentiment on the product could be positive or negative. The whole task of aspect-based sentiment analysis consists of several sub-tasks such as aspect extraction, entity extraction, and aspect sentiment classification. For example, from the sentence, "the voice quality of iPhone is great, but its battery stinks" the entity extractor has to identify "iPhone" as the entity, and the aspect extractor has to define "voice quality" and "battery" as two aspects. The aspect sentiment classifier has to classify the sentiment expressed on the voice quality of the iPhone as positive and on the battery of the iPhone as unfavorable. However, for simplicity, in most algorithms, aspect extraction and entity extraction are combined and are called aspect extraction or sentiment/opinion target extraction (Zhang et al. 2018; Liu 2012).

3 Arabic language processing: an overview

Arabic is the official language of 27 countries. It is spoken by more than 400 million speakers. Arabic is also recognised as the 4th most used language in Internet (Boudad et al. 2017; Guellil et al. 2018b). All the work in the literature (Habash 2010; Farghaly and Shaalan 2009; Harrat et al. 2017) classify Arabic in three main varieties: (1) Classical Arabic (CA) which is the form of the Arabic language used in literary texts, in which the Quran² is considered to be the highest form of CA text (Sharaf and Atwell 2012a), (2) Modern Standard Arabic (MSA) which is used for writing as well

¹ POS: Part Of Speech.

² The Quran is a scripture which, according to Muslims, is the verbatim words of Allah. It contains over 77,000 words revealed through Archangel Gabriel to Prophet Muhammad over 23 years beginning in 610 CE. It is divided into 114 chapters of varying sizes, where each section is divided into verses, adding up to a total of 6243 verses (Sharaf and Atwell 2012b).

as formal conversations, (3) Dialectal Arabic (DA) which is used in daily life communication and conversational level (Boudad et al. 2017). However, Arabic speakers on social media, discussion forums, and Short Messaging System (SMS) often use a non-standard romanization called “Arabizi” (Darwish 2013; Bies et al. 2014). For example, the Arabic sentence: رَاني فرحانة, which means “I am happy,” is written in Arabizi as “raani fer7ana”. Hence, Arabizi is an Arabic text written using Latin characters, numerals, and some punctuation (Darwish 2013). For handling Arabizi, almost all the works proposed a transliteration step converting Arabizi to Arabic. Guellil et al. (2017a, b, 2018c). In this paper, in addition to the three traditional varieties of Arabic, we also focused on the works targeting SA on Arabizi.

Different surveys have discussed the characteristics of Arabic language classifying and organizing the body of knowledge devoted to this subject (Habash 2010; Farghaly and Shaalan 2009; Shoufan and Alameri 2015). In the most recent survey in the field of Arabic natural language processing, presented in Shoufan and Alameri (2015), the authors propose a general classification of achievements for Arabic processing, which can be grouped into four categories: (1) Basic language analysis, (2) The building of lexical resources, (3) Dialect identification and recognition, and, (4) Semantic-level analysis and synthesis. Fundamental language analysis concentrates on morphological (e.g., POS tagging), syntactical, and orthographic tasks. The resources built could be lexicons comprehending monolingual or multilingual resources devoted to translation or sentiment analysis tasks. These resources consider also annotated corpora comprehending monolingual and pairwise lists. The identification is made on textual and phonetic data. The tasks on which most work has been conducted in the semantic analysis is the automatic translation of the text.

We propose this survey to highlight the work on Arabic sentiment analysis discussing methods, tools, and resources. We classify the body of knowledge devoted to Arabic sentiment analysis in two main categories: (1) Surveys work, presenting state of the art on Arabic sentiment analysis and, (2) Contribution work, offering the most recent advances in Arabic sentiment analysis. For the contribution works, we focus on the constructed resources (sentiment lexicons and corpora), we also focus on the methods proposed for validating the created resources and on the new tendencies in the context of Arabic SA.

4 Survey work

Korayem et al. (2012) identified work conducted on subjectivity analysis, i.e., how to classify a given sentence into objective or subjective. Also, for sentiment analysis tasks, the authors presented a set of corpora and lexicons available

for the research community. Arabic sentiment analysis in Twitter is presented in Harrag (2014), Alhumoud et al. (2015). While within Harrag (2014), the authors are focused on messages that are written in Arabic, in Alhumoud et al. (2015), the authors pay attention to Arabic dialects. Harrag (2014) discussed two main tasks of sentiment analysis: classification of subjectivity and polarity analysis. They presented algorithms used in sentiment classification tasks such as SVM and NB. These authors also give some of the most used applications of sentiment analysis, including online advertising. They finished by comparing some systems in term of the used classification algorithm, dataset, and performance. Alhumoud et al. (2015) discussed nine works presenting the following four aspects: (1) Type of the used corpus, (2) Number of tweets used, (3) Pre-processing techniques used, and (4) Classifications techniques used (lexicon or corpus-based approach).

The authors of the two works (Assiri et al. 2015; Kaseb and Ahmed 2016) exposed analytic state-of-the-art on sentiment analysis of Arabic and its dialects. Both surveys presented the proportion of articles that have been done by using a given approach by considering a given dialect. Moreover, Assiri et al. (2015) presented a detailed analysis of methods and results obtained in Arabic sentiment analysis. The authors pay attention to Arabic sentiment analysis and its variants in Arabic dialects. These authors relied on different classification of the presented work depending on the analysis level (document, sentence or aspect level), the used approach, the used dialect and finally the social media where the data were extracted. They finished by summarising the results obtained by each presented study. However, Kaseb and Ahmed (2016) presented a summary of each work studied, including among the aspects covered in the study the level of sentiment analysis addressed by each revised work, the language studied (Arabic language, a specific dialect or a combination of both), the pre-processing conducted and the results obtained. Biltawi et al. (2016) summarised thirty-two works according to the classification approach used. The authors of this work pay attention to the studies that have been done in the Arabic language and its dialects. Among the most used dialects in Arabic sentiment analysis, they highlighted the Egyptian, Jordanian, and Lebanese. Among the most used methods, the authors highlighted some methods based on lexicons and those based on support vector machines (SVM).

El-Masri et al. (2017) presented the most recent work that has been done on Arabic sentiment analysis. They presented the general process of Arabic Sentiment analysis by discussing the most commons challenges in each step of the process. For example, they commented about the lack of corpora for the step related to data gathering. They also discussed the lack of Arabic sentiment lexicons. They also highlighted challenges on the use of Arabizi. Guellil et al.

(2019) presented the most recent survey on Arabic Natural Language Processing (ANLP). These authors focused on the 90 most recent works that have been done on Arabic and its dialects in different NLP fields. They also dedicated a part (focusing on ten articles) for SA. However, as these surveys focused on all areas of NLP, the little works on SA were not presented in details.

Survey papers agree that the construction of lexical resources in Arabic is a significant problem. The lack of annotated data for supervised learning in Arabic sentiment analysis is also a relevant issue. In this survey, we first focus on the resource creation process (i.e., Arabic lexicon and corpus resources) and on recent works that make use of these resources. Then we also present the most recent advances and approaches that have been proposed for Arabic sentiment analysis including deep learning methods, an exciting research trend in Arabic sentiment analysis that has gained attention during the last years.

5 Constructed resources

5.1 Sentiment lexicon construction

Three trends are related to the construction of Arabic sentiment lexicons: (1) Manual construction, (2) Automatic construction, and (3) Semi-automatic construction. Only a few lexicons have been built manually (Abdul-Mageed and Diab 2012a; Mataoui et al. 2016; Mulki et al. 2019; Touahri and Mazroui 2019). Abdul-Mageed and Diab (2012a) describe the creation process of SIFAAT, a lexicon created manually comprising 3325 Arabic adjectives labeled with positive, negative, or neutral tags. The adjectives in SIFAAT belong to the domain of the news, and they were extracted from the first four parts of the Penn Arabic Treebank (Maamouri et al. 2004). In Mataoui et al. (2016), the authors focused on the problem of Algerian dialect by building three lexicons of sentiment words: (1) list of sentiment keywords, (2) list of sentiment words of denial, and (3) list of intensifying sentiment words. All these lexicons were manually built by using existing MSA and Egyptian resources. The translation from MSA and Egyptian to Algerian was done manually. The resulting lexicons contained 3093 words, of which 2380 were labeled as positive and the others 713 words were labeled as negative. In addition, to other used lexicons such as: NileULex (El-Beltagy 2016b), Mulki et al. (2019) constructed three others lexicons (LevLex for Levantine, GulfLex for Gulf and TunLex for Tunisian dialect) manually. LevLex contains 817 words, GulfLex comprises 100 words, and TunLex contains 5282 words. Touahri and Mazroui (2019) present HAPP, a manually constructed lexicon is containing 357 terms (where 221 are positives, and 136

are negatives). For building this lexicon, the authors relayed on a resource that treats happiness theme.³

Most of the lexicons introduced in the literature were built automatically. For the automatic creation of lexicons in Arabic three main trends have been dominant: (1) Lexicon construction based on automatic translation (Mohammad et al. 2016a, b; Salameh et al. 2015; Abdul-Mageed and Diab 2012a; Abdulla et al. 2014a; Guellil et al. 2017c); (2) Lexicon construction based on cross-lingual resources (Badaro et al. 2014; Gilbert et al. 2018; Eskander and Rambow 2015; Altrabsheh et al. 2017) and (3) Lexicon construction based on both automatic translation and cross-lingual resources. Mahyoub et al. (2014), Abdulla et al. (2014b), Abdulla (2013). The main idea behind the construction of lexicons based on automatic translation is to start with a sentiment lexicon in English (for example, the Bing Liu lexicon Ding et al. 2008, SentiWordnet Esuli and Sebastiani 2007 or SentiStrength Thelwall et al. 2010) and then translate it using an automated translator service as Google translate. Some translations are done using cross-lingual resources as Arabic/English dictionaries (Abdulla et al. 2014b). The main idea behind the use of cross-lingual resources is to combine different existing English/Arabic resources such as SentiWordnet, Arabic WordNet (Fellbaum et al. 2006) and Arabic Morphological Analyser (Graff et al. 2006; Buckwalter 2004) achieving better coverage on the Arabic language. To achieve the above, a reduced number of English sentiment words are translated into Arabic. Then these words are used as seeds for label propagation in Arabic, supporting the lexicon expansion task using semantic networks like the Arabic Wordnet.

The literature shows some works based on semi-automatic approaches (El-Beltagy 2016b; Abdul-Mageed and Diab 2016; Guellil et al. 2018a; Elshakankery and Ahmed 2019; Touahri and Mazroui 2019). El-Beltagy (2016b) presents NileULex, an Arabic sentiment lexicon composed of 45% of words in Egyptian dialect and 55% of words in MSA. The NileULex development began in the year 2013 in which the first version was introduced (El-Beltagy and Ali 2013). Later versions of NileULex were released including manually added words (ElSahar and El-Beltagy 2014) and re-validations (El-Beltagy 2016b). Re-validations were carried out to ensure that each entry in the lexicon was of high quality, limiting the effect of semantic ambiguity. The resulted lexicon contains 5953 distinct terms. Abdul-Mageed and Diab presented SANA (2016), a large-scale cross-lingual lexical resource for subjectivity and sentiment analysis of Arabic and its dialects. SANA comprises four communication contexts: chats, online news, YouTube comments, and tweets. In addition to MSA, SANA also covers both Egyptian and

³ <https://mawdoo3.com/>.

Levantine dialects along with short English descriptions. A significant fraction of SANA entries are augmented with part-of-speech tags, diacritics, gender, and other fine-grain word descriptors. SANA is developed both manually and automatically, and it comprises 224,564 entries.

Guellil et al. (2017c) presented SOCALALG, an Algerian dialect lexicon constructed automatically based on SOCAL English lexicon (Taboada et al. 2011). For constructing SOCALALG, the authors used Glosbe API⁴ for translating each entry of SOCAL. After the automatic translation, the same score was assigned to all the translated words. This score corresponds to the score of the English word from which they are translated. For example, all the translations of the English word 'excellent' with a score of +5, such as باهي (bAhy), لطيف (lTyf), and 'مليح' (mlyH), were assigned a score of +5. Six thousand seven hundred sixty-nine terms were obtained, including negative sentiment terms (labels ranging between - 1 and - 5) and positive terms (labels ranging between + 1 and + 5). The resulted DALG sentiment lexicon contains 2384 entries. This lexicon was manually reviewed in the work (Guellil et al. 2018a). The current version of SOCALALG lexicon contains 1745 entries where 968 are negative, 6 are neutral, and 771 are positives. Elshankery and Ahmed (2019) relied on Arabic sentiment corpus for automatically constructing the lexicon. Firstly, they constructed HILATSA, a lexicon based on many annotated sentiment corpus such as: ASTD (Nabil et al. 2015), ArTwitter (Mourad and Darwish 2013), etc. Afterward, the authors proposed a semi-automatic learning system that is capable of updating the lexicon to be up to date with language changes. In addition to HAPP, Touahri and Mazroui (2019) proposed ENG-AR, which was first constructed by automatically translate the Bing Liu lexicon (Ding et al. 2008) and then manually reviewed. The resulted ENG-AR lexicon contains 3504 words where 1278 representing positives words and 2226 representing negative ones.

Table 1 summarises the Arabic sentiment lexicons discussed in this section. Each lexicon is associated with the work presenting it, the number of its entries, the studies that have used the thesaurus, and if available the link to the resource.

5.2 Sentiment corpus construction

Three tendencies were observed in the creation of sentiment corpus in Arabic: (1) Manual construction, (2) Automatic construction, and (3) Semi-automatic construction.

A significant fraction of the corpus were manually built (Rushdi-Saleh et al. 2011; Abdul-Mageed and Diab 2012b; Medhaffar et al. 2017; Nabil et al. 2015;

Abdul-Mageed et al. 2014; Mourad and Darwish 2013; Rahab et al. 2019; Alahmary et al. 2019; Alharbi and Khan 2019). In most cases, the annotation was made by native annotators. Rushdi-Saleh et al. (2011) presented OCA (Opinion Corpus for Arabic), which contains 500 movie reviews collected from different Arabic blogs and web pages, comprising 250 positive reviews and 250 negative reviews. The reviews were processed manually. Abdul-Mageed and Diab (2012b) introduced AWATIF, a multi-genre corpus that comprises 10,723 Arabic sentences retrieved from three sources namely: the Penn Arabic Treebank (ATB) (Maamouri et al. 2004), a selection of web forums, and a list of Wikipedia talk pages. The sentences were manually annotated as objective or subjective, and the subjective sentences were annotated as positive or negative. Medhaffar et al. (2017) presented TSAC (Tunisian Sentiment Analysis Corpus). TSAC contains 17,060 Tunisian Facebook posts. These posts were manually annotated, including 8215 positive statements and 8845 negative statements. TSAC was collected from official pages of Tunisian radios and TV channels. Nabil et al. (2015) constructed ASTD (Arabic Sentiment Tweets Dataset). This corpus contains 10,000 Arabic tweets that were annotated using Amazon Mechanical Turk. The tweets were labeled as objective, subjective positive, subjective negative, or subjective mixed.

A multi-corpus was presented by Abdul-Mageed et al. (2014) including DARDASHA, a collection of 2798 chat messages retrieved from Maktoob⁵, TAGREED, a group of 3015 Arabic tweets, TAHRIR, a dataset composed by 3008 sentences from Wikipedia talk pages, and MONTADA, a dataset produced by 3097 sentences retrieved from web forums. This corpus was manually annotated by native speakers as positive or negative. Mourad and Darwish (2013) introduced a corpus for Arabic sentiment analysis that contains 2300 manually annotated tweets. A significant fraction of the corpus was annotated as positive or negative sentences by native annotators. ArTwitter (Abdulla 2013) is a manual annotated (by two human experts) sentiment corpus. This corpus contains 2000 labeled tweets (1000 positive tweets and 1000 negative ones) on various topics such as politics and arts, etc. After constructing and annotating the dataset, several treatments have applied on it such as: removing repetitions, stop words, and normalization, among others.

Rahab et al. (2019) presented SANA, a manually annotated corpus dealing with MSA and Algerian dialect. This corpus contains 513 comments extracted from a newspaper where 236 are positive, 194 are negative, and 83 are neutral. Two native speakers annotated it. Alahmary et al. (2019) manually annotated SDCT, a corpus containing

⁴ <https://glosbe.com/en/arq/excellent>.

⁵ chat.mymaktoob.com.

Table 1 Arabic sentiment lexicons

Name	Approach	Size	Used by	Resource link
SIFFAT Abdul-Mageed and Diab (2012a)	Manual	229,452	Abdul-Mageed and Diab (2016), Badaro et al. (2014)	–
Vernacular Arg. Mataoui et al. (2016)	Manual	3093	–	–
LevLex	Manual	817	–	–
GulfLex		100		
TunLex Mulki et al. (2019)		5282		
Arabic emoticon lexicon	Automatic	43,304	Htait et al. (2017), Kumar et al. (2016), Tellez et al. (2017)	All lexicons ^a
Arabic hashtag lexicon		21,964		
Arabic hashtag lexicon (dialect)		20,128		
NRC emoticon lexicon		26,740		
NRC hashtag lexicon Mohammad et al. (2016a, b), Salameh et al. (2015)		32,582		
Abdulla et al. lexicon (2014a)	Automatic	16,800	Abdul-Mageed and Diab (2016), Badaro et al. (2014)	–
ArSenL Badaro et al. (2014)	Automatic	33,995	Badaro et al. (2015), Al-Sallab et al. (2017), Farra and McKeown (2017), Eskander and Rambow (2015)	ArSenL ^b
SLSA Eskander and Rambow (2015)	Automatic	34,821	–	–
Altrabsheh et al. lexicon Altrabsheh et al. (2017)	Automatic	249,532	–	–
Mahyoub et al. lexicon Mahyoub et al. (2014)	Automatic	7400	–	–
Abdulla et al. lexicon (2014b)	Automatic	4815	–	–
ArSEL Gilbert et al. (2018)	Automatic	32,196	–	ArSEL ^d
SentiALG	Automatic	3408	Guellil et al. (2018a, b)	–
SOCALALG Guellil et al. (2017c)		2375		
NileULex El-Beltagy (2016b)	Semi-automatic	5953	El-Beltagy (2016a, 2017), El-Kilany et al. (2018)	NileULex ^c
SANA Abdul-Mageed and Diab (2016)	Semi-automatic	224,564	–	–
HILATSA Elshakankery and Ahmed (2019)	Semi-automatic	–	–	–
HAPP	Manual	357	–	–
ENG-AR Touahri and Mazroui (2019)	Semi-automatic	3504	–	–

^a<http://saifmohammad.com/WebPages/lexicons.html>

^b<http://www.oma-project.com/>

^c<https://github.com/NileTMRG/NileULex>

^d<http://oma-project.azurewebsites.net/>

32,063 written in Saudi dialect. The authors applied data annotation using crowdsourcing to classify the tweets into two classes consisting of 17,707 positive and 14,356 negative tweets. Afterward, the authors proposed a set of preprocessing steps such as: remove non-Arabic letters, normalization, removing repeated letters, etc., for cleaning their corpus. Alharbi and Khan (2019) proposed the manual construction of a corpus dedicated to distinguishing between comparative and non-comparative opinion. Three native Arabic speakers performed the annotation process. Two labelers annotated the sentences, and the third labeler made decisions about sentences that raised a conflict between the first and second labelers. The corpus contains 43% of comparative text and approximately 57% of non-comparative.

Some works have been conducted using automatic techniques for the construction of corpus. Two main techniques have been used for this purpose: (1) Automatic annotation based on rating reviews (Aly and Atiya 2013; ElSahar and El-Beltagy 2015), and (2) Automatic annotation based on sentiment lexicon or a set of expression (Guellil et al. 2018b; Gamal et al. 2019). In the context of annotations based on rating reviews, Aly and Atiya (2013) presented LABR (Large-scale Arabic Book Reviews), a collection of book reviews that contain 63,257 book reviews, each of them on a scale that ranges from 1 to 5 stars. Reviews with 4 or 5 stars were annotated as positive while reviews with 1 or 2 stars were annotated as negative. Reviews with three stars were annotated as neutral. ElSahar and El-Beltagy (2015) followed the same procedure proposed by

Table 2 Arabic sentiment corpus

Name	Approach	Size	Used by	Resource link
OCA Rushdi-Saleh et al. (2011)	Manual	500	Mahyoub et al. (2014), Rushdi-Saleh et al. (2011), Atia and Shaalan (2015), Rahab et al. (2017)	OCA ^a
AWATIF Abdul-Mageed and Diab (2012b)	Manual	10,723	–	–
TSAC Medhaffar et al. (2017)	Manual	17,060	–	TSAC ^b
ASTD Nabil et al. (2015)	Manual	10,000	Dahou et al. (2016), Vo and Zhang (2016)	ASTD + python code ^c
DARDASHA	Manual	2798	–	–
TAGREED		3015		
TAHRIR		3008		
MONTADA Abdul-Mageed et al. (2014)		3097		
Mourad and Darwish corpus Mourad and Darwish (2013)	Manual	2300	–	–
ArTwitter Abdulla (2013)	Manual	2000	Altowayan and Tao (2016), Alayba et al. (2018a)	ArTwitter ^f
SANA Rahab et al. (2019)	Manual	513	–	SANA ^g
SDCT Alahmary et al. (2019)	Manual	32,063	–	–
Comparative-corpus Alharbi and Khan (2019)	Manual	–	–	–
LABR Aly and Atiya (2013)	Automatic	63,257	Al Shboul et al. (2015), Dahou et al. (2016), Altowayan and Tao (2016)	LABR + python code ^d
ATT	Automatic	2154	Dahou et al. (2016)	All corpus + code ^e
HTL		15,572		
MOV		1524		
PROD		4272		
RES1		8364		
RES2		2642		
RES ElSahar and El-Beltagy (2015)		10,970		
SentiAlg Guellil et al. (2018b)	Automatic	8000	–	–
Twitter-Benchmark Gamal et al. (2019)	Automatic	151,548	–	–
Arasenti-tweet Al-Twairash et al. (2017)	Semi-automatic	17,573	–	–

^a<http://sinai.ujaen.es/oca-corpus-en/>^b<https://github.com/fbougares/TSAC>^c<https://github.com/mahmoudnabil/ASTD>^d<https://github.com/mohamedadaly/LABR>^e<https://github.com/hadyelsahar/large-arabic-sentiment-analysis-resources>^f<https://archive.ics.uci.edu/ml/datasets/>^g<http://rahab.e-monsite.com/medias/files/corpus.rar>

Aly and Atiya (2013) to annotate 7 collections of reviews named ATT, HTL, MOV, PROD, RES1, RES2, and RES. ATT and HTL are two datasets of tourist place reviews and hotel reviews extracted from TripAdvisor.com. Both datasets comprise 2154 reviews and 15,572 reviews, respectively. MOV is a dataset of movie reviews extracted from elcinema.com that contains 1524 reviews. PROD is a dataset of product reviews extracted from souq.com that contains 4,272 reviews. RES1 is a dataset of restaurant reviews extracted from qaym.com that contains 8364 reviews. RES2 is a dataset of restaurant reviews extracted from tripadvisor.com that contains 2642 reviews. Finally,

RES is a combination of RES1 and RES2 that contains 10,970 reviews.

Guellil et al. (2018b) based on their created sentiment lexicon (Guellil et al. 2017c), in Algerian to write down a large collection of messages. The corpus comprises messages written in Arabic and Arabizi, and it contains 8000 messages of which 4000 were written in Arabic and 4000 were written in Arabizi. Gamal et al. (2019) constructed a large sentiment corpus (containing 151,548 tweets where both positives and negative classes contain 75,774 tweets) dedicated to MSA and Egyptian dialect as well. For building their corpus, the authors first relied on the work of Aly and

Atiya (2013) for extracting and manually annotated 4404 phrases commonly used for expressing sentiment. AraSenTi-Tweet (Al-Twairsh et al. 2017) is a corpus created using a semi-automatic approach. This corpus contains 17,573 Saudi tweets that were manually annotated into four classes as positive, negative, neutral, or mixed tweets. To create this corpus, the authors recovered a collection of tweets using a list of sentiment words. After cleaning the recovered tweets, three native Arabic annotators reviewed the constructed corpus to validate the post tags inferred from the list of sentiment words.

Table 2 summarises the work discussed in this section.

6 Arabic sentiment analysis methods

6.1 Lexicon-based approach

Mataoui et al. (2016) proposed an approach based on four modules which are: (1) Common phrases similarity computation module, allowing to deal with common expressions before the word level handling. (2) Pre-processing module, principally based on a parser proposed by the authors for extracting the tokens (i.e., keywords, intensification words, negation words, and emoticons). (3) Language detection and stemming module, for distinguishing between the words which are written in Arabic and those written in Arabizi or other languages. The words written in Arabizi are transliterated, and those in other languages are translated using Google translate, and the words written in Arabic are stemmed using Khoja stemmer (Khoja and Garside 1999). (4) Polarity computation module for calculating the Semantic Orientation (SO) of each message by using the weights of the differently constructed lexicon (L1, L2, and L3). The authors applied their approach to 7698 comments extracted from Facebook and achieved accuracy up to 79.13%. The lexicons constructed by Mulki et al. (2019) was combined with Named Entity Recognition (NER) system. To show the importance of including NER in the process of SA, the authors carried out many experiments on different corpus presented in the literature such as TSAC (Medhaffar et al. 2017) by using and without using NER. The best results were obtained by using NER, and they are up to 82.8% (*F1*-score). The authors also investigated a corpus-based approach by using NER, but they conclude that NER were inefficient in the case of a corpus-based approach in contrast to a lexicon-based approach.

Mohammad et al. (2016a) applied their constructed lexicon in an Arabic sentence-level sentiment analysis system, which was built by reconstructing the NRC-Canada English system presented in Mohammad et al. (2013). The authors used a linear-kernel SVM (Chang and Lin 2011) classifier on the available training data. The results showed that the

combination of the differently constructed lexicons gives the best *F1*-score, which is up to 66.6%. The main idea presented by Salameh et al. (2015), Mohammad et al. (2016b) is to measure the effect of automatic translation on sentiment analysis. For calculating sentiment score, the authors used pointwise mutual information (PMI) between the term and the positive and negative categories. In these two works, the authors also used an Arabic sentence-level sentiment analysis system by reconstructing the NRC-Canada English system. The results obtained by these authors are up to 78.80%. The authors conclude that even with a non-perfect translation, SA system was able to compute sentiment scores. Abdulla et al. (2014a) proposed SA tool relying on different features for handling negation and intensification. The presented accuracy is up to 74.6%.

Badaro et al. (2014) relied on a nonlinear SVM, with the radial basis function (RBF) kernel, to evaluate different lexicons. The SVM classifier was trained using sentence vectors consisting of three numerical features that reflect the sentiments expressed in each sentence (i.e., positive, negative, or objective). The value of each feature is calculated by matching the lemmas in each sentence to each of the lexicons. The results showed that the union of the two techniques of lexicon construction gives better results by achieving an *F1*-score up to 64.5%. To validate their lexicon, Eskander and Rambow (2015) compared its results to those obtained by using ArSenL. They showed that SLSA results outperform the results obtained by using ArSenL where general *F1*-score for SLSA was up to 68.6%. The lexicon created by Mahyoub et al. (2014) was evaluated by incorporating it into a machine learning-based classifier. The experiments were conducted on several Arabic sentiment corpora, and the authors were able to achieve a 96% classification accuracy. Abdulla et al. (2014b) presented an algorithm looking for the polarity of each term of the message in the constructed lexicon. For handling Arabic, they first pretreated the messages. The pretreatment includes tokenization, repeat letters removal, normalization, light stemming, and stop word removing. The proposed algorithm is also able to handle both negation and intensification. The authors experimented their lexicon-based approach on two corpora (respectively extracted from Twitter and Yahoo!-Maktoob), each one containing 2000 messages (equally balanced between positives and negatives messages). The best accuracy was obtained on the Twitter corpus (up to 70.05%) where it was up to 63.76% for Yahoo!-Maktoob.

El-Beltagy (2016b) utilized a simple sentiment analysis task relies on two datasets (a set of tweets that were manually annotated), to illustrate that developed lexicon is useful. In all their experiments, dataset messages are represented using the bag of words model, with uni-gram and bi-gram TF-IDF weights. The weka workbench (Frank et al. 2009) was used for all tests. As a classifier, the authors used Naive Bayes

by combining the constructed lexicon (NileULex). The results showed that the integration of NileULex improved the results of classification where *F1*-score was up to 79%. Guellil et al. (2017c, 2018a) proposed a SA algorithm, to calculate the SO of a given message written in Algerian Arabic dialect. The algorithm includes four main steps: (1) Opposition, which is generally expressed in DALG with the keyword '<b.s.h>' (bSH—but). (2) Multi-word expressions (because the constructed lexicon contains multi-word entries). (3) Handling Arabic morphology by employing a simple rule-based light stemmer that handles Arabic prefixes and suffixes and (4) Negation handling by reversing the polarity. Negation in some Arabic dialect is usually expressed as an attached prefix, suffix, or a combination of both. To score a message, the sentiment scores are averaged of all words. The authors evaluated their algorithm on two datasets annotated manually (Facebook-corpus containing 426 posts and PADIC-corpus (Meftouh et al. 2015) containing 323 messages). The presented results are up to 64% for Facebook-corpus and up to 78% for PADIC-corpus.

6.2 Corpus-based approach

Rushdi-Saleh et al. (2011) have used cross-validation to compare the performance of two of the most widely used learning algorithms: SVM and NB. In their experiments, the tenfold cross-validation ($k = 10$) has been used to evaluate the classifiers. The authors used the Rapid Miner software⁶ with its text mining plug-in, which contains different tools designed to assist in the preparation of text documents for mining tasks (tokenization, stop word removal, and stemming, among others). Better accuracy was achieved with SVM classifier, and it is up to 91%. For validating their corpus (TSAC), Medhaffar et al. (2017) used three machine learning methods such as (SVM) and Naive Bayes (NB) and MultiLayer-Perceptron (MLP) classifier. All the experiments were conducted in Python using Scikit Learn for classification and gensim for learning vector representation. For extracting their features, the authors used Doc2vec toolkit (Le and Mikolov 2014), and messages were classified into positive or negative. MLP outperforms the other classifiers where both precision and recall, presented by these authors, were up to 78%. For validating their corpus (ASTD), Nabil et al. (2015) applied many classifiers such as SVM, Binomial NB(BNB), Logistic Regression (LR), etc. on both balanced and unbalanced dataset. The authors used TF-IDF and n -gram for extracting features. The best accuracy that they obtained was up to 69.1%, and it was on the unbalanced corpus, using TF-IDF and SVM classifiers.

Abdul-Mageed et al. (2014) developed a system called "SAMAR" that jointly classifies subjectivity of a text as well as its sentiments. This system is based on MSA and its dialects (mainly Egyptian dialect). In this work, the authors used an SVM classifier and achieved accuracy up to 71.3% for SA. Mourad and Darwish (2013) focused on both subjectivity and sentiment classification. The authors used tenfold cross-validation with 90/10 training/test splits. They used NB classifier for all their experiments. The authors employed stemming and POS tagging, leading to results outperforming the state-of-the-art results for MSA news subjectivity classification. The best results obtained by the authors were up to 71% for subjectivity classification and up to 66.4% for sentiment classification. These authors conclude by insisting on the necessity of integrating methods (bootstrapping or using machine translation) to increase the size of the training set automatically. For validating their lexicon, Abdulla (2013) proposed a SA algorithm for calculating SO by considering negation and intensification. For verifying this lexicon, they relied on ArTwitter corpus that they also annotated manually. For showing the importance of the size of the glossary on the results, they carried out different experiments where a more critical part of a lexicon is used in each one. The best results were obtained with the totality of glossary, and they are up to 59.6% (accuracy). However, the authors also showed that the results obtained using a corpus-based approach (relying on SVM and K -fold validation) outperforms the lexicon-based method with an accuracy of up to 87.2%.

For evaluating the constructed corpus, Rahab et al. (2019) carried out many experiments for showing the impact of word weighting approach, classification method, and of the light stemming. They used many word weighting algorithms such as Term frequency, Term occurrence, Term frequency, and Inverse document frequency (TF-IDF), and binary term occurrence BTO. The authors also relied on many classifiers such as SVM, NB, and k -nearest neighbors (KNN), and they opted for tenfold cross-validation. The best results were obtained using TF with NB classifier, and by relying on light stemming are they are up to 75% (accuracy). For evaluating their corpus, Alahmary et al. (2019) employed Word2vec model for learning vector representations of the words in an unsupervised way. Afterward, they applied two deep learning models which are long short-term memory (LSTM) and bidirectional long short-term memory (Bi-LSTM). They also compared the deep learning results to the results obtained from SVM, which is a well-known machine learning algorithm. LSTM classifier outperforms the other classifiers with an accuracy of up to 92%.

For validating the proposed corpus, Aly and Atiya (2013) firstly divided it to training and testing dataset. To avoid the bias of having more positive than negative reviews, the authors explored two settings, balanced (where the number

⁶ <https://rapidminer.com/>.

of reports from every class is the same) and unbalanced (where the number of reviews from each class is unrestricted). They used different features such as unigrams, bigrams, trigrams with/without TF–IDF weighting. For classifiers, they used Multinomial NB, Bernoulli NB, and SVM. The best $F1$ -score (up to 0.91) was achieved with the unbalanced dataset, with the combination of unigrams, bigrams, and trigrams, by using TF–IDF and with SVM classifiers. ElSahar and El-Beltagy (2015) explored the problem of sentiment classification as a two-class classification problem (positive or negative) and a three-class classification problem (positive, negative, and mixed). They used fivefold cross-validation. For extracting features TF–IDF and Delta TF–IDF (a derivative of TF–IDF in which each n -gram is assigned a weight equal to the difference of that n -gram's TF–IDF scores). They used different classifiers such as Linear SVM, Bernoulli NB, Logistic regression, stochastic gradient descent, and K -nearest-neighbour. Linear SVM outperforms the other classifiers with an accuracy up to 0.824 for two-class classification and up to 0.599 for three class classification.

Guellil et al. (2018b) used both Bag Of Words (BOW) and Doc2vec techniques for extracting features. For classification, they relied on different algorithms such as SVM, NB, LR, etc. where LR outperforms other classifiers by achieving an $F1$ -score up to 78%. However, these authors handled both Arabic and Arabizi (but $F1$ -score related to Arabizi classification was up to 68%). To improve the results associated with Arabizi sentiment analysis, Guellil et al. (2018a) extended their approach by proposing a transliteration component converting Arabizi into Arabic before SA. For validating their corpus, Gamal et al. (2019) applied different machine learning algorithms such as SVM, NB, Ridge Regression (RR), etc. For extracting features, they used TF–IDF. They used tenfold cross-validation and obtained their better results by using RR (which are up to 99.9% for either $F1$ -score and accuracy metrics). However, the authors carried out only intrinsic experiments; then this results concerns a part of the constructed corpus.

For validating their corpus, Al-Twairesh et al. (2017) conducted several experiments for multi-class sentiment classification. For two-class classification, they used only the positive and negative tweets, for three-class classification they used the positive, negative, and neutral tweets, and for four-class classification they used all the classes. These authors relied on term frequency and TF–IDF for extracting features. For classification, they used SVM with a linear kernel. The best results were obtained with a binary classifier and using term frequency. For validating their corpus (Alharbi and Khan 2019) proposed three classification approaches: (1) Linguistic Approach where the authors used MADAMIRA (Pasha et al. 2014) for extracting comparative adjectives that they used for classifying the comments, (2)

Machine Learning Approach where the authors used various supervised learning classifier, including NB, JRip rule-based classifier and C4.5 decision tree, and (3) Keywords approach by filtering the comments that contained only comparison relations using the keyword strategy. To do that, the authors manually identified a list of 83 keywords and key phrases. The decision tree classifier C4.5 (J48 implementation) outperform the other classifiers with an accuracy of up to 90%.

7 New trends in Arabic sentiment analysis

Elshakankery and Ahmed (2019) relied on the proposed lexicon for constructed features vectors of each analyzed tweets. Each feature vector contains necessary information about the tweet, such as the percentage of subjective words, the rate of positives/negatives words, whether the tweet includes emotions or not, etc. These vectors were used for the classification task, where the authors investigated the performance of three classifiers: SVM, LR, and Recurrent Neural Network (RNN). For showing the importance of updating the lexicon, the authors carried out two experiments (including After and before the lexicon update). The best results were up to 85% (for accuracy), and they were obtained using SVM with the updated lexicon. In the same context, Touahri and Mazroui (2019) relayed on supervised learning for constructing their model to classify a set of reviews. The authors associated each report to a vector of features, and then, they used an SVM classifier to build a model of the labeled vectors. The authors focused on four families of characteristics, which are: (1) Word-based features, focusing on the number of positives and negatives words, (2) Stem-based features, (3) Lemma based features, and finally (4) Semantic class-based features, organizing the lexicon into small classes such as love, optimism, hate and pessimism. Afterward, the authors constructed different vectors based on these features. They applied their approach to the different corpora proposed by ElSahar and El-Beltagy (2015) such as HTL, PROD, etc. The best results were obtained by combining various features, and they are up to 93.84% (an accuracy) obtained on HTL corpus.

BOW representation is commonly used to model text. BOW representations as TF–IDF are widely used in information retrieval mainly due to its simplicity as well as its efficiency in document recovery tasks (Alowaidi et al. 2017). Despite its popularity, this approach has two significant drawbacks for sentiment analysis purposes: (1) Loss of word order in sentences (word interchangeability), and (2) Insufficient representation of semantic at word level (Barhoumi et al. 2017). Furthermore, to use a BOW representation in sentiment analysis, an appropriate word feature extraction process is required (Al-Azani and El-Alfy 2017; Barhoumi et al. 2017). To address the limitations discussed above,

word and document embedding have begun to be used as alternative representations of the previous ones (Al-Azani and El-Alfy 2017; El Mahdaouy et al. 2016; Barhoumi et al. 2017; Altowayan and Tao 2016). Al-Azani and El-Alfy (2017) and Altowayan and Tao (2016) used a large-scale Arabic corpora. To train a skip-gram word2vec model of text (Mikolov et al. 2013). This corpus includes Quran-text (751,291 words), Arabic editions of international news networks such as CNN (24 million words) and BBC (20 million words), news articles based on a local Arabic newspaper (watan-2004 with 106 million words) and finally a set of consumer reviews (40 million words). The authors used the generated word vectors to train different sentiment classifiers showing excellent results in performance. Doc2vec sentence modelling (Le and Mikolov 2014) was used by Barhoumi et al. (2017), for sentiment classification on the corpus LABR (Aly and Atiya 2013). The system proposed by these authors composed with two parts: preprocessing part (to handle the input, light stemming) and classification part (to predict the polarity of the input). The authors used two classifiers, logistic regression (LR) and an MLP. The input vector of the classifier is the embedding obtained by learning paragraph vector. This vector was a concatenation of the two learned vectors, one from distributed memory version (DM) and the other from a distributed bag of words release (DBOW). Each model has 400 dimensions. This system was tested on LABR corpus. However, the results obtained with this system are lower than those obtained in Aly and Atiya (2013). The authors affirm that the complexity of Arabic morphology requires more treatment than for other languages such as English, for improving the results. El Mahdaouy et al. (2016) also find results that allow affirming that the use of doc2vec improves the performance of classifiers in sentiment analysis. Recently another algorithm for document representation known as Fastext has emerged (Joulin et al. 2016). The performance of Fastext is often compared with the performance obtained by word2vec in classification tasks (Tafreshi and Diab 2018; Schmitt et al. 2018).

Recently, deep learning algorithms such as Convolutional Neural Networks (CNN) (LeCun et al. 1995), Long Short-Term Memory networks (LSTM) (Hochreiter and Schmidhuber 1997) or Bidirectional LSTM networks (Bi-LSTM) (Graves et al. 2006) have taken an important place in sentiment classification. In this context, Dahou et al. (2016) introduce a method based on word2vec for Arabic sentiment classification, which evaluates polarity from product reviews. A convolutional neural network (CNN) model was trained on top of a set of pre-trained Arabic word embeddings. The authors used a multi-layer architecture defined by Kim (2014) to address this task. They applied their approach to different corpus presented in the literature studying the performance of LABR, ASTD, ATT, HTL, and MOV, among others. More recently, Attia et al. (2018) presented

a language-independent model for multi-class sentiment classification using a simple multi-layer neural network architecture. This model contains five layers. The first layer is a randomly-initialized word embedding layer that turns words in sentences into a feature map. The second layer is a convolution neural network (CNN) layer that scans the feature map. The third layer is a Global max-pooling, which is applied to the output generated by CNN layer to take the maximum score of each pattern. The purpose of the pooling layer is to reduce the dimensionality of the CNN representations by down-sampling the output and to keep the maximum value. The obtained scores are then fed to a single feed-forward (fully-connected) layer with ReLU activation. Finally, the outcome of that layer goes through a Softmax layer that predicts the output classes. As the proposed model is language-agnostic, it can be applied to multiple languages without the need to use lexical resources as dictionaries or ontologies. The authors applied their model to sentiment classification in three different languages, which are English, German, and Arabic. In the case of the Arabic language, the authors used the ASTD corpus (Nabil et al. 2015) to validate the model.

The literature shows that some works have been devoted to the study of sentiment analysis in Arabizi (Duwairi et al. 2016) or both in Arabic and Arabizi (Guellil et al. 2018b; Medhaffar et al. 2017). Duwairi et al. (2016) conducts a process of transliteration between Arabic and Arabizi before proceeding to the sentiment classification step. However, their approach presents some limitations: (1) They rely on a lexical resource for a transliteration from Arabizi to Arabic, which cannot handle polysemy. (2) They construct a small-scale manually annotated corpus that contains 3026 transliterated messages. Guellil et al. (2018b) and Medhaffar et al. (2017) are focused on both Arabic and Arabizi analysis. In Guellil et al. (2018b) the authors built SentiAlg, an automatically generated sentiment corpus devoted to the Algerian dialect. In Medhaffar et al. (2017), the authors built TSAC, a manually constructed sentiment corpus dedicated to the Tunisian dialect. Both works (Guellil et al. 2018b; Medhaffar et al. 2017) used doc2vec for sentence modelling. Sentiment classification based on SentiAlg reached a result of 68% in terms of *F1* measure while the results achieved using TSAC reached a 78% in performance.

8 Synthesis and discussion

We summarise in Table 3 the set of works that we have reviewed in this article. Table 3 shows that we have organized and classified ten survey works, and 57 contribution works focused on Arabic sentiment analysis. Among the solution research works, 32 are devoted to the creation of resources for Arabic sentiment analysis, of which 17 are

Table 3 Summary of the works devoted to Arabic sentiment analysis reviewed in this article

Language	Type of work	Work
Arabic	Survey	Korayem et al. (2012), Harrag (2014), Alhumoud et al. (2015), Assiri et al. (2015), Kaseb and Ahmed (2016), Biltawi et al. (2016), El-Masri et al. (2017), Al-Ayyoub et al. (2018), Sal-loum et al. (2018), Guellil et al. (2019)
	Contribution work	
	Creation of resources	Lexicon
		Abdul-Mageed and Diab (2012a), Mataoui et al. (2016), Mulki et al. (2019), Mohammad et al. (2016a, b), Salameh et al. (2015), Abdulla et al. (2014a, b), Badaro et al. (2014), Eskander and Rambow (2015), Mahyoub et al. (2014), El-Beltagy (2016b), Abdul-Mageed and Diab (2016), Gilbert et al. (2017c, 2018), Elshakankery and Ahmed (2019), Touahri and Mazroui (2019)
		Corpus
		Rushdi-Saleh et al. (2011), Abdul-Mageed and Diab (2012b), Medhaffar et al. (2017), Nabil et al. (2015), Abdul-Mageed et al. (2014), Guellil et al. (2018b), Gamal et al. (2019), Rahab et al. (2019), Alahmary et al. (2019), Alharbi and Khan (2019), Aly and Atiya (2013), Mourad and Darwish (2013), ElSahar and El-Beltagy (2015), Abdulla (2013), Al-Twairesh et al. (2017)
	Use of resources	Lexicon-based work
		Abdul-Mageed and Diab (2016), Badaro et al. (2014), Htair et al. (2017), Kumar et al. (2016), Tellez et al. (2017), Abdul-Mageed and Diab (2016), Badaro et al. (2014), Badaro et al. (2015), Al-Sallab et al. (2017), Farra and McKeown (2017), Eskander and Rambow (2015), El-Beltagy (2016a, 2017), El-Kilany et al. (2018), Guellil et al. (2018a, b)
		Corpus-based work
		Mahyoub et al. (2014), Rushdi-Saleh et al. (2011), Atia and Shaalan (2015), Rahab et al. (2017), Dahou et al. (2016), Vo and Zhang (2016), Al Shboul et al. (2015), Altowayan and Tao (2016), Alayba et al. (2018a)

dedicated to the creation of lexicons and the other 15 to the building of corpus. On the other hand, 25 works are dedicated to the use of these resources in Arabic sentiment analysis, of which 16 are devoted to the use of lexicons while the other nine are devoted to the use of corpus.

Our principal concern behind this paper was to survey the most recent works that have been done in Arabic sentiment analysis. In this spirit, 39 works (70%) from those that we have reviewed in this article were published between 2015 and 2019 (hence, during the last four years). The other work was published between 2010 and 2015. Also, Table 3 shows that there are more works devoted to the creation of lexicons than to the creation of corpus. Furthermore, more works make use of lexicons than works that make use of corpus. This fact is because many lexicons were automatically built, and consequently, they are composed of more lexical entries being more suitable for machine learning tasks due to its large size.

It can be deduced from our survey that some Arabic dialects suffer from the lack of works handling them (e.g., the Maghrebi dialects). Only five works have been conducted on Algerian dialect, of which two (Mataoui et al. 2016; Guellil et al. 2017c) are dedicated to the use of lexicons and three (Guellil et al. 2018a, b; Rahab et al. 2019) to the use of corpus. The reduced number of works in Algerian sentiment

analysis is a consequence of a lack of works devoted to natural language processing on this dialect.

It can be inferred from our survey that Arabic sentiment analysis has many open issues; among them, the most salient challenges are:

- The manual construction of resources gives accurate results in sentiment analysis. However, the manual construction of resources drives to small-scale resources not suitable for machine learning tasks due to its small size.
- Almost all the reviewed resources are not domain-oriented.
- Almost all the reviewed approaches work at the document level or the sentence level. Aspect level sentiment analysis in Arabic appears as an unexplored research line.

There are some challenges inherent in the processing of the Arabic language. Many difficulties are related to its rich morphology. We highlight the following challenges related to specific aspects of Arabic language processing:

- The agglutination of Arabic words could lead to errors when classifying sentiments. For example, the word "مأكلاً" (meaning "food") is translated to English as "He did not eat" but in the sentence "مأكلاً مآشرب" the mean-

ing is “he did not eat, and he did not drink”. When "مَا كَلَّ" is used as a word, it could not be separated and then it is equivalent to exactly one word and one meaning. However, in the sentence, the word "مَا كَلَّ" could be separated as: "مَا" + "كَلَّ" (where "مَا" represents a negative pronouns). Hence both use cases of the same word have different polarities depending on the context where the word is used.

- Regarding lexicon-based approaches, stemming is a crucial task because in a great proportion of the lexicons we found stems. However, stemming is a hard task in Arabic as different stems can be considered correct for the same word.
- Regarding corpus-based approaches, polysemic words as "مَا كَلَّ" may cause ambiguities.
- Arabic people in social media can switch from Arabic to Arabizi in the same sentence. Thus, the same word could be written in many different manners. Cotterell et al. (2014) argued that the word "ان شاء الله" (meaning “if the god willing”) could be written in 69 different manners. Hence, sentiment analysis in Arabizi represents a major challenge in terms of homonymy.

9 Conclusion

The principal purpose of this survey was to present and organize recent works in Arabic sentiment analysis. We classified the work into two main classes: survey work and contribution work. For contribution work, we distinguished between articles proposing new resources and papers that make use of resources for Arabic sentiment analysis tasks. Resources were organized as lexicon or corpus.

We observed that a significant proportion of the related work was recently published during the last four years. Also, we observed that there are more works related to the creation and use of lexicons than works devoted to the study of corpora. We noted that the Arabic morphology represents a significant challenge for sentiment analysis. Also, sentiment analysis in Arabizi represents a challenge. Hence, to analyze the sentiment of messages written in Arabizi, a lexical pre-processing step is crucial. Almost all work handling Arabizi suggest the use of a transliteration step to handle it.

This survey opens the door to many research questions: (1) Is it better to rely on manual construction of sentiment resources (lexicon and corpus) or is the automatic one suitable for this kind of purposes? (2) Is it crucial to transliterate to deal with Arabic and Arabizi sentiment analysis? (3) Are there efficient techniques based on word embeddings for Arabic sentiment analysis? (4) Is deep learning the solution for Arabic sentiment classification and if so, which algorithms are better for this specific task?

Undoubtedly, the next years will show us more advances in these promising lines, and some of these open questions will be successfully addressed.

References

- Abdul-Mageed M, Diab M (2012a) Toward building a large-scale Arabic sentiment lexicon. In: Proceedings of the 6th international global WordNet conference, pp 18–22
- Abdul-Mageed M, Diab MT (2012b) Awatif: a multi-genre corpus for modern standard Arabic subjectivity and sentiment analysis. In: LREC. Citeseer, pp 3907–3914
- Abdul-Mageed M, Diab M, Kübler S (2014) Samar: subjectivity and sentiment analysis for Arabic social media. *Comput Speech Lang* 28(1):20–37
- Abdul-Mageed M, Diab MT (2016) Sana: a large scale multi-genre, multi-dialect lexicon for Arabic subjectivity and sentiment analysis. In: LREC, pp 1162–1169
- Abdulla NA, Ahmed NA, Shehab MA, Al-Ayyoub M (2013) Arabic sentiment analysis: lexicon-based and corpus-based. In: 2013 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT). IEEE, pp. 1–6
- Abdulla N, Mohammed S, Al-Ayyoub M, Al-Kabi M et al (2014a) Automatic lexicon construction for Arabic sentiment analysis. In: 2014 international conference on future internet of things and cloud (FiCloud). IEEE, pp 547–552
- Abdulla NA, Ahmed NA, Shehab MA, Al-Ayyoub M, Al-Kabi MN, Al-rifai S (2014b) Towards improving the lexicon-based approach for Arabic sentiment analysis. *Int J Inf Technol Web Eng (IJITWE)* 9(3):55–71
- Akaichi J (2013) Social networks’ facebook’ statutes updates mining for sentiment classification. In: 2013 international conference on social computing (SocialCom). IEEE, pp 886–891
- Al-Ayyoub M, Khamaiseh AA, Jararweh Y, Al-Kabi MN (2018) A comprehensive survey of Arabic sentiment analysis. In: *Information processing & management*, pp 320–342
- Al-Azani S, El-Alfy ESM (2017) Using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short Arabic text. *Procedia Comput Sci* 109:359–366
- Al-Sallab A, Baly R, Hajj H, Shaban KB, El-Hajj W, Badaro G (2017) Aroma: a recursive deep learning model for opinion mining in Arabic as a low resource language. *ACM Trans Asian Low Resour Lang Inf Process (TALLIP)* 16(4):25
- Al Shboul B, Al-Ayyoub M, Jararweh Y (2015) Multi-way sentiment classification of Arabic reviews. In: 2015 6th international conference on information and communication systems (ICICS). IEEE, pp 206–211
- Al-Twairesh N, Al-Khalifa H, Al-Salman A, Al-Ohali Y (2017) Ara-senti-tweet: a corpus for Arabic sentiment analysis of Saudi tweets. *Procedia Comput Sci* 117:63–72
- Alahmary RM, Al-Dossari HZ, Emam AZ (2019) Sentiment analysis of Saudi dialect using deep learning techniques. In: 2019 international conference on electronics, information, and communication (ICEIC). IEEE, pp 1–6
- Alayba AM, Palade V, England M, Iqbal R (2018a) A combined CNN and LSTM model for Arabic sentiment analysis. In: International cross-domain conference for machine learning and knowledge extraction. Springer, pp 179–191
- Alayba AM, Palade V, England M, Iqbal R (2018b) Improving sentiment analysis in Arabic using word representation. In: 2018 IEEE 2nd international workshop on Arabic and derived script analysis and recognition (ASAR). IEEE, pp 13–18

- Alharbi FR, Khan MB (2019) Identifying comparative opinions in Arabic text in social media using machine learning techniques. *SN Appl Sci* 1(3):213
- Alhumoud SO, Altuwajri MI, Albuhairei TM, Alohaideb WM (2015) Survey on Arabic sentiment analysis in Twitter. *Int Sci Index* 9(1):364–368
- Alowaidi S, Saleh M, Abulnaja O (2017) Semantic sentiment analysis of Arabic texts. *Int J Adv Comput Sci Appl* 8(2):256–262
- Altowayan AA, Tao L (2016) Word embeddings for Arabic sentiment analysis. In: 2016 IEEE international conference on big data (big data). IEEE, pp 3820–3825
- Altrabsheh N, El-Masri M, Mansour H (2017) Combining sentiment lexicons of Arabic terms. In: Association for information and systems (AMICS)
- Aly M, Atiya A (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), vol 2, pp 494–498
- Antai R (2014) Sentiment classification using summaries: a comparative investigation of lexical and statistical approaches. In: 2014 6th computer science and electronic engineering conference (CEEC). IEEE, pp 154–159
- Assiri A, Emam A, Aldossari H (2015) Arabic sentiment analysis: a survey. *Int J Adv Comput Sci Appl* 6(12):75–85
- Atia S, Shaalan K (2015) Increasing the accuracy of opinion mining in Arabic. In: 2015 first international conference on Arabic computational linguistics (ACLing). IEEE, pp 106–113
- Attia M, Samih Y, El-Kahky A, Kallmeyer L (2018) Multilingual multi-class sentiment classification using convolutional neural networks. In: Proceedings of the eleventh international conference on language resources and evaluation (LREC), pp 635–640
- Badaro G, Baly R, Hajj H, Habash N, El-Hajj W (2014) A large scale Arabic sentiment lexicon for Arabic opinion mining. In: Proceedings of the EMNLP 2014 workshop on Arabic natural language processing (ANLP), pp 165–173
- Badaro G, Baly R, Akel R, Fayad L, Khairallah J, Hajj H, Shaban K, El-Hajj W (2015) A light lexicon-based mobile application for sentiment mining of Arabic tweets. In: Proceedings of the second workshop on Arabic natural language processing, pp 18–25
- Bahrainian SA, Dengel A (2013) Sentiment analysis using sentiment features. In: 2013 IEEE/WIC/ACM international joint conferences on web intelligence (WI) and intelligent agent technologies (IAT), vol 3. IEEE, pp 26–29
- Barhouni A, Aloulou YEC, Belguith LH (2017) Document embeddings for Arabic sentiment analysis. In: Conference on language processing and knowledge management (LPKM)
- Bhuta S, Doshi A, Doshi U, Narvekar M (2014) A review of techniques for sentiment analysis of Twitter data. In: 2014 international conference on issues and challenges in intelligent computing techniques (ICICT). IEEE, pp 583–591
- Bies A, Song Z, Maamouri M, Grimes S, Lee H, Wright J, Strassel S, Habash N, Eskander R, Rambow O (2014) Transliteration of Arabizi into Arabic orthography: developing a parallel annotated Arabizi–Arabic script sms/chat corpus. In: Proceedings of the EMNLP 2014 workshop on Arabic natural language processing (ANLP), pp 93–103
- Biltawi M, Etaiwi W, Tedmori S, Hudaib A, Awajan A (2016) Sentiment classification techniques for Arabic language: a survey. In: 2016 7th international conference on information and communication systems (ICICS). IEEE, pp 339–346
- Boudad N, Faizi R, Thami ROH, Chiheb R (2017) Sentiment analysis in Arabic: a review of the literature. *Ain Shams Eng J* 9(4):2479–2490
- Buckwalter T (2004) Buckwalter Arabic morphological analyzer version 2.0. linguistic data consortium, University of Pennsylvania, 2002. Ldc catalog no.: Ldc2004102. Tech. rep. ISBN 1-58563-324-0
- Chang CC, Lin CJ (2011) Libsvm: a library for support vector machines. *ACM Trans Intell Syst Technol (TIST)* 2(3):27
- Colace F, De Santo M, Greco L (2013) A probabilistic approach to tweets' sentiment classification. In: 2013 Humaine association conference on affective computing and intelligent interaction (ACII). IEEE, pp 37–42
- Cotterell R, Renduchintala A, Saphra N, Callison-Burch C (2014) An Algerian Arabic–French code-switched corpus. In: Workshop on free/open-source Arabic corpora and corpora processing tools workshop programme, p 34
- Dahou A, Xiong S, Zhou J, Haddoud MH, Duan P (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, pp 2418–2427
- Darwish K (2013) Arabizi detection and conversion to Arabic. arXiv preprint [arXiv:1306.6755](https://arxiv.org/abs/1306.6755)
- Ding X, Liu B, Yu PS (2008) A holistic lexicon-based approach to opinion mining. In: Proceedings of the 2008 international conference on web search and data mining. ACM, pp 231–240
- Duwairi RM, Alfaqeh M, Wardat M, Alrabadi A (2016) Sentiment analysis for Arabizi text. In: 2016 7th international conference on information and communication systems (ICICS). IEEE, pp 127–132
- El-Beltagy SR (2016a) Niletmrg at semeval-2016 task 7: deriving prior polarities for Arabic sentiment terms. In: Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), pp 486–490
- El-Beltagy SR (2016b) Nileulex: a phrase and word level sentiment lexicon for Egyptian and modern standard Arabic. In: Proceedings of tenth international conference on language resources and evaluation (LREC), pp 2900–2905
- El-Beltagy SR (2017) Weightednileulex: a scored Arabic sentiment lexicon for improved sentiment analysis. language processing, pattern recognition and intelligent systems. Special issue on computational linguistics, speech & image processing for Arabic language. World Scientific Publishing Co, Singapore
- El-Beltagy SR, Ali A (2013) Open issues in the sentiment analysis of Arabic social media: a case study. In: 2013 9th international conference on innovations in information technology (IIT). IEEE, pp 215–220
- El-Kilany A, Azzam A, El-Beltagy SR (2018) Using deep neural networks for extracting sentiment targets in Arabic tweets. In: Intelligent natural language processing: trends and applications. Springer, pp 3–15
- El Mahdaouy A, Gaussier E, El Alaoui SO (2016) Arabic text classification based on word and document embeddings. In: International conference on advanced intelligent systems and informatics. Springer, pp 32–41
- El-Masri M, Altrabsheh N, Mansour H (2017) Successes and challenges of Arabic sentiment analysis research: a literature review. *Soc Netw Anal Min* 7(1):54
- ElSahar H, El-Beltagy SR (2014) A fully automated approach for Arabic slang lexicon extraction from microblogs. In: International conference on intelligent text processing and computational linguistics. Springer, pp 79–91
- ElSahar H, El-Beltagy SR (2015) Building large Arabic multi-domain resources for sentiment analysis. In: International conference on intelligent text processing and computational linguistics. Springer, pp 23–34
- Elshakankery K, Ahmed MF (2019) Hilatsa: a hybrid incremental learning approach for Arabic tweets sentiment analysis. *Egypt Inform J*. <https://doi.org/10.1016/j.eij.2019.03.002>

- Eskander R, Rambow O (2015) Slsa: a sentiment lexicon for standard Arabic. In: Proceedings of the 2015 conference on empirical methods in natural language processing, pp 2545–2550
- Esuli A, Sebastiani F (2007) Sentiwordnet: a high-coverage lexical resource for opinion mining. *Evaluation* 17:1–26
- Farghaly A, Shaaan K (2009) Arabic natural language processing: challenges and solutions. *ACM Trans Asian Lang Inf Process (TALIP)* 8(4):14
- Farra N, McKeown K (2017) Smarties: Sentiment models for Arabic target entities. arXiv preprint [arXiv:1701.03434](https://arxiv.org/abs/1701.03434)
- Fellbaum C, Alkhalifa M, Black W, Elkateb S, Pease A, Rodriguez H, Vossen P (2006) Introducing the Arabic wordnet project
- Frank E, Hall M, Holmes G, Kirkby R, Pfahringer B, Witten IH, Trigg L (2009) Weka—a machine learning workbench for data mining. In: Data mining and knowledge discovery handbook. Springer, Boston, pp 1269–1277
- Gamal D, Alfonse M, El-Horbaty ESM, Salem ABM (2019) Twitter benchmark dataset for Arabic sentiment analysis. *Int J Mod Educ Comput Sci* 11(1):33
- Gilbert B, Hussein J, Hazem H, Wassim EH, Nizar H (2018) Arsel: a large scale Arabic sentiment and emotion lexicon. In: Proceedings of the eleventh international conference on language resources and evaluation (LREC)
- Graff D, Buckwalter T, Jin H, Maamouri M (2006) Lexicon development for varieties of spoken colloquial Arabic. In: Proceedings of the 5th edition of the international conference on language resources and evaluation (LREC), pp 999–1004
- Graves A, Fernández S, Gomez F, Schmidhuber J (2006) Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In: Proceedings of the 23rd international conference on machine learning. ACM, pp 369–376
- Guellil I, Boukhalifa K (2015) Social big data mining: a survey focused on opinion mining and sentiments analysis. In: 2015 12th international symposium on programming and systems (ISPS). IEEE, pp 1–10
- Guellil I, Azouaou F, Abbas M (2017a) Comparison between neural and statistical translation after transliteration of Algerian Arabic dialect. In: WiNLP: women & underrepresented minorities in natural language processing (co-located with ACL 2017)
- Guellil I, Azouaou F, Abbas M, Fatiha S (2017b) Arabizi transliteration of Algerian Arabic dialect into modern standard Arabic. In: Social MT 2017: first workshop on social media and user generated content machine translation (co-located with EAMT 2017)
- Guellil I, Azouaou F, Saâdane H, Semmar N (2017c) Une approche fondée sur les lexiques d'analyse de sentiments du dialecte algérien
- Guellil I, Adeel A, Azouaou F, Benali F, Hachani AE, Hussain A (2018a) Arabizi sentiment analysis based on transliteration and automatic corpus annotation. In: Proceedings of the 9th workshop on computational approaches to subjectivity, sentiment and social media Analysis, pp 335–341
- Guellil I, Adeel A, Azouaou F, Hussain A (2018b) Sentialg: Automated corpus annotation for Algerian sentiment analysis. arXiv preprint [arXiv:1808.05079](https://arxiv.org/abs/1808.05079)
- Guellil I, Azouaou F, Benali F, Hachani AE, Saadane H (2018c) Approche hybride pour la translittération de l'arabizi algérien : une étude préliminaire. In: Conference: 25e conférence sur le Traitement Automatique des Langues Naturelles (TALN), Rennes, France, 14–18 May 2018, pp 509–517. https://www.researchgate.net/publication/326354578_Approche_Hybride_pour_la_translittération_de_l%27arabizi_algerien_une_etude_preliminaire
- Guellil I, Saâdane H, Azouaou F, Gueni B, Nouvel D (2019) Arabic natural language processing: an overview. *J King Saud Univ Comput Inf Sci*. <https://doi.org/10.1016/j.jksuci.2019.02.006>
- Habash NY (2010) Introduction to Arabic natural language processing. *Synth Lect Hum Lang Technol* 3(1):1–187
- Harrag F (2014) Estimating the sentiment of Arabic social media contents: a survey. In: 5th international conference on Arabic language processing
- Harrat S, Meftouh K, Smaïli K (2017) Machine translation for Arabic dialects (survey). *Inf Process Manag* 56(2):262–273
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Htaït A, Fournier S, Bellot P (2017) Lsis at semeval-2017 task 4: using adapted sentiment similarity seed words for English and Arabic tweet polarity classification. In: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017), pp 718–722
- Joulin A, Grave E, Bojanowski P, Mikolov T (2016) Bag of tricks for efficient text classification. arXiv preprint [arXiv:1607.01759](https://arxiv.org/abs/1607.01759)
- Karamibekr M, Ghorbani AA (2012) Sentiment analysis of social issues. In: 2012 international conference on social informatics (SocialInformatics). IEEE, pp 215–221
- Kaseb GS, Ahmed MF (2016) Arabic sentiment analysis approaches: an analytical survey. *Int J Sci Eng Res* 7(10):712–723
- Khan AZ, Atique M, Thakare V (2015) Combining lexicon-based and learning-based methods for Twitter sentiment analysis. *Int J Electron Commun Soft Comput Sci Eng (IJECSCE)* 89:89–91
- Khoja S, Garside R (1999) Stemming Arabic text. Lancaster, UK, Computing Department, Lancaster University
- Kim Y (2014) Convolutional neural networks for sentence classification. arXiv preprint [arXiv:1408.5882](https://arxiv.org/abs/1408.5882)
- Korayem M, Crandall D, Abdul-Mageed M (2012) Subjectivity and sentiment analysis of Arabic: a survey. In: International conference on advanced machine learning technologies and applications. Springer, pp 128–139
- Kumar A, Kohail S, Kumar A, Ekbal A, Biemann C (2016) Iit-tuda at semeval-2016 task 5: beyond sentiment lexicon: combining domain dependency and distributional semantics features for aspect based sentiment analysis. In: Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), pp 1129–1135
- Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: International conference on machine learning, pp 1188–1196
- LeCun Y, Bengio Y et al (1995) Convolutional networks for images, speech, and time series. In: The handbook of brain theory and neural networks, vol 3361
- Liang PW, Dai BR (2013) Opinion mining on social media data. In: 2013 IEEE 14th international conference on mobile data management (MDM), vol 2. IEEE, pp 91–96
- Liu B (2012) Sentiment analysis and opinion mining. *Synth Lect Hum Lang Technol* 5(1):1–167
- Maamouri M, Bies A, Buckwalter T, Mekki W (2004) The Penn Arabic Treebank: building a large-scale annotated Arabic corpus. In: NEMLAR conference on Arabic language resources and tools, vol 27. Cairo, pp 466–467
- Maas AL, Daly RE, Pham PT, Huang D, Ng AY, Potts C (2011) Learning word vectors for sentiment analysis. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies-volume 1. Association for Computational Linguistics, pp 142–150
- Mahyoub FH, Siddiqui MA, Dahab MY (2014) Building an Arabic sentiment lexicon using semi-supervised learning. *J King Saud Univ Comput Inf Sci* 26(4):417–424
- Mataoui M, Zelmati O, Boumechache M (2016) A proposed lexicon-based sentiment analysis approach for the vernacular Algerian Arabic. *Res Comput Sci* 110:55–70
- Medhaffar S, Bougares F, Esteve Y, Hadrich-Belguith L (2017) Sentiment analysis of Tunisian dialects: Linguistic resources and

- experiments. In: Proceedings of the third Arabic natural language processing workshop, pp 55–61
- Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng J* 5(4):1093–1113
- Meftouh K, Harrat S, Jamoussi S, Abbas M, Smaili K (2015) Machine translation experiments on padic: a parallel Arabic dialect corpus. In: The 29th Pacific Asia conference on language, information and computation
- Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, pp 3111–3119
- Mohammad SM, Kiritchenko S, Zhu X (2013) NRC-Canada: building the state-of-the-art in sentiment analysis of tweets. arXiv preprint [arXiv:1308.6242](https://arxiv.org/abs/1308.6242)
- Mohammad S, Salameh M, Kiritchenko S (2016a) Sentiment lexicons for Arabic social media. In: Proceedings of tenth international conference on language resources and evaluation (LREC), pp 33–37
- Mohammad SM, Salameh M, Kiritchenko S (2016b) How translation alters sentiment. *J Artif Intell Res* 55:95–130
- Mourad A, Darwish K (2013) Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs. In: Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis, pp 55–64
- Mulki H, Haddad H, Gridach M, Babaoglu I (2019) Empirical evaluation of leveraging named entities for Arabic sentiment analysis. arXiv preprint [arXiv:1904.10195](https://arxiv.org/abs/1904.10195)
- Nabil M, Aly M, Atiya A (2015) Astd: Arabic sentiment tweets dataset. In: Proceedings of the 2015 conference on empirical methods in natural language processing, pp 2515–2519
- Pasha A, Al-Badrashiny M, Diab MT, El Kholly A, Eskander R, Habash N, Pooleery M, Rambow O, Roth R (2014) Madamira: a fast, comprehensive tool for morphological analysis and disambiguation of Arabic. *LREC* 14:1094–1101
- Rahab H, Zitouni A, Djoudi M (2017) Siaac: sentiment polarity identification on Arabic Algerian newspaper comments. In: Proceedings of the computational methods in systems and software. Springer, pp 139–149
- Rahab H, Zitouni A, Djoudi M (2019) Sana: sentiment analysis on newspapers comments in Algeria. *J King Saud Univ Comput Inf Sci*. <https://doi.org/10.1016/j.jksuci.2019.04.012>
- Rushdi-Saleh M, Martín-Valdivia MT, Ureña-López LA, Perea-Ortega JM (2011) Bilingual experiments with an Arabic–English corpus for opinion mining. In: Proceedings of the international conference recent advances in natural language processing 2011, pp 740–745
- Rushdi-Saleh M, Martín-Valdivia MT, Ureña-López LA, Perea-Ortega JM (2011) Oca: opinion corpus for Arabic. *J Assoc Inf Sci Technol* 62(10):2045–2054
- Salameh M, Mohammad S, Kiritchenko S (2015) Sentiment after translation: a case-study on Arabic social media posts. In: Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: human language technologies, pp 767–777
- Salloum SA, AlHamad AQ, Al-Emran M, Shaalan K (2018) A survey of Arabic text mining. In: Intelligent natural language processing: trends and applications. Springer, pp 417–431
- Schmitt M, Steinheber S, Schreiber K, Roth B (2018) Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. arXiv preprint [arXiv:1808.09238](https://arxiv.org/abs/1808.09238)
- Sharaf ABM, Atwell E (2012a) Qurana: corpus of the Quran annotated with pronominal anaphora. In: Proceedings of eighth international conference on language resources and evaluation (LREC), pp 130–137
- Sharaf ABM, Atwell E (2012b) Qursim: A corpus for evaluation of relatedness in short texts. In: LREC, pp 2295–2302
- Shoufan A, Alameri S (2015) Natural language processing for dialectical Arabic: a survey. In: Proceedings of the second workshop on Arabic natural language processing, pp 36–48
- Singh V, Piryani R, Uddin A, Waila P (2013) Sentiment analysis of movie reviews: a new feature-based heuristic for aspect-level sentiment classification. In: 2013 international multi-conference on automation, computing, communication, control and compressed sensing (iMac4s). IEEE, pp 712–717
- Taboada M, Brooke J, Tofiloski M, Voll K, Stede M (2011) Lexicon-based methods for sentiment analysis. *Comput Linguist* 37(2):267–307
- Tafreshi S, Diab M (2018) Emotion detection and classification in a multigenre corpus with joint multi-task deep learning. In: Proceedings of the 27th international conference on computational linguistics, pp 2905–2913
- Tellez ES, Miranda-Jiménez S, Graff M, Moctezuma D, Suárez RR, Siordia OS (2017) A simple approach to multilingual polarity classification in Twitter. *Pattern Recognit Lett* 94:68–74
- Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A (2010) Sentiment strength detection in short informal text. *J Am Soc Inf Sci Technol* 61(12):2544–2558
- Touahri I, Mazroui A (2019) Studying the effect of characteristic vector alteration on Arabic sentiment classification. *J King Saud Univ Comput Inf Sci*. <https://doi.org/10.1016/j.jksuci.2019.04.011>
- Vo DT, Zhang Y (2016) Don't count, predict! an automatic approach to learning sentiment lexicons for short text. In: Proceedings of the 54th annual meeting of the association for computational linguistics (Volume 2: short papers), vol 2, pp 219–224
- Wan X (2012) A comparative study of cross-lingual sentiment classification. In: Proceedings of the The 2012 IEEE/WIC/ACM international joint conferences on web intelligence and intelligent agent technology, vol 01. IEEE Computer Society, pp 24–31
- Zhang L, Wang S, Liu B (2018) Deep learning for sentiment analysis: a survey. *WIREs Data Min Knowl Discov* 8(4):e1253. <https://doi.org/10.1002/widm.1253>
- Zia SS, Fatima S, Mala I, Sadiq Ali Khan M, Naseem M, Das B (2018) A survey on sentiment analysis, classification and applications. *Int J Pure Appl Math* 119(10):1203–1211

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