



Sentiment analysis of Arabic comparative opinions

Mai M. Eldefrawi¹ · Doaa S. Elzanfaly^{1,2} · Marwa S. Farhan¹ · Ahmed S. Eldin³

© Springer Nature Switzerland AG 2019

Abstract

The rapid development of social media platforms allowed opinion mining research to increase significantly. Opinion mining/sentiment analysis is the process of extracting subjective information from opinions that imply single sentiment. Comparative opinion mining is a sub-field of opinion mining that deals with multi-sentiment opinions. Such opinions are expressed by comparing several entities to each other. The sentiment of a comparative relation is recognized by identifying the relation's direction and thus the preferred entity. This paper proposes an unsupervised sentiment analysis technique for Arabic comparative opinions to identify the preferred entity. The proposed technique considers three main elements when analyzing comparative opinions: The type of comparative keywords, the existence of features in the opinion, and the entities' position to the comparative keyword. Five main categories are proposed for classifying comparative keywords, which facilitates the analysis of each comparative sentence. The proposed technique limits the need for human interference to the initial steps of preparing the lexicons, collecting and categorizing comparative keywords. Furthermore, the proposed technique handles opinions that do not contain features at all. The results are very promising with a total average of 96.5% *f*-measure of correctly identified sentiment.

Keywords Sentiment analysis · Arabic comparative opinions · Comparative opinion mining · Comparative relation

1 Introduction

Opinion mining and sentiment analysis are terms used to describe the process of categorizing opinions based on the extracted sentiments using text mining techniques. Sentiment analysis is a research area that has been attracting attention because of the rapid development of web 2.0 where a wider social platform is provided for opinion holders to express their thoughts and ideas. Opinion mining usually considers two types of opinions: direct opinions and comparative opinions. Sentiment in direct opinions is directed toward a certain subject or entity, whereas comparative opinions imply sentiment that expresses differences or similarities of more than one entity. Acquiring facts from comparative opinions has a strategic benefit. Companies tend to market their products in comparison

with competitors to highlight their points of strength. Industries are continuously evolving and updating their products to attract more customers and consumers to their products away from competitors.

Leading work in the area of comparative opinion mining classifies comparative opinions into two main categories: gradable and non-gradable [1, 2]. Users compare entities through three main types of ordering relationships in *gradable comparative opinions*. The first type is **non-equal gradable** in which entities in the opinion are compared while grading them differently preferring one entity to another, ex: موبايل سوني افضل من موبايل سامسونج. The second type is **superlative** where one entity is compared to all other entities. The entity is graded the best or worst compared to all the rest, ex: موبايل سوني هو الافضل على الاطلاق. Finally, there is the **equative** type where all entities are

✉ Mai M. Eldefrawi, mai.eldefrawi@fci.helwan.edu.eg; Doaa S. Elzanfaly, doaa.saad@fci.helwan.edu.eg; Marwa S. Farhan, marwa.salah@fci.helwan.edu.eg; Ahmed S. Eldin, profase2000@yahoo.com | ¹Faculty of Computers and Information, Helwan University, Cairo, Egypt. ²Faculty of Informatics and Computer Science, The British University in Egypt, Cairo, Egypt. ³Faculty of Information Technology and Computer Science, Sinai University, Sinai, Egypt.



graded as equal based on their shared aspects; no entity is better or worse, ex: موبايل سوني نفس امكانيات سامسونج.

Non-gradable is a special type of comparative opinion that does not contain actual comparison. Entities are mentioned without specifying a preferred one or grading any of them. Non-gradable opinions can further be categorized into three main sub-types [3]. In the first type, the two entities are similar or different based on some shared features. The second type states that one entity has a certain feature and the other has a different feature. Finally, the last type states that one entity has a feature that the other entity does not have.

Comparative opinion mining includes three main tasks [1, 4]. The first is to identify the comparative opinions as direct opinions or comparative opinions. The second task is to extract the elements of the comparative relations from the identified opinions. This includes extracting the compared entities, features of the comparison and the comparative keyword. Entities are objects, people, services or anything being compared. Features are the criteria according to which entities are compared. For example, Uber's price is better than Careem. The first entity here is Uber/أوبر, the second entity is كريم/Careem and the feature is the price/السعر. The third and last task is to identify the directions of the comparative relation. The focus of this paper is the third task where preferred entities are to be identified.

Mainly comparative opinion research area applies its techniques on the English language; however, this paper introduces a technique for identifying the sentiment of Arabic comparative opinions. The proposed technique uses the available Arabic sentiment analysis lexicons [5–8] and enriches them by the lexicons presented in [9]. It introduces a sentiment measurement based on three main factors: the comparative keyword type, the existence of features and the position of entities in the opinion. It categorizes comparative keywords into five main categories for the sentiment calculation. The proposed technique limits human interference to the initial steps of preparing the data and categorizing comparative keywords. It also offers an approach for handling comparative opinions that contain comparative keywords without sentiments or features.

The rest of the paper is organized as follows. Section 2 provides a review of the work done in the area of comparative opinion's sentiment analysis followed by the proposed technique in Sect. 3, which first provides the definitions of the different types of comparative keywords used in this paper. Then, it demonstrates the processing steps carried out by each category and how the sentiment of different comparative opinions is calculated. The findings of the proposed technique are presented in Sect. 4 along with the used dataset and the collected lexicons. Finally,

a conclusion and recommendations for further research are offered in Sect. 5.

2 Related work

The approaches used with sentiment analysis can be categorized into three main categories: Supervised learning, unsupervised learning and a hybrid approach of the first two approaches. Supervised learning includes machine learning algorithms [10, 11] as support vector machines (SVM) and Naive based (NB). Such an approach uses a large labeled dataset and a set of predefined features to train the algorithm. This technique is mainly used for comparative opinions identification, comparative opinions classification and sentiment analysis of comparative opinions as in [1, 2, 4, 12]. Another popular machine learning algorithm that is used in comparative opinion is the conditional random field (CRF) algorithm. It is mainly used for comparative relation extraction, but it can also be used to identify the comparative relation direction as in [4]. Xu et al. [4] proposed a novel graphical approach for comparative relation extraction; namely, the use of a two-level CRF algorithm with interrelated dependencies. Their main contribution was in considering comparative sentences with more than one relation. They considered the unfixed interdependencies of relations to identify the better, worse, same and no comparison relations.

Another supervised machine learning approach was adopted in [13]. They retrained a semantic role labeling (SRL) system with a product review data, where comparative arguments and predicates were labeled. The comparative predicate of each sentence was labeled along with two entities and one feature as arguments. The sentiment identification was expected to have at least one or four in the form of (predicate, entity+, entity-, aspect) as any of the arguments could be empty. Their result of the argument classification as entity+ or entity- was relatively low; however, their approach can be improved by engineering more features for the training.

The unsupervised approach includes machine learning problems that do not require a predefined, labeled set as in the clustering problems. Lexicon-based sentiment analysis techniques are considered an unsupervised technique [14]. Little attention has been given to lexicon-based approach for comparative sentiment analysis. Work in [9] uses a lexicon approach by creating two lists and calculating the included sentiment terms through one-side association (OSA) measurement [11]. The two lists are the pros and cons they gathered from online reviews to provide external information for the sentiment calculation. They identify two main types of comparative keywords. Opinions were processed according to whether

the comparative keyword and features were opinionated or not. Their method is effective, but it cannot be applied directly to the Arabic language. The differences between the Arabic comparative expressions and the English ones require different categorization. The two lists created in this work are used in this paper to add more vocabulary to the final lexicon.

Another approach that uses deep neural network techniques is presented by Chen et al. [15]. It processes comparative sentences as part of multi-target sentences and determines an overall sentiment to the sentence. Their work does not specifically identify entities and features involved in the comparison, but it identifies several targets in the sentence. They used one-dimensional convolutional neural networks (1d-CNNs) for the sentiment classification. Sentences were categorized according to the number of targets in them. They trained 1d-CNNs on each type of sentences separately. Their results are good considering the dataset they used; however, they do not identify comparative opinions separately or specify the preferred entities.

Most work on comparative relations focuses on certain languages, mainly English and Chinese [4, 9, 10, 15–17]. Other languages such as Arabic have been given less attention [10]. The research done on comparative opinions in the Arabic language usually deals with preferences expressed using ‘elative’. The elative form in Arabic represents both comparative and superlatives nouns. It is generated by converting the triliteral adjective roots to a form that is equivalent to the elative form ‘افعل/afaal’ or ‘لافعل/alafaal’ to represent comparative and superlative nouns, respectively [18, 19]. The ‘افعل/afaal’ comparative keyword takes JJR tag in any Part of Speech (POS) tagger, e.g., ‘افضل/better’ where the ‘لافعل/al + afaal’ keyword takes DTJJR tag, e.g., ‘الافضل/the best’. They will be referred to as JJR comparatives and DTJJR superlatives for simplicity. These two forms are the most popular forms of expressing comparative opinions. However, there are other terms such as ‘مقارنه ب/compared to’, ‘يتفوق/يتميز/exceeds’, ‘اختار/i choose’, and ‘لكن/however’ can be used when comparing two or more entities.

Arabic comparative opinions were first addressed by El-Halees [12] where opinions were categorized into comparative and non-comparative opinions. They used POS techniques and machine learning algorithms for categorization. They further classified them into the four predefined comparative types, using manually created rules. Later in [20], Eldefrawi et al. proposed the use of CRF algorithm for the extraction of comparative relations; however, they did not identify the comparative relation’s direction.

Comparative opinion mining in the Arabic text is one of the long-neglected topics in sentiment analysis. Previous work that used linguistic forms for the analysis of comparative opinions in [9] was based on the English language

structure. Their focus was mainly on the two common types of comparative opinions which are the comparatives and superlatives that end with “er” and “est” and the ones use “most, least, less, more” without additional investigation for different comparative terms.

An approach for comparative opinions’ sentiment identification is proposed using lexicon-based techniques and a set of sentiment calculation rules. The proposed technique provides an analysis of the linguistic form of comparative opinions. To the best of our knowledge, calculating the sentiment of Arabic comparative opinions has not been addressed before. The proposed technique categorizes comparative keywords into five main categories. The categorization is made to fit the nature of the Arabic language and to allow the algorithm to calculate the sentiment for opinions based on the type of the comparative relation’s keyword. It also sheds light on certain comparative keywords other than traditional comparative ones. These different forms have separate processing steps to increase total accuracy.

3 The proposed technique

The proposed technique starts with the categorization of comparative keywords. This section introduces the five comparative categories adopted in the proposed technique. It then explains the processing steps carried out on each category in detail. Finally, it discusses how the sentiment and negation are calculated for the different features and terms in the opinion.

3.1 Comparative keywords’ categorization

Identifying the different types of comparative keywords is necessary, as every type has its own characteristics. This will better allow identifying the direction of each comparative relation where the processing of a certain comparative keyword type may not be applicable to other types. A list of comparative keywords is collected from the gathered dataset and through searching social platforms for comparative expressions. It is a straightforward categorical approach in which keywords are grouped into five categories based on their nature. This significantly automates the process and limits the need for human judgment to initial preparation of lexicons and dataset as discussed in Sect. 4.

Table 1 summarizes the five categories and provides examples for clarification. The first two categories are the JJR and the DTJJR comparative types. Though the Arabic language has an infinite number of JJR and DTJJR, there are certain common terms used by opinion holders. The commonly used ones were gathered and given a sentiment based on the observations of the dataset. The third

Table 1 The five main categories of comparative keywords

Comparative keywords Categories	Keywords	Example
Type A: JJR comparatives with sentiment	اسوأ / worse, احسن, افضل / better, اقذر, اوسخ / dirtier, اصعب / harder	مدرسه المنارات احسن كثير اكااديميا من مدرسه العروبه / Al Manarat school is much academically better than al Orouba
Type B: DTJJR suprelatives with sentiment	الافضل, الاحسن / the best, الاسوأ / the worst	نارمر طبعاً الافضل فيهم كلهم / Narmar of course is the best in all of them
Type C: JJR/ DTJJR comparative with no Sentiment	اعلى / higher, ادنى / lower, اكثر / more, اقل / less, الاعلى / the highest, الاقل / the lowest, الاكثر / the most	سوني بطاريته بتعيش اكثر من سامسونج / Sony's battery lives more than Samsung
Type D: Terms that show direct sentiment	اختر, انا اختار / I prefer, انا افضل / I choose, يتميز / distinguished, exceeds	الاتنين كويسين بس انا ارشح مطعم الصفوه عشان الخدمه / Both are good but I suggest Al Safwa restaurant for the service
Type E: Neutral Comparative	مقارنه ب, اما / as for, عكس / in contrary, لكن / however	سوني حلو و سريع اما سامسونج ميقاش زي الاول / Sony is nice and fast on the other hand Samsung is not like before

category is the JJR and the DTJJR comparative types, but with no sentiment. They are a special type of the first two and commonly used where comparative keywords have no obvious sentiment on their own. They show a decrease or increase of the features in an opinion and the sentiment of a keyword is identified based on the sentiment of the feature. The fourth category includes terms that show direct sentiment: they could be verbs, adjectives or a combination of terms. The existence of these terms facilitates the identification of the preferred or the non-preferred entity because their meaning is conclusive. The last category is the neutral comparatives which have the nature of separating the two entities into two different sentences. They have a neutral sentiment; therefore, to identify the preferred entity, the sentiment of the surroundings must be identified. Though the categorization is a one-time effort, the keyword list of each category can still be expanded by adding new comparative keywords whenever having new collected opinions.

The processing of the comparative opinions can be carried out after assigning the comparative keyword in the opinion to one of the five categories. In the next section, the processing of each category is discussed in detail.

3.2 Comparative opinions processing

Since all sentences are annotated and assigned to one of the five categories, they are processed with consideration of the other two elements (the existence of features and the position of entities to the comparative keywords). The pseudocode in Fig. 1 summarizes the rules guiding the

processing of each category. They are then detailed in the following sub-sections.

3.2.1 Type A: JJR comparatives with sentiment

This is the most dominant comparative form used while expressing preferences. It usually has the two entities on different sides of the comparative keyword. If the JJR keyword's sentiment is positive, the opinion's sentiment is positive toward the first entity, in case no negation is used. If the JJR keyword's sentiment is negative, the opinion's sentiment is negative toward the first entity and subsequently positive toward the second entity.

For instance: اسوأ بكثير من الغربي /The eastern hotel is much worse than the west. The comparative keyword here is اسوأ/worse. The keyword's sentiment is negative. The first entity which is الفندق الشرقي/eastern hotel takes the negative sentiment. In case, the two entities are before or after the comparative keyword, the entity right before or right after the comparative keyword gets its sentiment. This form is sometimes used to express the superlative sentiment as in اهل اهل احسن فريق /Al Ahly is a better team. In this case the one entity in the opinion takes the sentiment of the JJR comparative.

3.2.2 Type B: DTJJR superlatives with sentiment

It is a variant of JJR, mainly used for expressing superlatives. In this case, the sentiment of the DTJJR is given to that one entity. In some cases, DTJJR is used in a comparative way; it is used to express non-equal gradable

Fig. 1 Processing rules of each comparative type

<p>Input Comparative_keyword, Annotated_sentence</p> <p>Output Sentiment</p> <p>Case based Comparative_keyword</p> <ol style="list-style-type: none"> Case > Type A, JJR comparative or type B, DTJJR superlative with Sentiment Sentiment ← The first entity to the keyword takes the sentiment of the JJR/ DTJJR Case > Type B, DTJJR suprelative with sentiment (Expressing comparison) Check if pronouns like (Entity* +is/ are+ DTJJR) or (DTJJR+ is/ are/ than them+ Entity*) Then Sentiment← Entity* takes the sentiment of the DTJJR keyword. Case > Type C, JJR comparative/ DTJJR superlative without sentiment Calculate the sentiment of comparative keyword+ features IF (Positive/ neutral features+ increasing keyword) Sentiment← Positive_sentiment else if (Positive/ neutral features+ decreasing keyword) Sentiment← Negative_sentiment else if (Negative features+ increasing keyword) Sentiment← Negative_sentiment else if (Negative features+ decreasing keyword) Sentiment← Positive_sentiment Case > Type C, JJR comparative/ DTJJR superlative without sentiment or features Sentiment← total sentiment of terms around entities and keywords, then apply same rules in case 3 Case > Type D, terms that show Direct Emotions Sentiment← The entity following or preceding the term takes the sentiment Case > Type E, neutral comparatives with_features Calculate the sentiment of <i>features</i> in each sub-sentence Sentiment← Positive to the entity enclosed in the highest sub-sentences score Case > Type E, neutral_comparatives without_features Calculate the sentiment of <i>terms</i> in the sub-sentences Sentiment← Positive to the entity enclosed in the highest sub-sentences score <p>End Case</p>

comparatives. In such cases, the pronouns used 'هو/ هو and are/هم' with the DTJJR should be considered.

In the case of entity 1 + (هو, هي, هو, هم)/is, are)+ DTJJR, entity 1 takes the sentiment of the DTJJR keyword. Otherwise, if DTJJR+ (هو, هي, هو, هم, منها, هي, هو, هم, than them)+ entity 2, then entity 2 takes the sentiment of the DTJJR keyword.

For example, on DTJJR with a pronoun, 'الغربي هو الافضل ليس جيد الشرقي' /The eastern hotel is not good the best is the western'. Here 'the best' is followed by 'هو/is'. In this case, the second entity is the better 'الغربي /the western'

3.2.3 Type C with features: JJR comparatives, DTJJR superlatives with no sentiment

JJR comparatives that show an increase as 'اعلى/higher, اكثر/more' do not necessarily indicate positive sentiment. One example is 'اعلى سعرا /higher price', which indicates an obvious negative sentiment. Another example is 'اكثر تحملا /more durable'. 'سعرا/price' and 'تحملا/durable' are considered

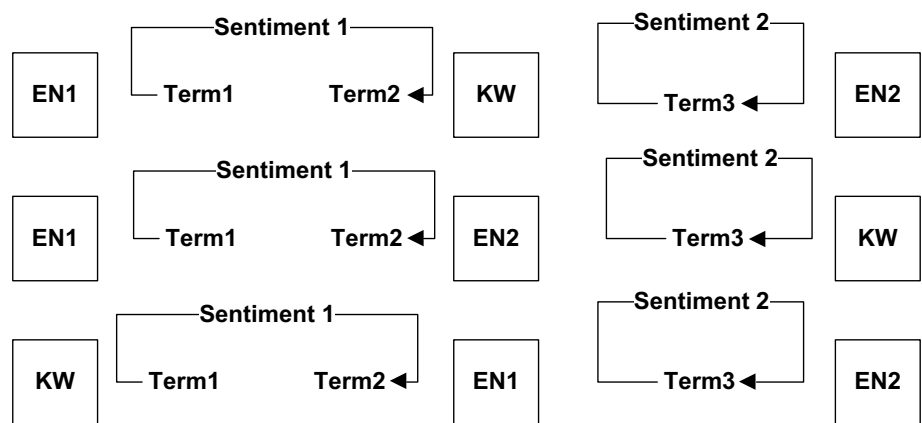
features in these opinions. Features need to be identified to calculate the sentiment of the sentence. The adopted approach here is similar to the one used in [9]. If a positive feature is associated with increased JJR, then the sentiment is positive. If a positive feature is associated with decreased JJR, then the sentiment is negative. The same approach applies to negative features with decrease and increase.

This approach also applies to DTJJR superlatives with no sentiment. However, the position of the pronouns should be considered to identify the direction of the sentiment as discussed with *type B* in the case of DTJJR expressing comparison between two entities.

3.2.4 Type C with no features: JJR comparatives, DTJJR superlatives with no sentiment

In case there are no explicit features in the sentence, then surrounding terms are used to calculate the sentiment of the comparative keyword. One example of this type is

Fig. 2 Positions of entities to the comparative keywords



موبايل سوني بيتحمل اكثر من سامسونج/Sony mobile bears more than Samsung. يتحمل اكثر/bears more here describes the durability feature, but using a verb. This type is considered an implicit feature. Figure 2 shows that there are three positions entities can exist in a comparative opinion. The entities could be either on different sides of the comparative keyword, or they both could be before or after the comparative keyword.

The total polarity of all terms in the positions of (sentiment 1 + sentiment 2) is calculated, positive features are given 1 and negative features are given - 1. In case only one entity exists in the opinion, then sentiment 1 is only considered. If (sentiment 1 + sentiment 2) ≥ 0, then the sentiment is positive. If (sentiment 1 + sentiment 2) < 0, then the sentiment is negative. Same rules apply here as type C with features, positive and increasing JJR/DTJJR then the total sentiment is positive to the first entity. If positive and decreasing JJR/DTJJR, then the total sentiment is negative to the first entity and vice versa for negative with increasing and negative with decreasing.

The positions of sentiment 1 and 2 in Fig. 2 tend to contain the most descriptive terms regarding the two entities. If the whole sentence is considered, more elaboration from the opinion holder could sometimes be misleading.

3.2.5 Type D: terms that show preferences

Verbs and adjectives that show clear preferences or dislikes are collected and processed accordingly. Some

terms as اأختار/I choose, اأفضل/I prefer, have the following entity as the preferred one. Other terms as اأبتتفوق/exceed, اأتميز, اأتميز, اأتميز/distinguished, have the preceding entity as the preferred one. This solution proves to be very effective, especially with this type that is found to be the second most common one used in the collected dataset as shown in Table 3.

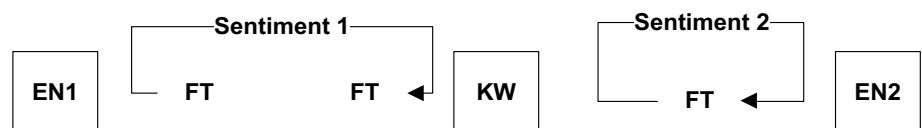
3.2.6 Type E with features: neutral comparative

The neutral comparative keywords do not imply any sentiment on their own. Identifying comparative relations direction with neutral comparatives is identifying the sentiment of two sub-sentences as shown in Fig. 3. The preferred entity is the one enclosed in the positive sub-sentence. For such keywords, the surrounding terms' sentiment needs to be identified.

The calculation is very simple but efficient. First, sentiment 1 of the first entity is calculated. If the total sum of features' sentiment > 0, then the sentiment of the sentence is positive for the first entity. If sentiment 1 < 0, then the sentiment is negative for the first entity.

If the total sum = 0, then sentiment 2 is calculated. If the polarity of the sentiment 2 > 0, then the sentiment is positive for the second entity. If the polarity of sentiment 2 < 0, then the sentiment is negative for the second entity. If both polarities are 0, then the first sentence is given the positive polarity and the first entity is the preferred one.

Fig. 3 Neutral comparative common case



3.2.7 Type E and no features: neutral comparative

In case no features exist in the sentence, then the total polarity of terms, including nouns, verbs and adjectives that are found in the lexicons are searched and their sentiment is calculated. Considering only terms in the same positions, see the features in Fig. 3.

The terms 'positive feature' and 'negative features' are used throughout the paper. Identifying the sentiment of features and terms is necessary to identify the total sentiment of the opinion. However, some factors must be considered as compound terms, negations, neutral terms etc. The next section discusses how to calculate the sentiment of features and terms within opinions.

3.3 Sentiment calculation

3.3.1 Features and terms sentiment calculations

In this research, features are explicitly stated nouns. Efforts are done with the collected features from the pros and cons lists and the four used Arabic sentiment lexicons. Features and terms identified in the lexicon as 'positive feature' and 'negative feature' could be single or compound keywords and terms. Neutral features are given a sentiment based on their surroundings of terms which give the feature its real sentiment.

When calculating the sentiment of the feature, compound terms/features from all the lexicons are always considered first. Assign the sentiment to the compound terms found in the opinion and move to the next feature. Once compound features are resolved, existing single features are resolved next.

Single term/feature that is not found in the final lexicons, is normalized, segmented and, if necessary, stemmed and its origin is searched again in the lexicon. If the feature is positive and associated with a negative verb/adjective or a negation, they are considered a negative feature. Examples: *دروس غالية*/expensive lessons, *يفتقد الامانه*/lacks honesty, *لااحترام يفتقد*/lacks respect. On the other hand, a negative feature associated with a positive verb/adjective or negation is identified as a positive sentiment. For example, *مش غالي*/not expensive, *لا تاخير*/no delays, *تتحمل الصدمات*/shock absorbers and so on.

Neutral features associated with a negative verb/adjective are considered negative features, whereas neutral features associated with a positive verb/adjective are considered positive features. In case of opinions without features, positive nouns associated with negative verb/adjectives are together given negative sentiment and vice versa for negative nouns associated with the positive verb/adjective. As adjectives describe the associated nouns and verbs identify a certain event or action that affects the nouns,

especially verbs derived from adjectives [18]. Adjectives associated with negation have their sentiment reversed. Other than that, every single term (noun, verb, adjective) is assigned its own sentiment. Negative feature/term is given -1 , while positive feature/term is given 1 .

3.3.2 Negation calculations

Negations are not only limited to 'لا/not', 'دون/no, عدم', 'مفیش/there is no, محصلش/didn't happen and so forth. When a negation is directly associated with the comparative keyword in the sentence, the sentence's sentiment is reversed. In case the negation is associated with a certain feature or term, then it only inverts the associated term or feature's sentiment. However, one must consider certain terminologies that have negations with a totally different meaning. For example, "مفیش احسن من العصير الطبيعي", there is no better than fresh juice", the negation "مفیش/there is no" has a positive meaning and does not inverse the sentiment of احسن/better.

There is a different type of negation that acts similar to the rules of case 3 in Fig. 1. It includes certain adjectives and verbs that show an increase or decrease like (يزود, يقال), they are given (+) or (-) sentiment, respectively. These terms are not assigned positive or negative sentiment on their own, but they reverse or preserve the associated sentiment. In the case of (-) with positive or negative features, the sentiment is reversed. In the case of (+) with the positive or negative features, they keep the sentiment the same.

The following example illustrates how sentiment and negations are calculated with neutral comparatives, using the neutral keyword 'اما/on the other hand'.

Example 1: *موبايل سامسونج شاشته ضعيفه اما سوني جميل*/mobile Samsung screen is weak on the other hand Sony is beautiful.

- Sentiment 1: a total of all features from entity 1 (موبايل سوني/mobile Samsung) to the comparative keyword (اما)
- Sentiment 2: a total of all features from the comparative keyword (اما) to entity 2 (موبايل سوني/mobile Samsung)

Sentiment 1: 'شاشته/screen' is the feature, if the features cannot be found in the lexicon directly, it is segmented, normalized and stemmed. After segmenting, the term will be 'شاشه'. Searching for this term in the lexicon, it will turn out to be neutral. The following word is a negative word (ضعيف/weak) associated with a neutral feature which results in a negative sentiment altogether. As there are no

other features, then sentiment 1 is -1 . This ends the calculation and the sentence is given a negative sentiment.

A similar example, but without the existence of any features in the opinion is: *موبايل سامسونج مش حلو اما سوني جميل*/mobile Samsung is not good; on the other hand, Sony is beautiful. There are no explicit features in this case **sentiment 1**: a total of all terms from entity 1 (*موبايل سوني*) to the comparative keyword (*اما*).

The following word is a negation (not/*مش*) associated with a positive word (*حلو*/good) which results in a negative sentiment altogether. According to the rules, positive term associated with negative or negation is a negative word. As there are no other terms, the sentiment 1 is -1 . This ends the calculation and the sentence is given a negative sentiment.

4 Evaluation and experiment results

For evaluating the proposed technique, a data set of comparative opinions was collected. The dataset consisted of all five types of comparative keywords. The processing rules explained in Sect. 3 were applied to the data. Then, the precision, recall and *f*-measure were calculated for positive and negative opinions separately.

4.1 Data preparation

The preparation of data includes the collection and cleaning of the dataset used for testing the proposed technique. It also includes the collection and preparation of the lexicons used for calculating the sentiment of features and terms in the sentences and the categorization of different comparative keywords.

4.1.1 Dataset

A total of 830 comparative opinions were collected for testing the proposed technique. These opinions were both Egyptian dialect opinions and Modern Standard Arabic (MSA). Some of the opinions were collected from the dataset used in [21]. More opinions were collected from public Facebook pages, Twitter and public blogs. The dataset size is close to the size of other datasets used in comparative sentiment analysis research; for instance, the work introduced in [9] used a dataset of 837 opinions. In [4], a dataset that consisted of 870 non-equal and superlative comparative opinions was used.

The sentences were annotated and given their sentiment by three Arabic native speakers. They were all non-equal gradable comparative and superlative sentences. This means there was always a preferred entity. In the case of non-equal gradable, a sentence was given a positive

Table 2 Distribution of positive and negative opinions in the dataset

Opinions	Positive	Negative	Total
Comparative	260	25	285
Superlative	497	48	545
Total	757	73	830

Table 3 The percentage of the different comparative opinions' types in the dataset

Comparative types	Opinions distribution (%)
Type A	77
Type B	6
Type C	6
Type D	7
Type E	4

sentiment if the first entity in the sentence was the preferred one. Otherwise, the sentiment of the sentence was considered negative, which means the second entity was the preferred one. In case of superlative, a sentence was considered positive if the sentiment was positive toward this single entity. Otherwise, the sentence is given negative if the sentiment is negative toward the single entity. The collected opinions included one relation and two entities or one entity in case of superlatives. Table 2 shows the distribution of the comparative opinions in the dataset regarding their sentiment. In Table 3, the comparative opinions of different comparative keywords types are displayed as a percentage of the total data.

4.1.2 Preprocessing of comparative opinions

Comparative opinions from the dataset were cleaned before any processing. The cleaning included removing special characters, noise, repeated letters and translating any foreign terms used. During the opinions processing, some language related tasks could be needed for identifying terms that were not directly found in the lexicon. AraNLP tool [22] was used for this purpose. This tool provided a wide range of language processing functions including segmentation and tagging, normalization, stemming and light stemming. It is an open source Java-based tool that is based on Stanford tagger [23]. In case, no matches were directly found with the lexicon, terms were first normalized, then segmented and lightly stemmed. After each step, the non-matched terms were submitted again until a match was found. The search then stopped and moved to the next term.

4.1.3 Lexicons used

Several sentiment Arabic lexicons were used and merged into one large final lexicon. The first one is ArSenl [5], it is a large-scale Arabic lexicon. ArSenl contains terms with their sentiment score and confidence. The second one is NileULex [6] is an MSA and Egyptian dialect lexicons consisting of single and compound terms. The third lexicon is the Arabic translation of Bing Liu's lexicon [7, 8] which is a translation of the English lexicon presented in [24]. Finally, the Arabic hashtag lexicon (dialectal) in [7, 8] was also used where it contained some dialectal Arabic terms. These lexicons contained Arabic and Egyptian terms with their polarity either positive or negative. They were scanned to remove redundancies. In addition to these lexicons, the pros and cons lists presented in [9] were also used. These two lists contained a combination of features and comparative keywords gathered from pros and cons opinions of various platforms. In [9], the two lists were used in a way that the features were considered negative if the combination of comparative keywords and features were more associated with the cons list than the pros list. The same applies to positive features to be more associated with the pros list than the cons list.

The two lists were used to further enrich the combined lexicon by extracting the adjectives and terms that are used to express negativity or positivity sentiment toward features. For instance, in the cons list 'جوده الصورة ضعيفه' /poor picture quality' if the word 'ضعيفه/poor' does not exist in the final lexicon, it is added as a negative word. 'جوده الصورة/picture quality' On the other hand, it can be viewed as a positive feature. Finally, the two terms: 'poor + picture quality' are added as a compound term with negative sentiment.

The reason for that is to solve the context-dependent problem. For example, the term 'متشدد/strict' alone is a negative term. However, the term 'متشدد/strict' + 'اخلاقيات السائق/driver's manners' has a positive sentiment. This allows the algorithm to assign 'متشدد/strict' + 'اخلاقيات السائق/driver's manners' together with a positive sentiment, and in another context, the term 'متشدد/strict' would still be assigned a negative sentiment.

The two lists were scanned for terms that occur frequently in the Arabic opinions; they were then collected and translated. The lexicon was also extended by considering some additional features that are not included in these lists. For instance, a feature like 'مكسبات الطعم/taste additives', 'مكسبات اللون/colour additives' was considered a negative feature and added to the lexicon. Terms that do not exist in any lexicon were also added to our lexicon, including additional MSA and Egyptian colloquial terms. For some other features that are usually part of an item like speakers

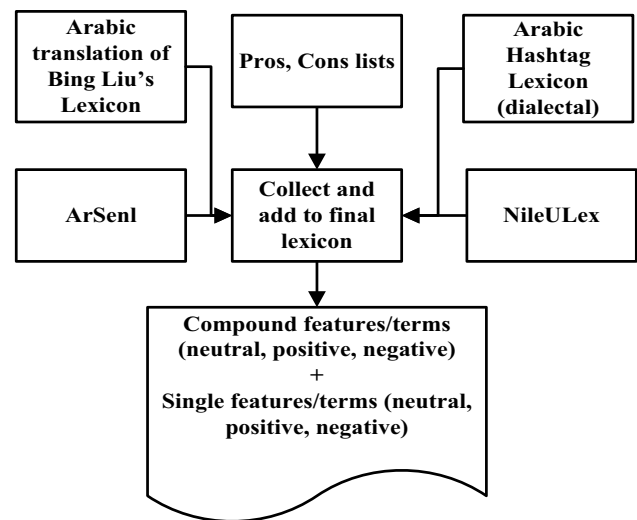


Fig. 4 Collected final lexicon

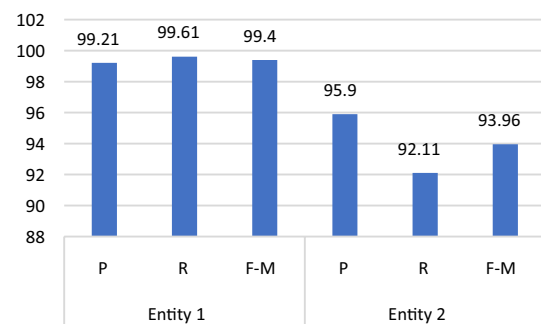


Fig. 5 Result in terms of precision, recall, f -measure

or a battery that have a neutral sentiment in the lexicon, their sentiment must be identified through supporting terms around them during the processing.

Figure 4 shows all used lexicons in this paper, including the two lists obtained from [9]. Identifying the sentiment for features from the two lists was done by three Arabic native speakers. Deciding the sentiment of a feature was done by observing how they were used in the collected data. Generally, if the more/high of a feature is positive, then this feature is positive and vice versa.

4.2 Experiment setting and results

Three main measures were used: precision, recall and f -measure. Precision (P) measures how much of the retrieved data is correct while Recall (R) measures how much is retrieved of the correct data. F -measure (F -M) is considered an approximate average between the two first measures.

Table 4 Precision, recall and *f*-measure of the all five comparative keywords' categories

	Entity 1			Entity 2		
	<i>P</i>	<i>R</i>	<i>F-M</i>	<i>P</i>	<i>R</i>	<i>F-M</i>
Type A	99.67	100	99.84	100	96.16	98.05
Type B	100	100	100	100	100	100
Type C	93.34	97.68	95.47	85.72	66.67	75.01
Type D	100	100	100	100	100	100
Type E	94.45	89.48	91.9	81.82	90	85.72

The best results are obtained with type A, B and D. Their results are added in bold for display

General observations show that approximately 90% of the gathered opinions have positive sentiment. In other words, opinion holders tend to favor the first entity. This reflects the human nature in comparing entities, by presenting their favorite one first. To avoid this problem, precision, recall and *f*-measure were calculated for the positive and negative sentiment sentences separately.

First the total sentiment of opinions, including all comparative keyword's types, is presented in Fig. 5. Results for entity 1 is very high, entity 2 is a little lower, but still a very good percentage. The reason for this is that it is usually easier to identify the positive sentiment, as negative sentiment could be implied in the opinion or not clearly stated by the opinion holder.

Regarding the validity of the proposed categories, the result of each type is displayed separately in Table 4. Results in Table 4 show that the highest results are achieved with type A, B and D. Type D approach is very effective: these terms have a very direct and strong sentiment. If someone says that they choose/hate/prefer entity 1, they state a clear sentiment about entity 1. Such opinions rarely have a different sentiment as in case of sarcasm. Type B is very similar to type A. The used approach is very efficient, especially that it is mainly used for expressing superlative opinions. In superlative, the sentiment is toward one entity and it makes it easier to identify the sentiment of the opinion. The results for Type C, JJR/DTJJR comparatives with no sentiment, and type E neutral comparatives are a little lower than the others because the comparative keywords are not decisive on their own. When the comparative keywords have a sentiment, the estimated sentiment is usually more accurate. For comparative keywords with no sentiment, looking for the surrounding features and terms offers a way to identify the sentiment. The proposed technique achieves high results in all five types. The quality of the used lexicon can affect the results; thus, in this paper a combination of lexicons is collected, and additional terms are added

to keep the result from being highly affected by unidentified terms.

5 Conclusion

This paper proposes a technique for sentiment analysis of comparative Arabic opinions. To the best of our knowledge, this is the first work to address this problem with the Arabic language. The proposed technique shows high results of 99% and 94% *f*-measure of correctly identified directions of the comparative relations. It uses available resources for the best results and limits the human interaction to the initial steps of collecting lexicons and categorizing comparative keywords, which shows potential for fully automating the process. It categorizes comparative keywords used to express opinions into five categories and processes them separately. Separating the analysis of each category achieves better results as each category has its own characteristics. The technique uses the linguistic structure of comparative opinions while considering how opinion holders tend to express their subjective opinions in a comparative manner. The technique also offers a way of addressing comparative opinions with no obvious sentiment and no features which form a good portion of comparative expressions found online.

The future work of this research is to address implicit features for the sentiment analysis. Implicit features are inferred from the text and not directly stated in the text. Studying how to analyze the sentiment of comparative opinions with more than one relation and more than two entities are set for future work. Moreover, addressing the continuous updating of the used lexicon by adding new features that do not exist in it is a way to keep the high results for all domains. The supervised technique can also be used for identifying the sentiment of comparative opinions guided by the work presented in this paper for selecting features to train algorithms. Finally, a larger comparative opinion dataset is to be collected for testing the proposed technique on a larger scale.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Human and animal rights This article does not contain any studies with animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Jindal N, Liu B (2006) Identifying comparative sentences in text documents. In: Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval—SIGIR'06, p 244
- Jindal N, Liu B, Bing L (2006) Mining comparative sentences and relations. In: Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval—SIGIR'06, vol 21, no 2, pp 1331–1336
- Liu B (2015) Sentiment analysis: mining opinions, sentiments, and emotions. Cambridge University Press, Cambridge
- Xu K, Liao SS, Li J, Song Y (2011) Mining comparative opinions from customer reviews for competitive intelligence. *Decis Support Syst* 50(4):743–754
- Badaro G, Baly R, Hajj H (2014) A large scale arabic sentiment lexicon for Arabic opinion mining. In: Arabic natural language processing workshop co-located with EMNLP 2014, Doha, Qatar, pp 176–184
- El-Beltagy SR (2016) NileULex: a phrase and word level sentiment lexicon for Egyptian and modern standard Arabic. In: Proceedings of LREC 2016, no. April, pp 2900–2905
- Mohammad SM, Salameh M, Kiritchenko S (2016) How translation alters sentiment. *J Artif Intell Res* 55:95–130
- Salameh M, Mohammad SM, Kiritchenko S, Canada C (2015) Sentiment after translation: a case-study on Arabic social media posts. In: *Naacl* 2015, pp 767–777
- Ganapathibhotla M, Liu B (2008) Mining opinions in comparative sentences. In: Proceedings of the 22nd international conference on computational linguistics—COLING'08, vol 1, no. August, pp 241–248
- Varathan KD, Giachanou A, Crestani F (2017) Comparative opinion mining: a review. *J Assoc Inf Sci Technol* 68(4):811–829
- Qazi A, Raj RG, Hardaker G, Standing C (2017) A systematic literature review on opinion types and sentiment analysis techniques: tasks and challenges. *Internet Res* 27(3):608–630. <https://doi.org/10.1108/IntR-04-2016-0086>
- El-Halees A (2012) Opinion mining from Arabic comparative sentences. In: 13th International Arab conference on information technology ACIT ASSESSMENT, pp 265–271
- Kessler W, Kuhn J (2013) Detection of product comparisons—how far does an out-of-the-box semantic role labeling system take you? In: Proceedings of the 2013 conference on empirical methods in natural language processing, no. October, pp 1892–1897
- Boudad N, Faizi R, Thami ROH, Chiheb R (2017) Sentiment analysis in Arabic: a review of the literature. *Ain Shams Eng J*. <https://doi.org/10.1016/j.asej.2017.04.007>
- Chen T, Xu R, He Y, Wang X (2017) Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Syst Appl* 72:221–230
- Wang W, Xin G, Wang B, Huang J, Liu Y (2017) Sentiment information extraction of comparative sentences based on CRF model. *Comput Sci Inf Syst* 14(3):823–837
- Gao S, Tang O, Wang H, Yin P (2018) Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *Int J Hosp Manag* 71:19–32
- Momani MM, Altaher AM (2015) A contrastive analysis of English and Arabic for translation students. *Int J Engl Lit Cult* 3(11):280–284
- Taine-cheikh C (2018) Expressiveness and evaluation in Arabic. *Stud Lang* 42(1):81–113. <https://doi.org/10.1075/sl.00004.tai>
- El Defrawi M, Salah M, Abd Al-Aziz A, Eldin AS (2017) Comparative relation extraction from Arabic opinions. *Int J Comput Sci Inf Secur* 15(10)
- 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, Cham, pp 23–34. https://doi.org/10.1007/978-3-319-18117-2_2
- Althobaiti M, Kruschwitz U, Poesio M (2014) AraNLP: a Java-based library for the processing of Arabic text. In: Proceedings of the ninth international conference on language resources and evaluation (LREC'14), pp 4134–4138
- Toutanova K, Klein D, Manning CD, Singer Y (2003) Feature-rich part-of-speech tagging with a cyclic dependency network. In: Proceedings of the 2003 conference of the North American chapter of the association for computational linguistics on human language technology—NAACL'03, vol 1, pp 173–180
- Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the 2004 ACM SIGKDD international conference on knowledge discovery and data mining—KDD'04, p 168

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