International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-3, February 2020

Twitter Sentiment Analysis Classification in the Arabic Language using Long Short-Term Memory Neural Networks

Wisam Hazım Gwad Gwad, Imad Mahmood Ismael Ismael, Yasemin Gültepe

Abstract*: The increasing use of social media and the idea of extracting meaningful expressions from renewable and usable data which is one of the basic principles of data mining has increased the popularity of Sentiment Analysis which is an important working area recently and has expanded its usage areas. Compiled messages shared from social media can be meaningfully labeled with sentiment analysis technique. Sentiment analysis objectively indicates whether the expression in a text is positive, neutral, or negative. Detecting Arabic tweets will help for politicians in estimating universal incident-based popular reports and people's comments. In this paper, classification was conducted on sentiments twitted in the Arabic language. The fact that Arabic has twisted language features enabled it to have a morphologically rich structure. In this paper we have used the Long Short Term Memory (LSTM), a widely used type of the Recurrent Neural Networks (RNNs), to analyze Arabic twitter user comments. Compared to conventional pattern recognition techniques, LSTM has more effective results in terms of having less parameter calculation, shorter working time and higher accuracy.

***Keywords*: Semantic analysis, Arabic language, classification, deep learning.**

I. INTRODUCTION

Nowadays, the widespread use of social media, blogging and online shopping has enabled users to access many interpretable ideas via virtual media. It is not possible to manually analyze such a large amount of user-generated data, so an effective and intelligent technique is required to be able to analyze and provide the polarity of this textual data. One of the first fields of study of Neuro Linguistic Programming (NLP) is sentiment analysis [1]. In recent years, algorithms with machine learning capabilities that use natural language processing technology have been developed and accuracy in sentiment analysis has been increasing.

Revised Manuscript Received on February 05, 2020.
* Correspondence Author
Wisam Hazım Gwad Gwad, Institute of Science and Technology, Kastamonu University, Kastamonu, Turkey. Email: wgwad@ogr.kastamonu.edu.tr
Imad Mahmoos Ismael Ismael, Institute of Science and Technology, Kastamonu University, Kastamonu, Turkey.
Yasemin Gültepe*, Department of Computer Engineering, Kastamonu University, Kastamonu, Turkey. Email: yasemin.gultepe@yahoo.com

Sentiment analysis is an active field of study where subjective short texts or images, such as product reviews, movie comments, and tweets, are classified according to their level of positivity or negativity. Although some studies have also been classified as neutral, it is not very common.

Although sentiment analysis is a commonly used term, it is also found in the field of sentiment classification and idea extraction, especially in usage areas of idea mining. Idea mining and sentiment analysis are generally used in the same sense, but some researchers say that people extract and analyze the idea of an entity related to idea mining, while emotion analysis finds and analyzes sentiment and sentiment expressions within the text [2] [3].

The most basic step of text-based problems is the conversion of text or speech into letters into a computer language consisting of numerical vectors. Expressing text based on character and word are two different ways of doing this. Although it is the first step, innovative and impressive contributions are constantly made [4].

The presence and activities of the sentiment terms glossary and natural language processing tools vary from language to language. English, the language on which most natural language processing is studied, has many powerful natural language processing tools and dictionaries of sentiment terms. Tools in other languages are not as numerous and powerful as English. A limited number of studies have also been conducted for Arabic. The spelling direction of the Arabic language is 28 letters from right to left. Most Arabic words are morphologically reproduced the literary root list, which is a triple, quadruple, or quintet literary [5, 6, 7, 8].

The purpose of sentiment analysis is to discover ideas, describe the emotions they explain and then classify their poles for later decision-making. In general, texts are polarized as a dual (positive or negative), which is a binary classification. Some of the other types except binary classification are as follows [9]:

- Fine-grained: In some problems, it may not be sufficient to evaluate it as positive or negative. For this reason, very positive and very negative expressions also make it possible to evaluate whether the speaker is angry or happy.
- Sentiment detection: Sentiment detection aims to detect emotions such as happiness, sadness, frustration and anger.

- Aspect-based: Aspect-based is aimed to measure the satisfaction of a person about a product rather than detecting positive or negative thoughts about a product.

Many tools and techniques are available today for automatic sentiment classification of data. Ahmad et al. (2017)

present three approaches used for this purpose: machine learning, lexicon based techniques and hybrid techniques [10].

The machine learning model is an algorithm that can be learned from data without relying on existing program applications when statistical models deal with low-quality, low-volume data, and there is a chance that compliance can occur. It is about finding relationships between variables to predict results. Learning strategies in machine learning methods; supervised, unsupervised and reinforced.

Lexicon based techniques are unsupervised learning. In unsupervised learning, it is aimed to observe the related samples as in clustering analysis and to define the classes based on the similarities between the characteristics of these examples.

The lexical methods that use simple algorithms use a predefined sensitivity dictionary where each word is related to a certain sensitivity. Sensitivity dictionary is a list of general lexical features classified as positive or negative according to their semantic orientations (polarity and power) [11, 12]. Unsupervised learning methods are often used for clustering. Uncontrolled learning methods such as K-Means, K-Medoids or CLARANS are used.

A dictionary of emotion analysis has been developed for many languages. Many emotion dictionaries such as SentiWordNet [13], SenticTweety [14], MPQA Opinion Corpus [15] have been created for English, which is one of the languages on which emotion analysis is performed most. The degree of emotion of the words in these dictionaries is combined with different methods and the sense of the whole text is revealed and classified.

Go et al. (2009) aim to perform sentiment analysis by using the twitter data using remote supervised learning method. The data set consisting of 1.600.000 tweets was used. The data set; Support Vector Machine (SVM), Naïve Bayes (NB), Maximum Entropy classification methods; unigram, bigram and unigram + bigram properties and all of them to compare the success rates have been tested. According the experiment result of this study, 82.2% accuracy rate was obtained according to SVM unigram, 81.6% accuracy rate was obtained compared to NB bigram and very good performance with 83% accuracy rate was obtained by using Maximum Entropi unigram and bi-gram together [16]. There are many studies on Turkish data sets related to sentiment analysis. Nizam and Akın (2014) investigated whether the distribution of data in the classes had an effect on the success rate of the classification algorithm. Two data sets, balanced and unbalanced, were used from the tweets of different products of some companies in the food sector. These two different data sets are divided into three different classes one by one, positive (+), negative (-) and neutral form. Tweets in the data sets were divided into three classes as hand, positive, negative and neutral. The first data set consists of a total of 2000 data: 1113 in the positive class, 277 in the negative class and 610 in the neutral class. The second data set consists of a total of 824 data, 257 in the positive class, 277 in the negative class and 290 in the neutral class. NB, Random Forest (RF), SVM, Decision Tree (J48) and k-Nearest Neighbors classification algorithms using two sets of accuracy, accuracy (precision), sensitivity (recall) and kappa statistics. According to the performance criteria and kappa statistic results of these four models, the balanced data set performed better than the unbalanced data set. In other words, one of the factors affecting the success of classification algorithms is the distribution of data in the classes. They obtained the best performance from SVM classification algorithm with 72.33% accuracy rate [17].

Abdulla et al. address both applications to semantic analysis: Collection-based and lexicon-based. They discuss how to create a manually labelled dataset and afterwards through the comprehensive paces of creating the dictionary. Tests were performed at different phases of this progression to carry out the progressions taken in the correctness of the system. The corpus-based tool using SVM for the classifying the data set has been found to provide the highest accuracy [18].

In this study; deep learning methods were used for sentiment analysis. Chapter II gives an overview of methodology; In the Chapter III, experimental results and basic findings of the approaches are given, in Chapter IV, the paper is summarized and the results are given.

II. METHODOLOGY

This section first provides information about preferred data set in the paper. Then, the details in the LSTM classifier training, which is one of the deep learning approaches, are explained.

A. Dataset

The basic data source of the paper is social media; it includes various platforms where users can share their comments and complaints about the products, express their political opinions and express their personal opinions, tastes or requests.

Twitter has a unique language and the data it has for reasons such as misspelled words is hard to understand with human perception. For this reason, various studies are carried out on natural language processing and data mining and sectoral products are brought to the industry with these methods. Sentiment analysis techniques are applied to user messages obtained from twitter and studies are conducted to determine the emotion that people feel when writing the message. While some of the sentiment analysis studies that attempted to reveal the emotions in the texts tried to determine only positive and negative emotions, some classified the emotions according to a scale of 0 to 10 (0-very negative, 10-very positive) [19].

The data set for this paper was obtained from the UCI Machine Learning Repository database [20]. Approximately 2000 tweets were collected for positive (1000 tweet) and negative (1000 tweet) class labels and a data set was created for Arabic sensitivity analysis.

The tweets in this dataset relate to politics and art. A tweet can be represented with a one-hot encoding vector representation. The twitter data set was split into two classes: Positive polarity and negative polarity [21].

The language used on the Internet in Arab countries is generally Modern Standard Arabic (MSA). Modern Standard Arabic is a simple Arabic language based on classical Arabic and is used as an official written language. MSA is used to provide news from most media sources. MSA also organizes conferences, religious advice and speeches for the purpose of mass media television broadcasting. Table 1 give sample tweets from the data set.

Table- I: Positive and negative tweets examples

Twitter	The sense of twitter
معلومه ممتازه Excellent information	Positive
واضح جدا إن أغلب الناس هاي عندها مشاكل نفسيه It is very clear that most people hai have psychological problems	Negative

The text of the tweet is also important in terms of being one of the most accumulated data types through social media and in terms of the sentiments it contains. One of the aims of data analytics is to analyze the texts written by people in order to express their feelings or express their opinions and to convert them into meaningful information and to create value.

Table II shows some statistical information about the twitter data set for the Arabic sentiment test data used in the paper. Total of tweets were (2000), total of words (16958), total of average word in each tweet (17.16), and total of average characters in each tweet (99.06) were in the data set.

Table- II: Total Number of Words in Dataset

Dataset	Number
Total Tweets	2000
Total Positive Tweets	1000
Total Negative Tweets	1000
Positive Word	7189
Negative Word	9769
Total Words	16.958

B. ANALYSIS OF SOCIAL MEDIA WITH TWITTER

Increased use of twitter in areas such as mass and perception management enables twitter to be defined as a social megaphone. Therefore, institutions and individuals who use social media channels correctly can benefit from their positive effects. Through social media platforms, millions of photos, articles and videos are shared on the internet every day. Through these shares, users can quickly comment on specific topics to relevant people or shares.

Initial studies of sentiment analysis are referred to as sentimental polarity and aim to classify the given text as positive, negative and neutral. Later studies also allowed analyses indicating different sentiment states. More fine-grained sentiment classification is possible, for example, using a sentiment score ranging from -10 (very negative) to 0 (neutral) to +10 (very positive) [22].

Automatic sentiment analysis takes on this task since it is impossible for political persons, firms or any unit that wants to reach any subjective information statistics to analyze without any tools.

There are three main stages of classification in sentiment analysis. These are; Document level, sentence level and appearance level. In document level, sentiment analysis studies, the whole document is considered as a whole. There is only one result of the document processing. The content of the document is positive or negative. At sentence level of sentiment analysis; each sentence written in the document or in the data text is evaluated separately and positive, negative or neutral results can be obtained on the basis of the sentence. Neutral thinking means that ideas cannot be obtained. The decision is made according to the main sentences. In each case, sentences do not consist of a single decision sentence. In addition to these sentences, it may also include side clauses. Appearance level is a more detailed sentiment analysis process than the other two levels. The appearance level first appeared in the literature as the property level. It is a process in which analysis is made according to the characteristics and details of the existing asset [23].

C. LSTM ARCHITECTURES

LSTMs are differentiated from RNNs in terms of long-term memory due to some differences in latency calculations. If an assessment is made on a word prediction problem, LSTM can make meaningful information about the current word after each word using the information of the previous words. Occasionally, the information required to clearly remove the searched word may be too far from the searched word, the distance between them may be too much. While RNN is inadequate in such cases, LSTM overcomes the problem of long-term dependence with strong latency calculations [24].

Hochreiter and Schmidhuber (1997) [25] suggested the LSTM architecture to solve the problem of transferring the dependencies between long-term sequence elements over state information. This architecture was later described by Gers et al. (1999) [26].

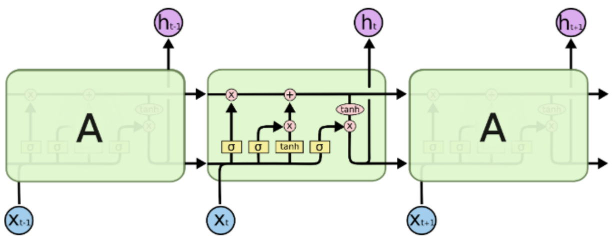

Fig. 1. Representation of the input and output vectors of an ordinary LSTM structure.

The LSTM structure is shown in “Fig. 1” [27]. The LSTM structure is calculated with the previous state h_{t-1} state information, h_t output x_t input as in RNN networks. Unlike conventional RNN structures, state information is stored in a separate artificial neural network called a memory cell. The task of the memory cell is to store the state information, to forget at a certain level at each step, and to continue to store the new state information by combining it with the previous state information.

In order to examine the LSTM structure step by step, the LSTM structure generally consists of three separate sections. These are Forget Gate, Input Gate and Output Gate. The purpose of the forgetting gate is to decide how much of the state information will be forgotten and how much of it will be transferred to the next stage. In the second step, there is a separate artificial neural network structure that calculates the information to be added to the information according to the x_t input at the time t and the output of the previous h_{t-1}. Finally, at the output gate, h_t output of the system is calculated. LSTM is a deep neural network algorithm commonly used in time series estimations. LSTM architectures have been successfully applied in different fields.

III. EXPERIMENTS

In natural language processing, it is much more tough to study sentiment analysis, especially in the field of target-based emotion analysis, in languages that are structurally different from the languages studied frequently.

The methods commonly used for converting text to numbers in developing an application on natural language processing are as follows:

- Calculating vector equivalence (WordVect) for words and giving these vectors in a number of ways.
- It is to use the embeding layer, which holds words in a sequence and translates them into vectors of a certain size during training.
- It is to fall as the string of text. To express each letter with a vector as long as the number of letters.

Sentiment Analysis is a method used to judge someone's sentiment or to make sense of a person's sentiment according to a particular thing. It is basically a text processing process and aims to determine the class that the given text wants to express emotionally. The data set used in this paper is about whether the tweets are positive or negative.

The proposed model in this paper was trained and evaluated using the twitter data set with respect to the Arabic sentiment analysis. We used dataset as 70% for training, 30% for testing. LSTM networks are implemented using the Keras deep learning library in the background of Tensorflow in Python. Many experiments have been performed with various LSTM configurations to achieve the best results with single-layer LSTM. This LSTM layer has 256 stored units and sigmoid is preferred as the activation function. In addition, because LSTM is much more efficient than other optimization methods, Adam optimization algorithm is used for parameter learning. The learning rate of Adam was set to 0.001. Binary cross entropy loss function was preferred for loss function. "Fig. 2" explains how the LSTM architecture is used.

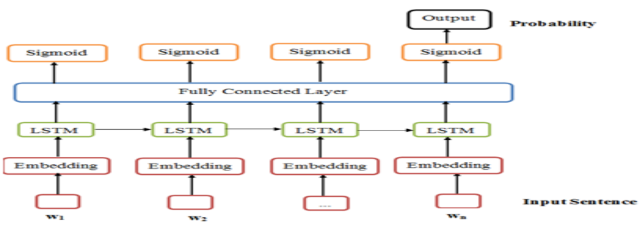

Fig. 2. LSTM architecture used in this paper.

So as to measure the accuracy of the data set used in the proposed LSTM model, another study using the same data set was compared with different learning and prediction algorithms [18]. SVM, NB, K-Nearest Neighbor (K-NN) and Decision Tree (D-Tree) methods were tested by Abdulla et al. (2013). As a result of the comparison, when the Table III is examined, it is seen that the offered LSTM model gives the highest success (accuracy=89.8% with GloVe embedding) on this dataset.

Table- III: Results obtained by other methods

Algorithm	Accuracy Results
SVM	84.7%
NB	80.4%
K-NN	51.3%
D-Tree	50%
LSTM	89.8%

As can be seen from the results presented in Fig. 3, LSTM results perform better by 89.8%.

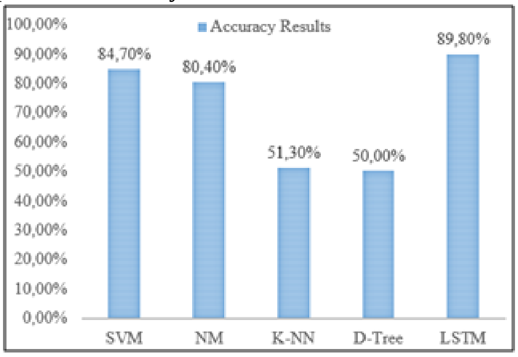

Fig. 3. Graphical representation of accuracy results.

IV. CONCLUSION

Following the developments in artificial intelligence in recent years, the importance and number of artificial intelligence applications based on natural language processing is increasing day by day. The applications developed in this regard will eliminate the human power in many jobs and enable us to use infrastructure that works faster and more accurately. One of the common examples developed with natural language processing is sentiment analysis.

Although Arabic is an important language widely spoken in the world, few studies have focued on the analysis of sentiment in Arabic. The rapid increae in Arabic texts as well as the recent political changes in the Arab countries that affect the whole world, show a clear need to support the extraction of Arabic sentiments. Thus LSTM neural network an aspect for sentiment polarity classification was adopted in this paper. As a result of the proposed model, training and test, an average performance of 89.8% was achieved. The same Arabic dataset was tested with traditional machine learning algorithms [18] and the proposed LSTM model achieved the highest performance.

Innovative architectures have emerged with each passing day as the studies in the domain of deep learning have progressed quite rapidly.

As an alternative to classification models, the research of different deep learning architectures can be carried out and tested one step further.

REFERENCES

1. Liu, B., "Sentiment analysis and opinion mining", Morgan & Claypool, 143, San Rafael.
2. R. Bhonde, B. Bhagwat, S. Ingulkar, and A. Pande, "Sentiment analysis based on dictionary approach", International Journal of Emerging Engineering Research and Technology, 2015, 3(1), pp. 51-55.
3. T. Mikalai, and P. Themis, "Survey on mining subjective data on the web", Data Mining and Knowledge Discovery., 2012, 24, pp. 478-514.
4. N. Pervan, "Semantic inference from turkish texts using deep learning approaches", Ankara University, Graduate School of Natural and Applied Sciences, 2019.
5. M. Korayem, D. Crandall, and M. Abdul-Mageed, "Subjectivity and sentiment analysis of Arabic: A survey", In Advanced Machine Learning Technologies and Applications, Springer Berlin Heidelberg, 2012, pp. 128-139.
6. M. Gridach, H. Haddad, and H. Mulki, "Empirical evaluation of word representations on Arabic sentiment analysis", In International Conference on Arabic Language Processing, 2017, pp. 147-158.
7. A. Alwehaibi, and K. Roy, "Comparison of pre-trained word vectors for Arabic text classification using deep learning approach", 17th IEEE International Conference on Machine Learning and Applications, 2017, pp. 1471-1474.
8. A. Fraghaly, and K. Shaalan, "Arabic natural language processing: Challenges and solutions", ACM Transactions on Asian Language Information Processing (TALIP0), 2009, 8(4).
9. T. Feilong, F. Luoyi, Y. Bin, and X. Wenchao, "Aspect based fine-grained sentiment analysis for online reviews", Information Sciences, 2019, 488, pp. 190-204.
10. M. Ahmad, S. Aftab, S. S. Muhammad, and S. Ahmad, "Machine learning techniques for sentiment analysis: A review", International Journal of Multidisciplinary Sciences and Engineering, 2017, 8(3).
11. B. Liu, "Sentiment analysis and subjectivity", Handbook of natural language processing, 2010, 2, pp. 627-666.
12. S. Almatarneh, and P. Gamallo, "A lexicon based method to search for extreme opinions", PloS ONE, 2018, 13(5).
13. SentiWordNet, 10.12.2019, [Online]. Available: http://swn.isti.cnr.it/.
14. SenticTweety, 10.12.2019, [Online]. Available: http://www.twitter.gelbukh.com/.
15. MPQA Opinion Corpus, 10.12.2019, [Online]. Available: https://mpqa.cs.pitt.edu/corpora/mpqa_corpus/.
16. A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision", 2009, Computer Science.
17. H. Nizam, and S. S. Akın, "Comparison of machine learning and sentiment analysis of balanced and unbalanced data sets in social media", XIX. Internet Conference in Turkey, 2014.
18. N. Abdulla, N. A. Ahmed, A., M. A. Shehab, and M. Al-Ayyoub, "Arabic sentiment analysis: Lexicon-based and corpus-based", in 2013 IEEE Jordan conference on applied electrical engineering and computing technologies, 2013.
19. M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strenght detection in short informal text", Journal of the American Society for Information Science and Technology, 2010, 61(12), pp. 2544-2558.
20. N. A. Abdulla, "Twitter data set for Arabic sentiment analysis data set", 04.11.2014, [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Twitter%2BData%2Bset%2Bfor%2BArabic%2BSentiment%2BAnalysis#.
21. V. A. Kharde, and S. S. Sonawane, "Sentiment analysis of twitter data: A survey of techniques", International Journal of Computer Application, 2016, 139(11).
22. A. Trilla, and F. Alías, "Sentence-based sentiment analysis for expressive text-to-speech", IEEE Transactions on Audio Speech and Language Processing, 2013, 21(2), pp. 223-233.
23. W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: Survey", Ain Shams Engineering Journal, 2014,5, pp. 1093-1113.
24. Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures", Neural Computation, 2019, 31(1), pp. 1-36.
25. S. Hochreiter, and J. Schmidhuber, "LSTM can solve hard long time lag problems", Advances in Neural Information Processing Systems, 1997, pp. 473-479.
26. F. A. Gers, Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM", International Conference on Artificial Neural Networks, 1999, 2, pp. 850-855.
27. C. Olah, "Understanding LSTM networks". 10.12.2019, [Online]. Available:http://colah.github.io/posts/2015-08-Understanding-LSTMs

AUTHORS PROFILE

Wisam Hazım Gwad Gwad Graduated B.Sc. in Computer engineering Department from University of Technology Baghdad /Iraq in 2004. He received his M.Sc. degree electrical and computer engineering, Altınbaş University/Turkey. He is currently pursuing institute of Science and Technology, Kastamonu University, Kastamonu, Turkey. His research interests include artificial intelligent, deep leering, natural language processing.

Imad Mahmood Ismael Ismael graduated B.Sc. in Medical Technology Faculty form Baghdad University, Iraq in 2004. Graduated B.Sc. Science and Natural Science Department form Life University, Tikrit/2012. His research interests include artificial intelligent, genetic algorithm.

Yasemin GÜLTEPE is currently working as Assistant Professor of Computer Engineering and Head of Computer Engineering Department at Kastamonu University. She has completed her B.Sc. and M.Sc. degree in Computer Engineering Department from Çanakkale Onsekiz Mart University, Çanakkale/Turkey. She has completed her doctor of philosophy in Institute of Sciences at Ege University, Izmir/Turkey. Her research is in the field of semantic web, machine learning, information management.