Journal of King Saud University – Computer and Information Sciences 29 (2017) 229–233

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Semantic Sentiment Analysis in Arabic Social Media

Samir Tartir *, Ibrahim Abdul-Nabi

Department of Computer Science, Philadelphia University, Amman, Jordan

ARTICLE INFO

Article history:
Received 4 June 2016
Revised 1 October 2016
Accepted 27 November 2016
Available online 5 December 2016

Keywords:
Arabic
Sentiment
Ontology
Semantic
Social
Twitter

ABSTRACT

Social media is a huge source of information. And is increasingly being used by governments, companies, and marketers to understand how the crowd thinks. Sentiment analysis aims to determine the attitudes of a group of people that are using one or more social media platforms with respect to a certain topic. In this paper, we propose a semantic approach to discover user attitudes and business insights from social media in Arabic, both standard and dialects. We also introduce the first version of our Arabic Sentiment Ontology (ASO) that contains different words that express feelings and how strongly these words express these feelings. We then show the usability of our approach in classifying different Twitter feeds on different topics.

1. Introduction

Many social media platforms (e.g. such as Facebook,[1] Twitter,[2] and Instagram[3]) allow people, businesses and organizations to share their information and thoughts online. Understanding public sentiment and concerns expressed on these different platforms is a crucial issue for both policy makers, business leaders, and the public (Vargas et al., 2016; Imran et al., 2015; Verma et al., 2011). For example, Google currently tracks political issues, and diseases to help policy makers better prepare for what will happen next (Google, 2016). Sentiment analysis (SA) is an automated task of rapidly determining the sentiment of large amounts of text or speech (Pang et al., 2008). It's worth mentioning that SA is deeper than hashtag counts that many social media platforms offer, as the earlier studies the meanings of post contents to come up with an overall mood instead of simply counting hashtags used by users.

Arabic is a major language, used (standard and its dialects) by around 422 million speakers (UNESCO, 2012). Still, while research on sentiment analysis has been done in other major languages (Agarwal et al., 2011; Schulz et al., 2013), little has been done in Arabic (Abdul-Mageed et al., 2014; Salem and Mourtada, 2012).

In this paper, we propose a semantic approach to discover user attitudes and business insights from Arabic social media by building an Arabic Sentiment Ontology that contains groups of words that express different sentiments in different dialects. We then use this ontology to analyze sentiments in different groups of Twitter posts (called tweets) as an example social media platform.

2. Related work

There is a large amount of work on using different techniques for sentiment analysis. The work in (Pang et al., 2008) is one of the early approaches that applied sentiment analysis on online movie reviews using machine learning. The results show that especially support vector machine and Naïve Bayes can be efficiently used to extract the sentiment from movie reviews when they compared their work to human analysis.

Other works, e.g. (Lai, 2010) brought sentiment analysis to the Twitter domain by applying similar machine learning techniques to classifying the sentiment of tweets.

The research in (Conover et al., 2011) showed that Twitter can be used as a platform for political deliberation. In addition, the work uses a database for sentiment lexicons and their results show that sentiment extraction can produce result similar to traditional

* Corresponding author.
E-mail addresses: startir@philadelphia.edu.jo (S. Tartir), abdulnabi.job@gmail.com (I. Abdul-Nabi)

[1] www.facebook.com.
[2] www.twitter.com.
[3] www.instagram.com.

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2016.11.011
1319-1578/

election polls, which mean if they expand the lexicon the result will be more efficient.

Only few studies have been performed on Arabic social media. For example, the work of (Abdul-Mageed et al., 2014) has focused on movie and product reviews. The work of (Abbasi et al., 2008) uses a genetic algorithm for sentiment detection in both English and Arabic Web forums on the document level. They exploit both syntactic and stylistic features, but do not use morphological features. The work of (Shoukry and Rafea, 2012) studies the effect of preprocessing on sentiment analysis of Egyptian dialect tweets.

3. Model

In this section we describe the details of our model for analyzing the sentiments of social media posts. In Fig. 1 we show the different components of our model. The first three steps shown in the figure are just to prepare the posts for classification. The majority of our contributions are in the following step, the "Classify" step which relies on the ASO. We therefore show the details of this step in Fig. 2. In the following, we will first introduce our work on the Arabic Sentiment Ontology, and then proceed to explain the steps we use to analyze posts.

3.1. Collection of posts

The first step of our model is building a collection of posts relevant to a certain topic or brand name. Several approaches exist for the different social media platforms. For example, Tweet Archivist[4] can be used to download Twitter posts tweets using a hashtag, a twitter user, a Boolean search, a complex query or simply a term. Similar tools exist for other platforms.

3.2. Filtering of posts

The following step is to filter out any irrelevant posts that were collected by the tool as the analyst might be aware of cases where the terms of interest might be used for other purposes by other groups.

3.3. Preprocessing

A major issue that is faced when dealing with social media is the informal style of the posts. Most posts are written informally and contain many errors and abbreviations and don't follow any grammatical rules. To minimize the effect of this informality on our classification we will pre-process posts before using them. For example, typos might be words that express sentiment but are misspelled, and therefore must be found and corrected to evaluate sentiment more accurately.

Following the work done on Egyptian Arabic in (Shoukry and Rafea, 2012), we use the following steps for preprocessing of posts:

- Remove any URLs.
- Shorten any elongated words (تكبيــــــــــــــــر → تكبير)
- Fix typos.
- Remove stop words (e.g. propositions).

For typos, we currently use the database at mo3jam[5] which is an Arabic database of words in Standard and its dialects.

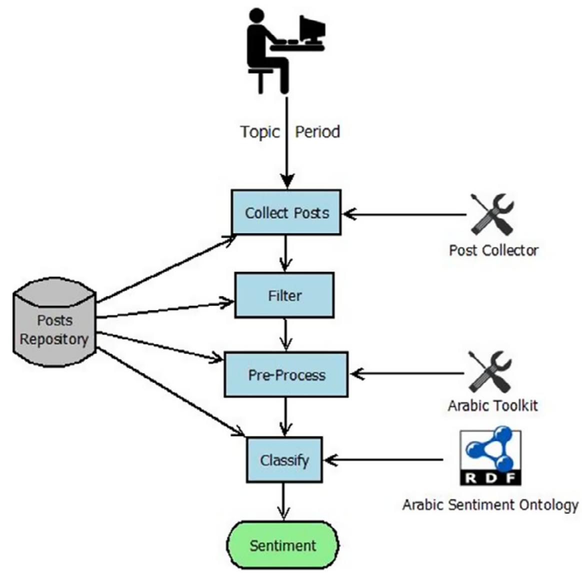

Figure 1. System components.

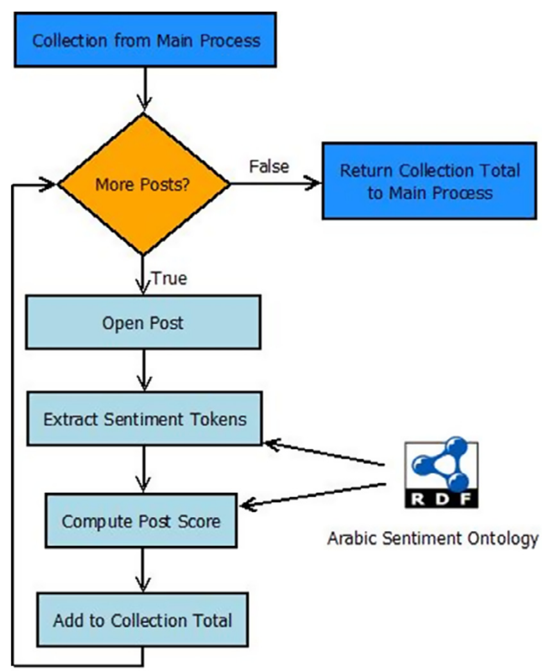

Figure 2. Sentiment analysis of posts.

3.4. Classification

Fig. 2 explains the steps we follow to classify groups of posts as positive or negative and the strength of the sentiment. Of course this all relies on the ASO, which is explained in Section 4 below

Table 1
Tweets sentiment classification.

Classification	Reason
Positive	• Post contains mostly positive words
Negative	• Post contains mostly negative words
Neutral	• No positive nor negative words
	• Equal weight for positive and negative words

[4] http://www.tweetarchivist.com/.
[5] http://en.mo3jam.com/.

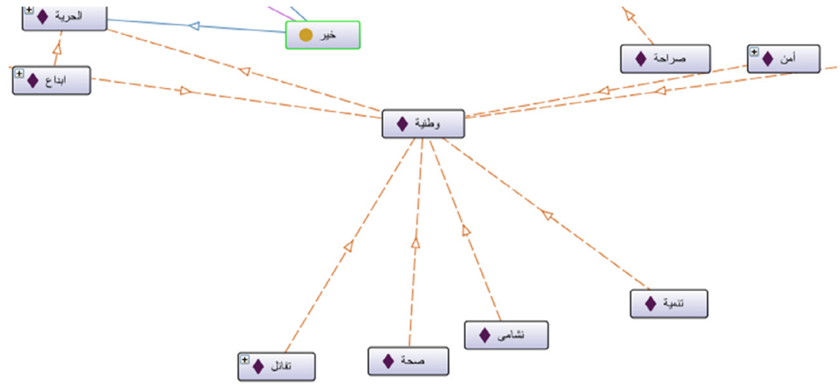

Figure 3. Positive sentiments in ASO.

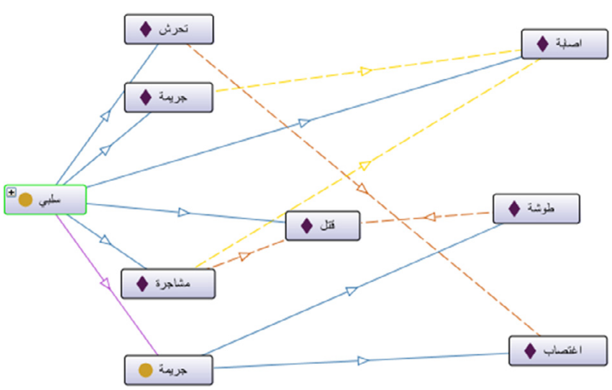

Figure 4. Negative sentiments in ASO.

Table 2
Positive words and their weights.

Word	Weight	Word	Weight
كويس	4	مليح	3
نشامى	3	فخور	4
حلو	4	عطف	3
صراحة	4	تنمية	1
مبروك	3	محترم	3
صح	2	عظيم	3

Table 3
Negative words and their weights.

Word	Weight	Word	Weight
أسعار	−1	فقر	−3
طوشة	−3	ضياع	−3
شكوى	−4	ملل	−4
قرف	−2	اغتصاب	−5
هبل	−1	تأخير	−4
وساخة	−4	عيب	−3

and as more sentiments are added to it, the more detailed the analysis will be.

In order to compute sentiment for any post we had to classify it into positive, negative, or neutral using Table 1.

Weights for each post are based on the weight of the word and also the total sum of parents in the ASO. For example: the weight of "مشاجرة" (fight) is −3, and it is connected to "قتل" (homicide) with a weight of −4, which is also connected to "جريمة" with a weight of −2. Therefore the total weight will be −9.

4. Arabic Sentiment Ontology (ASO)

To allow good SA of social media posts, we need a way to specify which phrases indicate different sentiments. We therefore built the first (up to our knowledge) ASO for Jordanian Arabic dialect. Our ontology focuses on semantic relations between sentiments and their instances. The ASO currently has two groups of sentiments: Positive & Negative. And under each group, there are some different types of sentiments, e.g. For example, we use the subtype relation to show that a certain sentiment, e.g. "جريمة" (crime) is a subtype of "سلبي " (negative) to indicate groupings of sentiments. We also defined two initial relationships between sentiments: تؤدي_إلى (causes), تنتج_من (a result of) and sameAs to link some of the sentiments together. Along with the sentiment classification, each instance is associated with a weight between 1 and 5 to indicate the strength of the sentiment it carries.

Below are two parts of our ontology. Fig. 3 shows words with positive sentiment, and Fig. 4 shows words with negative sentiment.

The current version of the ASO was built by collecting 24 words that express sentiments, and each word was given a weight by 20 evaluators, and the final weight we used was the average of the weights given by the 20 evaluators. Tables 2 and 3 show samples of positive and negative sentiment words and their weights defined in the ASO.

5. Experiments

To test the feasibility of our model, we will test it on Twitter. Twitter is a free microblogging service founded in 2006 by Jack Dorsey and Biz Stone. It has 332 million active users as of Jan/2016. At its heart are 140-character bursts of information called tweets. Registered users can read and post tweet (Dinerman, 2010).

To test the accuracy of our classification, we will compare our results with the manual classification of the tweets by 3 native Arabic speakers. Below are the results.

We used Tweet Archivist to collect tweets about our topics. Tweet Archivist is a Twitter analytics tool to search, archive, analyze, visualize, save and export tweets based on a search term or hashtag that is easy to crawl and collect data.

For our work we collected 1100 tweets from September/2013 using Tweet Archivist that talk about the following three major

Table 4
Manual annotation results.

Data Set	Jamalon		Khaberni		Ro'ya TV	
	Count	%	Count	%	Count	%
Positive	225	56%	85	42%	310	77%
Negative	109	27%	60	30%	60	15%
Neutral	66	17%	55	28%	30	8%
Total	400	100%	200	100%	400	100%

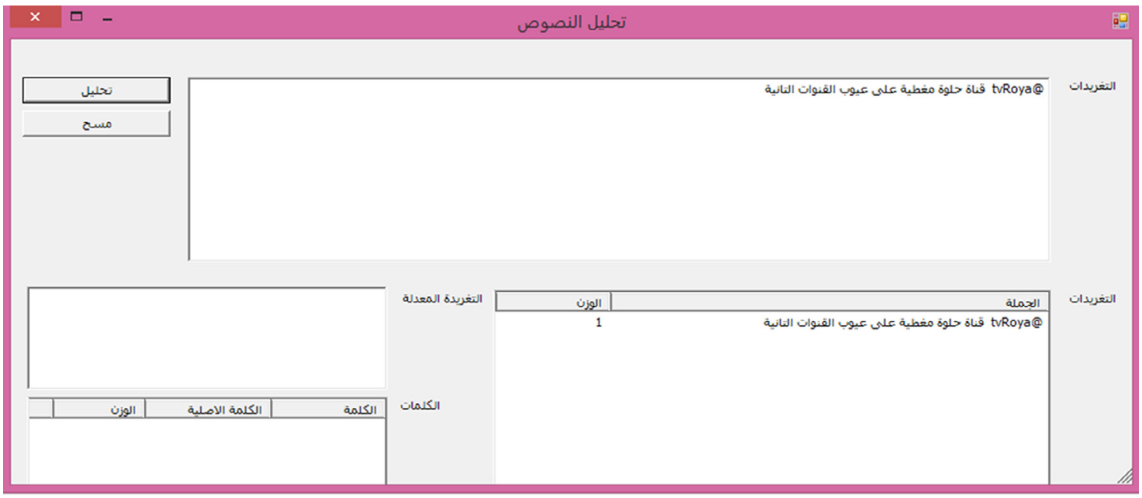

Figure 5. Arabic sentiment application.

Table 5
Automatic annotation results.

Data Set	Jamalon		Khaberni		Ro'ya TV	
	Count	%	Count	%	Count	%
Positive	203	50%	73	36%	292	73%
Negative	98	24%	70	35%	55	13%
Neutral	99	26%	57	29%	53	14%
Total	400	100%	200	100%	400	100%

organizations in Jordan. Below is a brief description of each of the organizations.

(1) Ro'ya TV[6]: A new and popular satellite TV station in Jordan that was launched in 2011.
(2) Jamalon[7]: An Arab bookstore based in Amman. Jamalon is now considered as one of the major sources for books in Arab world.
(3) Khaberni[8]: A Jordanian news website that is mainly concerned with local and regional issues.

We then removed about 100 English tweets, as our work is concerned with Arabic posts. From the remaining 1000 tweets, 400 tweets were for Ro'ya TV, 400 tweets were for Jamalon and 200 tweets were for Khaberni. Tweet Archivist provided tweets in Excel format that contains user id, tweet text, location and time.

5.1. Manual classification

We asked three native Arabic language speakers, to annotate the tweets. We asked them to manually classify each tweet as positive, negative, or neutral. We then averaged their classification for each tweet to get a single classification. Table 4 presents a summary of their classification.

We are currently limiting our classification to positive, negative, or neutral to simplify its initial implementation. Future implementations of our approach will have types of sentiments mentioned, and not simply grouped into positive, negative, or neutral.

5.2. Classification results

We built a sample application that implements our approach on the same tweets and we got the results shown in Table 4. The interface of the application is shown in Fig. 5.

Using our approach and the current version of the ASO we got the classification results for the same 1000 tweets we presented to the human evaluators, as shown in Table 5.

5.3. Results comparison

In this section we will compare manual and the automatic classification results. In this comparison we will calculate the precision and recall for positive, negative and neutral classifications using precision and recall that are defined using the following formulas,(Manning et al., 2008). Due to the lack of benchmarks in the Jordanian Arabic dialect, we consider the manual results our baseline

[6] http://www.roya.tv/.
[7] http://www.jamalon.com/.
[8] http://www.khaberni.com/.

Table 6
Precision and recall for Jamalon.

	Precision	Recall
Positive	180/225 = 80%	180/203 = 88%
Negative	60/109 = 55%	60/98 = 61%
Neutral	50/66 = 75%	50/99 = 50%
Average	70%	67%

Table 7
Precision and recall for Khaberni.

	Precision	Recall
Positive	67/85 = 78%	67/73 = 91%
Negative	50/60 = 83%	50/70 = 71%
Neutral	40/55 = 72%	40/57 = 70%
Average	78%	78%

Table 8
Precision and recall for Ro'ya TV.

	Precision	Recall
Positive	280/310 = 90%	280/292 = 95%
Negative	43/60 = 71%	43/55 = 78%
Neutral	19/30 = 63%	19/53 = 35%
Average	75%	70%

Table 9
Overall precision & recall.

	Precision	Recall
Jamalon	70%	67%
Khaberni	78%	78%
Ro'ya TV	75%	70%
Average	75%	72%

(true answer set) and we measure the efficiency of our approach (retrieved answer set) against this baseline.

$$\text{Precision} = \frac{|\text{Automatic Results} \cap \text{Manual Results}|}{|\text{Manual Results}|}$$

$$\text{Recall} = \frac{|\text{Automatic Results} \cap \text{Manual Results}|}{|\text{Automatic Results}|}$$

Based on above formulas we got the following results for each of the three domains we studied. Table 6 shows the precision and recall for our classification of Jamalon's tweets, with an average precision of 70% and average recall of 66%. Table 7 shows the precision and recall of our classification for Khaberni's tweets, with an average precision of 80% and average recall of 77%. Table 8 shows the precision and recall of our classification for Ro'ya TV's tweets, with an average precision of 74% and average recall of 69%.

Table 9 shows the overall precision and recall for our approach, with an average precision of 74.1% and average recall of 74.7%.

To compare our approach to similar approaches in Arabic dialects, we find the closest approach to be (Shoukry and Rafea, 2012). Our overall precision of 75% is very similar to the paper's results of 76%. And our overall recall of 72% is slightly less than the paper's recall of 75%.

6. Conclusion and future work

Understanding sentiment of a certain entity is crucial for decision makers to understand what their future actions need to be. Little work has been done so far to analyze sentiments in social media of Arab entities. Our approach and tests showed that a semantic approach even when dealing with limited sentiments to analyze posts can produce good understanding of the overall image of a certain entity.

In the future we are planning to extend our work in the following directions.

- Expand the ASO to include more sentiments.
- Expand the ASO to include more dialects, besides Jordanian.
- Enhance the implementation of our approach to report more accurate sentiments.
- Considering other algorithms besides decision support tree, e.g. Naïve Bayes, SVM. etc.

References

Abbasi, Ahmed, Chen, Hsinchun, Salem, Arab, 2008. Sentiment analysis in multiple languages: feature selection for opinion classification in web forums. ACM Trans. Inform. Syst. 26, 1–34.

Abdul-Mageed, Mohammad, Diab, Mona, Kübler, Sandra, 2014. SAMAR: subjectivity and sentiment analysis for Arabic social media. Comput. Speech Lang. 28 (1), 20–37.

Agarwal, Apoorv et al., 2011. Sentiment analysis of Twitter data. In: Proceedings of the Workshop on Language in Social Media. Association for Computational Linguistics, Portland, Oregon, pp. 30–38.

Conover, Michael et al., 2011. Political polarization on Twitter. International Conference on Web and Social Media – Barcelona, Spain, pp. 89–96.

Dinerman, Brad, 2010. Social Networking and Security Risks. GFI Software.

Google New Ways to Stay Informed About Presidential Politics [Online] – Google, 1 26, 2016. <https://googleblog.blogspot.com/2016/01/republican-debate-presidential-politics.html>.

Imran, Mohammad, 2015. Processing social media messages in mass emergency: a survey. ACM Comput. Surveys 47 (4), 67.

Lai, Patrick, 2010. Extracting strong sentiment trends from Twitter. Stanford Digital Library, Stanford University, Stanford, California, United States.

Manning, Christopher D., Raghavan, Prabhakar, Schütze, Hinrich, 2008. Introduction to Information Retrieval. Cambridge University Press, Cambridge, United Kingdom.

Pang, Bo, Lee, Lillian, 2008. Opinion mining and sentiment analysis. Found. Trends Inform. Retrieval 2 (2), 1–135.

Salem, Fadi, Mourtada, Racha, 2012. Social media in the Arab world: influencing societal and cultural change? Dubai School of Government, Dubai, UAE.

Schulz, Axel et al., 2013. A fine-grained sentiment analysis approach for detecting crisis related microposts. International Conference on Information Systems for Crisis Response and Management – Baden-Baden, Germany.

Shoukry, Amira, Rafea, Ahmed, 2012. Preprocessing Egyptian dialect tweets for sentiment mining. The Fourth Workshop on Computational Approaches to Arabic Script-based Languages – San Diego, CA, USA.

UNESCO World Arabic Language Day, 2012. UNESCO, 12 18, 05.02.2016 (Online). <http://www.unesco.org/new/en/unesco/events/prizes-and-celebrations/celebrations/international-days/world-arabic-language-day/>.

Vargas, Saúl et al., 2016. Comparing overall and targeted sentiments in social media during crises. International AAAI Conference on Web and Social Media – Cologne, Germany.

Verma, Sudha et al., 2011. Natural language processing to the rescue? Extracting "Situational Awareness" tweets during mass emergency. Fifth International Conference on Weblogs and Social Media – Barcelona, Spain. Association for the Advancement of Artificial Intelligence.