The construction of an accurate Arabic sentiment analysis system based on resources alteration and approaches comparison

Ibtissam Touahri
Department of Computer Science, Superior School of Technology, Moulay Ismail University, Meknes, Morocco

Received 22 December 2021
Revised 8 February 2022
29 April 2022
30 May 2022
Accepted 8 June 2022

Abstract

Purpose – This paper purposed a multi-facet sentiment analysis system.
Design/methodology/approach – Hence, This paper uses multidomain resources to build a sentiment analysis system. The manual lexicon based features that are extracted from the resources are fed into a machine learning classifier to compare their performance afterward. The manual lexicon is replaced with a custom BOW to deal with its time consuming construction. To help the system run faster and make the model interpretable, this will be performed by employing different existing and custom approaches such as term occurrence, information gain, principal component analysis, semantic clustering, and POS tagging filters.
Findings – The proposed system featured by lexicon extraction automation and characteristics size optimization proved its efficiency when applied to multidomain and benchmark datasets by reaching 93.59% accuracy which makes it competitive to the state-of-the-art systems.
Originality/value – The construction of a custom BOW. Optimizing features based on existing and custom feature selection and clustering approaches.

Keywords Sentiment analysis, Arabic language, Unsupervised approach, Classical supervised approach, Deep learning, Feature selection, Optimization

Paper type Research paper

1. Introduction

Opinions influence human activities, thus their analysis allows predicting the consequent behavior. However, the task of relevant information extraction from massive amounts of data remain a difficult challenge for humans which raises the need for Information technologies such as opinion mining and sentiment analysis.

Sentiment analysis is one of the most active research domains that deal with Web mining studies and data classification. Text analysis requires natural language processing tools and analysis approaches that will be applied to text. The main target of sentiment analysis is to identify the inferred polarity within reviews [1].

Training a model on a characteristic vector of considerable size is time consuming and makes the result analysis hard. As a result dimensionality reduction techniques are required. Feature selection is considered a pre-processing step for a machine learning-based system. Its primary target is to reduce data dimensionality. Dimensionality reduction approaches might be linear or non-Linear [2, 3]. Consequently, it reduces storage requirement as well as computation time [2, 4, 5] and helps to improve model readability and interpretation by

Applied Computing and Informatics
Vol. 22 No. 1/2, 2026
pp. 63-77
Emerald Publishing Limited
e-ISSN: 2210-8327
p-ISSN: 2634-1964
DOI 10.1108/ACI-12-2021-0338

reducing features number. It helps the model to train faster and it overcomes the overfitting challenge.

We can distinguish between the following types of feature selection approaches filter-based, wrapper-based, embedded, and hybrid. Each approach involves selecting a subset of features that performs the best based on a specific algorithm [4].

In this paper, we perform linguistic analysis on an opinionated multi domain corpus. Afterward, we go through the model characteristics in depth, how they are retrieved and how to select pertinent features and reduce their dimensionality. Then, we build a classifier from it. The feature vector may be shrunk not only by using the existing methods but also by introducing custom clustering approaches of characteristics. We follow a custom approach to reduce dimensionality and perform lexicon semantic clustering by defining a set of sentiment clusters, where each lexicon word is added to the relevant cluster. Moreover, we use a Part of speech (POS) tagger [6] to cluster the lexicon by defining noun, adjective and verb clusters. The selected features are evaluated and compared to the generated ones in term of size and performance. Furthermore, the system performance is compared afterward with state-of-the-art systems.

2. Previous work

Sentiment analysis is considered a subcomponent technology for other decision-making systems [7] that help to understand person attitudes and gender expressions [8], improvise features, find out strengths and weaknesses based on the online reviews of potential users and identify problems in the world of social networking sites such as Facebook and Twitter [9, 10]. It is also used to predict sentiment changes over time [11].

Subjectivity, sentiment analysis levels, and opinion types, in addition to SA resources, have been outlined in important studies in the realm of sentiment analysis. They emphasized multiple classification approaches and investigated cross-domain and cross-language variations as well as the impact of summarization [1, 12].

Sentiment analysis systems rely on linguistic resources, namely sentimental corpora and lexicons. Subjective tweets may be positive, negative, neutral or mixed in the labeled corpora. SA systems require sentiment corpora annotation and lexicons extraction [13]. Opinion Corpus for Arabic (OCA) is a collection of Arabic movie reviews, from which, the English version EVOCA, is generated [14]. LABR has almost 63,000 book reviews scored from 1 to 5 stars [15]. ArSAS and ArSentD-LEV are respectively Arabic speech-act and Levantine Multi-Topic sentiment analysis corpora [16]. The baptized ArSAS [17], SemEval 2017 [18] and ASTD [19] corpora contain tweets annotated as positive, negative or neutral. The sentimental lexicons contain sentimental terms that are verified manually, or obtained automatically by machine translation [20]. Sentiwordnet is a lexical resource that assigns numerical scores to each wordnet synset based on objectivity, positivity, and negativity [21]. ArSEL is a comprehensive Arabic Sentiment and Emotion Lexicon [16]. Gold tags and existing lexicons such as SentiWordNet 3.0 help to expand and evaluate other polarity lexicons [22]. Different Bag-of-Words (BOW) aspects have been investigated to detect sentiments [23]. Many studies have been conducted to determine lexicon words domain and to disambiguate their meaning based on fuzzy lexico-semantic and word meaning similarities [24–26].

Sentiment classification can be bipolar (positive, negative), tripolar (positive, negative, mixed) or fine grained that considers the strength of positive and negative polarities [10, 19]. Moreover, it can be applied to document, sentence, phrase and aspect levels based on grammatical and semantic orientation approaches [27–29]. Text categorization techniques based on subjectivity summarization can be applied to subjective documents [30].

Sentiment classification approaches are either unsupervised based on dictionaries [31] and apply rules [32] or supervised that build a model from a labeled corpus [33]. Since the supervised approaches are domain dependent, algorithms that address domain independence

have been proposed [34, 35]. Moreover, deep learning models for multidomain Arabic sentiment analysis have been performed [36]. Attention-based Bidirectional CNN-RNN Deep Model that extracts both past and future contexts [37], as well as the Convolutional LSTM model, has been used [38]. Word Embedding Parameters variation and Hyperparameter Tuning for Machine Learning Algorithms have been undertaken to assess their impact on Arabic Sentiment Analysis performance [39, 40]. Many studies provided systems based on word2vec, CNN and LSTM as well as a collection of open-source tools for Arabic natural language processing tasks such as sentiment analysis using AraBERT and mBERT [41, 42]. BERT post-training has been performed for aspect-based sentiment analysis [43]. A powerful comparison of effective approaches [44] and deep learning frameworks [45] for Arabic sentiment analysis has been performed. Different valuable tools [46] as well as challenges and trends of sentiment analysis [45] have been presented. The semi-supervised learning algorithms develop patterns with great generalizability from a limited labeled sample [47]. Semi-supervised learning may be used to predict users' personality traits which may improve personalized service and human psychology research [48]. Besides the semi-supervised approaches, there are clustering approaches that segment data into different classes without the need for annotated data and pre-trained models [49, 50].

The linguistic content generated by Web users is multi-lingual since it may contain various languages, combine different dialects or languages or switch between them within the same expression. Aside from multilingual sentiment analysis, adaptation of English resources and sentiment classification approaches to other languages have been conducted [51–53]. Besides MSA sentiment analysis, many studies have focused on Arabic dialects [54, 55] and integrated stem and lemma lexicon morphologies [56]. Other studies carried out an in-depth study of Arabic and multi-lingual sentiment analysis, and presented their approaches and tools as well as their challenges [57, 58]. Sentiment analysis is faced with many challenges among which, spam, polarity fuzziness, sarcasm, domain dependency, fake news, Arabic varieties, language morphology and code-switching [1, 57, 59, 60].

Since sentiment analysis is considered a classification domain, feature selection has gained researchers interest who presented feature selection algorithms, their applications and categories [2, 3, 5, 61] and addressed their strengths and challenges [62] besides ranking fundamental algorithms used to reduce dimensionality [2] according to relevance [63], computation time [4, 64], and the matching degree between the algorithm and the known optimal solution [65–67].

3. Linguistic resources

3.1 Corpus collection

The employed corpora that vary in terms of domain and length were extracted from different websites by the authors of [68] and cover various domains Hotels (HTL), Products (PROD), Movies (MOV) and Restaurants (RES). Table 1 summarizes the statistics of the used corpora that were preprocessed by removing all non Arabic characters, namely, Latin letters, punctuation marks, and digits.

3.2 Lexicon

3.2.1 Manual lexicon. The domain-specific lexicon which statistics are given in Table 1 was established by Ref. [68], re-checked, altered and cleaned by Ref. [56]. We created two lexicons by browsing the investigated corpora, negation words (NW) that contain 167 negation indicators that reverse terms polarity, and a set (SW) of 558 stop words that keep the same meaning regardless to the context. Manual lexicon extraction and adjustment is a difficult and time-consuming task. Hence, we describe in the next section the followed steps to perform the Bag-of-Words construction.

Domain	Tag	Comments		Manual lexicon	Initial BOW	Th	First F_O		Second F_O		Third F_O		Forth F_O	
		Size	AVG				Size	Reduction rate	Th	Size	Th	Size	Th	Size
HTL	Positive	10764	93	115	51016	1.85	14452	71.67	4	3139	9.39	870	17.45	245
	Negative	2644		75	16188	1.27	2590	84.00	2.74	866	4.23	199	7.38	65
PROD	Positive	2802	12	180	7219	1.56	1718	76.20	3.36	445	6.48	139	10.38	42
	Negative	793		117	1923	1.17	221	88.51	2.54	58	4.08	9	8.55	2
MOV	Positive	966	341	44	37233	1.61	10286	72.37	3.23	2655	6	715	9.65	250
	Negative	383		34	11192	1.19	1486	86.72	2.48	385	3.88	154	5.20	45
RES	Positive	7925	38	281	28842	1.56	7038	75.60	3.30	1781	6.35	509	10.54	165
	Negative	2646		236	10904	1.20	1453	86.67	2.51	408	3.82	163	5.06	43

Table 1.
Statistics of sentiment corpora and pretreated lexicons

3.2.2 Bag-of-words. We aim to improve the classical Bag-of-Words extraction by addressing automation, domain dependency and semantic disambiguation. For this, we opt for a custom approach to generate an automatic BOW by performing many filtering and threshold decision steps. Figure 1 describes the BOW construction process.

We construct a custom BOW that weighs terms based on their occurrences. After pretreatment, we tokenize positive and negative reviews into raw positive and negative lexicon terms using a space delimiter; and removed stop words, negation words, redundancy and intersection within the positive and negative lexicons. We automatically define the threshold (Th), used to obtain the BOW size, as the average of lexicon terms occurrences in either the positive or the negative corpus. The reduced BOW, after each filtering operation (F_O), consists only of terms whose occurrences are greater than the threshold. We present in Table 1 the different initial and reduced BOW sizes; and the reduction rate obtained following the computed thresholds for each domain.

4. System methodology

We aim to construct a sentiment analysis system that addresses the characteristic of the research topic and improves the optimization approaches. It can serve as a roadmap for many classification domains whose main target is the separation of data with similar characteristics besides their interpretation. We present in Figure 2 the system architecture.

4.1 Classification approaches

Machine learning algorithms prefer well defined fixed-length inputs and outputs. In the following, we describe the extracted features and how models are generated from the data using machine learning approaches.

4.1.1 Unsupervised. The unsupervised approach is based on the criteria that consider the major score of sentimental terms within an expression, hence a review is labeled with the polarity of the major score.

4.1.2 Classical supervised. We represent comments by a vector V_W based on lexicon terms. The characteristic vector $V_W = (P_1, \cdots, P_p, N_1, \cdots, N_n, P_w, N_w, \overline{P}_w, \overline{N}_w)$ of a review W

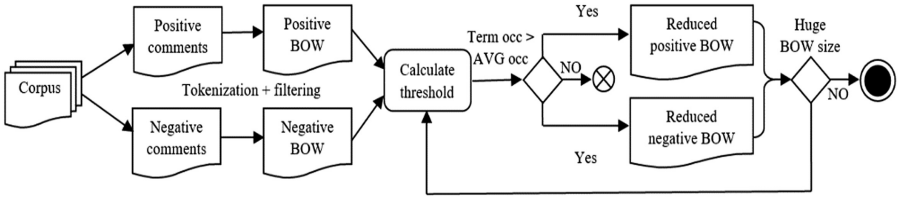

Figure 1.
BOW construction process

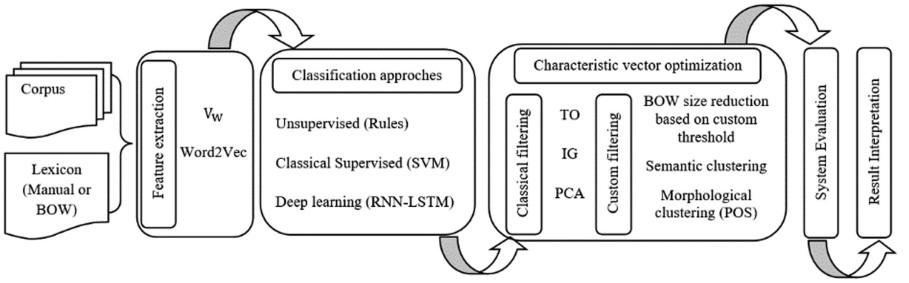

Figure 2.
System architecture

is composed of terms occurrences from the positive P_i ($1 \leq i \leq p$) and negative N_j ($1 \leq j \leq n$) lexicons respectively, as well as their sum P_W, N_W; and the number of times they have been preceded by a negation term $\overline{P}_W$ and $\overline{N}_W$ [56]. Afterward, we build a model V_W - SVM using 80% for training from each corpus and 20% for testing.

4.1.3 Deep learning. From the training corpus we create a Word2Vec model that transforms the words of the corpus, with a frequency greater than 5 and windows size equals to 10, which are an empirical choice, into a set of numeric vectors with a size of 300. We employ padding arrays to provide a consistent representation for all reviews of different lengths. The mask matrix contains 1 if data is present and 0 otherwise. The entered corpus is represented by a characteristic matrix, labels either positive or negative as well as their masks. For the second model, we use the vector V_W described in section 4.1.2 to represent the corpus. Subsequently, the set of vectors is fed to a neural network for weight estimation. The neural network is made up of layers, the input layer, and the inner LSTM and RNN layers, which helps to have Word2Vec – RN and V_W – RN models. For the LSTM layer, we initialize weights using Xavier, use their update program Adam, and Tanh as the activation function. Finally, the RNN layer has a softmax activation function that gives a probability distribution over the classes, and defines loss using MCXENT function.

4.2 Characteristic vector optimization

We optimize the characteristic vector based on existing approaches such as term occurrence filter (TO), information gain (IG) and PCA; and custom approaches based on BOW reduction, semantic and morphological clustering which helps to maintain high accuracy while reducing feature number, execution time, and storage requirement; and improving model interpretation.

4.2.1 Classical filtering. We use term occurrence based on the characteristic vector V_W, information gain and PCA to perform data filtering. Information gain of an attribute is measured with respect to the class. PCA enables the transformation of a dataset into a new dataset of lower dimensionality based on the identification of correlations within it.

4.2.2 Custom filtering. 4.2.2.1 BOW size reduction. In this paper, we replace the manual lexicon construction and semantic verification with a BOW constructed using a custom automatic approach. We feed the characteristic vector V_W based on the BOW lexicons into an SVM classifier to build a supervised sentiment analysis system.

4.2.2.2 Lexicon semantic clustering. In order to reduce the characteristic vector size V_W, we diminish the number of lexicon segments to twelve $V_{W\_S} = (P_1, \cdots, P_6, N_1, \cdots, N_6, P_W, N_W, \overline{P}_W, \overline{N}_W)$. The positive segments are Love, Optimism, Joy, Satisfaction, Entertainment, and Relief; whereas the negative segments include Hatred, Pessimism, Sadness, Dissatisfaction, Boredom and Fear.

4.2.2.3 Lexicon POS clustering. We accomplish classification using a model built from an optimized characteristic vector $V_{W\_P} = (P_V, P_N, N_V, N_N, P_W, N_W, \overline{P}_W, \overline{N}_W)$ composed of the occurrences of the four POS classes terms as well as the last four features of V_W. The vector is based on POS clustering where each positive and negative lexicon is segmented using the POS Tagger [6] into two classes V and N, where V is the class of verbs and N is that of adjectives and nouns. The segmentation is followed by a slight manual check that proved the efficiency of the automatic tagging. Adjectives and nouns are confused within the same category since they can be used to tag the same term in some cases, for instance, the term سعيد /sEyd/ (happy). Moreover, the adjectives are abundant in the employed lexicon which was already confirmed by many previous studies where they were considered the most significant class for sentiment analysis as they are the most clues for subjectivity.

5. Experimental work and result interpretation

We choose the best performing approach and classifier according to the size and characteristics of the data, then we perform experiments on the characteristic vector optimization.

5.1 Performance according to classification approaches

The results of the comparison between classification approaches, namely the unsupervised, classical supervised and deep learning, described in section 4 are given in Table 2.

From the results, we can point out that the best performance is achieved in the HTL domain. The poor results recorded in the MOV domain can be explained by the nature of the reviews, and their length (Table 1).

According to Table 2, the supervised approach gave better results than the unsupervised and deep learning approaches. The degraded results of deep learning can be mainly due to the limited size of the used corpora. In addition, when comparing the two deep learning models, the extracted vector V_W outperforms Word2Vec which shows the relevance of the extracted sentimental terms. Hence, we opt for the supervised approach as well as the vector V_W to perform the remaining classification tests that aim to optimize the characteristic vector using various approaches.

5.2 Characteristic vector optimization

We based dimensionality reduction on the classical filtering operations TO, IG and PCA (section 4.2.1) and three custom approaches which are BOW lexicon reduction and the segmentation of the lexicon using manual semantic clustering or automatic morphology clustering (Sections 4.2.2.1, 4.2.2.2 and 4.2.2.3) based on POS tagger [6]. In Table 3, we give the experimental results of each optimization approach and we compare them with the result obtained using the raw lexicon. We also define the number of features as well as the execution time (ET) for HTL domain only in order to lighten the paper.

5.2.1 Classical filtering. Table 3 shows that we have comparable results related to the classical filtering operations TO, IG and PCA whose accuracies are very close to each other. Moreover, from a reduced set of features, we have obtained pertinent results. The PCA filter gives degraded results at PROD and MOV domains which may be explained by the fact that the principal components calculated for these domains are not easily separable by the SVM classifier.

5.2.2 Custom approaches. 5.2.2.1 BOW size reduction. The results show that the execution time is optimized whereas the accuracy degrades when the Bag-of-Words size is reduced, which may be caused by the removal of relevant features when passing from a BOW to a reduced one. The strength of the BOW lexicon besides the result relevance, since it has an advantage over the manual lexicon, lies in that it weighs corpus terms based on their occurrences according to an automatic process.

5.2.2.2 Lexicon semantic and morphological clustering. From Table 3, we have comparable results between words; semantic and POS classes models. Furthermore,

Domain	Unsupervised	Supervised	Deep learning	
		V_W – SVM	V_W – RN	Word2Vec - RN
HTL	91.18	93.48	89.3	66.97
PROD	84.02	84.70	84.28	74.48
MOV	82.96	83.33	82.59	71.15
RES	83.01	84.44	82.97	69.23

Table 2. Classification accuracies according to approaches

Domain		Lexicon				BOW				Semantic classes	POS
		word	TO	IG	PCA	F_O_1	F_O_2	F_O_3	F_O_4		
HTL	Approach	Manual		Automatic			Automatic			Manual	Automatic
	Features	213	194	160	176	17046	4009	1073	314	16	8
	Execution time (s)	3.96	2.94	1.6	2.01	2679.89	347.85	113.09	7.77	0.87	0.53
	Accuracy	93.59	93.48	92.24	93.10	94.42	91.03	88.11	84.89	92.05	92.13
PROD		86.79	84.70	85.95	83.73	87.48	82.89	78.99	77.33	84.42	84.42
MOV		82.96	83.33	82.96	80.74	88.89	87.78	96.67	81.85	82.96	82.96
RES		84.77	84.44	84.29	84.15	86.01	84.20	81.03	78.09	83.49	83.73

Table 3.
Accuracies of models based on filtering operation

Interpretation can be easier when using semantic and POS vectors since their size is limited and not proportional to lexicon terms which means that the extension of the lexicon will not affect the characteristic vectors size which is not the case for the word model. The semantic and POS models have achieved the same accuracy in PROD and MOV domains which is an advantage for the automatic segmentation using POS tagging in comparison to the manual semantic segmentation.

5.2.2.3 Custom approaches comparison. We compute the information gain of each characteristic based on which the words; semantic and POS classes models are constructed.

5.2.2.3.1. Feature significance. For the word model, the features N_w, P_w and $\overline{P}_w$ are respectively ranked from the first to the third followed by $\overline{N}_w$ that is of low significance in comparison to the best ranked features. However, the performance of our system is improved using this feature since it inverts the polarity of the negative terms preceded by a negation word.

For the semantic segmentation (Figure 3, left), the feature with low significance is $\overline{N}_w$ that is ranked 13 out of 16 features and the most significant features are N_W, P_w and $\overline{P}_W$ that are ranked the first, the third and fifth out of 16 respectively, which shows the rarety of the negative terms $\overline{N}_w$ preceded with a negation term and also the importance of positive terms P_w and negative terms (true negative N_W and false positive $\overline{P}_W$) in the classification. The most significant segment relates to the category of satisfaction and dissatisfaction P_4 and N_4 that are ranked the fourth and the second out of 16 respectively.

Using POS segmentation (Figure 3, right), we identify the feature with low significance, namely the one with a low information gain which is $\overline{N}_w$ that is ranked in the seventh position out of 8, and the most significant feature N_W that is ranked the first out of 8 features. As a result, there aren't many terms preceded by a negation word ($\overline{N}_w$), and the presence of negative terms N_W is discriminant in identifying the class of each analyzed review. Moreover, P_N and N_N namely the positive and negative nouns, that are ranked the third and the second respectively, are the most subjectivity indicators that convey more information than other remaining features which proves that the identification of sentiment is basically related to the presence of this category within comments.

The feature ranking related to word, semantic classes and POS clustering when analyzing the four domain specific models is almost the same with a slight difference in the position.

Using semantic classes and analyzing the model on which it has been created will help to define emotion categories and also to detect hate speech by adding related lexicon categories as the model clearly define the best features based on their information gain. For the segmentation based on the POS it will help in defining which category of the lexicon has to be added to improve the accuracy of our system.

The lexicon clustering turns out to be pertinent since it summarizes the information dispersed when using the word model and it may be extended by defining which categories

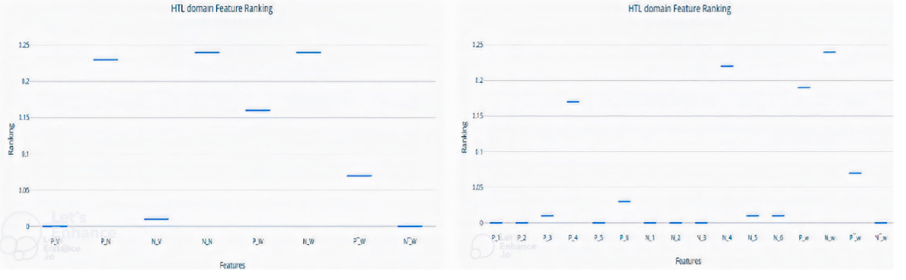

Figure 3.
Semantic and POS feature ranking

we are interested in and which features are the most pertinent based on a reasonable size of features, making the model interpretation and error analysis easier.

5.2.2.3.2. Model characteristics. From the results of Table 3, semantic classes give comparable results to the word lexicon with a gain in execution time and storage requirement. POS segmentation, in turn, gives comparable results to semantic classes and hence helps to overcome the same challenges with which the word model is faced.

In the case of our paper, where the lexicon is of limited size (190 terms), there isn't a significant gain since the word lexicon is small, however, the segmentation will be helpful when it comes to lexicons with huge size, for instance, the BOW with 2679.89 ET that is characterized by 17,042 terms, which largely exceeds the size of semantic and POS classes categories that are fixed to 12 and 4 respectively. Moreover, the representation based on lexicon segmentation helps to augment the interpretation of a model and preserves the consistency of each feature significant which is lost when using the number of features proportional to the number of lexicon terms.

6. Systems comparison

After measuring the performance of our system based on various configurations, we give in Table 4 examples of correctly classified and misclassified comments.

From Table 4, the annotation subjectivity may be the cause for the misclassification since a comment mixed can be annotated as positive or negative according to its context. Moreover, the lack of comment sentimental terms within the lexicon can lead to its misclassification, and hence the need for the lexicon extension that enlarges the characteristic vector and this was the major cause behind the raised questions, how to keep information and thus high accuracy, to which we have responded in this paper by the optimization of the characteristic vector.

In order to state our approach with the previous works, we compare our system to Mazajak, CAMel Tools, SemEval 2017; and Abu Farha and Magdy systems [18, 41, 42, 44] based on deep learning and pretrained models. The system results are given in Table 5 using accuracy, precision, recall, and F-measure metrics.

The comparison is performed on the same datasets which are ArSAS [17], SemEval 2017 [18], and ASTD [19] in order to reach a fair comparison between the different systems. The results prove the efficiency of our system that shows an improvement. Moreover, the obtained low F-measure measured by 69.95% is since there are few data to train on in comparison to testing data. Hence, we inversed the training and testing portions and obtained a 93.55% F-measure.

Table 4.
Examples of correctly classified and misclassified comments

Correctly classified	Misclassified
اعجبني جيدا هذا المنتج (I really like this product),1	السلعه جيده الى حد ما ولكن لم اجد بعض المميزات التى فى الاعلان ولكنى سوف انصح بعض اصدقائى به على امكانيت هذا الجهاز.- (The product is good to some extent, but I did not find some of the features in the advertisement, but I will recommend it to some of my friends for the possibility of this device) **The comment has a mixed sentiment and it was annotated manually as negative by the corpus owner, whereas our system annotated it as positive since the major score when reading the comment is positive**
ممتازة فرح بها أولادى فرح لا يوصف (Excellent, the joy of my children with it is indescribable),1	
ممتازة و مريحة للالم بالرقبة حيث اني كنت اعاني من تصلب فى عضلات الرقبة و لكن الحمدلله بعد استخدام هذة السلعة ساعدتنى فى التخلص من الالم (Excellent and comfortable for neck pain, as I was suffering from stiffness in the neck muscles, but thank God, after using this product, it helped me get rid of the pain),1	
السعر غالى جدا لضعف الامكانيات (The price is very high due to the weakness of the capabilities),-1	غير عملية بالمرة وغير مطابق للمواصفات وبطيئة جداجدا كما لم تاتي بالملحقات الخاصة بها (It is completely impractical, does not meet the specifications, and is very slow, as it did not come with its own files) **The comment is negative and it was annotated as positive because of the lack of its sentimental terms within the used lexicon**
يوجد عيوب فى التصنيع خامة رديئة ولا انصح بشرائها (There are manufacturing defects, poor material, and I do not recommend buying it),-1	

System	Corpus	Approach	Accuracy	Metrics			Our system improvement
				Precision	Recall	*F*-measure	
Mazajak [41]	ArSAS [17]	CNN followed by an LSTM	92	90	90	90	2.9%
Our system		BOW + SVM	94.6	91.4	94.65	92.9	
CAMel Tools [42]	ASTD [19]	AraBERT	–	–	–	73	0.75%
Our system		BOW + SVM	77.26	73.7	73.85	73.75	
SemEval 2017 [18]	3,555 training and 6,100 testing tweets [18]	Deep learning	58.1	63.96	58.3	61	8.95%
System [44]		AraBERT	68	71.12	67	69	0.95%
Our system		BOW + SVM	71.23	70.15	69.8	69.95	
	6,100 training and 3,555 testing tweets [18]		93.85	93.6	93.5	93.55	Custom test

Table 5.
Systems comparison

7. Conclusion and further work

In this paper, we have aimed to optimize the components of a sentiment analysis system, we first collected multi-domain datasets and lexicons. Since the manual construction and verification of a lexicon is time consuming, we have constructed a custom BOW whose size is diminished following a custom threshold. We have performed classification based on the unsupervised; and the classical and deep neural supervised approaches. The execution time, and storage requirement, as well as model interpretation, have gained the interest in data analytics, thus we have opted for existing and custom methods to optimize the characteristic vector of the opinionated reviews. Moreover, to make the interpretation of our models and results easier, the reduced characteristic vector was based on semantic and morphological lexicon segmentation to give significance to its components. The current system proved efficient in comparison to the enhanced state-of-the-art models. As further work, we intend to apply the described approaches to a wide range of classification areas to prove their efficiency since we believe that the sole requirement is the adaptation of domain categories. Moreover, the automatic annotation of corpora will be one of the main focuses. We intend also to extract cross-domain and cross-lingual features. In order to minimize the effort when performing sentiment analysis, we will base the task on transfer learning.

References

1. Liu B. Sentiment analysis and opinion mining. Synth Lectures Hum Lang Tech. 2012; 5(1): 1-167.
2. Anowar F, Sadaoui S, Selim B. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). Computer Sci Rev. 2021; 40: 100378.
3. Fazili S, Grover J, Wazir S, Mehta I. Recent trends in dimension reduction methods. ICIDSSD. 2021; 2020: 68.
4. Zebari R, Abdulazeez A, Zeebaree D, Zebari D, Saeed J. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. J Appl Sci Technology Trends. 2020; 1(2): 56-70.
5. Ayesha S, Hanif MK, Talib R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. Inf Fusion. 2020; 59: 44-58.
6. Ababou N, Mazroui A. A hybrid Arabic POS tagging for simple and compound morphosyntactic tags. Int J Speech Technol. 2016; 19(2): 289-302. doi: 10.1007/s10772-015-9302-8.
7. Cambria E, Das D, Bandyopadhyay S, Feraco A. Affective computing and sentiment analysis. In: A practical guide to sentiment analysis. Cham: Springer; 2017. 1-10.
8. Wolf A. Emotional expression online: gender differences in emoticon use. Cyberpsychology Behav. 2000; 3(5): 827-33.
9. Liu B, Hu M, Cheng J. Opinion observer: analyzing and comparing opinions on the web. In: Proceedings of the 14th International Conference on World Wide Web; 2005. 342-51.
10. Yang Q, Rao Y, Xie H, Wang J, Wang FL, Chan WH. Segment-level joint topic-sentiment model for online review analysis. IEEE Intell Syst. 2019; 34(1): 43-50.
11. McIntyre-Bhatty YT. Neural network analysis and the characteristics of market sentiment in the financial markets. Expert Syst. 2000; 17(4): 191-8.
12. Liu B. Sentiment analysis and subjectivity. Handbook Nat Lang Process. 2010; 2: 627-66.
13. Abdul-Mageed M, Diab M, Korayem M. Subjectivity and sentiment analysis of modern standard Arabic. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies; 2011. 587-91.
14. Rushdi-Saleh M, Martín-Valdivia MT, Lopez LAU, Perea-Ortega JM. Bilingual experiments with an Arabic-English corpus for opinion mining. In: Proceedings of the International Conference Recent Advances in Natural Language Processing 2011; 2011. 740-5.

15. Aly M, Atiya A. Labr: a large scale Arabic book reviews dataset. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (2 Short Papers); 2013. 494-8.
16. Al-Khalifa H, Magdy W, Darwish K, Elsayed T, Mubarak H. Proceedings of the 4th workshop on open-source Arabic corpora and processing tools, with a shared task on offensive language detection. In: Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection; 2020.
17. Elmadany A, Mubarak H, Magdy W. Arsas: an Arabic speech-act and sentiment corpus of tweets. OSACT. 2018; 3: 20.
18. Rosenthal S, Farra N, Nakov P. SemEval-2017 task 4: sentiment analysis in twitter. arXiv preprint arXiv:1912.00741. 2019.
19. Nabil M, Aly M, Atiya A. Astd: Arabic sentiment tweets dataset. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing; 2015. 2515-19.
20. Touahri, I and Mazroui, A. Opinion and sentiment polarity detection using supervised machine learning. In 2018 IEEE 5th International Congress on Information Science and Technology (CiSt), IEEE; 2018. 249-53.
21. Esuli A, Sebastiani F. Sentiwordnet: a publicly available lexical resource for opinion mining. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06); 2006.
22. Abdul-Mageed M, Diab M. Toward building a large-scale Arabic sentiment lexicon. In: Proceedings of the 6th International Global WordNet Conference; 2012. 18-22.
23. Cummins N, Amiriparian S, Ottl S, Gerczuk M, Schmitt M, Schuller B. Multimodal bag-of-words for cross domains sentiment analysis. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE; 2018. 4954-8.
24. Wang Y, Yin F, Liu J, Tosato M. Automatic construction of domain sentiment lexicon for semantic disambiguation. Multimedia Tools Appl. 2020; 79: 22355-73.
25. Goularte FB, Sorato D, Nassar SM, Fileto R, Saggion H. MSC+: language pattern learning for word sense induction and disambiguation. Knowledge-Based Syst. 2020; 188: 105017.
26. Rudkowsky E, Haselmayer M, Wastian M, Jenny M, Emrich Š, Sedlmair M. More than bags of words: sentiment analysis with word embeddings. Commun Methods Measures. 2018; 12(2-3): 140-57.
27. Farra N, Challita E, Abou Assi R, Hajj H. Sentence-level and document-level sentiment mining for Arabic texts. In: 2010 IEEE International Conference on Data Mining Workshops, IEEE; 2010. 1114-19.
28. Cambria E. An introduction to concept-level sentiment analysis. In: Mexican International Conference on Artificial Intelligence, Berlin, Heidelberg: Springer; 2013. 478-83.
29. Cambria E, Schuller B, Liu B, Wang H, Havasi C. Knowledge-based approaches to concept-level sentiment analysis. IEEE Intell Syst. 2013; 28(2): 12-14.
30. Pang, B and Lee, L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. arXiv preprint cs/0409058, 2004.
31. Kodipaka RR, Polepaka S, Rafeeq M. Design of sentiment analysis system using polarity classification technique. Int J Computer Appl. 2015; 125(15): 22-24.
32. Abdulla NA, Ahmed NA, Shehab MA, Al-Ayyoub M. Arabic sentiment analysis: lexicon-based and corpus-based. In: 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), IEEE; 2013. 1-6.
33. Purpura A, Masiero C, Silvello G, Antonio Susto G. Supervised lexicon extraction for emotion classification. In: Companion Proceedings of the 2019 World Wide Web Conference; 2019. 1071-8.
34. Assiri A, Emam A, Al-Dossari H. Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis. J Inf Sci. 2018; 44(2): 184-202.

35. Chetviorkin I, Loukachevitch N. Two-step model for sentiment lexicon extraction from twitter streams. In: Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis; 2014. 67-72.

36. El-Affendi MA, Alrajhi K, Hussain A. A novel deep learning-based multilevel parallel attention neural (MPAN) model for multidomain Arabic sentiment analysis. IEEE Access. 2021; 9: 7508-18. doi: 10.1109/ACCESS.2021.3049626.

37. Basiri ME, Nemati S, Abdar M, Cambria E, Acharya UR. ABCDM: an attention-based bidirectional CNN-RNN deep model for sentiment analysis. Future Gener Comput Syst. 2021; 115: 279-94. doi: 10.1016/j.future.2020.08.005.

38. Behera RK, Jena M, Rath SK, Misra S. Co-LSTM: Convolutional LSTM model for sentiment analysis in social big data. Inf Process Manag. 2021; 58(1): 102435. doi: 10.1016/j.ipm.2020.102435.

39. Alnawas, A and Arici, N. Effect of word embedding variable Parameters on Arabic sentiment analysis performance, 2018. 6.

40. Elgeldawi E, Sayed A, Galal AR, Zaki AM. Hyperparameter tuning for machine learning algorithms used for Arabic sentiment analysis. Informatics. 2021; 8(4): 79. doi: 10.3390/informatics8040079.

41. Farha IA, Magdy W. Mazajak: an online Arabic sentiment analyser. In: Proceedings of the Fourth Arabic Natural Language Processing Workshop; 2019. 192-8.

42. Obeid O, Zalmout N, Khalifa S, Taji D, Oudah M, Alhafni B, Inoue G, Eryani F, Erdmann A, Habash N. CAMeL tools: an open source python toolkit for Arabic natural language processing. In: Proceedings of the 12th Language Resources and Evaluation Conference; 2020. 7022-32.

43. Xu H, Liu B, Shu L, Yu PS. BERT post-training for review reading comprehension and aspect-based sentiment analysis. arXiv preprint arXiv:1904.02232. 2019.

44. Abu Farha I, Magdy W. A comparative study of effective approaches for Arabic sentiment analysis. Inf Process Manag. 2021; 58(2): 102438. doi: 10.1016/j.ipm.2020.102438.

45. Zahidi Y, Younoussi YE, Al-Amrani Y. A powerful comparison of deep learning frameworks for Arabic sentiment analysis. Int J Electr Comput Eng IJECE. 2021; 11(1): 745. doi: 10.11591/ijece.v11i1.pp745-752.

46. Zahidi Y, El Younoussi Y, Al-Amrani Y. Different valuable tools for Arabic sentiment analysis: a comparative evaluation. Int J Electr Comput Eng IJECE. 2021; 11(1): 753. doi: 10.11591/ijece.v11i1.pp753-762.

47. Han Y, Liu Y, Jin Z. Sentiment analysis via semi-supervised learning: a model based on dynamic threshold and multi-classifiers. Neural Comput Appl. 2020; 32(9): 5117-29.

48. Zheng H, Wu C. Predicting personality using Facebook status based on semi-supervised learning. In: Proceedings of the 2019 11th International Conference on Machine Learning and Computing; 2019. 59-64.

49. Akmal S, Asif HMS. Sentiment analysis based on soft clustering through dimensionality reduction technique. Mehran Univ Res J Eng Technology. 2021; 40(3): 630-44.

50. Al-saqqa, S and Al-naymat, G, Unsupervised sentiment analysis approach based on clustering for Arabic text; 2025, 13.

51. Denecke K. Using sentiwordnet for multilingual sentiment analysis. In: 2008 IEEE 24th International Conference on Data Engineering Workshop, IEEE; 2008. 507-12.

52. Brooke J, Tofiloski M, Taboada M. Cross-linguistic sentiment analysis: from English to Spanish. In: Proceedings of the International Conference RANLP-2009; 2009. 50-4.

53. Abbasi A, Chen H, Salem A. Sentiment analysis in multiple languages: feature selection for opinion classification in web forums. ACM Trans Inf Syst (TOIS). 2008; 26(3): 1-34.

54. Oussous A, Lahcen AA, Belfkih S. Improving sentiment analysis of Moroccan tweets using ensemble learning. In: International Conference on Big Data, Cloud and Applications, Cham: Springer; 2018. 91-104.

55. Harrat S, Meftouh K, Smaïli K. Maghrebi Arabic dialect processing: an overview. J Int Sci Gen Appl. ISGA. 2018; 1. hal-01873779.

56. Touahri I, Mazroui A. Studying the effect of characteristic vector alteration on Arabic sentiment classification. J King Saud University-Computer Inf Sci. 2021; 33(7): 890-898.

57. Oueslati O, Cambria E, HajHmida MB, Ounelli H. A review of sentiment analysis research in Arabic language. Future Generation Computer Syst. 2020; 112: 408-30.

58. Lo SL, Cambria E, Chiong R, Cornforth D. Multilingual sentiment analysis: from formal to informal and scarce resource languages. Artif Intelligence Rev. 2017; 48(4): 499-527.

59. Touahri I, Mazroui A. Enhancement of a multi-dialectal sentiment analysis system by the detection of the implied sarcastic features. Knowledge-Based Syst. 2021; 227: 107232.

60. Liu B, Zhang L. A survey of opinion mining and sentiment analysis. In: Mining text data. Boston, MA: Springer; 2012. 415-63.

61. AlNuaimi N, Masud MM, Serhani MA, Zaki N. Streaming feature selection algorithms for big data: a survey. Appl Comput Inform. 2022; 18(1/2): 113-135. doi: 10.1016/j.aci.2019.01.001.

62. Sutha K, Tamilselvi JJ. A review of feature selection algorithms for data mining techniques. Int J Computer Sci Eng. 2015; 7(6): 63.

63. Hashim BM, Amutha R. Human activity recognition based on smartphone using fast feature dimensionality reduction technique. J Ambient Intelligence Humanized Comput. 2021; 12(2): 2365-74.

64. Kalaivani, KS, Uma, S and Kanimozhiselvi, CS. A review on feature extraction techniques for sentiment classification. In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), IEEE; 2020. 679-83.

65. Madasu A, Elango S. Efficient feature selection techniques for sentiment analysis. Multimedia Tools Appl. 2020; 79(9): 6313-35.

66. Gokalp O, Tasci E, Ugur A. A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. Expert Syst Appl. 2020; 146: 113176.

67. Tubishat M, Abushariah MA, Idris N, Aljarah I. Improved whale optimization algorithm for feature selection in Arabic sentiment analysis. Appl Intelligence. 2019; 49(5): 1688-707.

68. ElSahar H, El-Beltagy SR. Building large Arabic multi-domain resources for sentiment analysis. In: International Conference on Intelligent Text Processing and Computational Linguistics, Cham: Springer; 2015. 23-34.

Corresponding author
Ibtissam Touahri can be contacted at: ibtissamtouahri555@gmail.com