

# Sentiment Analysis on Arabic Tweets: Challenges to Dissecting the Language

Malak Abdullah<sup>(✉)</sup> and Mirsad Hadzikadic

College of Computing and Informatics, University of North Carolina at Charlotte,  
Charlotte, NC, USA  
{mabdul15,mirsad}@uncc.edu

**Abstract.** Sentiment analysis and opinion mining are designed to detect peoples' emotions and opinions. Many public and private sectors are interested in extracting information regarding opinions that consist of subjective expressions across a variety of products or political decisions. Recently, the Arab region has had a significant role in international politics and the global economy that has grasped the attention of political and social scientists. Detecting Arabic tweets will be helpful for politicians in predicting global events based on the popular news and people's comments. Still, the Arabic language has not received a proper attention from modern computational linguists. This literature review provides a comprehensive study of sentiment analysis on Twitter data and analyzes the existing work that has been done in order to detect and analyze Arabic tweets.

**Keywords:** Twitter · Sentiment analysis · Opinion mining · Natural language processing · Machine learning · Classification · Arabic language

## 1 Introduction

Sentiment analysis and opinion mining are considered hot topics where researchers are extracting information regarding emotions and viewpoints. It is believed that these concepts consist of subjective expressions across a variety of products or political decisions [6, 42]. The terms sentiment analysis and opinion mining are not exactly the same. The meaning of the term “opinion” is broader than the term “sentiment”. Prior researchers have used these two terms interchangeably. In this literature review, the term sentiment analysis has been used to refer to both of them. Sentiment analysis is used to track attitudes and opinions on the web and determines if the audience positively or negatively receives these ideas. This helps companies determine strategies for improving the quality of their products or to assist decision makers. Sentiment analysis of data involves building a system by using natural language processing, statistics, and machine learning methods to examine opinions or sentiments in a text unit [6].

Microblogging services, such as Twitter and Facebook, are considered important communication tools for people to share their opinions or spread information. The nature of these microblogs encourages people in their daily lives to post

real-time messages about their opinions on current events. People are sharing their daily life activities on these microblogging tools [30].

The Twitter microblog was launched in July 2006, and since then it has gained worldwide popularity. Many scholars hold the view that the use of Twitter is playing a vital role in spreading information and influencing people's opinions in a specific direction. Statistics from Statista website show that the Twitter in 2016 has more than 317 million active users. Many users tweet their opinions on a variety of subjects, discuss many political topics or marketing issues, and express their views on many aspects of their lives. Every tweet has a maximum length of 140 characters. Due to the shortness of the messages, people convey their opinions and thoughts openly most of the time. Therefore, Twitter is considered a rich data bank and one of the largest platforms that is full of sentiments [32].

According to recent reports, the fastest growing language on Twitter between 2010 and 2011 was Arabic [24]. While there is a great need for natural language analysis of large amounts of Arabic language text, the reality shows little work has been done in this area. Most of the sentiment analysis resources and systems built so far are tailored to English and other Indo-European languages. Reasons for the lack of research in this area include the complexity and the variety of dialects of the Arabic language that make it harder to build one system that is applicable to all of its dialects [10, 45].

The Arabic language belongs to the Semitic language family. It is recognized as the fifth most widely spoken language in the world and is considered the official or native language for 22 countries (approximately more than 300 million people) [25, 29]. The Arab region has a large, growing population and has become an important player in international politics and the global economy. Furthermore, the Arabic language is in the top ten of the most used languages to create Internet content [45].

This study primarily aims to review the efforts of building sentiment analysis systems for the Arabic language and lists some applications and systems that have been built to analyze Arabic Twitter data. This research also presents a general study of sentiment analysis and explores some of the machine learning algorithms and natural language processing classification techniques.

The remainder of this literature review is organized as follows. Section 2 gives the reader general background material on Twitter sentiment analysis by describing techniques and vital features that have been used in this area. Section 3 examines the techniques that have been used to analyze Arabic tweets and to summarize the key findings of recent research in this field. This literature review concludes with future directions of research in Sect. 4.

## 2 Background on Sentiment Analysis

In general, tweets generated by users can be categorized as objective or subjective tweets. Objective tweets contain facts that refer to the nature of entities, events, and attributes [46]. An example of an objective tweet is: *Election Day in the United States of America is the Tuesday following the first Monday in November.*

While subjective tweets express users' opinions regarding entities, events, and attributes. Subjectivity classification seeks tweets that contain users opinions. Some examples of subjective tweets include:

- *I'm happy election day is almost here.* (positive tweet)
- *I hate this election. Everything about it makes me miserable.* (negative tweet)
- *I don't care who wins the upcoming presidential election.* (neutral tweet)

Sentiment analysis is considered a part of the Natural Language Processing (NLP) field. It was first explored in 2003 by Nasukawa and Yi [35]. In sentiment analysis of Twitter data, the researchers focus their studies on subjective, not objective, tweets. They are interested mainly in classifying tweets as positive and negative [27]. The researchers studied sentiment analysis through three levels. The first level is the document level that classify and analyze sentiments for the whole document [36, 49]. Analyzing sentences is considered a second level. And finally, the phrase level is when the researchers are analyzing sentiments in phrases [5, 50]. They also investigated the utility of linguistic features for detecting the sentiment of Twitter posts.

## 2.1 Sentiment Analysis Work-Flow

The process of performing sentiment analysis for a micro-blogging tool usually goes through multiple phases [10, 12]:

- Phase 1: Data-gathering (crawling data). In this phase, the required amount of tweets that are related to a specific topic are retrieved. This data is filtered according to a particular time frame and keywords/users.
- Phase 2: Data-preprocessing (text normalization). This is an important step in the data mining field. The retrieved data from the first phase will be tokenized by converting the sentences into words. These words will be cleaned to remove any irrelevant and redundant information.
- Phase 3: Building-a-classifier. In this phase, a classifier model will be selected. Subject. 3.2 discusses the classification techniques that can be used to analyze peoples' sentiments more deeply.
- Phase 4: Visualization. This phase focuses on visualizing the results of sentiments attached to a particular topic and follows opinion changes over time. This can be performed by a graphical representation in several forms.

## 2.2 Sentiments Classification Algorithms

There are many techniques that perform sentiment analysis on Twitter data. According to Boiy [14], Symbolic techniques and Machine Learning techniques are the two basic methodologies used in sentiment analysis for text [45]. Symbolic technique, which is also called Semantic Orientation, uses sentiment lexicons which are lists of words or phrases associated with positive and negative sentiments. Some of these lexicons add other features and provide a score to

specify the strength of its class. This approach works to extract the score of its words and sum them up to show an overall positive or negative sentiment. Turney [49] used bag-of-words approach in which the document is treated as a collection of words regardless of the relationship between the words. Turney gave every word a value and combined all the values by using aggregation functions. Turney's technique is used to figure out the overall value for the whole document. On the other hand, Kamps [26] developed a distance metric on wordNet which is a database consisting of words and their relative synonyms. Another simple classifier model is the k-nearest neighbor algorithm that uses distance measure to assign a class label  $y$  to  $x$  if  $y$  is the nearest label to  $x$  [47].

Many classifier models have been built to classify tweets as positive, negative, or neutral according to their training data sets. These are grouped under the machine learning umbrella. The term machine learning was first coined by Samuel in the 1950s and was meant to encompass many intelligent activities that could be transferred from human to machine. The research in this field focuses on finding relationships in data [20]. Machine learning modeling methods can be supervised or unsupervised. In the supervised learning classification model, a training labeled set of data are used to predict the class of a search query. While in the unsupervised learning classification models, there is no labeled training data and the model will classify the corpus to specific classes based on some clustering computations. Labeling data in many applications is an expensive process and sometimes it may be labeled with errors and that may reflect the classification results. The unsupervised learning model is used frequently to predict the topic for a page or a text. Out of the sentiment analysis models that are using the supervised modeling in this survey, most of them have been built by using one of three standard algorithms: Naive Bayes classification, Maximum Entropy classification, and Support Vector Machines classification [6].

The efficiency of a classifier depends on the type of engineering feature associated with it. Feature extraction is the process of creating a representation for, or a transformation from, the original data. Numerous feature extraction algorithms have been proposed and successfully applied in many classifier models. Features can be binary, categorical, or continuous. Some of these powerful features are [6, 12]:

1. Term Presence vs. Term Frequency: It has been proven experimentally that the presence of a term is more important than counting the term frequencies.
2. Term Position: The term position can determine the sentiment for a tweet which plays an important role in sentiment analysis.
3. Part-of-Speech: Many articles show that this feature plays an important role in all Natural Language Processing tasks. This feature concentrates on the adjective and adverb words in the text.
4. Unigram: In this feature, a single word can be considered as a feature by itself. The results showed that unigram presence taken as feature turns out to be the most efficient.

Results show that n-grams features are the most widely used features for Twitter sentiments analysis [6].

The performance of sentiment classification system can be evaluated by using a well-known table called (Error Matrix) or (Confusion Matrix) [48]. Each column of the matrix represents predicted classifications and each row represents actual defined classifications. Based on the Confusion Matrix, four indexes can be computed to reflect the performance. These are Accuracy, Precision, Recall and F1-score.

### 3 Arabic Sentiment Analysis

#### 3.1 Arabic Language Aspects and Challenges

Arabic is the mother tongue of 22 countries with more than 300 million people speaking that language [25]. It is also the language of more than 1.4 billion Muslims around the world. It has been used for more than 2000 years [28]. The Arabic alphabet consists of 28 letters with no upper or lower cases and the orientation of writing is from right to left. Its letters can be written with different shapes according to their position in the word. According to [22,25], the Arabic language is classified into two main categories: Standard Arabic (SA) and Dialectal Arabic (DA). SA consists of two forms: Classical Arabic (CA), and Modern Standard Arabic (MSA). CA is the standard poetic language and the language of the Qur'an (Holy Islamic Book). While MSA language, which is a simplified form of CA, is used in most current printed Arabic publications such as books, newspapers, and also used in news broadcasts or formal speeches [44]. Although MSA is the primary language of the media and education in Arab countries, it is not spoken as a native language in people's informal daily communication. In contrast to MSA, DA is spoken but not written in books or taught in schools. DA has a strong presence in texting SMS on cellular phones, commenting on microblogging networks or in emails, blogs, discussion forums, and chats. Each dialect is spoken by a specified geographical area for daily verbal communication. Therefore, there is only one MSA language for all Arabic speakers but several dialects with no formal written form [18]. According to [28], the dialects are affected by many factors such as: which Arab tribe has lived in this geographical area and which foreign language was the source of loanwords. Also, if the geographical area is a village or countryside, or if the people are bedouin or sedentary. Arabic Dialects are greatly varied, and are classified into five main groups according to [51]: Egyptian, Levantine, Iraqi, Gulf, and Maghribi.

#### 3.2 Classification Techniques for Arabic Tweets

Minimal work has been done in Arabic sentiment analysis area. Several reasons may have explained the lack of studies in this area. Assiri in [12] mentioned two main reasons: 1- limited research funding in this area, 2- Arabic has a very complex morphology relative to the morphology of other languages. The complexity and variety of Arabic dialects require advanced pre-processing and lexicon-building procedures [10,12,45].

Working in this area needs a full understanding of Arabic standard layer-based structure of linguistic phenomena such as phonology, morphology, syntax and semantics [43]. Arabic is a highly inflectional and derivational language with many word forms and diacritics. Several suffixes, affixes, and prefixes in Arabic words make it harder for lexicon or morphological analyzers to extract the root of words correctly [19].

MSA has more studies and analysis as compared to DA. Numerous tools for detecting sentiments on short or long texts in MSA have been built. Knowing that applying NLP tools designed for MSA directly to DA yields significantly lower performance. This led a group of researchers to build other resources and tools for analyzing DA [15, 39, 45].

Many researchers have applied Machine Translation (MT) in their studies by translating Arabic statements into English and then applying sentiment analysis tools on the translated materials [33, 40]. This approach has been explored widely for other foreign languages by performing sentiment analysis on the English translation [9, 13]. The problem of this approach was the loss of nuance after translating the source to English. It is shown in [11] that finding an Arabic MT that meets human requirements is a difficult task. This field still needs more efforts to be improved. Most of the previous work focused on the translation of news and official texts. Much work has been done on MSA; however, research on DA is still lacking in MT [41].

A prior important step for analyzing sentiments in any language is Building Resources (BR). This step aims at creating lexica, corpora with annotated expressions or opinions. There is a need for large scale of annotated resources for the Arabic language in order to do sentiment analysis. Some efforts have been paid to build Arabic Treebanks that contain collections of manually-annotated syntactic analyses of sentences [21, 23, 31]. Researchers focus mainly on building corpus/corpora that contain annotated data for MSA and less attention is paid towards DA. Most of these resources are either of limited size or not available for public. Recently, a study [17] addressed this problem and generated a large multi-domain dataset for sentiment analysis in Arabic. The study scraped 33K annotated reviews for movies, hotels, restaurants and products. Then, the researchers built multi-domain lexicons from the generated datasets and tested the classifier models on this data. Another research published in 2014 [37] with a dataset of 8,868 multi-dialectal Arabic annotated tweets. They employed morphological features, simple syntactic features, such as n-grams, as well as semantic features. Other research studies can be found in Table 1 which summarizes the recent work on building resources for Arabic language.

Most of the sentiment analysis tools perform three main data pre-processing steps before applying the classification techniques in order to prepare the Arabic texts, which are: 1- Normalization, 2- Stemming, and 3- Stop word removal. Once data pre-processing has been applied to the text, it will be ready for the feature extraction step. Several text features are considered for the Arabic sentiment analysis: n-grams, term presence or its frequency, part-of-speech, or emoticon symbols. The goal of feature extraction step is to select which text features are

**Table 1.** Building resources for Arabic

Ref.	Name	Year	Description
[38]	OCA	2011	Arabic corpus contains 500 movie reviews collected from different web pages and blogs in Arabic with 250 positive reviews and 250 negative reviews. It has limited size and for specific domain (movies)
[1]	AWATIF	2012	Multi-genre corpus of Modern Standard Arabic labeled for subjectivity and Sentiment Analysis. It is only dedicated for MSA and not available for public
[34]	LABR	2014	Over 63,000 book reviews rated on a scale of 1 to 5 stars. It is only for specific domain (books)
[2]	SANA	2014	A large scale multi-genre sentiment lexicon (more than 200K) of MSA and some Arab dialects. It is also not public
[37]	NA	2014	A dataset of 8,868 multi-dialectal Arabic annotated tweets
[17]	NA	2015	A large multi-domain dataset (33K annotated reviews for movies, hotels, restaurants and products) for sentiment analysis in Arabic

best to be applied in sentiment analysis tool. Most of the features in Arabic sentiment analysis are classified into three types [7]: (1) Syntactic, which includes: word/POS tag n-grams, phrase patterns, punctuation, (2) Semantic, this type includes: polarity tags, appraisal groups, semantic orientation, and (3) Stylistic, which is concerned with lexical and structural measures of style.

A considerable amount of previous work has been published on Arabic sentiment analysis. This literature review is focused on the studies that categorized the Arabic tweets into specific domains using different classification techniques. Table 2 presents and summarizes the latest work in this area according to the classification techniques and extracted features. It also states whether the study has been applied to MSA or DA.

In 2012, a study [45] proposed a model that used two machine learning approaches, NB and SVM. The researchers used a list of stop words from Egyptian dialect in the preprocessing step. They selected 1000 tweets that hold only one opinion, not sarcastic, subjective and from different topics. SAMAR, is another proposed tool in the same year [3]. It is also a machine learning system for Arabic social media texts. The researchers tested their tool in four different genres: chat, Twitter, Web forums, and Wikipedia talk pages. For Twitter, a corpus of 3015 Arabic tweets has been collected that has a mixture of MSA and DA.

Next year, 2013, a new study [7] annotated 4000 tweets from different popular topics: technology, politics, religion, and sports, respectively. The study found that it is better to use unigrams with tweets. Another study has also been presented in 2013 [33] that built a baseline system for performing subjectivity and sentiment analysis for Arabic news and tweets. MT has been employed to translate an existing English subjectivity lexicon to build large coverage lexicons

**Table 2.** Analysis of previous work on Arabic sentiment analysis for Twitter data.

Ref.	Year	Tweets	Features	Classification techniques	Performance	MSA or DA
[45]	2012	1000	Unigrams and bigrams	NB and SVM	SVM Acc 72.6%	Egyptian DA
[3]	2012	3015	Morphological, POS tags, and adjective polarity lexicon	SVM-light	Acc of 71.85%	MSA and DA
[7]	2013	4000	N-Grams and unigrams	SVM, NB, MaxEnt, Bayes Net, and J48 D-tree	SVM Acc 86.38%	-
[33]	2013	2300	Stem-level, sentence-level, and tweet specific	NB	Acc 80.6%	MSA and DA
[4]	2013	2000	Unigram	Supervised (SVM, NB, KNN, and D-tree) and unsupervised ML	Acc 87.5%	MSA and Jordanian DA
[16]	2013	500	POS tag with weight	Unsupervised approach	Acc 83.8%	Egyptian DA
[15]	2014	25000	-	NB, k-NN, SVM	NB (76.78%)	MSA and Jordanian DA
[39]	2014	340,000	Opinions-oriented words extraction	D-tree and SVM	Prec 76%, recall 61%	Kuwaiti DA
[8]	2015	900	Giving weights to words	Unsupervised	Acc 86.89%	MSA
[25]	2015	1000	Linguistically and syntactically motivated	Semi-supervised with SVM	Acc 95%	MSA and DA

in Arabic. Another study in 2013 [4] addressed both approaches; supervised and unsupervised, for sentiment analysis for Arabic twitter data. The researchers in this study collected and labeled 2000 tweets in both MSA and Jordanian dialect. One of the key finding of this study was that the unsupervised approach gives much lower accuracy compared to the supervised approach. A group of researchers constructed a lexicon-based tool to analyze sentiments of egyptian dialectal tweets in 2013 [16]. Every word in the lexicon has been assigned weights that determined semantic orientation based on the sentiment lexicon.

In 2014, an Arabic sentiment analysis tool was presented in [15] which contains a lexicon that maps Jordanian Dialect to MSA, a lexicon that maps Arabizi words to MSA, and a lexicon of emoticons. In the same year, SVM classifier has also been tested on a corpus of 340,000 tweets in Kuwait [39]. This system handled Kuwaiti dialect which used Opinions-Oriented words extraction features to extract the opinion-oriented words through language resources that they have been developed for the Kuwaiti dialect.



Recently in 2015, another tool that used an unsupervised (lexicon-based) approach has been introduced [8]. This tool has access to a sentiment lexicon that contains a set of words along with their sentiment values. A sentiment lexicon of about 120,000 Arabic terms has been constructed through three steps: collect Arabic stems, translate them into English, and use online English sentiment lexicons to determine the sentiment value of each word. They stated that the proposed tool performed better than the keyword-based approach.

Finally, a research [25] studied an Arabic idioms/saying phrases lexicon to improve the sentiment polarity in Arabic sentences has gained a high accuracy around 95%. This study used semi-supervised approach with using SVM classifier to analyze MSA and Egyptian dialectal Arabic tweets and microblogs, such as hotel reservation, and product reviews.

## 4 Conclusion

This paper has presented the challenging task of sentiment analysis and opinion mining on Twitter data in the domain of the Arabic language. We reviewed numerous studies that analyzed people's opinions in English and other Indo-European languages. However, we found few studies that analyzed people's opinions in the Arabic language.

This current investigation examined the prior studies to determine how the sentiment analysis was applied to a high volume of Arabic tweets. This study aimed to help newcomers to this field understand the different aspects posed by the research within the past few years. A sophisticated categorization of a large number of recent articles has been reviewed in this study to cover a wide variety of sentiment analysis in the Arabic language.

One of the main findings of this review shows that there is still a great need for extensive research to gain a better understanding of Arabic dialects in addition to further MSA studies. Up to the time of writing this literature review, no single system existed that could handle all Arabic dialects and MSA with high accuracy. This has created a wide gap in this field for researchers to address in subsequent investigations.

This study demonstrated a need for building and publishing additional lexicon Arabic resources with different genres and various dialects for both the public and research community. Assembling all lexicons for Arabic dialects from different geographical areas in the Middle East in one lexicon repository is a worthy goal.

Recently, growing Internet usage has produced a new written form called Arabizi. This type of the Arabic language is derived from the spoken Arabic dialects and is written using Latin letters and numbers. Detecting and analyzing tweets written in Arabizi has not been thoroughly studied. Knowing that Arabizi has been used widely by teen-agers, it is important to conduct future studies on this type of language and to include young researchers and annotators to bridge the gap.

## References

1. Abdul-Mageed, M., Diab, M.T.: AWATIF: a multi-genre corpus for modern standard Arabic subjectivity and sentiment analysis. In: LREC, pp. 3907–3914. Citeseer (2012)
2. Abdul-Mageed, M., Diab, M.T.: SANA: a large scale multi-genre, multi-dialect lexicon for Arabic subjectivity and sentiment analysis. In: LREC, pp. 1162–1169 (2014)
3. Abdul-Mageed, M., Kübler, S., Diab, M.: SAMAR: a system for subjectivity and sentiment analysis of Arabic social media. In: Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, pp. 19–28. Association for Computational Linguistics (2012)
4. Abdulla, N.A., Ahmed, N.A., Shehab, M.A., Al-Ayyoub, M.: Arabic sentiment analysis: lexicon-based and corpus-based. In: 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), pp. 1–6. IEEE (2013)
5. Agarwal, A., Biadys, F., Mckeown, K.R.: Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, pp. 24–32. Association for Computational Linguistics (2009)
6. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.: Sentiment analysis of Twitter data. In: Proceedings of the Workshop on Languages in Social Media, pp. 30–38. Association for Computational Linguistics (2011)
7. Ahmed, S., Pasquier, M., Qadah, G.: Key issues in conducting sentiment analysis on Arabic social media text. In: 2013 9th International Conference on Innovations in Information Technology (IIT), pp. 72–77. IEEE (2013)
8. Al-Ayyoub, M., Essa, S.B., Alsmadi, I.: Lexicon-based sentiment analysis of Arabic tweets. *Int. J. Soc. Netw. Min.* **2**(2), 101–114 (2015)
9. Alexandra, B., Marco, T.: Multilingual sentiment analysis using machine translation. In: Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, pp. 52–60 (2012)
10. Alhumoud, S.O., Altuwajri, M.I., Albuhaire, T.M., Alohaideb, W.M.: Survey on Arabic sentiment analysis in Twitter. *Int. Sci. Index* **9**(1), 364–368 (2015)
11. Alqudsi, A., Omar, N., Shaker, K.: Arabic machine translation: a survey. *Artif. Intell. Rev.* **42**(4), 549–572 (2014)
12. Assiri, A., Emam, A., Aldossari, H.: Arabic sentiment analysis: a survey. *Int. J. Adv. Comput. Sci. Appl.* **1**(6), 75–85 (2015)
13. Bautin, M., Vijayarenu, L., Skiena, S.: International sentiment analysis for news and blogs. In: ICWSM (2008)
14. Boiy, E., Hens, P., Deschacht, K., Moens, M.F.: Automatic sentiment analysis in on-line text. In: ELPUB, pp. 349–360 (2007)
15. Duwairi, R., Marji, R., Sha’ban, N., Rushaidat, S.: Sentiment analysis in Arabic tweets. In: 2014 5th International Conference on Information and Communication Systems (ICICS), pp. 1–6. IEEE (2014)
16. El-Beltagy, S.R., Ali, A.: Open issues in the sentiment analysis of Arabic social media: a case study. In: 2013 9th International Conference on Innovations in Information Technology (IIT), pp. 215–220. IEEE (2013)
17. ElSahar, H., El-Beltagy, S.R.: Building large arabic multi-domain resources for sentiment analysis. In: Gelbukh, A. (ed.) CICLEing 2015. LNCS, vol. 9042, pp. 23–34. Springer, Cham (2015). doi:[10.1007/978-3-319-18117-2\\_2](https://doi.org/10.1007/978-3-319-18117-2_2)

18. Eviatar, Z., Ibrahim, R.: Why is it hard to read Arabic? In: Saiegh-Haddad, E., Joshi, R.M. (eds.) *Handbook of Arabic literacy*, pp. 77–96. Springer, Netherlands (2014)
19. Farra, N., Challita, E., Assi, R.A., Hajj, H.: Sentence-level and document-level sentiment mining for Arabic texts. In: 2010 IEEE International Conference on Data Mining Workshops, pp. 1114–1119. IEEE (2010)
20. Guyon, I., Elisseeff, A.: An introduction to feature extraction. In: Guyon, I., Nikravesh, M., Gunn, S., Zadeh, L.A. (eds.) *Feature Extraction*, pp. 1–25. Springer, Heidelberg (2006)
21. Habash, N., Roth, R.M.: CATiB: The Columbia Arabic treebank. In: *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pp. 221–224. Association for Computational Linguistics (2009)
22. Habash, N.Y.: Introduction to Arabic natural language processing. *Synth. Lect. Hum. Lang. Technol.* **3**(1), 1–187 (2010)
23. Hajic, J., Smrz, O., Zemánek, P., Šnaidauf, J., Beška, E.: Prague Arabic dependency treebank: development in data and tools. In: *Proceedings of the NEMLAR International Conference on Arabic Language Resources and Tools*, pp. 110–117 (2004)
24. Harrag, F.: Estimating the sentiment of Arabic social media contents: a survey. In: *5th International Conference on Arabic Language Processing* (2014)
25. Ibrahim, H.S., Abdou, S.M., Gheith, M.: Sentiment analysis for modern standard Arabic and colloquial. *arXiv preprint, arXiv:1505.03105* (2015)
26. Kamps, J., Marx, M., Mokken, R.J., De Rijke, M.: Using wordnet to measure semantic orientations of adjectives. In: *LREC*, vol. 4, pp. 1115–1118. Citeseer (2004)
27. Khairnar, J., Kinikar, M.: Machine learning algorithms for opinion mining and sentiment classification. *Int. J. Sci. Res. Publ.* **3**(6), 1–6 (2013)
28. Khrisat, A.A., Alharthy, Z.A.: Arabic dialects and classical Arabic language. *Adv. Soc. Sci. Res. J.* **2**(3), 254–260 (2015)
29. Korayem, M., Crandall, D., Abdul-Mageed, M.: Subjectivity and sentiment analysis of Arabic: a survey. In: Hassaniien, A.E., Salem, A.-B.M., Ramadan, R., Kim, T. (eds.) *AMLTA 2012. CCIS*, vol. 322, pp. 128–139. Springer, Heidelberg (2012). doi:[10.1007/978-3-642-35326-0\\_14](https://doi.org/10.1007/978-3-642-35326-0_14)
30. Kouloumpis, E., Wilson, T., Moore, J.: Twitter sentiment analysis: the good the bad and the OMG!. *ICWSM* **11**, 538–541 (2011)
31. Maamouri, M., Bies, A., Buckwalter, T., Mekki, W.: The penn Arabic treebank: building a large-scale annotated Arabic corpus. *NEMLAR Conf. Arabic Lang. Res. Tools* **27**, 466–467 (2004)
32. Mohammad, S.M., Kiritchenko, S., Zhu, X.: NRC-Canada: building the state-of-the-art in sentiment analysis of tweets. In: *Second Joint Conference on Lexical and Computational Semantics (\*SEM)*, vol. 2, pp. 321–327 (2013)
33. Mourad, A., Darwish, K.: 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 (2013)
34. Nabil, M., Aly, M.A., Atiya, A.F.: LABR: a large scale Arabic book reviews dataset. *CoRR*, abs/1411.6718 (2014)
35. Nasukawa, T., Yi, J.: Sentiment analysis: capturing favorability using natural language processing. In: *Proceedings of the 2nd International Conference on Knowledge Capture*, pp. 70–77. ACM (2003)

36. Pang, B., Lee, L.: A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, p. 271. Association for Computational Linguistics (2004)
37. Refaee, E., Rieser, V.: An Arabic Twitter corpus for subjectivity and sentiment analysis. In: LREC, pp. 2268–2273 (2014)
38. Rushdi-Saleh, M., Martín-Valdivia, M.T., Ureña-López, L.A., Perea-Ortega, J.M.: OCA: Opinion corpus for Arabic. *J. Am. Soc. Inf. Sci. Technol.* **62**(10), 2045–2054 (2011)
39. Salamah, J.B., Elkhilfi, A.: Microblogging opinion mining approach for Kuwaiti dialect. In: The International Conference on Computing Technology and Information Management (ICCTIM2014), The Society of Digital Information and Wireless Communication, pp. 388–396 (2014)
40. Salameh, M., Mohammad, S.M., Kiritchenko, S.: 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 (2015)
41. Salloum, W., Habash, N.: Dialectal Arabic to English machine translation: pivoting through modern standard Arabic (2013)
42. Selmer, O., Brevik, M.: Classification and visualisation of Twitter sentiment data (2013)
43. Shaalan, K., Habash, N.Y.: Introduction to Arabic natural language processing (synthesis lectures on human language technologies). *Mach. Transl.* **24**(3), 285–289 (2010)
44. Shaalan, K.: Rule-based approach in Arabic natural language processing. *Int. J. Inf. Commun. Technol. (IJICT)* **3**(3), 11–19 (2010)
45. Shoukry, A., Rafea, A.: Sentence-level Arabic sentiment analysis. In: 2012 International Conference on Collaboration Technologies and Systems (CTS), pp. 546–550. IEEE (2012)
46. Singh, S.K., Paul, D.S., Kumar, D.: Sentiment analysis approaches on different data set domain: survey. *Int. J. Database Theor. Appl.* **7**(5), 39–50 (2014)
47. Smola, A., Vishwanathan, S.: Introduction to Machine Learning, vol. 32. Cambridge University, Cambridge (2008)
48. Stehman, S.V.: Selecting and interpreting measures of thematic classification accuracy. *Remote Sens. Environ.* **62**(1), 77–89 (1997)
49. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417–424. Association for Computational Linguistics (2002)
50. Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., Patwardhan, S.: Opinionfinder: a system for subjectivity analysis. In: Proceedings of HLT/EMNLP on Interactive Demonstrations, pp. 34–35. Association for Computational Linguistics (2005)
51. Zaidan, O.F., Callison-Burch, C.: Arabic dialect identification. *Comput. Linguist.* **40**(1), 171–202 (2014)