Sentiment Analysis System for Arabic Articles News (SASAAN)

Fawzia Zaki Farahat*, Alaa Hamouda*, Ali Mahmoud Rashed*

Department of Systems and Computers Engineering
Al_Azhar University, Cairo, Egypt

Eng.fawzia@gmail.com
Alaa_hamouda@azhar.edu.eg
rashed@etcp.edu.eg

Abstract: ***Sentiment analysis (also known as opinion mining) identifies and analyzes opinions and emotions in many domains (e.g. news, articles, product reviews, blogs, forum posts). Opinion mining is very important for companies, governments and every one interested to know opinion about special subject. This research discusses the problem of identifying opinion in Arabic news and Arabic articles. Most previous researches focused on extracting opinion from direct sentiments at the level of the article. Considering that an article contains large number of sentences, and some of these sentences may be about different topics and may be not opinion sentence, we propose a new methodology for sentiment analysis for Arabic articles. It starts with identifying opinion sentence related to the target of the article. Machine learning and Typed Dependency Relations (TDR) are used to identify the opinion sentences. Sentences that contain one word of high frequency nouns or adjectives are classified as target sentences. Then opinion lexicon is built using machine learning based on dataset that was collected from different domains (e.g. politics, economy, government, sports, and art). Three methods are used to identify opinion mining in articles. A method that depends on Opinion Lexicon achieved F-score of 62.8%. Machine learning (SVM) method achieved F-score 42.63%. whereas, our method that identifies opinion sentences that are related to the target of article then using opinion lexicon achieved the best results (F-score of 73.25%). So we recommended to identify opinion sentences that are related to the target of the article, then use the opinion lexicon to know the opinion.***

Keywords***: Opinion Sentences, Arabic Grammar, Target Sentences, Opinion Lexicon, Machine Learning (SVM).***

1 INTRODUCTION

Nowadays people have the chance to express their opinions and sentiments and make it available in electronic news. Opinion mining is concerned with the analysis of people's opinions and sentiments towards companies, governments and individuals automatically. Manually extracting opinions consumes time and effort.

Most efforts for opinion mining deal with English texts, some new works deal with other languages and a few works are in Arabic. These Arabic works deal with social networks (tweets, comments or posts). This research is the first research analyzing Arabic article news about a certain topic. Ref. [1] identifies citizen opinions from comments about governmental decisions by using three machine learning (K-Nearest Neighbor, Support Vector Machine (SVM) and Naive Bayes. Citizen opinions about Portuguese Politics were identified from weblog posts, comments to those posts, and comments to news [2]. Ref. [3] recognized opinion in English quotations (reported speech) in newspaper by using large lexicons, as well as specialized training and testing data. Ref. [4] tool can determine if the comment or review is: (subjective or objective), (positive or negative), and (strong or weak). Most works in the field of opinion mining concentrate on extracting knowledge from direct opinions. Directly opinion mining from an article is not effective because article contains large number of sentences, some of these sentences may be written in other topic or do not contain opinion. Opinion mining for Arabic article news is very important for government, different establishments and every one interested in knowing the opinion about a special subject.

This paper presents different techniques to identify opinion sentences and target sentences then use opinion lexicon to identify sentiments for the words. This paper is divided into five sections, section two describes related work, section three deals with identifying opinion sentences, target sentences and building opinion lexicon, section four is the system evaluation and results. Finally, the conclusion and future work are presented in section five.

2 RELATED WORK

There are many Opinion Mining and Sentiment systems for analysing social networks. Many approaches have been used in opinion mining of which machine learning and lexicon based are the widely used ones. Polarity classification is used to classify a document as positive or negative [5] whereas, Machine Learning (SVM, NB) identifies opinion in Egyptian tweets [6]. Although Machine Learning methods have high performance, article contains large number of sentences and requires an annotated corpus to train a classifier which is not easy to obtain from Arabic corpus. Opinion lexicon is used for opinion in English quotations (reported speech) in newspaper articles [3]. It is a collection of words and sentiments regardless of the relations between the individual words. Another method for opinion mining is the utilization of three successive methods (lexicon based method, maximum entropy model and k-nearest model). This method is used to extract opinion from multi Arabic documents [7]. It works on posts in domains of education, sports and Politics rather than articles. Ref. [8] identifies opinion in English news by using lexicon and POS-tagging as features to extract contextual polarity. Ref. [9] filters out negative articles and serves only good news after classifying news articles using sentiment analysis. This method will enable customers to focus only on the good news which will help spread positivity around and would allow people to think positively.

3 PROPOSED MODULE

The proposed methodology includes three steps; identifying opinion sentences, discovering target sentences and sentiment analysis using reliable lexicon dictionary.
The methodology three steps are explained as following:

A. Opinion Sentences Identification

To extract opinion sentences from an article, we used two methods: Type Dependency Relations and Machine Learning.

1) *Type Dependency Relations*: identifies the opinion sentences by selecting and utilizing relevant dependency relations between words by using three relations [10] that indicate a probable presence of general opinions in sentences. These relations from Stanford Typed Dependencies give definition to 55 binary grammatical relations between a governor and a dependent that can possibly be present in a sentence.

R1: Any form of verbs then any form of adjectives Example:

- وتابع مدير الكرة **كنت حريصا** خلال الساعات الماضية على توصيل رسالة للمسئولين عن الكرة المصرية بضرورة الحفاظ على جميع اللاعبين، مؤكداً أنه **ليس منطقياً** أن يُشارك لاعب واحد مع ٣ منتخبات فضلاً عن مشاركاته مع ناديه محلياً وقارياً.

The sporting director said I **was careful** during the past few hours to deliver a message to Egyptian officials about the need to preserve all players, stressing that it was **not logical** for a single player to participate in three teams as well as his participation with his club locally and continental

كنت فعل حريصا صفة ، ليس فعل منطقيا صفة .

R2: Any form of adjectives then any form of verbs
Example:

- الاعتماد على مصادر **مجهولة يضرب** الاقتصاد المصرى، بعد زيارات ناجحة لزعماء العالم للقاهرة، الأيام الماضية وتوقيعهم على عقود استثمار بالمليارائكد الخبير الاقتصادى، والمستشار القانونى لقناة السويس الجديدة، أن مصر ستعبر أزمتها الاقتصادية مليون%، ولديها فرصة **ذهبية تحقق** النمو، نظراً لكونها أكبر أسواق الشرق الأوسط وموقعها الجغرافى ومقوماتها الجبارة.

Dependency on **unknown resources hit** the Egyptian economy, after successful visits to Cairo by the world leaders, in the last few days they signed on investment contracts by millions. Economic expert and legal adviser to the new Suez Canal make sure that Egypt will cross the economic crisis million % , and has a golden chance to achieve development , due to it being the biggest Middle East market, geographical position and its resistance mighty.

مجهولة صفة يضرب فعل، ذهبية صفة تحقق فعل.

R3: Adverb then verb or adjective, adjective then adverb, verb then adverb
Example:

- وأكد "سرى الدين"، أن الإرادة السياسية موجودة بجانب الرغبة والعزيمة على تحقيق النجاح، لكن **هناك ضعف** فى بعض الأجهزة التنفيذية، معتبراً أن حلول الكهرباء جبارة ومعجزة بكل المقاييس.

Sari EL-din make sure that political will existent beside the desire and determination to achieve success, but **there is weakness** in the executive devices, considering that electricity solutions are a powerful and miracle by any standards.

هناك ظرف ضعف صفة .

- العلاج الذي ينبغي أن يكون عاجلاً وحاسماً، وشافياً ، ومطمئناً ومريحاً، لا ينبغي أن **يقتصر فقط** على الاجراءات العقابية، والوسائل العنفية.

The treatment, which should be an urgent, decisive, panacea, reassuring and comforting, should not be **limited only** on punitive measures, and violent means.

يقتصر فعل ، فقط ظرف .

In Arabic language adjective often comes after noun so we add these relations.

R4: Any form of verbs then noun then adjective
Example:

- قال حسام غالى قائد الأهلى، أن المباريات الثمانية المتبقية فى الدورى **تعد لقاءات حاسمة** ولا مجال فيها للتفريط فى أى نقطة، حتى **يحافظ الفريق الأحمر** على تفوقه ويحصل على درع البطولة.

Hussam Ghali Al-Ahly captain , said that the remaining eight league games **are crucial meetings** where there is no room for negligence at any point, even the **red team maintains** its superiority and gets Shield tournament.

تعد فعل لقاءات موصوف حاسمة صفة ، يحافظ فعل الفريق فاعل الأحمر صفة.

- **شهد الاقتصاد المصرى**، اليوم احداثا هامة ومتنوعة، جاء على رأسها، إعلان الجهاز المركزى للتعبئة العامة والإحصاء زيادة قيمة الاستثمارات المالية للبنوك بقيمة ٢٣.١%.

The **Egyptian economy had** an important and variety events, it came at first that, announcement from Central device for Public Mobilization and Statistics increasing value of financial investment banks to 23.1 %.

شهد فعل الاقتصاد فاعل المصري صفة .

R5: Adverb then noun then adjective

اكدت الجرايد العامة للدولة والرسمية ان **هناك ايدى خفية** متسببة فى هذا الهلاك والدمار الذى كلف اصحاب تلك المحلات المحروقة الكثير

Public Journals state official confirmed that **there are hidden hands**, causing the death and destruction that cost the owners of those scorched shops a lot.

هناك ظرف ايدي اسم خفية صفة.

1) *Machine Learning:* we collected 2000 opinion and non-opinion sentences from electronic newspaper articles in different domains(700inPolitics, 450 in sports, 350 in economy, 200 in electricity and 300 in government) for training by supervised machine learning Support Vector Machine(SVM) using the following 5 features [11]:

1- Sentence position: The position of each sentence in article.
2- Sentence length: The number of words in each sentence.
3- Subject: A subject describes the action of the verb in a sentence.
4- Frequencies of each part of speech in the sentence.
5-Tenses: Tense in sentence present, past or future.

To be able to use feature #3 (subject), we have a challenge of extracting the subject. We used Arabic Grammar relation [12] and Stanford part of speech tagger [13]. POS tagging can detect past and present tense only. To identify future tense, we add a rule:

- If a verb starts withس or comes after سوف, then it is in the future tense.

Arabic Grammar uses relations between words to recognize Grammatical Relations (GRS).

"الفاعل" Subject: Arabic language has multi representation of subject. The subject describes the action of the verb and always, comes after it. We used the following rules:

R1: Verb + NP {(Proper-Noun) + complement}→The subject is the proper noun

Ex: Ahmed wrote an article.مثال: كتب أحمد مقالة

كتب: فعل، احمد: فاعل، مقالة: مفعول به. "Ahmed: subject, write: verb, object: article".

R2: VP + NP {(SN) +Proper-Noun}→ The subject is the Singular Noun

Ex: Doctor Mohammed confirmed the need for commitment to work.مثال: اكد الدكتور محمد ضرورة الالتزام بالعمل

اكد: فعل، الدكتور: فاعل ، محمد ضرورة الالتزام بالعمل "Doctor Mohammed: subject, confirmed: verb, the need for commitment to work"

R3: VP {V + PRP} + NP | PP{(Noun)+ complement}→The subject is the noun

Ex: The minister brought decision to me.مثال: وجه الي الوزير قرارا

وجه فعل ، إلي حرف ،الوزير فاعل، قرارا مفعول به.

"The minister: subject, brought: verb, object: decision, to: preposition, me."

R4: VP {V} + NP {(SN) + ADJ}→ The subject is the Singular Noun

Ex: industrious student succeeded.مثال: نجح الطالب المجتهد

نجح فعل ، الطالب فاعل ، المجتهد صفة

"industrious: adjective, student: subject, succeeded: verb"

B. *Target Sentences*

To determine if the opinion sentence is talking about the target of the article, we have a challenge of extracting the main topics in the sentence. A sentence is Noun Phrase (NP) or Verb Phrase (VP). The Noun Phrase usually contains topic and/or object in the sentence, or in simple words, this is what the sentence is talking about, while Verb Phrase describes some action between the objects in the sentence.To extract the main topics of the sentence, we consider the following words as targets of the sentences [14]:

- Every noun,
- Noun after noun and
- Adjective after noun
- Then we take the sentence related to article target by taking sentences that include the most frequent words. We made several experiments with different percentages of high frequency target words to find the best percentages.

Example:

-سجلت **أسعار العملات** أمام **الجنيه المصرى**، فى نهاية تعاملات يوم الأحد، وفقًا لآخر تحديثات البنك المركزى، استقرارًا مقارنة بأسعار **التعاملات الصباحية**.

Exchange rates recorded in front of the **Egyptian pound**, at the end of trading on Sunday, according to the latest central bank updates, are stable compared to prices of morning trade.

- قارئ يشكو من ارتفاع **أسعار الكهرباء** وأرسل شكوى من ارتفاع **أسعار الكهرباء** هذا الشهر، وأرفق صورة للإيصال الخاص به توضح هذا الارتفاع.

Reader complains of **rising electricity** prices and sent a complaint from **rising electricity** prices this month with a copy of its receipt explain this rise .

C. *Opinion Lexicon*

In order to create the Opinion lexicon, we collected tweets, news and articles about different domains and classify them manually into positive, negative or neutral, then we used machine learning to train data and get weight for each word to determine the sentiment level of each word.

1) *Data Collection and Preparation*:
From tweeter:
-We collected dataset from different domains then classified manually into (negative, positive and neutral), negative and positive data used only for training by ML.
-They were split into very simple tokens such as numbers, punctuations and words of different types. Then stemming analysis process was applied to produce the stem of tokens.

2) *Lexicon classification*: We use this tweets prepared data for training using support vector machine (SVM) based on frequency of words in all tweets as a feature. SVM was used to infer the values of words implicitly from the training data. The generated model from this training, which has many tokens with positive or negative weight values includes 11000words. Tweets have been written in Egyptian Dialect, so we use Transformation of Egyptian Dialect into Modern Standard Arabic system [15] to transform these words to standard Arabic language to be suitable for articles and neglect the words which contains alphabetical mistakes. This process resulted in the removal of 1219 words and only 9781 correct words were used.
We use these news and articles prepared data for training by (SVM) based on frequency of words in all articles as a feature.
-The generated model from this training contains (11887) words with positive or negative weight values.

Example:

0.653	انجاز 'achievement'	0.539	جيد 'good'	0.5124	ارتق 'imperforate'	0.5	يكرم 'honors'
-0.665	يلعن 'damned'	-0.37	سيئ 'bad'	-0.593	حرام 'ban'	-0.31	الم 'pain'

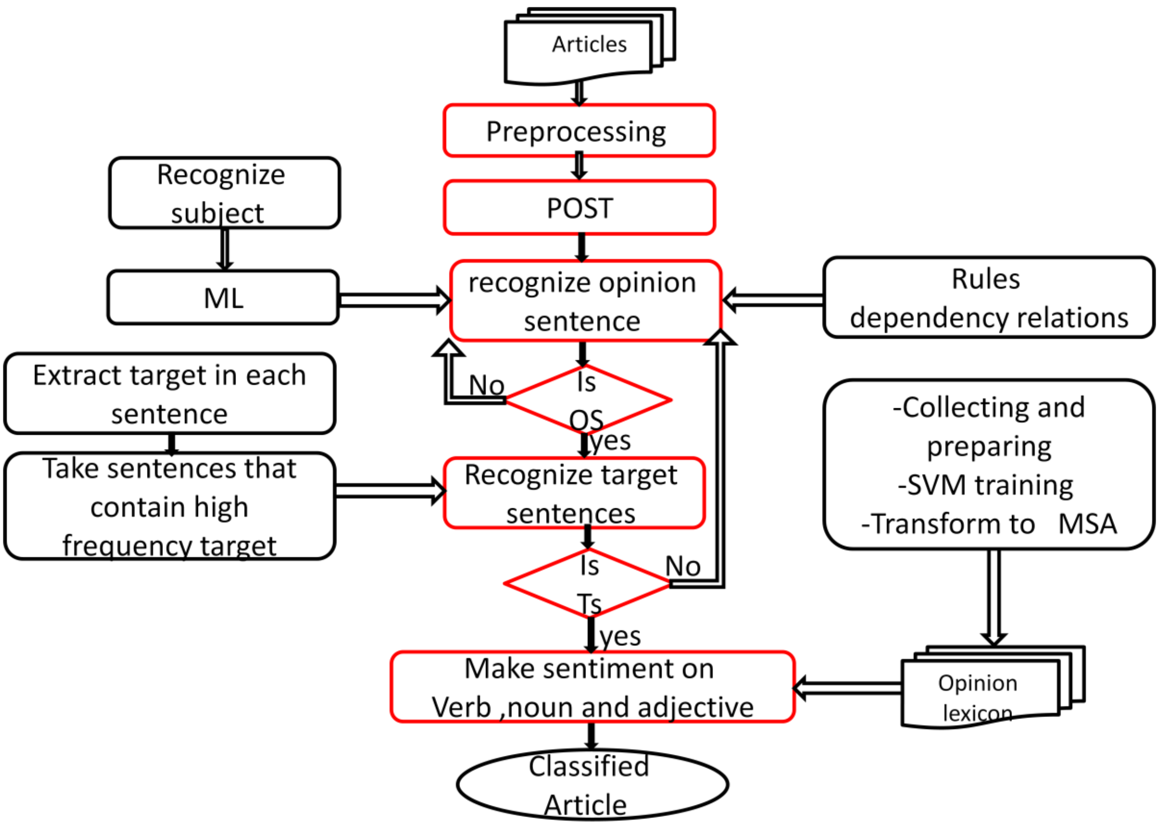

Figure 1: SASAAN Block Diagram

After the recognition of the opinion sentences related to article target, POS tagging was used to select verb, adjective and noun words only from these sentences then we used the opinion lexicon to know positive or negative weight values for these words [16]. A final positive score leads to the classification of the article or news as positive, whereas a final negative score leads the system to classify the article as a negative one. If the score is from 0.02 to - 0.008, this article or news is considered neutral. The description of the System is shown in Fig. 1.

4 SYSTEM IMPLEMENTATION AND EVALUATION

A. Data Collection

We generated our OL by collecting tweets, news and articles from the internet. These tweets, news and articles address general topics such as political, education, economy, government, art, sports, and communication.

From twitter:

-A total of 18,764 tweets from different domains (4,970 for celebrities data, 6197in communication and 7597 in government) were collected.

-These tweets were classified manually into (13037 negative, 3593 positive and 2134 neutral). Then we used only 8700 tweets (5100 negative and 3600positive) to reduce the difference between positive and negative tweets and we neglected the neutral ones.

From electronic news web sites:

- A total of 350 articles and 500 news from different domains(e.g. politics, economy, government, sports, and art) were collected.

- This data was classified manually into (313 negative, 293 positive and 246 neutral). We neglected the neutral articles and news.

To evaluate the proposed system (SASAAN), we evaluate each stage. The first stage is opinion sentences extraction. The second is target sentences extraction. The third is to develop the opinion lexicon to evaluate sentiment analysis. Each stage is evaluated in subsection B, C and D. Finally the system as a whole is evaluated. We use two metrics to evaluate the efficiency of the proposed system (SASAAN).

1. Accuracy $= \frac{TP+TN}{TP+FP+TN+FN}$
2. Precision, Recall and F-score

B. Opinion Sentences Extraction

We identify opinion sentences by three methods:

1) *Type Dependency Relations*: the results of identifying opinion sentence using TDR on a dataset of 250 sentences are presented in Table 1. It shows the results of each relation. We get the best results when we use all these relations in this technique.

TABLE 1
IDENTIFYING OPINION SENTENCES BY TDR

Relation	**Precision**	**Recall**	**F-score**
R1: Any form of verb then adjective	90%	12%	21.17%
R2: Any form of adjective then verb	80%	47%	59.2%
R3: Adverb then verb, adverb or adjective	80%	14%	23.8%
R1+R2+R3	82%	40%	53.8%
R4: Any form of verb then noun then adjective	73%	65%	68.8%
R5: Adverb then noun then adjective	70%	63%	66.3%
R1+R2+R3+R4 +R5	75%	65%	69.7%

Regarding the effect of R4 and R5 on the OS extraction performance, in Arabic language, the adjective often comes after noun. So we can note that there was an improvement of 15.9% in the F-score.

2) *Machine Learning*: we use dataset of 250 sentences for testing the identification of opinion sentences by machine learning (SVM), we got this result:-
Precision=70% Recall=67% F-score=68.46%

3) *Combining TDR and SVM:* we used dataset of 250 sentences to evaluate the identification of the opinion sentences with the following methods:
- When the sentence is identified as opinion by two methods (TDR and SVM), we got this result:-
Precision=75% Recall=67% F-score=70.8%
-When the sentence is identified as opinion by any one of methods (TDR or SVM), we got this result:-
Precision=65% Recall=96% F-score=77.5%

It is apparent that the best results for opinion sentences extraction are obtained when using TDR or SVM.

C. Target Sentences Extraction

We extract the sentence related to the target of the article by making several experiments with different high frequency target words, as shown in Fig. 2. The best result is obtained when we take quarter of high frequency words.

We used a dataset of 100 articles to evaluate identifying sentences related to the target of article. Then we use opinion lexicon. We get this result:-

Precision=71% Recall=68.8% F-score=70.15%

When we analyze a dataset of news to evaluate identifying target sentences, we find most sentences in the news are related to its target due to the small number of news existing in the news compared to the articles.

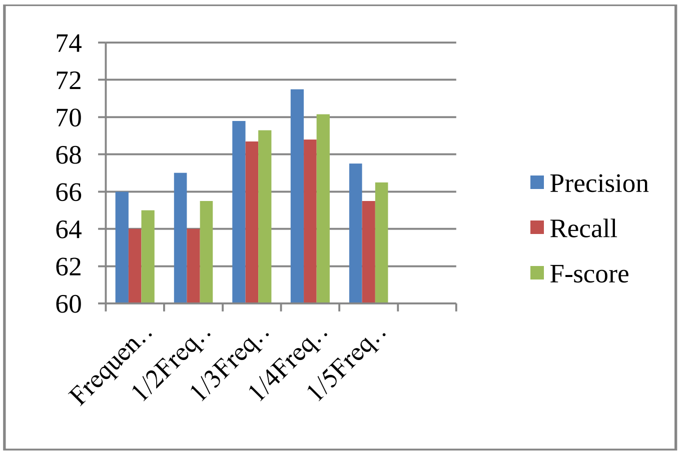

Figure 2: Target Sentences By High Frequency Target Words

D. Opinion Lexicon

We get the following results when we use the published lexicon sentiment [17] on our data set:-

Accuracy positive =50%

Accuracy negative=47%

So we decided to build two opinion lexicon:

First one by using (18,764 tweets). When we use our tweets opinion lexicon, we get the results in Table 2.

TABLE 2
TWEETS OPINION LEXICON RESULT

Field	Type	Precision	Recall	F-Score
Tweets	positive	78%	77%	77,5%
	negative	80%	80%	80%
News		66%	65%	65.5%
Articles		61%	62.8	61.9%

Second by using (350 articles and 500 news).When we use our tweeters opinion lexicon, we get the results in Table 3.

TABLE 3
NEWS AND ARTICLES OPINION LEXICON RESULT

Field	Precision	Recall	F-Score
News	63.14%	63.66%	63.4%
Articles	61%	43.35	50.7%

E. System As All

1) *Machine learning:* We have collected 428 articles from deferent domain (politics, economy, government, sports, and art) and classified them manually into negative, positive and neutral. Segment the words in each article and then use these words features for training by support vector machine (SVM). When we identify opinion in article by using machine learning (SVM) we get this result:-
   Precision =37%   Recall =48%   F-score=41.78%
   Opinion lexicon and SVM classifier are used for opinion in English quotations (reported speech) in newspaper articles [3]. This method achieves low F-measure 33% for positive and 66% for negative.

2) *Opinion Sentences then Lexicon:* we used dataset of 400 news (97 negative, 179 positive) to evaluate identifying opinion sentences then use:
   -Tweets opinion lexicon to know the opinion. We get these results shown in Table 4.

TABLE 4
OPNION SENTENCES THEN OPINION LEXICON ON NEWS RESULT

Case	Precision	Recall	F_ score
When use TDR OR ML to identify opinion sentence then use lexicon	69%	73.4%	71.2%
When use TDR and ML to identify opinion sentence then use lexicon	66%	67%	66.7%
When use TDR to identify opinion sentence then use lexicon	65%	64%	65.5%
When use ML to identify opinion sentence then use lexicon	61.3%	62.5%	61.9%

- News and articles opinion lexicon to know the opinion. We get these results shown in Table 5.

TABLE 5
OPNION SENTENCES THEN OPINION LEXICON ON NEWS RESULT

Case	Precision	Recall	F_ score
When use TDR OR ML to identify opinion sentence then use lexicon	77.42%	67.4%	72.5%
When use TDR and ML to identify opinion sentence then use lexicon	75.64%	61.7%	67.97
When use TDR to identify opinion sentence then use lexicon	68.1%	63.5%	65.77%
When use ML to identify opinion sentence then use lexicon	60.7%	60.5%	60.6%

We used dataset of 100 articles (64 negative, 36 positive) to test identifying opinion sentences then used:

- Tweets opinion lexicon to know the opinion in articles. We get these results shown in Table 6.

TABLE 6
OPNION SENTENCES THEN OPINION LEXICON ON ARTICLES RESULT

Case	Precision	Recall	F_ score
When use ML to identify opinion sentence then use lexicon to know the opinion in an article	71.25%	68.5%	69.87%
When use TDR to identify opinion sentence then use lexicon to know the opinion in an article	71.5%	70%	70.8%
When use TDR and ML to identify opinion sentence then use lexicon to know the opinion in an article	72%	71%	71.5%
When use TDR OR ML to identify opinion sentence then use lexicon to know the opinion in an article	69.8%	68.77%	69.5%.

- News and articles opinion lexicon to know the opinion in articles. We get these results shown in Table 7.

TABLE 7
OPNION SENTENCES THEN OPINION LEXICON ON ARTICLES RESULT

Case	Precision	Recall	F_ score
When use ML to identify opinion sentence then use lexicon to know the opinion in an article	54.28%	62.63%	58.5
When use TDR to identify opinion sentence then use lexicon to know the opinion in an article	56%	63.4%	59.5
When use TDR and ML to identify opinion sentence then use lexicon to know the opinion in an article	54.28%	62.63%	58.5
When use TDR OR ML to identify opinion sentence then use lexicon to know the opinion in an article	57.3%	64.49%	60.8%

3) *Opinion Sentences related to target then Lexicon:* we use dataset of 400 news to test identifying sentences that are related to target of news then using opinion lexicon we get this result:-

Precision=63.4% Recall=61.8% F-score=65.6%

Finally we use dataset of 100 articles to test identifying opinion sentences that are related to target of article then using opinion lexicon we get this result:-

Precision=74% Recall=72.5% F-score=73.25%

We can notice that the result of using lexicon after identifying opinion sentences by union of two methods that are related to the target of article gives the best results of articles. Whereas, the best results for news is identifying opinion sentence by union of two methods then use lexicon to know the opinion. Results seems to be logically accepted because the article contains large number of sentences, some of which may be about another topic or not opinion sentence while most sentences in news are related to its target.

5 CONCLUSION

An Arabic system to identify opinion in Arabic articles and Arabic news has been presented. We proposed a methodology with three sequential steps:

1) Identifying opinion sentences by two methods: machine learning and Type Dependency Relations, 2) Extracting the opinion sentences related to the target of the article, 3) Building opinion lexicon based on training data by SVM classifier. From results we concluded that using lexicon opinion on specific words (verb, adjective and noun) is better than identifying opinion in article by using machine learning (SVM). Identifying opinion sentences by machine learning (SVM) and TDR then opinion lexicon achieved better results. Moreover, when we use opinion lexicon for opinion sentences related to the target of the article, it gives the best results (F-score 73.25**%)**. In the future, we will use SVM classification based on new set of features such as position, syntactic features and meta-discourse features for extracting topic sentences. Also, in the future, we plan to increase opinion lexicon size by increasing the training dataset.

REFERENCES

[1] G. Stylios, D. Christodoulakis, J. Beshara, M. Vonitsanou, I. Kotrotsos, A. Koumpouri, and S. Stamou, "Public Opinion Mining for Governmental Decisions," *Electronic Journal of e-Government, vol. 8, Issue 2, pp203-214*, 2010.

[2] M. J. Silva, P. Carvalho, L. Sarmento, P. Magalhaes and E. Oliveira, The Design of Optimism, an Opinion Mining System for Portuguese Politics, supported by grants SFRH from FCT-Portugal, co-financed by POSI, 2009.

[3] A. Balahur, R. Steinberger, E. V. D. Goot, B. Pouliquen and M. Kabadjov, Opinion Mining on Newspaper Quotations, *International Joint Conferences on Web Intelligence and Intelligent Agent Technologies*,(IEEE), 2009.

[4] M. N. Al-Kabi, A. H. Gigieh, I. M. Alsmadi, H. A. Wahsheh and M. M. Haidar, "Opinion Mining and Analysis for Arabic Language," *International Journal of Advanced Computer Science and Applications,* (IJACSA), Vol. 5, No. 5, pp.181-195, 2014.

[5] B. Pang, And L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," *Foundations and Trends in Information Retrieval*, Vol. 2, No. 1, pp.1-135, 2008.

[6] A. M. Shoukry and A. Rafea, "Arabic Sentence Level Sentiment Analysis," Master dissertation, American University, Cairo, 2013.

[7] A. El-Halees, "Arabic Opinion Mining Using Combined Classification Approach, Islamic University of Gaza," in proc. *of International Arab Conference on Information Technology*, (ACIT), vol.13, pp. 265-271, 2012.

[8] T. Wilson, J. Wiebe and P. Hoffmann, "Recognizing contextual polarity in Phrase-Level Sentiment Analysis," *In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 347–354, 2005.

[9] K. S. Doddi, Y. V. Haribhakta and P. Kulkarni, "Sentiment Classification of News Articles," *International Journal of Computer Science and Information Technologies*, (IJCSIT), vol.5, no. 3, pp. 4621-4623, 2014.

[10] A. Qadir, "Detecting Opinion Sentences Specific to Product Features in Customer Reviews using Typed Dependency Relations," *Proceedings of the Workshop on Events in Emerging Text Types*, (eETTs), pp. 38–43, 2009.

[11] D. Kobayashi, H. Nanba and T. Takezawa, "Extraction of Opinion Sentences using Machine Learning: Hiroshima City University," *Proceedings of NTCIR-7 Workshop Meeting*, Tokyo, Japan, 2008.

[12] O. I. Hammadi and M. J. Abdel-Aziz, "Grammatical Relation Extraction in Arabic Language," *Journal of Computer Science*, Vol. 8, No. 6, pp. 891-898, 2012.

[13] Stanford POS Tagger v. 3.1.1, Available from: http://retroweaver.sourceforge.net/, (accessed 1 August 2013).

[14] The Tokenizer, Available from: https://thetokenizer.com/2013/05/09/efficient-way-to-extract-the-main-topics-of-a-sentence/, (accessed 1 January 2016).

[15] F. E. El-taher, A. A. Hammouda and S. Abdel-Mageid, "Automation of Understanding Social Networks Text Contents", *IEEEE MoWnet*, Cairo, 2016.

[16] Soo-Min Kim and Eduard Hovy, "Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text," Report USC Information Sciences Institute, 2006.

[17] M. Rohaim and A. Hamoda, "Design of Sentiment Analysis Lexicon for Opinion Mining", *Journal of Advances in Information Technology*, vol. 2, no. 4, pp.199-203, November 2011.

BIOGRAPHY

Ali Mahmoud Rashed obtained his Ph.D. degree in electronics and communications engineering from the Ain Shams university (1982). He is lecturer in Systems & Computer Dept (1986), Associate Prof.(1990), Prof. of Computers Engineering since 1996 till now. He Supervised more than 50 Ph.D and M.Sc. thesis. He is a member in the researches reviewers committee for Teacher Higher Staff Position and Engineering Research Journal – Al Azhar University. He is holding a technical consultant position in ETCP. He is teaching VLSI digital electronics, Microprocessors, Microcontroller architecture, Interfacing and Computer Architecture.

Alaa Hamouda is Associate Professor, Faculty of Engineering, Al-Azhar University, Department of Systems and Computers Engineering. He earned a master's degree in the field of intelligent system design for agents in the field of e-commerce, and then a PhD in the design of intelligent system to prevent the spread of the worm spreading. His research interests focus on artificial intelligence, data mining, text mining, summarization of Arabic texts, opinion mining and trends analysis, software engineering and management in accordance with the quality and management models like CMMI and PMI. He has led several research projects in the field of Arabic language processing.

Fawzia Zaki Farahat Born in Egypt, in December 20, 1989. She obtained her bachelor degree in computer and systems from Faculty of Engineering, Alazhar University – Cairo in 2012. She started working as Developer and tester at Horizons Software (2-2013:5-2014). Now, She works in Damietta port as control engineer. Her main areas of research include: Data Mining, Natural Language Processing and Software Engineering.

نظام لتحليل المقالات العربية الإخبارية

فوزية ذكي فرحات[١]، علاء حمودة[٢]، علي محمد راشد[٣]
قسم النظم والحاسبات هندسة الأزهر
[1]eng.fawzia@gmail.com
[2]Alaa_hamouda@azhar.edu.eg
[3]rashed@etcp.edu.eg

ملخص—تحليل الشعور (يعرف ايضا بتحليل الآراء) يميز ويحلل الآراء والعواطف في كثير من المجالات (الأخبار والمقالات وتقارير المنتجات والتعليقات والتويتات....). تحليل الآراء مهم جدا للشركات والحكومات وأي شخص مهتم بمعرفة الرأي حول موضوع معين . تناقش هذه الورقة مشكلة التعرف علي الرأي في المقالات والأخبار العربية ، معظم الأبحاث في مجال تحليل الآراء تعتمد علي أخذ الرأي من جميع الجمل علي مستوى المقالة رغم أن المقالة تحتوي علي جمل كثيرة قد يكون بعضها متعلقا بموضوع آخر وقد تكون بعض الجمل اعتراضية لا تحتوي علي رأي . لذالك قمنا بالتعرف علي جمل الرأي المتعلقة بهدف المقال ، وذلك باستخدم نموذج تعلم الآلة والعلاقات المعتمدة المتنوعة للتعرف علي جمل الرأي. تم أخذ الجمل الذي تحتوي علي كلمة اوأكثرمن الأسماء أو الصفات الأكثر تكرارا علي أنها جمل متعلقة بهدف المقالة . بعد ذلك تم بناء معجم للآراء يعتمد علي تعليم الآلة علي البيانات المجمعة من مجالات مختلفة. استخدمت ثلاث طرق لتحليل الآراء في المقالات. باستخدام معجم الآراء كانت دقة التعرف على رأي وتوجه المقالة هي ٦٢.٨% و عند استخدام تعليم الالة كانت النتيجة ٤٢.٦ % . أما عند التعرف علي الرأي باستخدام معجم الآراء علي جمل الرأي والمتعلقة بهدف المقالة فقد أعطت افضل النتائج٧٣.٢٥%. لذا ننصح بالتعرف علي جمل الرأي المتعلقة بهدف المقالة ثم استخدام معجم الآراء.