



# Journal of Applied Sciences

ISSN 1812-5654

**science**  
alert

**ANSI***net*  
an open access publisher  
<http://ansinet.com>



## Research Article

# A Hybrid Method for Arabic Educational Sentiment Analysis

A. Eesee and N. Omar

Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor Darul Ehsan, Malaysia

### Abstract

**Background:** Subjectivity detection is the process of identifying whether a sentence contains an opinion or not. Subjectivity detection is considered as a prior task before conducting sentiment analysis in which the error rate could be decreased. Education is one of the domains that brought many attentions where the student reviews are being analysed to enhance the learning process. Usually, student reviews contain several objective sentences in which the students are expressing facts about their activities. Therefore, it is necessary to accommodate subjectivity detection before classifying polarity of these reviews. **Objective:** This study proposes a subjectivity detection method using specific features for identifying opinionated sentences. **Methodology:** Consequentially, a sentiment classification based on lexicon-based and Support Vector Machine (SVM) classifier is being conducted to classify the polarity. **Results:** Results showed an outperformance of classifying polarity when the proposed subjectivity detection method is used prior. **Conclusion:** It concluded that superior performance of sentiment classification when using subjectivity detection as a prior task.

**Key words:** Educational sentiment analysis, subjectivity detection, lexicon-based approach, support vector machine

**Received:** March 07, 2016

**Accepted:** March 25, 2016

**Published:** April 15, 2016

**Citation:** A. Eesee and N. Omar, 2016. A hybrid method for arabic educational sentiment analysis. J. Applied Sci., 16: 216-222.

**Corresponding Author:** N. Omar, Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor Darul Ehsan, Malaysia

**Copyright:** © 2016 A. Eesee and N. Omar. This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

**Competing Interest:** The authors have declared that no competing interest exists.

**Data Availability:** All relevant data are within the paper and its supporting information files.

## INTRODUCTION

Companies and organizations nowadays have an essential demand in verifying its services or products. This demand relies on the consumer's perspectives toward such services or products. Thus, knowing the consumer's opinions has become a significant task to improve the quality of services or products by collecting customer's feedback and classifying it into classes such as negative, positive or neutral. With the dramatic expansion of the World Wide Web, there is a significant opportunity to discover opinion over the internet by analyzing user's reviews. Such reviews lie on social media, blogs and website's comments in a textual mode. The process of analyzing these reviews in order to gain opinions called sentiment analysis<sup>1</sup>. Sentiment analysis consists of two kinds of classifications first is called the binary classification where an opinion lies on one of two categories (e.g., positive or negative), while the second one is called multiclass classification where an opinion lies on one of multiple categories (e.g., strongly agree, agree, fair, disagree and strongly disagree)<sup>2</sup>. On other hand, there is a sub-task within sentiment analysis called subjectivity classification where documents are analysed in order to identify the document that hold opinion<sup>3</sup>.

The traditional sentiment analysis approaches are mainly depend on classifying polarity of opinions<sup>4</sup>. This means that these approaches consider handling opinions only. However, dealing with large documents or so-called document-level sentiment analysis is being limited in this case where numerous sentences could be objective or factual sentence which means that they are not emphasizing opinions<sup>5</sup>. Therefore, sentence-level sentiment analysis appears to be more suitable in this case, in which every single sentence is being analysed in terms of subjectivity<sup>6</sup>. The process of identifying whether the sentence is subjective or objective is called subjectivity detection.

In fact, many researchers have addressed the process of identifying subjective sentences<sup>7-9</sup>. According to Mihalcea *et al.*<sup>10</sup>, the problem of identifying subjective sentences is considered more difficult than the problem of identifying its polarity in which any attempts to enhance the identification of subjectivity would significantly enhance the process of sentiment classification. In their studies, researchers have utilized several features such as POS tagging which has the ability to identify adjectives. However, there is still room for improvements in terms of utilizing new features.

Several sentiment analysis approaches have been presented for several domains including social media<sup>11</sup>, movie reviews<sup>12</sup> and business reviews<sup>13</sup>. Presently, some researchers

tend to use sentiment analysis for educational domain. Educational domain appears to be more challenging in which students can express several objective sentences. Therefore, there is a vital demand to accommodate subjectivity detection prior to the sentiment polarity classification for the educational domain. In addition, there is an essential need to provide an automatic approach to attain annotated data for supervised approach. Lexicon-based approach has been proposed for this purpose where a lexicon that contains numerous words with their class labels (i.e., positive or negative) could be used to label the data.

This study proposes a subjectivity detection approach using specific features for identifying opinionated sentences in Arabic student reviews. Consequently, a sentiment analysis based on lexicon-based approach and SVM classifier is being performed to classify the reviews into their polarities.

Many researchers have addressed the problem of identifying subjective sentences. For example, Abdul-Mageed *et al.*<sup>14</sup> have proposed a system called SAMAR which aims to perform subjectivity and sentiment analysis tasks for Arabic language using social media data. In fact, the authors have addressed the problem of Arabic language which represented on the two kinds of Arabic; formal and informal Arabic language. Even the informal Arabic contains several dialects. This makes the process of subjectivity detection and sentiment analysis for Arabic language is a challenging task. In particular, the authors have utilized POS tagging feature with Support Vector Machine (SVM) classifier in order to classify the subjective sentences as well as, sentiment analysis.

On the other hand, other researchers have concentrated on educational sentiment analysis for instance, Guitart *et al.*<sup>15</sup> have proposed an automatic approach for analysing posts forum of communities for students and teachers in open University of Catalonia (UOC) in order to classify opinions toward enhancing the process of learning. The proposed method has intended to utilize multiple language resources including WordNet, FreeLing and adopted dictionary that has been produced by the authors. Basically, POS tagging has been used in order to identify the tag for each word. Then, based on the fact that most opinions contain adjectives and adverbs, the proposed method will extract such tags and identify its polarity using the external resources. The external resources provide synonyms and homonym for each adjective and adverb.

Other languages such as Arabic have also targeted the study efforts of educational sentiment analysis. For instance, El-Halees<sup>16</sup> has proposed a mining method in order to extract and classify opinion of students toward enhancing

course evaluation. In fact, the author used a collection of reviews that have been collected from forums, discussion groups and blogs. Such reviews contain student’s opinion toward specific courses. The proposed method contains two main tasks. First is document-level sentiment classification, which aims to classify the polarity of documents. Second is sentence-level, which aims to identify the polarity of certain opinion whether positive or negative for each sentence.

**MATERIALS AND METHODS**

The proposed method consists of five main phases as shown in Fig. 1 namely; dataset, pre-processing, subjectivity detection, lexicon-based approach and classification.

**Corpus collection:** The first task aims to identify the source for the data collection. According to Guitart *et al.*<sup>15</sup>, there is no benchmark dataset for student reviews which makes most of the researchers, who concentrate on educational sentiment analysis, tend to collect student’s reviews manually from blogs or social media posts. Therefore, the data has been collected from reviews (<http://www.guidetoonlineschools.com/>) which is an online web service that collect vast amount of online student reviews from 1,723 colleges. In order to highlight the details of dataset in terms of reviews quantity, sentence quantity and other information (Table 1).

The second task which is translation task in which the reviews are being translated from English to Arabic. In fact, similar to English reviews, there is no a benchmark dataset for Arabic student reviews<sup>16</sup>. In addition, there is no an available source that enables reviewing students’ posts in Arabic. This is due to the limited access of Arabic colleges or universities that provide access only for their students. For this manner, the collected review, which are in English language, are being translated into Arabic using Google translator.

**Pre-processing:** In this phase, three sub-tasks are being performed in order to turn the data into suitable form. First task is sentence splitting, since the reviews consist of multiple sentences thus, each sentence is isolated separately in order to treat them properly. In fact, this task is crucial for the subjectivity detection in which the sentence-level sentiment classification is being applied to classify the opinionated sentences. Second task is normalization in which the irrelevant data is being eliminated in order to obtain pure text. Such data consists of three types; numbers, punctuation and stopwords. Third task is stemming which aims to turn the words into their roots by removing the derivational inflections<sup>17</sup>.

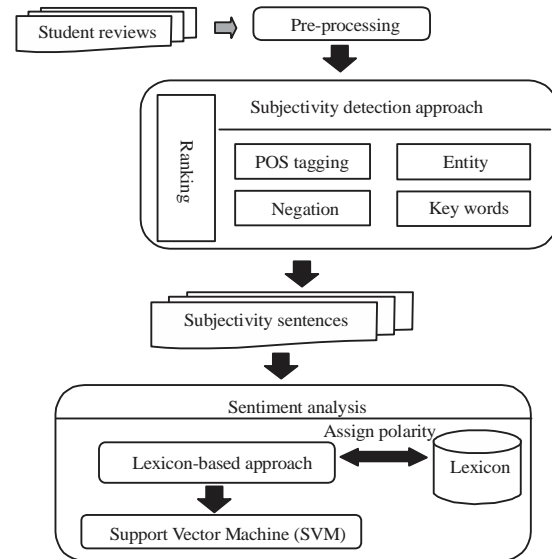


Fig. 1: Proposed method

Table 1: Dataset details

Characteristics	Quantity
No. of colleges	1,723
No. of reviews	900
No. of sentences	5436
No. of words	32616

**Subjectivity detection:** As mentioned earlier, this phase is very crucial due to it reflects the main objective of this study in which the sentences are being classified into subject (i.e., opinionated) sentences and objective (i.e., not opinionated) sentences. For this manner, four features have been developed in order to detect the subjectivity sentences. These features consist of POS tagging, entity, key words and negation which are described in the following sub-sections.

**Pos tagging:** One of the common feature that has been widely used for this purpose in the literature is the Parts Of Speech (POS) tagging. The POS tagging has the ability to identify the adjectives, adverbs and pronouns which usually yield an opinion<sup>12</sup>.

**Entity feature:** Another feature is the entity where such feature aims to identify entities such as person name (e.g., Prof. Adam), organization name (e.g., Stanford University) and location (e.g., convocation hall). In fact, these entities usually represent the targeted objects that the opinion expressed about them for example, ‘Abilene university is amazing’ the word ‘Abilene’ is an object that has been targeted by an opinion which is ‘amazing’.

**Key words:** The key words feature which aims to utilized key words for specific domain in order to identify the opinionated sentences such as the word ‘stock’ in financial domain and ‘scenario’ in Movie reviews. Similarly, the education domain contains several key words such as ‘assignment’, ‘university’, ‘college’, ‘faculty’ and others.

**Negation:** Furthermore, the negation feature is a useful feature to detect subjectivity sentences where it usually refers to an opinion such as ‘I don’t like physics course’.

**Ranking:** Finally, a ranking approach is being applied in order to provide a weight for each sentence. This weight refers to the presence of the mentioned features. In order to understand the ranking mechanism, assume a student review obtained from the dataset be as in Table 2.

After applying the sentence splitting, normalization and stemming the sample review can be shown as in Table 3.

Hence, feature can be represented for each sentence by annotating each word as in Table 4.

As shown in Table 4, each word has been annotated with the exact feature. In order to determine which sentence yield an opinion, a ranking approach has been used in order to rank the sentences with a value that refers to the probability of yielding an opinion. Table 5 depicts the values of each feature.

As shown in Table 5, each feature has been assigned with a value that indicates its importance. These values are adjusted based on priority assumption. For example, the entity feature has the highest value due to it represents the object (e.g., Data Mining) that would be targeted by the opinion. Hence, there is a minimum threshold to decide whether a given sentence is opinionated or not. This study has determined the threshold as 0.5 in which the sentence that has a ranking value that equal or greater than the threshold will be considered as an opinionated sentence. For example, a sentence with entity and adjective features will get a rank value that equivalent to the threshold and it identically represent an opinion (e.g., Data mining bad). Whilst, a

sentence with key word and entity features (e.g., Data structure course) or key words and adjective features (e.g., bad grade), is below the minimum threshold. Therefore, it cannot be represented as an opinion regarding the ambiguity. Finally, negation feature has been assigned with 0.0 value because it neither increase nor decrease the probability of containing an opinion. However, Table 6 apply the ranking approach on the sentences.

As shown in Table 6, each sentence has been ranked. Note that, the first sentence is below the threshold 0.5 thus, it would be eliminated. Whereas the other sentence is higher than the threshold thus, it will be remained as opinionated sentence. Table 7 shows these results.

As a shown in Table 7, the opinionated or subjective sentence is being identified. The following section illustrates the next phase which is lexicon-based in which the opinionated sentences will be labelled.

**Lexicon-based approach:** Generally, any supervised machine learning technique rely on a predefined or annotated data in which the classes are being declared (i.e., positive and negative)<sup>18</sup>. This is because, the classifier has to be trained on the such predefined data in order to discriminate the instances. Since, the data has been collected manually thus, the class label is not provided. Therefore, the lexicon-based approach will be used in order to provide the class label for each sentences. According to Feldman<sup>19</sup> the sentiment lexicon is the most sensitive resource for most sentiment analysis algorithms. For this purpose, an Arabic lexicon has been used which has been introduced by Khalifa and Omar<sup>20</sup>. Such lexicon contains vast amount of Arabic adjectives and adverbs with their synonyms and polarity.

**Classification:** After acquiring the subjective sentences which have been labelled using the lexicon-based, the process of classifying these sentences into their polarities is being performed using Support Vector Machine (SVM) classifier. Using unigram and bigram representation for the features,

Table 2: Sample review

Sample review	Meaning
I am a postgraduate student. In the last semester I took Data Mining course it was very interesting	

Table 3: Pre-processed review

Sentence No.	Review	Meaning
1		Graduate student
2		Last semester take Data Mine course interest

Table 4: Annotating the words based on the proposed features

Sentence No.	Review	Translation
1		Graduate/JJ student/KY
2		Last/JJ semester/KY take/VB Data/EN Mine/EN course/KY interest/JJ

SVM is being trained on the data by dividing such data into 80% training and 20% testing. One of the key features behind SVM lies on its ability to handle high dimensionality of features<sup>21</sup>.

**RESULTS**

Basically, evaluating results are divided into two parts; subjectivity detection evaluation and sentiment classification evaluation. First, to evaluate the results of subjectivity detection method, a group of expert in Arabic language has been consulted to evaluate the resulted sentences. The evaluation aimed to identify subjective sentences, objective sentences and unsure sentences. Hence, the precision can be computed as follow:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

where, TP is the No. of correctly identified subjective sentences, FP is the No. of objective sentences. Whilst, recall can be computed as follow:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

where, TP is the No. of correctly identified subjective sentences and FN is the unsure sentences. Now it is possible to compute F-measure as follow:

$$F1 = \frac{2Pr \times Re}{Pr + Re} \tag{3}$$

Table 8 depicts the results of subjectivity detection method based on precision, recall and F-measure.

As shown in Table 8, the proposed subjectivity detection has attained 614 correct subjective sentences out of 740 total No. of sentences. This leads to 85% of precision, 96% of recall and 90% of F-measure. Such results demonstrate the use of the proposed features in which POS tagging, negation, entity and key words features have been combined with a rank approach.

Second, to evaluate the sentiment classification results, precision, recall and F-measure are being used too in which TP is the No. of correctly classified instances, FP is the No. of incorrectly classified instances and FN is the No. of instances that were not being classified. Thus, using Eq. 1, 2 and 3, the precision, recall and F-measure can be computed. Table 9 depicts the results of sentiment classification using SVM and SVM with subjectivity detection method.

As shown in Table 9, in both applications of SVM, unigram has the superior results compared to bigram. In this manner, SVM using unigram has obtained 78% of F-measure and SVM

Table 5: Ranking weights

Feature	Value
EN	0.3
JJ or RB	0.2
KY	0.1
NEG	0.0

Table 6: Sentence ranking

Sentence	Translation	Patterns	Rank
	Graduate/JJ student/KY	KY (0.1)+JJ (0.2)	0.3
	Last/JJ semester/KY take/VB	KY (0.1)+JJ (0.2)	0.6
	Data/EN Mine/EN course/KY interesting/JJ	+ EN (0.3)	

Table 7: Subjectivity detection

Sentence	Review	Translation	Class
1		Graduate student	Objective
2		Last semester took Data Mining course interesting	Subjective

Table 8: Results of subjectivity detection method

No. of subjective (TP)	No. of objective (FP)	No. of unsure (FN)	Total instances	Precision	Recall	F1
614	107	19	740	0.85	0.96	0.90

Table 9: Classification results

Feature representation	Precision	Recall	F1
<b>SVM results</b>			
Unigram	0.85	0.75	0.78
Bigram	0.78	0.75	0.76
<b>SVM with subjectivity detection results</b>			
Feature representation			
Unigram	0.90	0.82	0.84
Bigram	0.86	0.78	0.80

with subjectivity detection using unigram has obtained 84% of F-measure. This can be justified by the larger possibility of single word to occurred in a sentence than the two words.

On the other hand, the results of SVM with subjectivity detection has outperformed the application of SVM without subjectivity detection by achieving 90, 82 and 84% of precision, recall and F-measure. This can demonstrate the use of subjectivity detection which aims to eliminate the factual sentences. Such factual sentences are being usually classified incorrectly. This is because the class of these sentences are neutral while the data could be labelled into positive and negative or using numeric labels (1, 2, 3, 4 and 5). In such case, these factual sentences will be incorrectly classified which can increase the error rate.

## DISCUSSION

In terms of educational sentiment polarity classes, the study of El-Halees<sup>16</sup> has obtained 76% of F-measure for classifying the polarity of student reviews in Arabic language using SVM classifier. In comparison, the proposed SVM with subjectivity detection method achieved 84% F-measure. This can demonstrate the usability of the subjectivity detection method with other approaches and datasets.

A comparison with other previous studies is not possible due to the different dataset and the language used. However, in terms of subjectivity detection for English, the study of Pang and Lee<sup>22</sup> has obtained 87.15% of accuracy in which SVM classifier has been employed to classify subjective sentences from movie reviews. In addition, the study of Barbosa and Feng<sup>7</sup> has obtained 81.9% of accuracy for detecting subjective sentences from tweets using SVM. Furthermore, the study of Lin *et al.*<sup>9</sup> has gained 75.6% of F-measure for detecting subjective sentences from new documents in which a Bayesian model is being used for such classification. On the other hand, for the Arabic language, the study of Mourad and Darwish<sup>23</sup> has obtained nearly 76.43% of F-measure when classifying subjective sentences in Twitter using POS tagging and stemming features. Finally, Abdul-Mageed *et al.*<sup>14</sup> has achieved 84.36% of accuracy for detecting Arabic subjective sentences in social media using SVM. Apparently, the proposed subjectivity detection appears competitive as compared to the baseline studies by achieving 90% of F-measure.

## CONCLUSION

This study proposed a subjectivity detection method for student reviews in Arabic based on specific feature set

including POS tagging, key words, entity feature and negation. In addition, a ranking approach has been used to provide weight for each sentence in order to judge whether the sentence is objective or subjective. Consequently, the sentiment classification has been performed using lexicon-based approach with SVM classifier. Results showed a superior performance of sentiment classification when using subjectivity detection as a prior task. However, the main drawback of this study lies on the dataset used which was translated from English language into Arabic where the translation may not accurate. Thus, examining the subjectivity detection method with a real-time Arabic student reviews could be addressed in future study.

## ACKNOWLEDGMENT

This study is supported by the University Kebangsaan Malaysia (UKM) and funded by research grant DPP-2015-FTSM.

## REFERENCES

1. Pang, B. and L. Lee, 2008. Opinion mining and sentiment analysis. *Found. Trends Inform. Retrieval*, 2: 1-135.
2. Tan, S., X. Cheng, Y. Wang and H. Xu, 2009. Adapting Naive Bayes to Domain Adaptation for Sentiment Analysis. In: *Advances in Information Retrieval*, Boughanem, M., C. Berrut, J. Mothe and C. Soule-Dupuy (Eds.). Springer, Berlin, Heidelberg, ISBN: 978-3-642-00957-0, pp: 337-349.
3. Wiebe, J.M., R.F. Bruce and T.P. O'Hara, 1999. Development and use of a gold-standard data set for subjectivity classifications. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, June 20-26, 1999, University of Maryland, College Park, Maryland, USA., pp: 246-253.
4. Stavrianou, A. and J.H. Chauchat, 2012. Opinion mining issues and agreement identification in forum texts. *Atelier FODOP'08*, pp: 51-58.
5. Vinodhini, G. and R.M. Chandrasekaran, 2012. Sentiment analysis and opinion mining: A survey. *Int. J. Adv. Res. Comput. Sci. Software Eng.*, Vol. 2.
6. Prabowo, R. and M. Thelwall, 2009. Sentiment analysis: A combined approach. *J. Inform.*, 3: 143-157.
7. Barbosa, L. and J. Feng, 2010. Robust sentiment detection on twitter from biased and noisy data. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, August 2010, Stroudsburg, PA., pp: 36-44.
8. Kim, S.M. and E. Hovy, 2005. Automatic detection of opinion bearing words and sentences. *Proceedings of the 2nd International Joint Conference on Natural Language Processing*, October 11-13, 2005, Jeju Island, Korea, pp: 61-66.

9. Lin, C., Y. He and R. Everson, 2011. Sentence subjectivity detection with weakly-supervised learning. Proceedings of the 5th International Joint Conference on Natural Language Processing, November 8-13, 2011, Chiang Mai, Thailand, pp: 1153-1161.
10. Mihalcea, R., C. Banea and J. Wiebe, 2007. Learning multilingual subjective language via cross-lingual projections. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, June 2007, Prague, Czech Republic, pp: 976-983.
11. Gigieh, A., M. Al-Kabi, I. Alsmadi, H. Wahsheh and M. Haidar, 2008. Building and evaluating an opinion analysis tool for standard and colloquial Arabic language. Yarmouk University, Jordan.
12. Abbasi, A., H. Chen and A. Salem, 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. ACM Trans. Inform. Syst., Vol. 26. 10.1145/1361684.1361685
13. Elhawary, M. and M. Elfeky, 2010. Mining Arabic business reviews. Proceedings of the IEEE International Conference on Data Mining Workshops, December 13, 2010, Sydney, NSW, pp: 1108-1113.
14. Abdul-Mageed, M., M. Diab and S. Kubler, 2014. SAMAR: Subjectivity and sentiment analysis for Arabic social media. Comput. Speech Lang., 28: 20-37.
15. Guitart, I., J. Conesa, L. Villarejo, A. Lapedriza, D. Masip, A. Perez and E. Planas, 2013. Opinion mining on educational resources at the open university of Catalonia. Proceedings of the 7th International Conference on Complex, Intelligent and Software Intensive Systems, July 3-5, 2013, Taichung, pp: 385-390.
16. El-Halees, A., 2011. Mining Opinions in User-Generated Contents to Improve Course Evaluation. In: Software Engineering and Computer Systems, Zain, J.M., W.M. bt Wan Mohd and E. El-Qawasmeh (Eds.). Springer, New York, ISBN: 9783642221910, pp: 107-115.
17. Alemayehu, N. and P. Willett, 2002. Stemming of Amharic words for information retrieval. Literary Linguistic Comput., 17: 1-17.
18. Abdul-Mageed, M., M.T. Diab and M. Korayem, 2011. Subjectivity and sentiment analysis of modern standard Arabic. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers, Volume 2, June 19-24, 2011, Stroudsburg, PA., pp: 587-591.
19. Feldman, R., 2013. Techniques and applications for sentiment analysis. Commun. ACM., 56: 82-89.
20. Khalifa, K. and N. Omar, 2014. A hybrid method using lexicon-based approach and naive Bayes classifier for Arabic opinion question answering. J. Comput. Sci., 10: 1961-1968.
21. Huang, J., J. Lu and C.X. Ling, 2003. Comparing naive Bayes, decision trees and SVM with AUC and accuracy. Proceedings of the 3rd IEEE International Conference on Data Mining, November 19-22, 2003, Canada, pp: 553-556.
22. Pang, B. and A.L. Lee, 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. Proceedings of the 42nd Annual Meeting of the Association for Computational, November 16-24, 2004, Kimberly Patch, pp: 271-278.
23. Mourad, A. and K. Darwish, 2013. Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs. Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, June 14, 2013, Atlanta, Georgia, pp: 55-64.