TeamUNCC at SemEval-2018 Task 1: Emotion Detection in English and Arabic Tweets using Deep Learning

Malak Abdullah **Samira Shaikh**
College of Computing and Informatics
University of North Carolina at Charlotte
North Carolina, U.S
mabdull5, sshaikh2@uncc.edu

Abstract

Task 1 in the International Workshop SemEval 2018, Affect in Tweets, introduces five subtasks (El-reg, El-oc, V-reg, V-oc, and E-c) to detect the intensity of emotions in English, Arabic, and Spanish tweets. This paper describes TeamUNCC's system to detect emotions in English and Arabic tweets. Our approach is novel in that we present the same architecture for all the five subtasks in both English and Arabic. The main input to the system is a combination of word2vec and doc2vec embeddings and a set of psycholinguistic features (e.g. from AffectiveTweets Weka-package). We apply a fully connected neural network architecture and obtain performance results that show substantial improvements in Spearman correlation scores over the baseline models provided by Task 1 organizers, (ranging from 0.03 to 0.23). TeamUNCC's system ranks third in subtask El-oc and fourth in other subtasks for Arabic tweets.

1 Introduction

The rise and diversity of social microblogging channels encourage people to express their feelings and opinions on a daily basis. Consequently, sentiment analysis and emotion detection have gained the interest of researchers in natural language processing and other fields that include political science, marketing, communication, social sciences, and psychology (Mohammad and Bravo-Marquez, 2017; Agarwal et al., 2011; Chin et al., 2016). Sentiment analysis refers to classifying a subjective text as positive, neutral, or negative; emotion detection recognizes types of feelings through the expression of texts, such as anger, joy, fear, and sadness (Agarwal et al., 2011; Ekman, 1993).

SemEval is the International Workshop on Semantic Evaluation that has evolved from SensEval. The purpose of this workshop is to evaluate semantic analysis systems, the SemEval-2018 being the 12th workshop on semantic evaluation. Task 1 (Mohammad et al., 2018) in this workshop presents five subtasks with annotated datasets for English, Arabic, and Spanish tweets. The task for participating teams is to determine the intensity of emotions in text. Further details about Task 1 and the datasets appear in Section 3.

Our system covers five subtasks for both English and Arabic. The input to the system are word embedding vectors (Mikolov et al., 2013a), which are applied to fully connected neural network architecture to obtain the results. In addition, all subtasks except the last one, use document-level embeddings doc2vec (Le and Mikolov, 2014) that are concatenated with different feature vectors. The models built for detecting emotions related to Arabic tweets ranked third in subtask El-oc and fourth in the other subtasks. We use both the original Arabic tweets as well as translated tweets (to English) as input. The performance of the system for all subtasks in both languages shows substantial improvements in Spearman correlation scores over the baseline models provided by Task 1 organizers, ranging from 0.03 to 0.23.

The remainder of this research paper is organized as follows: Section 2 gives a brief overview of existing work on social media emotion and sentiment analyses, including for English and Arabic languages. Section 3 presents the requirements of SemEval Task1 and the provided datasets. Section 4 examines the TeamUNCC's system to determine the presence and intensity of emotion in text. Section 5 summarizes the key findings of the study and the evaluations. Section 6 concludes with future directions for this research.

Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval-2018), pages 350–357
New Orleans, Louisiana, June 5–6, 2018. ©2018 Association for Computational Linguistics

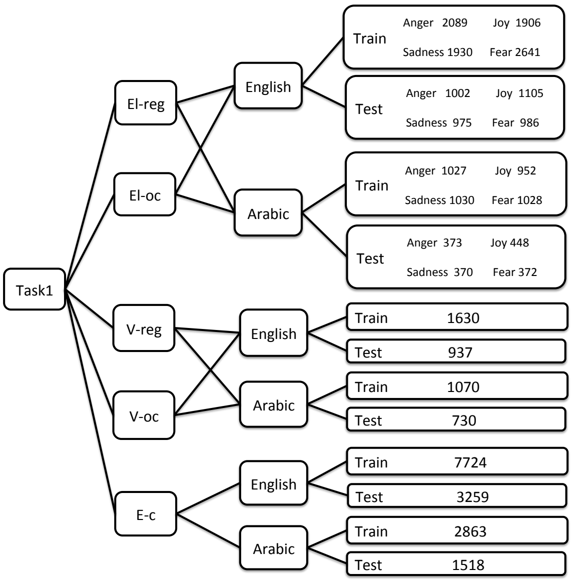
2 Related work

Sentiment and Emotion Analysis: Sentiment analysis was first explored in 2003 by Nasukawa and Yi (Nasukawa and Yi, 2003). An interest in studying and building models for sentiment analysis and emotion detection for social microblogging platforms has increased significantly in recent years (Kouloumpis et al., 2011; Pak and Paroubek, 2010; Oscar et al., 2017; Jimenez-Zafra et al., 2017). Going beyond the task of mainly classifying tweets as positive or negative, several approaches to detect emotions were presented in previous research papers (Mohammad and Kiritchenko, 2015; Tromp and Pechenizkiy, 2014; Mohammad, 2012). Researchers (Mohammad and Bravo-Marquez, 2017) introduced the WASSA-2017 shared task of detecting the intensity of emotion felt by the speaker of a tweet. The state-of-the-art system in that competition (Goel et al., 2017) used an approach of ensembling three different deep neural network-based models, representing tweets as word2vec embedding vectors. In our system, we add doc2vec embedding vectors and classify tweets to ordinal classes of emotions as well as multi-class labeling of emotions.

Arabic Emotion Analysis: The growth of the Arabic language on social microblogging platforms, especially on Twitter, and the significant role of the Arab region in international politics and in the global economy have led researchers to investigate the area of mining and analyzing sentiments and emotions of Arabic tweets (Abdullah and Hadzikadic, 2017; El-Beltagy et al., 2017; Assiri et al., 2016). The challenges that face researchers in this area can be classified under two main areas: a lack of annotated resources and the challenges of the Arabic language's complex morphology relative to other languages (Assiri et al., 2015). Although recent research has been dedicated to detect emotions for English content, to our knowledge, there are few studies for Arabic content. Researchers (Rabie and Sturm, 2014) collected and annotated data and applied different preprocessing steps related to the Arabic language. They also used a simplification of the SVM (known as SMO) and the NaiveBayes classifiers. Another two related works (Kiritchenko et al., 2016; Rosenthal et al., 2017) shared different tasks to identify the overall sentiments of the tweets or phrases taken from tweets in both English and Arabic. Our work uses the state-of-the-art approaches of deep learning and word/doc embedding.

3 Task Description and Datasets

SemEval-2018 Task 1, Affect in Tweets, presents five subtasks (EI-reg, EI-oc, V-reg, V-oc, and E-c.) The subtasks provide training and testing for Twitter datasets in the English, Arabic, and Spanish languages (Mohammad and Kiritchenko, 2018). Task 1 asks the participants to predict the intensity of emotions and sentiments in the testing datasets. It also includes multi-label emotion classification subtask for tweets. This paper focuses on determining emotions in English and Arabic tweets. Figure 1 shows the number of tweets for both training and testing datasets for individual subtasks. We note that subtasks *EI-reg* and *EI-oc* share the same datasets with different annotations, and the same for subtasks *V-reg* and *V-oc*.



Task1
El-reg
El-oc
V-reg
V-oc
E-c
English
Arabic
Train Anger 2089 Joy 1906 Sadness 1930 Fear 2641
Test Anger 1002 Joy 1105 Sadness 975 Fear 986
Train Anger 1027 Joy 952 Sadness 1030 Fear 1028
Test Anger 373 Joy 448 Sadness 370 Fear 372
English
Train 1630
Test 937
Arabic
Train 1070
Test 730
English
Train 7724
Test 3259
Arabic
Train 2863
Test 1518

Figure 1: Datasets of SemEval-2018 Task 1.

The description of each subtask is:

EI-reg: Determine the intensity of an emotion in a tweet as a real-valued score between 0 (least emotion intensity) and 1 (most emotion intensity).

EI-oc: Classify the intensity of emotion (anger, joy, fear, or sadness) in the tweet into one of four ordinal classes (0: no emotion, 1, 2, and 3 high emotion).

V-reg: Determine the intensity of sentiment or valence (V) in a tweet as a real-valued score between 0 (most negative) and 1 (most positive).

V-oc: Classify the sentiment intensity of a tweet into one of seven ordinal classes, corresponding to various levels of positive and negative sentiment intensity (3: very positive mental state can be inferred, 2, 1, 0, -1, -2, and -3: very negative mental state can be inferred)

E-c: Classify the tweet as 'neutral or no emotion' or as one, or more, of eleven given emotions (anger, anticipation, disgust, fear,joy, love, optimism, pessimism, sadness, surprise, and trust).

4 The TeamUNCC System

Our team, TeamUNCC, is the only team that participated in all subtasks of Task 1 of SemEval-2018 for both English and Arabic tweets. Subtasks *El-reg* and *V-reg* are considered similar because they determine the intensity of an emotion or a sentiment (respectively) in a tweet as a real-valued score. While subtasks *El-oc* and *V-oc* classify the intensity of the emotion or the sentiment (respectively) to ordinal classes. Our system, designed for these subtasks, shares most features and components; however, the fifth subtask, *E-c*, uses fewer of these elements. Figure 2 shows the general structure of our system. More details for the system's components are shown in the following subsections: Section 4.1 describes the system's input and prepocessing. Section 4.2 lists the feature vectors, and Section 4.3 details the architecture of neural network. Section 4.4 discusses the output details.

4.1 Input and Preprocessing

EngTweets: The original English tweets in training and testing datasets have been tokenized by converting the sentences into words, and all uppercase letters have been converted to lowercase. The preprocessing step also includes stemming the words and removal of extraneous white spaces. Punctuation have been treated as individual words (".,?!:;()[]#@'), while contractions (wasn't, aren't) were left untreated.

ArTweets: The original Arabic tweets in training and testing datasets have been tokenized, white spaces have been removed, and the punctuation marks have been treated as individual words (".,?!:;()[]#@').

TraTweets: The Arabic tweets have been translated using a powerful translation tool written in python (translate 3.5.0)[1]. Next, the preprocessing steps that are applied to EngTweets are also applied on TraTweets.

4.2 Feature Vectors

AffectTweets-145: Each tweet, in either EngTweets or TraTweets, is represented as 145 dimensional vectors by concatenating three vectors obtained from the AffectiveTweets Weka-package (Mohammad and Bravo-Marquez, 2017; Bravo-Marquez et al., 2014), 43 features have been extracted using the TweetToLexiconFeatureVector attribute that calculates attributes for a tweet using a variety of lexical resources; two-dimensional vector using the Sentiment strength feature from the same package, and the final 100 dimensional vector is obtained by vectorizing the tweets to embeddings attribute also from the same package.

Doc2Vec-300: Each tweet is represented as a 300 dimensional vector by concatenating two vectors of 150 dimensions each, using the document-level embeddings ('doc2vec') (Le and Mikolov, 2014; Lau and Baldwin, 2016). The vector for each word in the tweet has been averaged to attain a 150 dimensional representation of the tweet.

Word2Vec-300: Each tweet is represented as a 300 dimensional vector using the pretrained word2vec embedding model that is trained on Google News (Mikolov et al., 2013b), and for Arabic tweets, we use the pretrained embedding model that is trained on Arabic tweets (Twt-SG) (Soliman et al., 2017).

PaddingWord2Vec-300: Each word in a tweet is represented as a 300 dimensional vector. The same pretraind word2vec embedding models that are used in Word2Vec-300 are also used in this feature vector. Each tweet is represented as a vector with a fixed number of rows that equals the maximum length of dataset tweets and a standard 300 columns using padding of zero vectors.

4.3 Network Architecture

Dense-Network: The input 445 dimensional vector feeds into a fully connected neural network with three dense hidden layers. The activation function for each layer is RELU (Maas et al., 2013), with 256, 256, and 80 neurons for each layer, respectively. The output layer consists of one sigmoid neuron, which predicts the intensity of the emotion or the sentiment between 0 and 1. Two dropouts are used in this network (0.3, 0.5) after the first and second layers, respectively. For optimization, we use SGD (Stochastic Gradient Descent) optimizer (lr=0.01, decay=1×10^{-6},

[1] https://pypi.python.org/pypi/translate

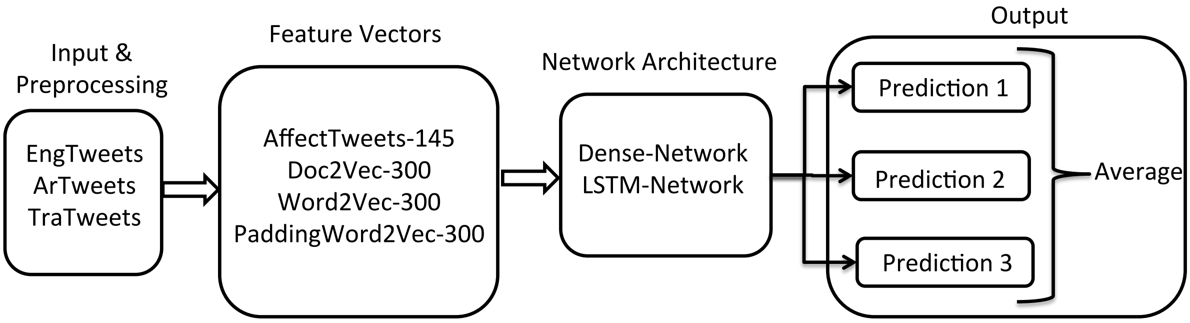


Figure 2: The structure for our system.

and momentum=0.9) [2], optimizing for 'mse' loss function and 'accuracy' metrics. Early stopping is also applied to obtain best results.

LSTM-Network: The input vector feeds an LSTM of 256 neurons that passes the vector to a fully connected neural network of two hidden layers and two dropouts (0.3, 0.5). The first hidden layer has 256 neurons, while the second layer has 80 neurons. Both layers use the RELU activation function. The output layer consists of one sigmoid neuron, which predicts the intensity of the emotion or the sentiment between 0 and 1. For optimization, we use SGD optimizer (lr=0.01, decay=1 × 10^{-6}, and momentum=0.9), optimizing for 'mse' loss function and 'accuracy' metrics as well as early stopping to obtain the best results.

4.4 Output

Subtasks El-reg, El-oc, V-reg, and V-oc: These four subtasks for each language (English and Arabic) share the same structure as shown in Figure 2, the only difference is in the output stage. Each subtask passes the tweets to three different models that produces three predictions. See Table 1 and Table 2 for more comprehensive details on how each prediction with English and Arabic language is produced, respectively. The average of the predictions for each tweet is a real-valued number between 0 and 1. This output is considered the final output for both subtasks *El-reg* and *V-reg*, while subtasks *El-oc* and *V-oc* classify this real-valued number to one of the ordinal classes that are shown in Section 3. We note that *El-reg* and *El-oc* shares the same datasets. We also noticed that V-reg and V-oc shares the same dataset. Therefore, we found the ranges of values for each ordinal class by comparing the datasets. Table 3 shows the range of values to obtain the ordinal classes for *El-oc* subtask in English, Table 4 shows the same for *El-oc* subtask in Arabic, and Table 5 shows the for *V-oc* in both English and Arabic.

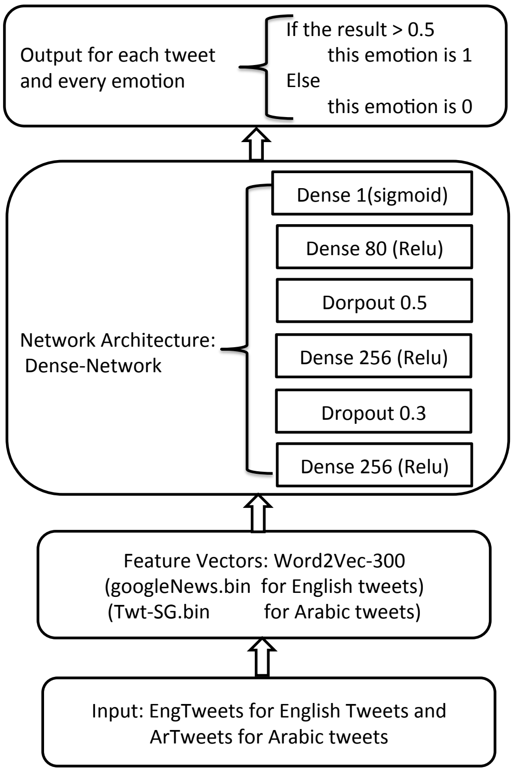


Figure 3: The detailed structure for our system related to subtask E-c.

Subtask E-c: In this subtask, our system makes only one prediction. See Figure3 for more details on the process of predicting the results. The input is *EngTweets* for English language and ArTweets for Arabic language. We use *Word2Vec-300* as the feature vector with *GoogleNews* for English tweets and *Twt-SG* for Arabic tweets. The network architecture is Dense-Network. This process is applied for each emotion of the eleven emotions:

[2] https://keras.io/optimizers/

-	**Prediction 1**	**Prediction2**	**Prediction3**
Input	EngTweets	EngTweets	EngTweets
Feature Vectors	AffectTweets-145 Doc2Vec-300	AffectTweets-145 Word2Vec-300	PaddingWord2Vec-300
Neural Network	Dense-Network	Dense-Network	LSTM-Network

Table 1: The Architecture details for English subtasks El-reg, El-oc, V-reg, and V-oc.

-	**Prediction 1**	**Prediction2**	**Prediction3**
Input	TraTweets	ArTweets TraTweets	ArTweets
Feature Vectors	AffectTweets-145 Doc2Vec-300	AffectTweets-145 Word2Vec-300	PaddingWord2Vec-300
Neural Network	Dense-Network	Dense-Network	LSTM-Network

Table 2: The Architecture details for Arabic subtasks El-reg, El-oc, V-reg, and V-oc.

Output class	**Angry**	**Joy**	**Fear**	**Sadness**
0: no emotion can be inferred	0-0.42	0-0.36	0-0.57	0-0.44
1: low amount of emotion can be inferred	0.42-0.52	0.36-0.53	0.57-0.69	0.44-0.54
2: moderate amount of emotion can be inferred	0.52-0.7	0.53-0.69	0.66-0.79	0.54-0.7
3: high amount of emotion can be inferred	0.7-1	0.69-1	0.79-1	0.7-1

Table 3: Classify the output to ordinal classes for English El-oc.

Output class	**Angry**	**Joy**	**Fear**	**Sadness**
0: no emotion can be inferred	0-0.40	0-0.31	0-0.45	0-0.47
1: low amount of emotion can be inferred	0.40-0.55	0.31-0.51	0.45-0.56	0.47-0.54
2: moderate amount of emotion can be inferred	0.55-0.64	0.51-0.75	0.56-0.76	0.54-0.67
3: high amount of emotion can be inferred	0.64-1	0.75-1	0.76-1	0.67-1

Table 4: Classify the output to ordinal classes for Arabic El-oc.

Output class	**English Sentiment**	**Arabic Sentiment**
-3: very negative emotional state can be inferred	0-0.23	0-0.20
-2: moderately negative emotional state can be inferred	0.23-0.38	0.20-0.37
-1: slightly negative emotional state can be inferred	0.38-0.43	0.37-0.43
0: neutral or mixed emotional state can be inferred	0.43-0.61	0.43-0.56
1: slightly positive emotional state can be inferred	0.61-0.70	0.56-0.69
2: moderately positive emotional state can be inferred	0.70-0.78	0.69-0.81
3: very positive emotional state can be inferred	0.78-1	0.81-1

Table 5: Classify the output to ordinal classes for English and Arabic V-oc.

anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust. The output of each individual tweet is a real-valued number between 0 and 1. This output is normalized to either 1 (contains an emotion) if it is greater than 0.5 or 0 (no emotion) if it is less than 0.5.

5 Evaluations and Results

Each participating system in the subtasks *El-reg*, *El-oc*, *V-reg*, and *V-oc*, has been scored by using Spearman correlation score. The subtask *E-c* has been scored by using accuracy metric. Table 6 shows the performance of our system in *E-reg* and *El-oc* with each emotion and the average score for both English and Arabic. Table 7 shows the results for subtasks *V-reg*, *V-oc*, and *E-c*. The performance of our system beats the baseline model's performance, which is provided by the Task's organizers, see Figure 4 to capture the difference between the two performances. Our system ranks third in the subtask *El-oc* for Arabic language, and Fourth in the subtasks *El-reg*, *V-reg*, *V-oc*, and *E-*

Task	Angry	Joy	Fear	Sadness	Average
El-reg (English)	0.722	0.698	0.692	0.666	**0.695**
El-reg (Arabic)	0.524	0.657	0.576	0.631	**0.597**
El-oc (English)	0.604	0.638	0.544	0.610	**0.599**
El-oc (Arabic)	0.459	0.538	0.483	0.587	**0.517**

Table 6: The Spearman correlation scores for subtasks El-reg and El-oc.

Task	Spearman score
V-reg (English)	**0.787**
V-reg (Arabic)	**0.773**
V-oc (English)	**0.736**
V-oc (Arabic)	**0.748**

Task	Accuracy score
E-c (English)	**0.471**
E-c (Arabic)	**0.446**

Table 7: The results for subtasks V-reg, V-oc, and E-c.

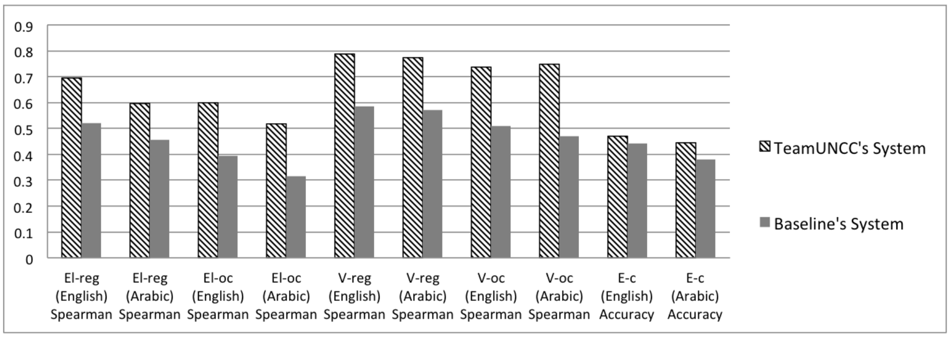


Figure 4: Comparing performances of the TeamUNCC and the baseline systems.

c for Arabic language too. It is worth mentioning that these results have been obtained by using the task datasets without using any external data.

6 Conclusion

In this paper, we have presented our system that participated in Task 1 of Semeval-2018. Our system is unique in that we use the same underlying architecture for all subtasks for both languages - English and Arabic to detect the intensity of emotions and sentiments in tweets. The performance of the system for each subtask beats the performance of the baseline's model, indicating that our approach is promising. The system ranked third in El-oc for Arabic language and fourth in the other subtasks for Arabic language too.

In this system, we used word2vec and doc2vec embedding models with feature vectors extracted from the tweets by using the AffectTweets Weka-package, these vectors feed the deep neural network layers to obtain the predictions.

In future work, we will add emotion and valence detection in Spanish language to our system by applying the same approaches that have been used with Arabic. We also want to investigate the Arabic feature attributes in order to enhance the performance in this language.

References

Malak Abdullah and Mirsad Hadzikadic. 2017. Sentiment analysis on arabic tweets: Challenges to dissecting the language. In *International Conference on Social Computing and Social Media*, pages 191–202. Springer.

Apoorv Agarwal, Boyi Xie, Ilia 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.

Adel Assiri, Ahmed Emam, and Hmood Al-Dossari. 2016. Saudi twitter corpus for sentiment analysis. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 10(2):272–275.

Adel Assiri, Ahmed Emam, and Hmood Aldossari. 2015. Arabic sentiment analysis: a survey. *Interna-*

tional Journal of Advanced Computer Science and Applications, 6(12):75–85.

Felipe Bravo-Marquez, Marcelo Mendoza, and Barbara Poblete. 2014. Meta-level sentiment models for big social data analysis. *Knowledge-Based Systems*, 69:86–99.

Delenn Chin, Anna Zappone, and Jessica Zhao. 2016. Analyzing twitter sentiment of the 2016 presidential candidates. *News & Publications: Stanford University*.

Paul Ekman. 1993. Facial expression and emotion. *American psychologist*, 48(4):384.

Samhaa R El-Beltagy, Talaat Khalil, Amal Halaby, and Muhammad Hammad. 2017. Combining lexical features and a supervised learning approach for arabic sentiment analysis. *arXiv preprint arXiv:1710.08451*.

Pranav Goel, Devang Kulshreshtha, Prayas Jain, and Kaushal Kumar Shukla. 2017. Prayas at emoint 2017: An ensemble of deep neural architectures for emotion intensity prediction in tweets. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 58–65.

Salud Maria Jimenez-Zafra, M Teresa Martin Valdivia, Eugenio Martinez Camara, and Luis Alfonso Urena-Lopez. 2017. Studying the scope of negation for spanish sentiment analysis on twitter. *IEEE Transactions on Affective Computing*.

Svetlana Kiritchenko, Saif Mohammad, and Mohammad Salameh. 2016. Semeval-2016 task 7: Determining sentiment intensity of english and arabic phrases. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 42–51.

Efthymios Kouloumpis, Theresa Wilson, and Johanna D Moore. 2011. Twitter sentiment analysis: The good the bad and the omg! *Icwsm*, 11(538-541):164.

Jey Han Lau and Timothy Baldwin. 2016. An empirical evaluation of doc2vec with practical insights into document embedding generation. *arXiv preprint arXiv:1607.05368*.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International Conference on Machine Learning*, pages 1188–1196.

Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Saif M Mohammad. 2012. # emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 246–255. Association for Computational Linguistics.

Saif M Mohammad and Felipe Bravo-Marquez. 2017. Wassa-2017 shared task on emotion intensity. *arXiv preprint arXiv:1708.03700*.

Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.

Saif M Mohammad and Svetlana Kiritchenko. 2015. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, 31(2):301–326.

Saif M. Mohammad and Svetlana Kiritchenko. 2018. Understanding emotions: A dataset of tweets to study interactions between affect categories. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference*, Miyazaki, Japan.

Tetsuya Nasukawa and Jeonghee Yi. 2003. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, pages 70–77. ACM.

Nels Oscar, Pamela A Fox, Racheal Croucher, Riana Wernick, Jessica Keune, and Karen Hooker. 2017. Machine learning, sentiment analysis, and tweets: an examination of alzheimers disease stigma on twitter. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 72(5):742–751.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10.

Omneya Rabie and Christian Sturm. 2014. Feel the heat: Emotion detection in arabic social media content. In *The International Conference on Data Mining, Internet Computing, and Big Data (BigData2014)*, pages 37–49. The Society of Digital Information and Wireless Communication.

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-2017)*, pages 502–518.

Abu Bakr Soliman, Kareem Eissa, and Samhaa R El-Beltagy. 2017. Aravec: A set of arabic word embedding models for use in arabic nlp. *Procedia Computer Science*, 117:256–265.

Erik Tromp and Mykola Pechenizkiy. 2014. Rule-based emotion detection on social media: putting tweets on plutchik's wheel. *arXiv preprint arXiv:1412.4682*.