



Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks

Abubakr H. Ombabi¹ · Wael Ouarda² · Adel M. Alimi²

Received: 18 August 2019 / Revised: 28 March 2020 / Accepted: 19 June 2020 / Published online: 5 July 2020
© Springer-Verlag GmbH Austria, part of Springer Nature 2020

Abstract

Recently, the world has witnessed an exponential growth of social networks which have opened a venue for online users to express and share their opinions in different life aspects. Sentiment analysis has become a hot-trend research topic in the field of natural language processing due to its significant roles in analyzing the public's opinion and deriving effective opinion-based decisions. Arabic is one of the widely used languages across social networks. However, its morphological complexities and varieties of dialects make it a challenging language for sentiment analysis. Therefore, inspired by the success of deep learning algorithms, in this paper, we propose a novel deep learning model for Arabic language sentiment analysis based on one layer CNN architecture for local feature extraction, and two layers LSTM to maintain long-term dependencies. The feature maps learned by CNN and LSTM are passed to SVM classifier to generate the final classification. This model is supported by FastText words embedding model. Extensive experiments carried out on a multi-domain corpus demonstrate the outstanding classification performance of this model with an accuracy of 90.75%. Furthermore, the proposed model is validated using different embedding models and classifiers. The results show that FastText skip-gram model and SVM classifier are more valuable alternatives for the Arabic sentiment analysis. The proposed model outperforms several well-established state-of-the-art approaches on relevant corpora with up to + 20.71% accuracy improvement.

Keywords Arabic sentiment analysis · Opinion mining · Text classification · Deep learning neural networks

1 Introduction

Recently, the world has witnessed the explosion of social networks (SNs) such as Twitter, Facebook, Instagram, etc, which have attracted a wide portion of internet users to interactively collaborate and globally communicate with each other, in SNs people can express and share their opinions and experiences using different types of social data such as textual data (e.g., comments, tweets, reviews, etc.), visual data (e.g., shared and liked images), in addition to multimedia data (e.g., videos and sounds). A huge volume of data is generated out of SNs on a daily basis, this data reflect the sentiment tendencies of the audience towards different life aspects such as political, business, social subjects, etc. For

the researchers, this data contain valuable information that can be used on products/services quality improvement and adaptation, predicting upcoming marketing trends, changing sales strategies, etc (Birjali et al. 2017). Social data are described as informal, unstructured, and rapidly evolving contents, therefore, processing and analyzing this data using conventional analysis methods is a very time-consuming and resource-intensive task (Elouardighi et al. 2017). Natural Language Processing (NLP) is theory-motivated computational technique that enables the computers to smartly understand, analyze, and derive meaning from human's natural languages (Tarwani and Edem 2017) which is complicated in its sequential and hierarchical structure. NLP algorithms enable to perform different natural language-related tasks such as part-of-speech (POS) tagging, parsing, machine translation, and dialogue systems. According to Al-ayyoub and Nuseir (2016), sentiment analysis (SA) is a hot-trend research area of NLP which is concerned with classifying the opinions or emotions towards a product, service, topic, etc, into certain sentiment label. SA-based textual data aim to use text mining, linguistics, and statistical knowledge

✉ Abubakr H. Ombabi
abubakr.h.ombabi@gmail.com

¹ Computer Science, Sudan University of Science and Technology, Khartoum, Sudan

² REGIM-Lab, National School of Engineers, University of Sfax, Sfax, Tunisia

techniques to automatically assign predefined sentiment labels (e.g., negative, positive, or neutral) to the text generated by online users (Alowaidi et al. 2017). However, labels vary according to the context of sentiment analysis. Sentiment analysis provides different subtasks such as polarity classification, subjectivity detection, humor detection, etc, which can be conducted either at sentence-level, document-level or aspect-level (Mostafa 2017; Lu et al. 2018).

For decades, many machine learning algorithms such as SVM and logistic regression have been proposed to address different NLP problems. Recently, neural networks based on dense vector representations have achieved state-of-the-art performances in every NLP-related task (Sze et al. 2017; Haydar et al. 2018) due to their effectiveness and automatic learning capabilities (Ain et al. 2017). Deep learning neural networks have achieved impressive advances in pattern recognition and computer vision. Following this trend, several complex deep learning algorithms have been introduced to perform difficult NLP-related tasks particularly sentiment analysis. Sentiment analysis has gained considerable research attention. Many researches have been conducted on the English since it is the dominant language of sciences, besides other Indo-European languages. Recently, Arabic language has recorded an explosive growth rate in the number of internet users (population) (Boudad et al. 2017; Alsmearat et al. 2015). Figure 1 illustrates the top ten languages based on the percentage of internet users, according to the Internet World Stats ranking, Arabic language ranks fifth among top five internet using languages with more than 168.1 million native speakers (Alowaidi et al. 2017).

However, very few researches have investigated sentiment analysis on Arabic text compared with other languages due to the challenging nature of the Arabic language (Alowaidi et al. 2017; Guellil et al. 2019) such as the dialectal varieties and morphological complexities that require heavy preprocessing and advanced dictionaries (lexicons) more than other languages (Altrabsheh et al. 2017). According to Al-kabi et al. (2014), one Arabic sentence can have several inflectional and derivation forms, for instance, the positions of the

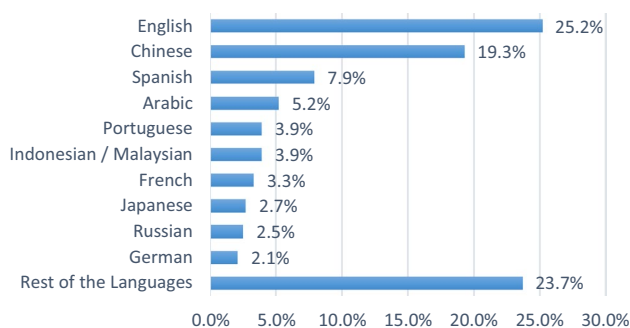


Fig. 1 Top ten internet using languages

words in the sentence and the type of sentence itself whether it is verbal or nominal may change the transitional meanings of the words. Therefore, Arabic text opinion mining is subjective to the context and the domains, also one word can be used to express different polarity classes for different contexts. The Arabic language is diversified in terms of words suffixing, prefixing, and affixing which have a direct impact on words and sentence representation (Boudad et al. 2017). Common spelling mistakes and lack of available corpora are additional challenges in the Arabic language. Therefore, efficient algorithms and tools are required to perform effective and automated features extraction.

The main contribution of this paper is to propose a novel deep learning model based on convolution neural network and long short-term memory for Arabic sentiment analysis based on user's generated textual contents, and also this study aims to demonstrate comparative evaluation using FastText (Skip-gram and CBOW), Word2Vec and AraVec words embedding models on Arabic text classification.

The rest of this paper is organized as follows: Sect. 2 presents the related works done in SA. Section 3 introduces our proposed approach. Section 4 presents the experimental settings. Section 5 presents the experimental results and the evaluation of the proposed model. Section 6 concludes the paper and provides some future works.

2 Related works

Mainly, there are two mainstream solutions for sentiment analysis: supervised (corpus-based) and unsupervised (lexicon-based) approaches (Ravi and Ravi 2015). This section presents different SA approaches in different languages.

2.1 Unsupervised based approach

Clustering-based approach depends on calculating the TF-IDF criterion for features extraction, TF is proportional to the frequency of terms in a document, and IDF is used as a weighting factor. Potential features are the terms with the highest TF-IDF values (Hemmatian and Sohrabi 2017). Claypo and Jaiyen (2014) used *K*-means clustering algorithm and MRF feature selection for SA. MRF was utilized to select only the most relevant features, then *K*-means was used for the final classification. *K*-means achieved the best performance against Hierarchical Clustering and Fuzzy C-Means. Taj et al. (2019) utilized TF-IDF to determine the frequently used terms and their weights, then WordNet was employed to assign sentiment scores to the keywords, and an operator was used to predict the final sentiment label. Huang et al. (2017a) presented a multi-modal which joins sentiment and topic classification tasks based on latent Dirichlet

allocation, and this model was evaluated on a multifarious dataset.

On the other hand, lexicon-based approach (e.g., Keyvanpour et al. 2020; Elhawary and Elfeky 2010) is a popular practical approach to perform sentiment analysis, and this approach utilizes a weighted dictionary to detect the semantic polarity of the words. Lu et al. (2010) evaluated the sentiment polarity strength of the reviews by multiplying the strength of an adjective and adverb words, the strength of an adverb was calculated manually, then the strength of an adjective was determined using progressive relation rules of adjectives and propagation algorithm. Eirinaki et al. (2012) presented High Adjective Count algorithm to identify the nouns and their respective scores which are the number of adjectives associated with that noun, and Max Opinion Score Algorithm to rank the nouns according to their scores, nouns with highest values are selected as potential features. Sasmita et al. (2017) performed an aspect extraction using indicator words constructed using seed words, and the extracted pronoun or noun is compared against the indicator words. An opinion lexicon was used to determine the sentiment orientation of a particular opinion term. Also, Blair et al. (2017) performed SA using lists of positive and negative seed words and the number of topics. Three functions were introduced: objective topics detection, positive and negative sentiment detection, and sentiment classification functions. Pawar and Deshmukh (2015) proposed a hybrid SA approach. N-gram and POS features were extracted using rule-based learning. After calculating the sentiment scores, a threshold was used for the final classification. Also, NB, QDA, and RF ML classifiers were used to classify the tweets into their respective class. ML approach achieved the best results.

2.2 Supervised based approach

According to Kim (2014), Deep Learning has presented remarkable contributions in named-entity recognition (e.g., Chiu and Nichols 2015), computer vision (e.g., Krizhevsky et al. 2012), and speech recognition (e.g., Graves et al. 2013). Unlike conventional machine learning-based NLP models, DL models can perform multi-layers automatic features representation (Young et al. 2018; Chen and Zhang 2018), which makes a simple DL model achieve superior performance over the state of the art in AI tasks (Sohangir et al. 2018). Inspired by the humans' brain, DL model is a complex neural network or machine learning architecture composed of several layers of perceptron (Glorot et al. 2011). DL algorithms are effective in extracting the implicit semantic features which would help in transferring across domains. The application of these algorithms on the SA tasks has reduced the human intervention, computation time, and feature engineering processes (Vateekul and Koomsubha

2016). Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) are the most commonly used DL models for features representation and classification (Hassan and Mahmood 2017).

CNN is a type of feed-forward NNs that requires less training data (Shickel et al. 2018). CNN has presented remarkable performance in different NLP-related tasks due to its ability to capture the syntactic and semantic features for a specific task. CNN applies convolutional operations over the input layer to automatically extract the local features, and the learning capabilities of the CNN is increased due to the weight sharing across all neurons (Ombabi et al. 2017). According to Ravuri and Stoicke (2016) CNN can achieve superior classification performance over strongly competitive neural networks including FNN, RNN, and LSTM. Wint et al. (2018) utilized two parallel CNN layers with BLSTM for sentence-level SA. The feature maps of the two layers are interleaved at a pooling layer, and sigmoid function was used to classify the reviews into (bullied/no bullied) and (positive/negative) labels. This model outperformed different baseline models. Also, Huang et al. (2017b) proposed deep learning model based on CNN and LSTM, and the proposed model was supported by a pre-trained word representation model. Ouyang et al. (2015) incorporated Word2Vec over 7-layers CNN which contains 3 pairs of convolutional and pooling layers, with adopted PReLU, normalization and dropout functions; this model achieved the best classification accuracy over RNN and MV-RNN models.

Recurrent Neural Network is mainly used in text data classification due to its ability to capture long-term dependencies and to maintain the sequence of variable length data. RNN maintains a connection between the current hidden states and the output of the previous hidden layer, i.e., RNN takes the output of the previous hidden layer as inputs to the current hidden layer. However, RNN suffers from vanishing gradient problem (Al-Smadi et al. 2018). Preethi and Krishna (2017) explored the application of RNN in sentiment analysis. They aimed to provide optimized place recommendations services based on SA, and experiments on Amazon dataset showed the improved classification performance of this approach.

Li et al. (2017) proposed a hybrid neural network architecture based on BTM, RSM, and Latent semantic machines. A regularized transfer learning model was used to incorporate the semantic domain knowledge into the NN and to boost the classification performance. Wang and Cao (2017) proposed SA approach based on LSTM with L2 and Nadam optimizer for Chinese text SA, and this model was evaluated on online-shopping reviews. Results indicated that the adopted loss and optimization functions can improve the classification accuracy. Ghosh et al. (2016) proposed a deep learning model for SA based on Probabilistic Neural Network with two-layered Restricted Boltzmann (RBM).

TF-IDF was used for the data representation, and PNN was employed for the predict the final sentiment class. Lalji and Deshmukh (2016) proposed a hybrid model based on lexicon and machine learning approaches for sentiment analysis. Tree Tagger and POS Tagging techniques were used for features extraction. NB, RF, SVM, and LDA ML classifiers were used to classify the tweets into a certain label.

There are very few researches have investigated the application of deep learning techniques in the Arabic NLP, particularly SA, Dahou et al. (2016) proposed CNN and neural words embedding architectures for Arabic sentiment analysis. The proposed architecture outperformed different existing approaches. Al-Smadi et al. (2018) introduced an aspect-based SA approach that contains two implementations: aspect opinion target expression extraction (OTEs) using character-level BLSTM with (CRFs) classifier, and aspect sentiment polarity classification using aspect-based LSTM. The proposed approaches achieved significant performance improvement over the baseline approaches. Alayba et al. (2018) integrated CNN with SemEval-2016 Arabic Twitter, and the Arabic Health Twitter Lexicons to perform Arabic SA. Word2Vec (CBOW) was utilized as the embedding model, and this approach obtained promising results. Hassan and Mahmood (2018) described a joint CNN and RNN framework stacked over unsupervised words embedding model, and in this framework, the former information was combined with the feature sets extracted using convolutional layer. This approach outperformed several existing approaches in terms of accuracy. Table 1 shows a summary of supervised and unsupervised Learning related works presented in this study.

Unsupervised approaches are commonly used in SA. However, keywords vagueness and ambiguity can decrease the accuracy of predictions. These approaches cannot consider the semantic relationships between words in the sentences. For Arabic sentiment analysis, unsupervised approaches cannot be effective due to the numerous words from several dialects to be included in the lexicons. Also, it is observed that using only CNN or using only LSTM is inadequate to achieve the desired results on Arabic sentiment analysis (Huang et al. 2017b), this is because CNN fails to maintain long-term dependencies, and LSTM is weak to capture local features. Unlike other deep learning approaches, in this work, we propose a new architecture based on deep learning of features representation and features classification. The proposed architecture uses the recent FastText model which can generate the corresponding vectors for the Out-Of-Vocabulary words (OOV) and rare words. Convolutional neural network architecture is used for n-gram local-region features and information extraction. The performance of the CNN is improved using two stacked LSTM layers to address the difficulties of training CNN to capture long-term dependencies. Finally, the feature maps

learned by CNN and LSTM are passed to the SVM classifier to generate the final sentiment labels.

3 Proposed approach

In this study, we propose a novel deep learning model for Arabic SA namely (Deep CNN–LSTM Arabic-SA) which joins FastText words representation model over one layer CNN architecture which inspired by Kim's work (Kim 2014). Due to the locality of the convolutional and pooling layers, CNN cannot capture long-distance dependencies the input sentences, however one single recurrent layer can effectively overcome this limitation (Hassan and Mahmood 2017), therefore, we propose to utilize two LSTM layers to minimize the local information loss. Finally, SVM classifier is used to classify the sentences into a certain sentiment label (positive or negative). Figure 2 illustrates the overall processes of Deep CNN–LSTM Arabic-SA, Fig. 3 visualizes the fundamental architecture and the information flow, and Fig. 4 presents the architecture of the CNN and LSTM used in Deep CNN–LSTM Arabic-SA.

3.1 Word embedding

This architecture takes advantage of using the recent FastText model proposed by Mikolov et al. (2017) for word embedding, FastText is trained on a wide range of languages including English and Arabic. As in word2vec, FastText provides two models: Skip-gram model which is used to predict a target word using the closed neighboring words, while CBOW uses the surrounding words in the context to predict the target word, both methods generate a text file which contains numerical representation (vectors) of the learned words. In this study, the FastText skip-gram model is used in which each word is represented as a bag of character n-grams; sentences are then represented as a summation of their words vectors. FastText is run in its default configurations: 100-dimension vector space, sub-word size is 3-6 characters which is appropriate for Arabic text because any Arabic word has three letters root Altowayan (2017).

3.2 Convolutional neural network

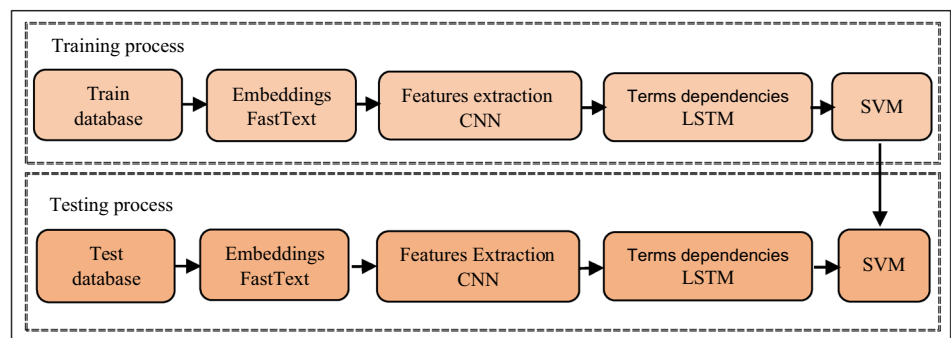
let $X_i \in R^k$ denotes the k-dimension word vector equivalent to the i th word in a sentence with length (n) which is represented as a concatenation of its words vectors see Eq.(1), zero padding is applied to the sentences with length less than (n).

$$X_{1:n} = X_1 \oplus X_2 \cdots \oplus X_n \quad (1)$$

Table 1 Summary of supervised and unsupervised related works

Learning method	Study	Embedding & Features representation	Features classification	Dataset	Accuracy (%)
Supervised	Wint et al. (2018)	BiLSTM-CNN	Sigmoid function	Amazon, movie reviews and Yelp	94.38
	Ouyang et al. (2015)	Word2Vec-CNN	Softmax function	10.662 movie reviews	45.41
	Preethi and Krishna (2017)	RNN	Naive Bayes	Restaurant and movie reviews	90.47
	Li et al. (2017)	Word2vec, Semantic LSMs, RSM and BTM	RSM	SemEval , ISEAR and SinaNews	90.47
	Wang and Cao (2017)	BiLSTM with L2 and Nadam optimizer function	BiLSTM	13.000 products reviews	96.01
	Ghosh et al. (2016)	Two layered restricted Boltzmann machine	A probabilistic neural network	Multi-domain and movie reviews	80.81
	Dahou et al. (2016)	Word2vec, CNN	Sigmoid function	LABR, ASTD, Arabic Gold-Standard Twitter	89.61
	Al-Smadi et al. (2018)	FastText, BLSTM	CRFs	SemEval-ABSA 2016 Task-5	82.61
Unsupervised	Alayba et al. (2018)	Word2Vec, TF, TFIDF, POS, two lexicons	NB, Bernoulli NB, SVM, LR, SGD, RDG	2.026 Arabic health sentiment dataset	95.01
	Pawar and Deshmukh (2015)	Opinion Lexicons and rule based learning	NB, QDA, and RF classifiers	Twitter Sentiment version 0.2	88.65
	Blair et al. (2017)	Lists of positive and negative seed words, bag-of-words	LDA	Stanford movie review, Sentiment1401	67.5
	Sasmita et al. (2017)	Opinion lexicon, Top-N similar words function	Semantic orientation function	17.273 online reviews	72.9
	Eirinaki et al. (2012)	High adjective count algorithm, POS tagging	Max opinion score algorithm	Products reviews	87.01
	Lu et al. (2010)	POS tagging, lexicon	Review strength function	4.000 hotel review	78.6
	Huang et al. (2017a)	A multimodal joint sentiment topic model	LDA	18.000 microblogging messages	70.75
	Taj et al. (2019)	TF-IDF, WordNet, Senti-WordNet	Extract sentiment operator	2.225 BBC news articles	72.2
	Claypo and Jaiyen (2014)	MRF feature selection and fuzzy C-means	K-means clustering	1.060 restaurant reviews	75.5
	Lalji and Deshmukh (2016)	MPQA Lexicon, Tree Tagger POS	NB, RF, SVM, and LDA	28.000 tweets	63.23

Fig. 2 Overview of the proposed Deep CNN–LSTM Arabic-SA model



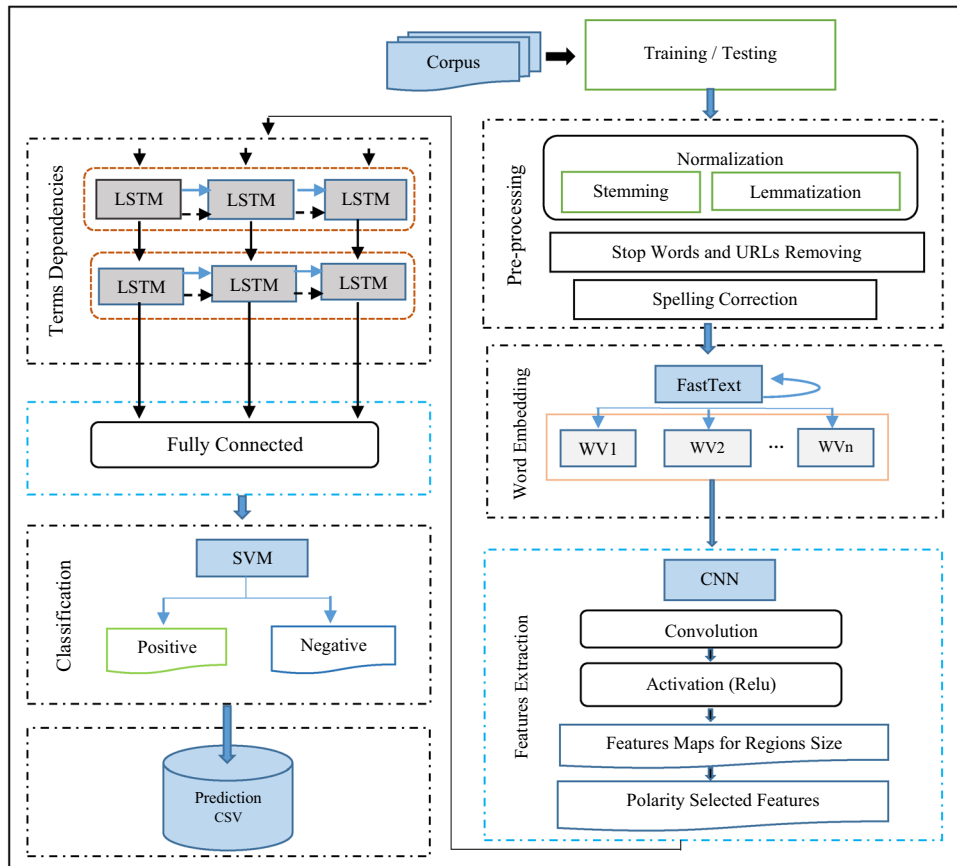


Fig. 3 Flow diagram of the proposed Deep CNN-LSTM Arabic-SA

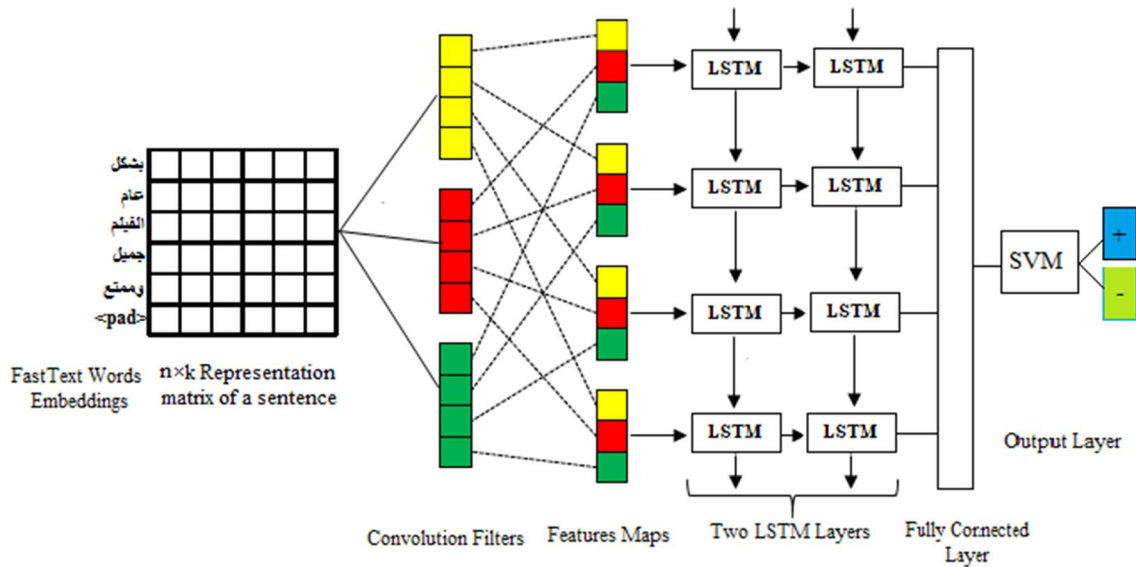


Fig. 4 Proposed Deep CNN-LSTM Arabic-SA architecture

\oplus is a concatenation operator. Let $X_{i:i+j}$ denotes the concatenation of the words $X_i, X_{i+1}, \dots, X_{i+j}$, the convolution filter $W \in R^{hk}$ is applied to a window of h words in a sentence representation matrix of a shape $n \times k$ to generate a new features matrix, $X_{i:i+j}$ is the basic element from the i th to the $(i + j)$ th which represent the local feature matrix from the i th line to the $(i + j)$ th line of the current sentence vector. A feature C_i (i -th feature value) can be generated from a window of words $X_{i:i+h-1}$ using Eq. (2).

$$C_i = f(W.X_{i:i+h-1} + b). \tag{2}$$

b refers to bias term, where $b \in R$, f is a nonlinear activation function such as sigmoid and hyperbolic tangent. b and W are learned during the training. The filter is convoluted on every window of words in the input sentence $X_{1:h}, X_{2:h+1}, X_{n-h+1:n}$ to produce a features map using Eq.(3).

$$C = [C_1, C_2, C_{n-h+1}] \tag{3}$$

with $C \in R^{n-h+1}$.

We just explained the process of producing one feature map that is captured from one filter. Note that, convolution layer with m filters will produce $m(n - h + 1)$ features. Max-overtime pooling is not applied over the feature maps, because features sampling can effect the sequence organization before the LSTM layers. The feature maps are directly fed into the LSTM layers to encode the temporal patterns.

3.3 Capturing long-term dependencies

LSTM can efficiently control the information by preventing vanishing gradient and capturing long-term correlations in sequences with arbitrary length (Yuan et al. 2018). As shown in Fig. 5, LSTM architecture contains a newly added memory cell to selectively maintain the information for a

longer time without degeneration. In addition to input, output, forget gates.

To process the input vectors, LSTM applies recursive execution of the current cell block using the old hidden state (h_{t-1}) and the current input x_t , where (t) and ($t - 1$) refer to the current time and the former time, respectively. Now i_t, f_t , and (o_t) are the input gate, forget gate, and output gate, respectively, and \tilde{C}_t refers to the current memory cell state at a time (t) in cell block, the operational principle of the LSTM can be described as follows: Using Eqs. (4) and (5) the values of i_t , and \tilde{C}_t are computed for the memory cells states at a time (t).

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{4}$$

$$\tilde{C} = \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{5}$$

Equation (6) calculates the activation value f_t of the forget gate at time (t):

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{6}$$

Equation (7) calculates the new state C_t of the memory cell at a time (t):

$$C_t = i_t * \tilde{C} + f_t * C_{t-1} \tag{7}$$

Memory cells output gates values are computed for the new state using C_t as in Eqs. (8) and (9).

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o c_t + b_o) \tag{8}$$

$$h_t = o_t * \tanh(C_t) \tag{9}$$

where x_t refers to the input of the memory cell at t . $W_i, W_c, W_f, U_i, W_o, U_c, U_f, U_o$, and V_o are the weights matrices. b_i, b_f, b_c, b_o are the bias vectors. σ is a logistic sigmoid function, o is an element-wise multiplication. During the training the model learns the values of W_i and U_i . The values of f_t, i_t and o_t are in $[0, 1]$. In this architecture, the output of the first LSTM layer is passed to the second LSTM layer which produces deep representation of the original sentence. The final outputs of the LSTM layers are merged into one matrix; this matrix is passed to a fully connected layer.

4 Experiments

This section presents the experimental settings and configurations of Deep CNN–LSTM Arabic-SA on different datasets. The experiments are conducted on the TensorFlow framework running on Python.

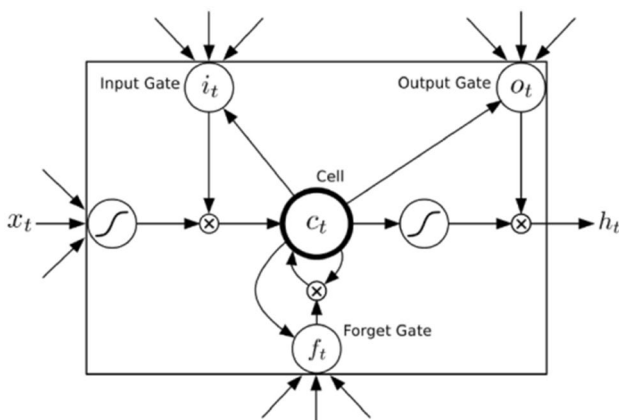


Fig. 5 Long-short term memory architecture with memory cell

Table 2 Constructed training set statistics

Dataset	Labels distribution		
	Positive	Negative	Total
Hotel reviews	1800	1800	3600
Restaurant reviews	2000	2000	4000
Movie reviews	1900	1900	3800
Product reviews	1600	1600	3200
LABR	1500	1500	3000
Total	7550	7550	15.100

4.1 Datasets

Data acquisition and annotation are the most difficult tasks in the Arabic sentiment analysis as presented in Sect. 1, therefore, we relied on the previously published works to construct a multi-domain sentiment corpus which contains positive and negative reviews on five topics. We have sampled subsets from: the corpus collected by ElSahar and El-Beltagy (2011) which scrapped from different websites, also the corpus collected by Aly and Atiya (2013) which is the largest sentiment corpus for Arabic text; it contains 63,000 books reviews. As presented in Table 2, the total size of the constructed training set is 15.100 of equally size positive and negative reviews. An Arabic NLTK was used to automatically correct the misspelled words, remove the stop-words and the duplicated letters (e.g., Beauuutiffllll=جميبييللل). Then letters such as: alif (أ) is normalized to | and ta'a (آ) is normalized to (ا), also non-Arabic contents are filtered out. For testing and validation, we used a dataset of 4.000 reviews distributed as 2.000 positive and 2.000 negative.

Table 3 shows sample of positive sentiment reviews from hotels domain with its English translation, and Table 4 shows sample of negative sentiment reviews.

Table 3 Sample of positive reviews

Arabic Comment	فندق رائع يتميز بالهدوء ومستوى النظافة ممتاز، يتعامل لطيف من العاملين بالغرف رائعة بكل ما تبحث عنه تجده في هذا الفندق..
English Translation	wonderful hotel, it is quiet and clean, the staff is friendly, rooms are wonderful, you can find anything you looking for in this hotel..

Table 4 Sample of negative reviews

Arabic Comment	اسوء مكان ممكن تزوره اوتقيم فيه غرف سيئة وخدمة متدنية - بصورة عامة مخيب للأمال
English Translation	It is the worst place to stay or visit, bad rooms and services. Generally, it is disappointing.

Predicted/Actuals	Positive (2000)	Negative (2000)
Positive	89.10%	10.90%
Negative	7.60%	92.40%

Fig. 6 Confusion matrix

4.2 Model hyper-parameters

Empirically, different hyper-parameters and settings have been tested. For the CNN configurations, the adopted convolutional layer used multiple filters of width (3, 4, 5), 256 feature maps, (ReLus) as activation function, dropout was set to 0.5 before the recurrent layer to minimize the overfitting. Padding was set to zero when needed. For the LSTM configurations, the hidden state dimensionality was configured to 128, and sigmoid function was used as an activation function. The number of epochs was set to (5–10) in the entire architecture.

5 Results and discussion

Deep CNN–LSTM Arabic-SA was trained on a multi-domain sentiment corpus in Table 2, then the classification performance is evaluated using the testing set. The confusion matrix is a measure that is used to assess the correctness of classification. The obtained confusion matrix of this experiment is presented in Fig. 6 where 89.10% of the positive reviews are correctly classified as positive, with only 10.90% which misclassified as negative. 92.40% of the negative reviews are correctly classified as negative, with only 7.60% which misclassified as positive by Deep CNN–LSTM Arabic-SA.

We followed the conventions to report the classification performance of Deep CNN–LSTM Arabic-SA using precision, recall, *F1*-score, and accuracy measures. As presented

Table 5 Classification performances with different classifiers

Classifier	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
SVM	89.10	92.14	92.44	90.75
NB	88.02	87.60	87.56	88.75
Softmax	87.81	86.63	87.39	87.43
KNN	86.72	86.94	86.73	86.83

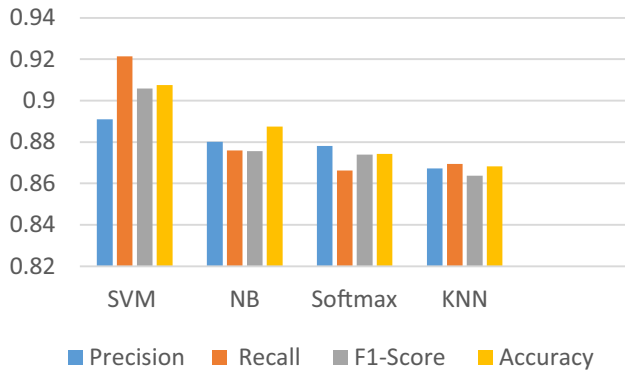


Fig. 7 Classification performances using different classifiers

in Table 5, Deep CNN–LSTM Arabic-SA achieved competitive classification performance with 89.10%, 92.14%, and 90.44% of precision, recall, and *F1*-score respectively. Deep CNN–LSTM Arabic-SA achieved 90.75% which is significant accuracy improvement over only CNN models in the Arabic sentiment classification.

5.1 Best performing classifier

The classifier performance eventually determines the quality of the word embedding and features extraction approaches, therefore, we intensively evaluated the performance of Deep CNN–LSTM Arabic-SA using Naive Bayes (NB), *K*-Nearest Neighbor (*KNN* (*K* = 10)) classifiers, in addition to Softmax as classification function after the fully connected layer against SVM using the same training parameters and dataset splits, as presented in Table 5 and Fig. 7, SVM achieved superior performance over NB, Softmax, and KNN. Based on these results SVM classifier is more reliable for Arabic text classification which is consistent with (Nabil et al. 2015; Aly and Atiya 2013).

5.2 Optimal number of LSTM layers

However, LSTMs are considered as a deep feed-forward neural network architecture, we have validated the effect of the number of LSTM layers on the classification performance. We have experimented Deep CNN–LSTM Arabic-SA using one LSTM layer compared with two LSTM layers, where

Predicted/Actuals	Positive (2000)	Negative (2000)
Positive	86.1 %	13.9 %
Negative	10.15 %	89.85 %

Fig. 8 Confusion matrix of Deep CNN–LSTM Arabic-SA using single LSTM layer

Table 6 Effects of the number of LSTM layers

N. of LSTM layers	Precision (%)	Recall (%)	Accuracy (%)
1	86.10	89.45	87.98
2	89.10	92.14	90.75

Table 7 Classification accuracy using different embedding models

Embedding model	Classification accuracy (%)
FastText(CBOW)-CNN–LSTM	88.90
FastText(Skip-gram)-CNN–LSTM	90.75
Word2Vec-CNN–LSTM	86.84
AraVec-CNN–LSTM	81.35

each LSTM layer has 128 units. The confusion matrix of this experiment is presented in Fig. 8. As shown in Table. 6, two stacked LSTM layers can help to improve the classification performance with + 2.77% in terms of accuracy, in addition to + 3% and + 2.69% in terms of precision and recall, respectively over one layer LSTM. Therefore, two LSTM layers are appropriate for producing more higher-order feature representations of Arabic sentences to be more easily separable into different classes. This result is consistent with Pal et al. (2018) which deduced that stacking LSTM layers one upon another can increase the classification accuracy.

5.3 Best performing embedding model

The classification performance of Deep CNN–LSTM Arabic-SA is examined using two other pre-trained word representations models: word2Vec which is introduced by Mikolov et al. (2013), it uses two-layers Neural Network to construct distributed representation of words. It contains 3 million words represented in a 300-dimensional vector space. AraVec which is introduced by Soliman et al. (2017), it is a pre-trained distributed word representation model for Arabic language, it provides two architectures: CBOW and Skip-gram with 300 dimension vector space. To gain more insight into the performance of Deep CNN–LSTM

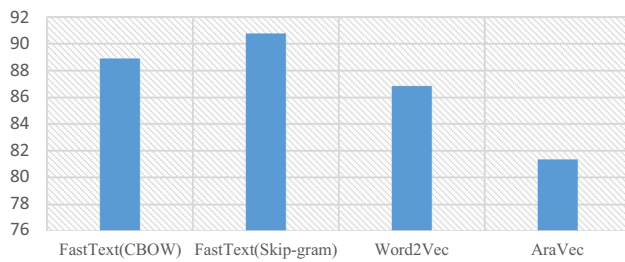


Fig. 9 Classification performances using different embedding models

Arabic-SA, Table 7 and Fig. 9 show the test accuracy comparisons of Word2Vec-CNN-LSTM and AraVec-CNN-LSTM against FastText-CNN-LSTM models, according to the results, FastText (Skip-gram and CBOW) based methods achieved superior performance with an accuracy of 90.75% and 88.90% respectively, which outperformed Word2Vec and AraVec with up to + 3.3% and + 8.8% accuracy improvements, respectively. On the other hand, the FastText Skip-gram model achieved the best classification accuracy, which is consistent with Bojanowski et al. (2017) that FastText skip-gram can produce high quality vectors representations using the semantic and syntactical information from the texts, also it can cover the out-of-vocabulary words.

5.4 Comparison with the state-of-the-art

To validate the performance of Deep CNN-LSTM Arabic-SA against the state of the art, we performed different experiments on different datasets: Large Scale Arabic Book Reviews (LABR) dataset constructed by Aly and Atiya (2013), it contains 63,000 book reviews that have been collected from Goodreads. Arabic Sentiment Tweets Dataset (ASTD) collected by Nabil et al. (2015), it contains 10,000 Arabic tweets. Arabic sentiment analysis Twitter dataset collected by Abdulla et al. (2013), it contains 2,000 positive and negative tweets. The performance of Deep CNN-LSTM Arabic-SA is compared with: Dahou et al. (2016) used one layer CNN Architecture over Word2Vec model. Altowayan (2017) experimented FastText with SVC and Logistic Regression classifiers on LABR and ASTD datasets. Altowayan and Tao (2016) incorporated POS tags and word stemming features with Logistic Regression on both LABR and ASTD datasets. ElSahar and El-Beltagy (2011) utilized three feature representation techniques: Delta-TF-IDF, TF-IDF and Count, with Linear SVM for features selection and classification. Abdulla et al. (2013) proposed a lexicon-based, and supervised-based approach (SVM classifier). And Nabil et al. (2015) which used token counts and the TF-IDF with SVM classifier. And Nabil et al. (2015) which used token counts and the TF-IDF with SVM classifier. Table 8 and Fig. 10 show the classification accuracy of

Table 8 Accuracy comparison with the existing methods

Studies	LABR	ASTD	Ar-Twitter
Dahou et al. (2016)	89.60	79.07	85.01
Altowayan (2017)	84.97	87.01	–
Altowayan and Tao (2016)	78.60	80.21	–
Nabil et al. (2015)	–	69.01	–
ElSahar and El-Beltagy (2011)	80.20	–	–
Abdulla et al. (2013)	–	–	87.20
Deep CNN-LSTM Arabic-SA	90.20	89.72	88.52

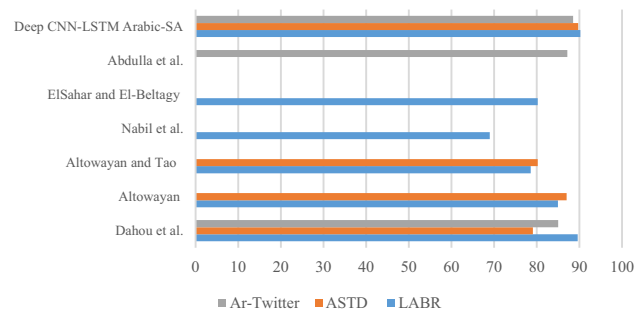


Fig. 10 Comparison of accuracy with other methods

Table 9 Performance comparison with the existing methods

Dataset	Study	Precision (%)	Recall (%)
LABR	Altowayan (2017)	81.27	79.63
	Altowayan and Tao (2016)	73.17	88.74
	Dahou et al. (2016)	87.31	82.53
ASTD	Deep CNN-LSTM Arabic-SA	88.23	83.50
	Altowayan (2017)	86.00	83.25
Ar-Twitter	Deep CNN-LSTM Arabic-SA	89.36	86.65
	Abdulla et al. (2013)	85.10	74.91
	Deep CNN-LSTM Arabic-SA	89.97	85.92

Deep CNN-LSTM Arabic-SA in each dataset against other approaches listing their best classification accuracy.

On LABR dataset, Deep CNN-LSTM Arabic-SA achieved 90.20% classification accuracy which outperformed the baseline results by up to + 11.6% in terms of accuracy. Deep CNN-LSTM Arabic-SA reached its highest accuracy on this dataset due to the sufficient dataset size and the balanced distribution of labels. On ASTD dataset, Deep CNN-LSTM Arabic-SA achieved significant accuracy increase of + 10.65% over only CNN model proposed by Dahou et al. (2016), and up to + 20.71% accuracy improvement over the three other approaches. For tweeter dataset (Ar-Twitter) Deep CNN-LSTM Arabic-SA as the best

performing model achieved an accuracy of 88.52% with + 3.51% and + 1.32 % of accuracy improvement over Dahou et al. (2016) and Abdulla et al. (2013) respectively.

Table 9 presents deeper details about the performance of Deep CNN–LSTM Arabic-SA in each dataset with the performances of the other approaches in terms of precision and recall. Deep CNN–LSTM Arabic-SA achieved the best performance in all datasets with 89.79% and 85.92% of precision and recall respectively, this evaluation proved the reliability of the proposed deep learning model for Arabic text sentiment analysis. According to the obtained results, one convolutional layer CNN architecture supported by two layers LSTMs can improve the process of Arabic features representation and classification as confirmed by Hassan and Mahmood (2017). Moreover, these results confirmed that generating word vectors using FastText acts better than using word2vec and AraVec models on the word-level as it helps to better learn the hidden features about the language, and the out-of-vocabulary words.

Due to the large dataset size and balanced destitution of data, Deep CNN–LSTM Arabic-SA reached its highest performance on LABR dataset with up to + 15.60% and + 3.87% of improvement over the baseline performance. On ASTD dataset, Deep CNN–LSTM Arabic-SA ranks top the list of recent works and achieved + 3.36% and + 3.40% of performance enhancement. On Ar-Twitter dataset, Deep CNN–LSTM Arabic-SA increase the performance baseline with + 4.87% and + 11.01%.

6 Conclusion

Recently, social media have witnessed exponential growth in user-generated content which contains enormously valuable information for different applications. Sentiment analysis is concerned with analyzing social data to identify the inclinations of the public audience. For Arabic, it is challenging to perform sentiment analysis regardless of deep considerations of semantic and syntactic rules, in addition to terms dependencies of the input sentence. Thus, this paper proposed a deep learning model for Arabic sentiment analysis, and this model skillfully joint one-layer CNN architecture with two LSTM layers. This architecture is supported by FastText word embedding model as the input layer. The experiments on a multi-domain corpus showed the remarkable performance of this model with 89.10%, 92.14%, 92.44%, and 90.75% in terms of precision, recall, *F1*-Score, and accuracy, respectively. This study extensively validated the effect of the words embedding techniques on the Arabic sentiment classification and deduced that the FastText model is a more relevant alternative to learn semantic and syntactic information. Furthermore, the performance of the proposed model

is evaluated using NB and KNN classifiers. The results showed that SVM is the best performing classifier with up to + 3.92% accuracy improvement. Due to the efficiency of the CNN in features extraction and the recurrent nature of LSTM, the proposed model achieved encouraging results and outperformed state-of-the-art methods on several benchmarks with up to + 11.6% of accuracy improvement.

For future research, it is worth investigating the application of deep learning architectures in the user's interests discovery and recommendation, and to improve the quality of word embedding by integrating WordNet lexical database with the input layer.

References

- Ain QT, Ali M, Riaz A, Noureen A, Kamran M, Hayat B, Rehman A (2017) Sentiment analysis using deep learning techniques: a review. *Int J Adv Comput Sci* 8(6):424
- Al-ayyoub M, Nuseir A (2016) Hierarchical classifiers for multi-way sentiment analysis of Arabic reviews. *Int J Adv Comput Sci Appl* 7(2):531–539
- Al-kabi MN, Gigieh AH, Alsmadi IM, Wahsheh HA (2014) Opinion mining and analysis for Arabic language. *Int J Adv Comput Sci Appl* 5(5):181–195
- Al-Smadi M, Talafha B, Al-Ayyoub M, Jararweh Y (2018) Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *Int J Mach Learn Cybern*. <https://doi.org/10.1007/s13042-018-0799-4>
- Alayba AM, Palade V, England M, Iqbal R (2018) Improving sentiment analysis in Arabic using word representation. In: 2nd international workshop on Arabic and derived script analysis and recognition (ASAR), pp 13–18
- Alowaidi S, Saleh M, Abulnaja O (2017) Semantic sentiment analysis of Arabic texts. *Int J Adv Comput Sci Appl* 8(2):256–262
- Alsmearat K, Shehab M, Al-Ayyoub M, Al-Shalabi R, Kanaan G (2015) Emotion analysis of Arabic articles and its impact on identifying the author's gender
- Altowayan AA (2017) Improving Arabic sentiment analysis with sentiment-specific embeddings. In: IEEE international conference on big data (BIGDATA) improving, pp 4314–4320
- Altowayan AA, Tao L (2016) Word embeddings for Arabic sentiment analysis. In: IEEE international conference on big data (big data) word, pp 3820–3825. <http://tanzil.net>
- Altrabsheh N, El-Masri M, Mansour H (2017) Successes and challenges of Arabic sentiment analysis research: a literature review. *Soc Netw Anal Min*. <https://doi.org/10.1007/s13278-017-0474-x>
- Aly M, Atiya A (2013) LABR: a large scale Arabic book reviews dataset. In: Proceedings of the 51st annual meeting of the association for computational linguistics, pp 494–498. <https://doi.org/10.13140/2.1.3960.5761>. <https://www.aclweb.org/anthology-new/P/P13/P13-2088.pdf>
- Birjali M, Beni-hssane A, Erritali M (2017) Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. *Procedia Comput Sci* 113:65–72. <https://doi.org/10.1016/j.procs.2017.08.290>
- Blair SJ, Bi Y, Mulvanna MD (2017) Unsupervised sentiment classification: a hybrid sentiment-topic model approach. <https://doi.org/10.1109/ICTAI.2017.00076>

- Bojanowski P, Grave E, Joulin A, Mikolov T (2017) Enriching word vectors with subword information, vol 5, no 3, pp 729–734. [arXiv:160704606](https://arxiv.org/abs/160704606). <https://doi.org/10.1039/c2ay25919b>
- Chen Y, Zhang Z (2018) Research on text sentiment analysis based on CNNs and SVM. In: Proceedings of the 13th IEEE conference on industrial electronics and applications, ICIEA 2018, pp 2731–2734. <https://doi.org/10.1109/ICIEA.2018.8398173>
- Chiu JPC, Nichols E (2015) Named entity recognition with bidirectional LSTM-CNNs (2003). <https://doi.org/10.3115/1119176.1119204>. [arXiv:1511.08308](https://arxiv.org/abs/1511.08308)
- Claypo N, Jaiyen S (2014) Opinion mining for Thai restaurant reviews using neural networks and mRMR feature selection. In: 2014 international computer science and engineering conference. ICSEC 2014, pp 394–397. <https://doi.org/10.1109/ICSEC.2014.6978229>
- Dahou A, Xiong S, Zhou J, Haddoud MH, Duan P (2016) Word embeddings and convolutional neural network for Arabic sentiment classification. In: Proceedings of the COLING 2016, 26th international conference on computational linguistics :technical papers, pp 2418–2427. <https://www.aclweb.org/anthology/C/C16/C16-1228.pdf>
- Dellavale D, Urdapilleta E, Cámpora N, Velarde OM, Kochen S, Mato G (2020) Prediction of epileptic seizures based on mean phase coherence. [BioArXiv](https://arxiv.org/abs/2001.01111), pp 1–60
- Eirinaki M, Pisal S, Singh J (2012) Feature-based opinion mining and ranking. *J Comput Syst Sci* 78(4):1175–1184. <https://doi.org/10.1016/j.jcss.2011.10.007>
- Elhawary M, Elfeky M (2010) Mining Arabic business reviews. In: Proceedings—IEEE international conference on data mining, ICDM, pp 1108–1113. <https://doi.org/10.1109/ICDMW.2010.24>
- Elouardighi A, Maghfour M, Hammia H, Aazi Fz (2017) Analysis in the standard or dialectal Arabic. In: 2017 3rd international conference of cloud computing technologies and applications (CloudTech)
- Elsahar H, El-Beltagy SR (2011) Building large Arabic multi-domain resources for sentiment analysis, vol 9042. Springer, Berlin, pp 23–34. https://doi.org/10.1007/978-3-319-18117-2_2
- Ghosh R, Ravi K, Ravi V (2016) A novel deep learning architecture for sentiment classification. In: 3rd international conference on recent advances in information technology—RAIT-2016—A, vol 27, pp 1102–1111. https://doi.org/10.1007/978-3-319-68195-5_122. [arXiv:1707.05809](https://arxiv.org/abs/1707.05809)
- Glorot X, Bordes A, Bengio Y (2011) Domain adaptation for large-scale sentiment classification: a deep learning approach. In: Proceedings of the 28th international conference on machine learning, no 1, pp 513–520. http://www.icml-2011.org/paper/s/342_icmlpaper.pdf
- Graves A, Rahman Mohamed A, Hinton GE (2013) Speech recognition with deep recurrent neural networks. In: 2013 IEEE international conference on acoustics, speech and signal processing, no 6, pp 6645–6649. <https://doi.org/10.1109/ICASSP.2013.6638947>
- Guellil I, Azouaou F, Mendoza M (2019) Arabic sentiment analysis: studies, resources, and tools. *Soc Netw Anal Min* 9(1):1–17. <https://doi.org/10.1007/s13278-019-0602-x>
- Hassan A, Mahmood A (2017) Deep learning approach for sentiment analysis of short texts. In: 2017 3rd international conference on control, automation and robotics (ICCAR), pp 705–710. <https://doi.org/10.1109/ICCAR.2017.7942788>. <http://ieeexplore.ieee.org/document/7942788/>
- Hassan A, Mahmood A (2018) Convolutional recurrent deep learning model for sentence classification. *IEEE Access* 6:13949–13957. <https://doi.org/10.1109/ACCESS.2018.2814818>
- Haydar MS, Helal MA, Hossain SA (2018) Sentiment extraction from bangla text: a character level supervised recurrent neural network approach. In: 2018 international conference on computer, communication, chemical, material and electronic engineering (IC4ME2), pp 1–4. <https://doi.org/10.1109/IC4ME2.2018.8465606>
- Hemmatian F, Sohrabi MK (2017) A survey on classification techniques for opinion mining and sentiment analysis. *Artif Intell Rev*. <https://doi.org/10.1007/s10462-017-9599-6>
- Huang F, Zhang S, Zhang J, Yu G (2017a) Multimodal learning for topic sentiment analysis in microblogging. *Neurocomputing* 253:144–153. <https://doi.org/10.1016/j.neucom.2016.10.086>
- Huang Q, Chen R, Zheng X, Dong Z (2017b) Deep sentiment representation based on CNN and LSTM. In: Proceedings—2017 international conference on green informatics, ICGI 2017, pp 30–33. <https://doi.org/10.1109/ICGI.2017.45>
- Keyvanpour M, Karimi Zandian Z, Heidarypanah M (2020) OMLML: a helpful opinion mining method based on lexicon and machine learning in social networks. *Soc Netw Anal Min*. <https://doi.org/10.1007/s13278-019-0622-6>
- Kim Y (2014) Convolutional neural networks for sentence classification, pp 1746–1751. <https://doi.org/10.3115/v1/D14-1181>. [arXiv:1408.5882](https://arxiv.org/abs/1408.5882)
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1–9. <https://doi.org/10.1016/j.protcy.2014.09.007>
- Lalji TK, Deshmukh SN (2016) Twitter sentiment analysis using hybrid approach. *Int Res J Eng Technol (IRJET)* 3:2887–2890
- Li X, Rao Y, Xie H, Lau RYK, Yin J, Wang FL (2017) Bootstrapping social emotion classification with semantically rich hybrid neural networks. *IEEE Trans Affect Comput* 3045(c):1–16. <https://doi.org/10.1109/TAFFC.2017.2716930>
- Lu Y, Kong X, Quan X, Liu W, Xu Y (2010) Exploring the sentiment strength of user reviews. *LNCS*, vol 6184. Springer, Berlin, pp 471–482. https://doi.org/10.1007/978-3-642-14246-8_46
- Lu Y, Rao Y, JYang, JYin (2018) Incorporating lexicons into LSTM for sentiment classification. In: 2018 international joint conference on neural networks (IJCNN), p 1. <http://mendeley.csuc.cat/fitxers/0f093fd3fff6230dab142add74997c48>
- Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality, pp 1–9. [arXiv:1310.4546](https://arxiv.org/abs/1310.4546). <http://arxiv.org/abs/1310.4546>
- Mikolov T, Grave E, Bojanowski P, Puhresch C, Joulin A (2017) Advances in pre-training distributed word representations *Tomas*, vol 28, no 7, pp 2114–2118. [arXiv:171209405](https://arxiv.org/abs/171209405)[csCL]. <https://doi.org/10.1589/jpts.28.2114>. <http://arxiv.org/abs/1712.09405>
- Mostafa AM (2017) An evaluation of sentiment analysis and classification algorithms for Arabic textual data. *Int J Comput Appl* 158(3):975–8887
- Nabil M, Aly M, Atiya A (2015) ASTD: Arabic sentiment tweets dataset. In: Proceedings of the 2015 conference on empirical methods in natural language processing (September), pp 2515–2519. <https://doi.org/10.18653/v1/D15-1299>. <http://aclweb.org/anthology/D15-1299>
- Na A, Na A, Ma S, Al-ayyoub M (2013) Arabic sentiment analysis. In: Jordan conference on applied electrical engineering and computing technologies (AEECT13), vol 6, no 12, pp 1–6. <https://doi.org/10.1109/AEECT.2013.6716448>
- Ombabi AH, Lazzez O, Ouarda W, Alimi AM (2017) Deep learning framework based on Word2Vec and CNN for users interests classification. In: 2017 Sudan conference on computer science and information technology (SCCSIT), pp 1–7
- Ouyang X, Zhou P, Li CH, Liu L (2015) Sentiment analysis using convolutional neural network. In: 2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing, pp 2359–2364. <https://doi.org/10.1109/CIIT.2015.7392359>

- doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.349. <http://ieeexplore.ieee.org/document/7363395/>
- Pal S, Ghosh S, Nag A (2018) Sentiment analysis in the light of LSTM recurrent neural networks. *Int J Synth Emot* 9(1):33–39. <https://doi.org/10.4018/ijse.2018010103>
- Pawar KK, Deshmukh RR (2015) Twitter sentiment classification on sanders data using hybrid approach. *IOSR J Comput Eng* 17(4):118–123. <https://doi.org/10.9790/0661-1741118123>
- Preethi G, Krishna PV (2017) Application of deep learning to sentiment analysis for recommender system on cloud
- Ravi K, Ravi V (2015) A survey on opinion mining and sentiment analysis: tasks, approaches and applications, vol 89. Elsevier, Amsterdam. <https://doi.org/10.1016/j.knosys.2015.06.015>
- Ravuri S, Stoicke A (2016) A comparative study of neural network models for lexical intent classification. In: 2015 IEEE workshop on automatic speech recognition and understanding, ASRU 2015—proceedings, vol 2, pp 368–374. <https://doi.org/10.1109/ASRU.2015.7404818>
- Sasmita DH, Wicaksono AF, Louvan S, Adriani M (2017) Unsupervised aspect-based sentiment analysis on Indonesian restaurant reviews. In: Proceedings of the 2017 international conference on Asian language processing, IALP 2017 2018-Janua, pp 383–386. <https://doi.org/10.1109/IALP.2017.8300623>
- Shickel B, Tighe PJ, Bihorac A, Rashidi P (2018) Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform* 22(5):1589–1604. <https://doi.org/10.1109/JBHI.2017.2767063>
- Sohangir S, Wang D, Pomeranets A, Khoshgoftaar TM (2018) Big data: deep learning for financial sentiment analysis. *J Big Data*. <https://doi.org/10.1186/s40537-017-0111-6>
- Soliman AB, Eissa K, El-Beltagy SR (2017) AraVec: a set of Arabic word embedding models for use in Arabic NLP. *Procedia Comput Sci* 117:256–265. <https://doi.org/10.1016/j.procs.2017.10.117>
- Sze V, Chen YH, Yang TJ, Emer JS (2017) Efficient processing of deep neural networks: a tutorial and survey. *Proc IEEE* 105(12):2295–2329. <https://doi.org/10.1109/JPROC.2017.2761740>
- Taj S, Shaikh BB, Fatemah Meghji A (2019) Sentiment analysis of news articles: a lexicon based approach. In: 2019 2nd international conference on computing, mathematics and engineering technologies, iCoMET 2019, pp 1–5. <https://doi.org/10.1109/ICOMET.2019.8673428>
- Tarwani KM, Edem S (2017) Survey on recurrent neural network in natural language processing. *Int J Eng Trends Technol* 48(6):301–304
- Vateekul P, Koomsubha T (2016) A study of sentiment analysis using deep learning techniques on Thai Twitter data. In: 2016 13th international joint conference on computer science and software engineering, JCSSE 2016, pp 1–6. <https://doi.org/10.1109/JCSSE.2016.7748849>. <http://ieeexplore.ieee.org/document/7748849/>
- Wang J, Cao Z (2017) Chinese text sentiment analysis using LSTM network based on L2 and Nadam, pp 1891–1895
- Wint ZZ, Manabe Y, Aritsugi M (2018) Deep learning based sentiment classification in social network services datasets. In: 2018 IEEE international conference on big data, cloud computing, data science & engineering (BCD), pp 91–96. <https://doi.org/10.1109/BCD2018.2018.00022>. <https://ieeexplore.ieee.org/document/8530698/>
- Young T, Hazarika D, Poria S, Cambria E (2018) Recent trends in deep learning based natural language processing [review article]. *IEEE Comput Intell Mag* 13(3):55–75. <https://doi.org/10.1109/MCI.2018.2840738>
- Yuan S, Wu X, Xiang Y (2018) Incorporating pre-training in long short-term memory networks for tweet classification. *Soc Netw Anal Min*. <https://doi.org/10.1007/s13278-018-0530-1>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.