

ARB-SEN at SemEval-2018 Task1: A New Set of Features for Enhancing the Sentiment Intensity Prediction in Arabic Tweets

El Moatez Billah Nagoudi

Laboratoire d'Informatique et de Mathématiques LIM, Laghouat, Algeria
Echahid Hamma Lakhdar University, El Oued, Algeria

e.nagoudi@lagh-univ.dz

Abstract

This article describes our proposed Arabic Sentiment Analysis system named *ARB-SEN*. This system is designed for the International Workshop on Semantic Evaluation 2018 (SemEval-2018), Task1: Affect in Tweets. *ARB-SEN* proposes two supervised models to estimate the sentiment intensity in Arabic tweets. Both models use a set of features including sentiment lexicon, negation, word embedding and emotion symbols features. Our system combines these features to assist the sentiment analysis task. *ARB-SEN* system achieves a correlation score of 0.720, ranking 6th among all participants in the valence intensity regression (V-reg) for the Arabic sub-task organized within the SemEval 2018 evaluation campaign.

1 Introduction and Related Work

According to [Mohammad \(2016\)](#) the Sentiment Analysis (SA) task is used to refer to the “*task of automatically determining the valence or polarity of a piece of text, whether it is positive, negative, or neutral*”.

Nowadays, social media platforms like Twitter, Facebook, LinkedIn, and Quora are widely used ([Lenze, 2017](#)). For instance, [Ranginwala and Towbin \(2017\)](#) estimate that Twitter has 320 million active monthly users. These social media platforms allow people to communicate not only the sentiment they are feeling (positive or negative) but also the intensity of this sentiment. For example, from the tweet of your friend, you can estimate that: he is very happy (most positive), slightly angry (slightly negative), absolutely sad (most negative) or neutral ([Mohammad et al., 2017](#)).

Automatically determining the sentiment intensity is an important task in several application fields, such as public health, intelligence gathering, commerce and social welfare ([Mohammad and Bravo-Marquez, 2017](#)).

In the Semantic Evaluation (SemEval), sentiment analysis in Twitter task has been proposed for the first time in *SemEval-2013* by ([Nakov et al., 2013](#)). Since *SemEval-2013*, this task has become a principal task in SemEval : *SemEval-2014* ([Rosenthal et al., 2014](#)), *SemEval-2015* ([Rosenthal et al., 2015](#)), *SemEval-2016*, ([Nakov et al., 2016](#)) and *SemEval-2017* ([Rosenthal et al., 2017](#)).

In the Arab world, [Salem \(2017\)](#) estimates that the number of monthly active Arabic users was 11.1 million in March 2017, which makes Arabic an emergent language for sentiment analysis in Twitter ([Rosenthal et al., 2017](#)). Sentiment analysis task in Arabic is particularly a challenging research task ([Rosenthal et al., 2017](#)) due to its complex morphological and syntactic structure ([Habash, 2010](#)).

Many Arabic sentiment analysis tools and studies have been proposed in order to overcome this challenge. For example in Arabic newswire, the most relevant works are: ([Abdul-Mageed et al., 2011](#)), ([Elarnaoty et al., 2012](#)). In Arabic reviews we find ([Elhawary and Elfeky, 2010](#)), ([Elnagar, 2016](#)), ([Altowayan and Elnagar, 2017](#)). In Arabic Twitter many researches are focused on the sentiment analysis task such as ([Mourad and Darwish, 2013](#)), ([Abdul-Mageed et al., 2014](#)), ([Refaee and Rieser, 2014](#)), , ([Salameh et al., 2015](#)) and ([Al-dayel and Azmi, 2016](#)).

In this article we present our *ARB-SEN system* devoted to enhancing the detection of sentiment intensity in Arabic tweets. *ARB-SEN system* proposes two methods to measure this valence. Our best submitted method achieves a correlation of 0.720, ranking 6th in the Arabic Detecting Sen-

timent Intensity shared task (Mohammad et al., 2018a), SemEval-2018.

2 System Description

The sentiment intensity detection in ARB-SEN system relies on a set of features. In what follows we describe the considered features:

2.1 Sentiment Lexicon Features (SLF)

We employed the following four sentiment lexicons to extract the SLF features:

Arabic Sentiment (Valence) Lexicons

Created as part of *SemEval-2016* by Kiritchenko et al. (2016), this Arabic sentiment lexicon is a list of 1,168 single words and 198 simple phrases and their associations with positive and negative sentiment. The lexicon include both standard and dialectal Arabic terms.

Arabic Sentiment (Valence) Lexicons

This is a annotated Arabic sentiment lexicon that is created by (Saif M. Mohammad and Kiritchenko, 2016). These lexicons were created by measuring the extent to which the words in a tweets corpus co-occurred with a set of seed positive and seed negative terms. This lexicon includes about 43k entries (23k positive and 20k negative).

ArabSenti sentiment lexicon

ArabSenti is a manually annotated Arabic sentiment lexicon of 14k words that was created and by Abdul-Mageed et al. (2011). Each word in *ArabSenti* is associated with a positive/negative sentiment label.

Dialectal sentiment lexicon

This is a freely available Arabic sentiment lexicon with more than 480 dialectal Arabic words. The lexicon is proposed by Refaee and Rieser (2014) and it is manually annotated by native Arabic speakers. Using these sentiment lexicons, we extract for each tweet four features:

- 1) **Sum Score** The sum of sentiment scores of all the words in the tweet.
- 2) **Average Score** This feature computes the average of sentiment scores of all the words in the tweet.
- 3,4) **Min and Max Score** Represent the minimum and maximum sentiment score of words in the tweet.

For each of these features, if one word in the tweet does not exist in a sentiment lexicon, its corresponding sentiment score is not considered.

2.2 Negation Feature (NF)

Negation refers to words that reverse the sentiment of the word/phrase coming after them. For example: أنا لست سعيد (I'm not happy), in this example the word سعيد (happy) has a *positive* sentiment, however, due to the negation word لست (I'm not) the sentiment of expression becomes *negative*. This feature is used by *ENCU* system (Wang et al., 2016) the best system in SemEval-2016 (Sentiment Intensity Task). Wang et al. (2016) they showed that the sentiment of the phrase can be reversed by adding a negation. Thus, for this binary feature, we have used a list of five main negation word in the Modern Standard Arabic (MAS) { ليس , ما , لن , لا , لم } proposed by (Abdulla et al., 2013). If the tweet contains at least one negation, this feature is set to 1, else 0.

2.3 Word Embedding Feature (WEF)

One of the main advantages of word embedding model is the fact that it allows for the retrieval of a list of words that are used in the same contexts with respect to a given word (Mikolov et al., 2013). In fact , we use the Arabic CBOW model (Zahran et al., 2015) to construct a list of 5-*closet* words for each word in the tweet as described in (Nagoudi et al., 2017). Then, we extract for each tweet the same features described in the section 2.1, with the difference that we compute the sentiment score for each word based on their 5-*closet* in word embedding:

- 1) **Sum Score** The sum of the average sentiment scores of all the 5-*closet* words in tweet.
- 2) **Average Score** This feature computes the average of sentiment scores of the 5-*closet* words in the tweet.
- 3,4) **Min and Max Score** Represent the minimum and maximum average sentiment score of the 5-*closet* words in tweet.

To compute these features, we have used the same sentiment lexicons presented in the section 2.1.

2.4 Emoticons and Emojis Features (EEF)

The emoticons and emojis are already used in the sentiment analysis task in twitter (Read, 2005)

Positive	
Emoticons	Emojis
:-) :) :D :o) =]	😄 😊 😁 😂 😃
:] :3 :c) 8) :-]	😁 😎 😊 😍 😜
Negative	
Emoticons	Emojis
:[:-(:(:-c :c	😡 😠 😞 😓 😔
:-C :C :-[:[:{	😞 😡 😠 😞 😔
Neutral	
Emoticons	Emojis
:/ :-/ :-, :/ :t :s	😐 😐 😐 😐 😐
=/ =I :L =L :Z	😐 😐 😐 😐 😐

Table 1: A sample of the positive, negative and neutral emoticons and emojis.

and (Wolny, 2016). Therefore, we have used the Emoticons and Emojis as an indicator to predict the sentiment intensity of the tweet. We have used 3 set of emoticon and emoji positive, negative and neutral. Table 2 shows a sample of the positive, negative and neutral of emoticons and emojis.

2.5 Models Construction

The previously described features are fed into two different regression classifiers : Linear Regression (*LR*) and Support Vector Regression (*SVR*). We have used the python-based machine learning *scikit-learn* library¹ to trained these classifiers on the training and development data set of SemEval 2018 (Mohammad et al., 2018b), along with the previously discussed features to predict the sentiment intensity score for each tweet. Figure 1 illustrates an overview of the *ARB-SEN* system.

3 Experiments And Results

3.1 Training Data

The organisers of SemEval 2018 provided a training and development data set, which contained 933 and 139 Arabic tweets respectively. Thus, the trial and development are used as training data for our supervised models.

3.2 Data Pre-processing

In order normalize tweets, many pre-processing techniques have been proposed in the literature, such as:(Agarwal et al., 2011), (Ahmed et al.,

¹<http://scikit-learn.org>

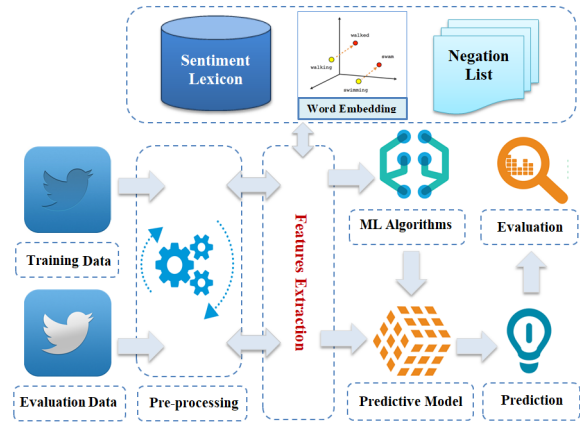


Figure 1: Architecture of the *ARB-SEN* system.

2013), and (Rosenthal et al., 2014). Therefore, we normalize our tweets using the following pre-processing steps:

1. Removing @user_names, RTs, and URLs;
2. Removing diacritics and non-alphanumeric characters;
3. Tokenizing the #hashtags of each tweet by breaking them into words, e.g: #very_nice_day becomes very, nice and day;
4. Normalizing the exchangeable Arabic letters as described in (Darwish et al., 2012), e.g: normalizing اَ, اِ, اِ to ا and replacing final ع followed by ء with ع.

3.3 Tests and Results

To evaluate the performance of our system, our two supervised models were assessed based on their accuracy on the 731 tweets in the Arabic Sentiment Intensity Evaluation Set². In addition, we studied the impact of sentiment lexicon, negation, word embedding and emotion symbols features on the prediction efficiency.

We calculate the Pearson correlation between our assigned Sentiment Intensity scores and the gold labels. The results are presented in Table 2.

These results demonstrate that *SVR* classifier with all features succeed in predict the sentiment intensity in Arabic tweets with a Pearson correlation score of 0.720. However, the *LR* classifier with all features achieves a correlation score of 0.617. Thus, we can easily observe that *SVR* classifier with all features outperforms the *LR* classi-

²<http://saifmohammad.com/WebDocs/AIT-2018/AIT2018-DATA>

Methods	Features	Correlation
LR	SLF	0.523
	SLF+NF	0.524
	SLF+NF+WEF	0.561
	SLF+NF+WEF+EEF	0.617
SVR	SLF	0.647
	SLF+NF	0.649
	SLF+NF+WEF	0.683
	SLF+NF+WEF+EEF	0.720
Baseline	-	-0.052

Table 2: Correlation results

fier with a gain of +11%. Regarding the impact of the extracted features, all of them improve the results of the sentiment intensity prediction. Interestingly, we notice that the word embedding and emotion symbols features play a key role in improving the performance of the prediction accuracy in both classifiers with a mean of +3.5% and +4.7% respectively.

4 Conclusion and Future Work

In this article, we have presented two supervised models to predicate the sentiment intensity in Arabic tweets. Both classifiers are trained along with a set of Arabic tweets characterised by a set of features including: sentiment lexicon, negation, word embedding and emotion symbols features. The performance of our proposed system was confirmed through the Pearson correlation between our assigned sentiment scores and the golden labels. As future work, we are going to extend our features by using an Arabic Combined-Sentiment Word Embedding model. We would also like to further investigate the Arabic sentiment analysis task with more recent classifiers, namely Neural Deep learning.

References

- Muhammad Abdul-Mageed, Mona T Diab, and Mohammed Korayem. 2011. Subjectivity and sentiment analysis of modern standard arabic. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 587–591. Association for Computational Linguistics.
- Muhammad Abdul-Mageed, Mona Diab, and Sandra Kübler. 2014. Samar: Subjectivity and sentiment

analysis for arabic social media. *Computer Speech & Language*, 28(1):20–37.

- Nawaf A Abdulla, Nizar A Ahmed, Mohammed A Shehab, and Mahmoud Al-Ayyoub. 2013. Arabic sentiment analysis: Lexicon-based and corpus-based. In *Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on*, pages 1–6. IEEE.
- Apoorv Agarwal, Boyi Xie, Iliia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the workshop on languages in social media*, pages 30–38. Association for Computational Linguistics.
- Soha Ahmed, Michel Pasquier, and Ghassan Qadah. 2013. Key issues in conducting sentiment analysis on arabic social media text. In *Innovations in Information Technology (IIT), 2013 9th International Conference on*, pages 72–77. IEEE.
- Haifa K Aldayel and Aqil M Azmi. 2016. Arabic tweets sentiment analysis—a hybrid scheme. *Journal of Information Science*, 42(6):782–797.
- A Aziz Altowayan and Ashraf Elnagar. 2017. Improving arabic sentiment analysis with sentiment-specific embeddings. In *Big Data (Big Data), 2017 IEEE International Conference on*, pages 4314–4320. IEEE.
- Kareem Darwish, Walid Magdy, and Ahmed Mourad. 2012. Language processing for arabic microblog retrieval. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 2427–2430. ACM.
- Mohamed Elarnaoty, Samir AbdelRahman, and Aly Fahmy. 2012. A machine learning approach for opinion holder extraction in arabic language. *arXiv preprint arXiv:1206.1011*.
- Mohamed Elhawary and Mohamed Elfeky. 2010. Mining arabic business reviews. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pages 1108–1113. IEEE.
- Ashraf Elnagar. 2016. Investigation on sentiment analysis for arabic reviews. In *Computer Systems and Applications (AICCSA), 2016 IEEE/ACS 13th International Conference of*, pages 1–7. IEEE.
- Nizar Y Habash. 2010. Introduction to arabic natural language processing. *Synthesis Lectures on Human Language Technologies*, 3(1):1–187.
- Svetlana Kiritchenko, Saif M. Mohammad, and Mohammad Salameh. 2016. Semeval-2016 task 7: Determining sentiment intensity of english and arabic phrases. In *Proceedings of the International Workshop on Semantic Evaluation, SemEval ’16*, San Diego, California, June.

- Nele Lenze. 2017. Social media in the arab world: Communication and public opinion in the gulf states. *European Journal of Communication*, 32(1):77–79.
- Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic regularities in continuous space word representations. In *Hlt-naacl*, volume 13, pages 746–751.
- Saif M Mohammad and Felipe Bravo-Marquez. 2017. Wassa-2017 shared task on emotion intensity. *arXiv preprint arXiv:1708.03700*.
- Saif M Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3):26.
- Saif M Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018a. Semeval-2018 task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.
- Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018b. Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.
- Saif M Mohammad. 2016. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier.
- Ahmed Mourad and Kareem Darwish. 2013. Subjectivity and sentiment analysis of modern standard arabic and arabic microblogs. In *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 55–64.
- El Moatez Billah Nagoudi, Jérémy Ferrero, Didier Schwab, and Hadda Cherroun. 2017. Word embedding-based approaches for measuring semantic similarity of arabic-english sentences. In *The 6th International Conference on Arabic Language Processing*, pages 19–33. Springer.
- Preslav Nakov, Zornitsa Kozareva, Alan Ritter, Sara Rosenthal, Veselin Stoyanov, and Theresa Wilson. 2013. Semeval-2013 task 2: Sentiment analysis in twitter.
- Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2016. Semeval-2016 task 4: Sentiment analysis in twitter. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1–18.
- Saad Ranginwala and Alexander J Towbin. 2017. The power of promotion: using social media to promote a radiology department. *Academic Radiology*, 24(4):488–496.
- Jonathon Read. 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL student research workshop*, pages 43–48. Association for Computational Linguistics.
- Eshrag Refaee and Verena Rieser. 2014. An arabic twitter corpus for subjectivity and sentiment analysis. In *LREC*, pages 2268–2273.
- Sara Rosenthal, Alan Ritter, Preslav Nakov, and Veselin Stoyanov. 2014. Semeval-2014 task 9: Sentiment analysis in twitter. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 73–80, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.
- Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. Semeval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 451–463.
- Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval '17*, Vancouver, Canada, August. Association for Computational Linguistics.
- Mohammad Salameh Saif M. Mohammad and Svetlana Kiritchenko. 2016. Sentiment lexicons for arabic social media. In *Proceedings of 10th edition of the the Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia.
- Mohammad Salameh, Saif Mohammad, and Svetlana Kiritchenko. 2015. Sentiment after translation: A case-study on arabic social media posts. In *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 767–777.
- F. Salem. 2017. The arab social media report 2017: Social media and the internet of things: Towards data-driven policymaking in the arab world. Vol. 7.
- Feixiang Wang, Zhihua Zhang, and Man Lan. 2016. Ecnu at semeval-2016 task 7: An enhanced supervised learning method for lexicon sentiment intensity ranking. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 491–496.
- Wiesław Wolny. 2016. Sentiment analysis of twitter data using emoticons and emoji ideograms. *Studia Ekonomiczne*, 296:163–171.
- Mohamed A Zahran, Ahmed Magooda, Ashraf Y Mahgoub, Hazem Raafat, Mohsen Rashwan, and Amir Atyia. 2015. Word representations in vector space and their applications for arabic. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 430–443. Springer.