Journal of Theoretical and Applied Information Technology
15th September 2019. Vol.97. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A FRAMEWORK FOR ARABIC SENTIMENT ANALYSIS USING MACHINE LEARNING CLASSIFIERS

AHMAD HAWALAH
College of Computer Science and Engineering, Taibah University, Saudi Arabia
E-mail: ahawalah@taibahu.edu.sa

ABSTRACT

In recent years, the use of Internet and online comments, expressed in natural language text, have increased significantly. However, it is difficult for humans to read all these comments and classify them appropriately. Consequently, an automatic approach is required to classify the unstructured data. In this paper, we propose a framework for Arabic language comprising of three steps: pre-processing, feature extraction and machine learning classification. The main aim of the proposed framework is to exploit the combination of different Arabic linguistic features. We evaluate the framework using two benchmark Arabic tweets datasets (ASTD, ATA), which enable sentiment polarity detection in general Arabic and Jordanian dialects. Comparative simulation results show that machine learning classifiers such as Support Vector Machine (SVM), Naive Bayes, MultiLayer Perceptron (MLP) and Logistic Regression-based produce the best performance by using a combination of n-gram features from Arabic tweets datasets. Finally, we evaluate the performance of our proposed framework using an Ensemble classifier approach, with promising results.

Keywords: *Arabic, Sentiment Analysis, NLP*

1. INTRODUCTION

Over the past decade or so, Internet users have been increasingly contributing to discussions on different topics such as sports, politics, brands etc, by adding their own comments and opinions into web pages and social media. Web 2.0 allows Internet surfers to share their views via social media such as Facebook, Twitter, Instagram etc. This has resulted in massive amounts of unstructured big data. Automated sentiment analysis is required to extract valuable information from the data [2] [3] [20] [27] [28] [34].

The aim of sentiment analysis is to classify the data based on its polarity and automatically extract users opinion relating to relevant issues, events and services. User views are ex- pressed in a variety of ways, such as articles, reviews, forums, posts, comments, and tweets etc. The automatic extracting of these opinions is a very challenging task [1] [9] [18] [30].

There are currently two main approaches for sentiment classification such as machine learning and lexicon-based approach. The machine learning approach consists of supervised, unsupervised and semi-supervised techniques. The supervised technique contains a set of labeled data along with their class such as positive and negative. Based on these labeled data, machine learning classifiers are trained. The semi-supervised approach consists of labeled and unlabeled data. The machine learning classifiers are trained using the labeled data and predict the labeled for unlabeled data. The unsupervised technique is based on unlabeled data. The lexicon-based approach uses lexicons to identify the polarity of the sentence [4] [16] [17] [24] [25].

Arabic is the official language of 22 countries, spoken by around 400 million people. It is considered as one of the fastest growing languages in the world. There are about 65 million Arabic-speaking users online, which make up of around 18.8% of the global Internet population. Arabic is a Semitic language and consists of different dialects. These dialects are used in everyday communication and are not taught in schools. There is one formal standard writing in Arabic, termed as Modern Standard Arabic (MSA). There is a large difference between MSA and most Arabic dialects. Further, MSA is not a native language of any Arabic country, and is syntactically, morphologically and phonologically based on Classic Arabic [31].

The automatic identification polarity of Arabic text is difficult due to various factors. First, the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

complexity of the language with regards to both morphology and structure creates lots of problems and has resulted in limited tools and resources for the aim of sentiment analysis. Second, Arabic consists of different dialects and every region has its own dialect. Arabic is one of the top 10 languages used on the Internet, with various websites and weblogs specialized in Arabic reviews. However, no valuable research is available in Arabic [5] [52].

Sentiment analysis on social media has been the interest of a growing number of researchers world-wide. Many have used machine learning approaches to classify sentiments of English tweets. However, there is a less work carried out on the sentiment analysis of tweets for other languages [6] [45].

In this paper, we propose a new framework to extract Arabic features such as N-grams, from Arabic tweets datasets. Machine learning classifiers are used to comparatively evaluate the performance of selected features for Arabic tweets datasets. In the literature, extensive research has been carried out to determine the optimum features in English and Chinese languages in particular. However, to the best of our knowledge, Arabic feature selection for sentiment analysis is not yet well researched.

This paper is organized as follows: Section 2 outlines related work in Arabic and other languages. Section 3, presents the proposed framework. Section 4 describes the comparative experimental results. Section 5 provides the discussion of the experiments. Finally, section 6 concludes the work and outlines some future work directions.

2. RELATED WORK

In the literature, extensive research has been carried out to model novel sentiment analysis for English, Arabic and other languages.

2.1 Arabic language

Shoukry et al. [51] have proposed an approach to detect polarity in Arabic tweets. These tweets are collected, pre- processed and converted into vectors. Finally, machine learning classifiers such as SVM and Nave Bayes are used to evaluate the performance of the approach. The experimental results show that SVM classifier (72.1%) has achieved better performance as compared to Nave Bayes classifier (65.4%). Abdulla et al. [7] have proposed an approach to identify polarity in Arabic reviews. In this approach large datasets are collected and manually annotated. Afterwards, two different classifiers are used to evaluate the performance of the approach. The experimental results show the SVM classifier (64.1%), which outperforms Naive Bayes classifier (55.9%). Ibrahim et al.[32] have proposed a feature-based approach to detect polarity in Arabic product reviews. The proposed approach introduces Arabic lexicon to detect idioms in Arabic language. Further- more, they develop lexicon to automatically detect polarity in Arabic reviews. Experimental results show the SVM classifier (95.12%), which outperforms Naive Bayes classifier (83.64%).

Al-Moslmi et al. [11] have introduced an Arabic sentilexicon for generating feature vectors and developed multi- domain Arabic corpus. They have evaluated the performance of the approach by using four different machine learning classifier. Experimental results show that SVM classifier (69%) outperforms Nave Bayes (53.87%), KNN (41.87%), and Logistic regression classifiers (45.87%). Al-Smadi et al. [13] have proposed an approach to detect polarity in Arabic hotel reviews. The long short-term memory (LSTM) and neural network are used to evaluate the performance of the approach. The hotel reviews are converted into character embedding. Experimental results have showed the proposed approach (82.7%) outperforms to CNN (76.4%). Ahmad et al. [10] have proposed a framework to detect polarity in the financial news in Arabic and Chinese. In order to evaluate the performance of the framework, the corpus is built, which consists of 8000 financial headline news. Finally, the data has been analyzed using grid-based analysis. The main disadvantage of the proposed framework is lack of machine learning classification.

2.2 Various languages

Sharma et al. [50] have proposed an approach for visualization of Twitter messages. Online users are able to measure the sentiment. The main motivation is to build this application by providing an automated platform, which serves as complete end-to-end system for sentiment analysis of Twitter messages along with their visualization. Saif et al. [49] have developed a lexicon to detect polarity for sentiment analysis on Twitter. The proposed approach detects sentiment in entity-level and tweet-level. It has been evaluated using three different Twit- ter datasets. Experimental results show the proposed lexicon (81.03%) outperforms as compared to MPQA (63.79%) and SentiStrength (62.07%). However, the proposed lexicon only

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

covers English tweets, and its not able to detect polarity in different languages such as Arabic.

Dashtipour et al. [19] have proposed a lexicon to detect polarity in Persian sentences. The lexicon consists of 1500 words along with their parts-of-speech tags and their polarity. The Persian headline news are used to evaluate the performance of the approach. The experimental results show the SVM classifier (67.03%) outperforms Naive Bayes classifier (61.2%). Grbner et al. [26] have proposed an approach to detect polarity in English customer reviews. The approach consists of three steps such as (1) building a lexicon, (2) apply lexicon to customer reviews and (3) apply machine learning to evaluate the performance of the approach. Experimental result shows the proposed approach has achieved accuracy of 90%.

Li et al. [39] have proposed an approach to detect IKEA post. Similarly, they have developed a lexicon to detect polarity of the IKEA posts and evaluated it by using machine learning classifiers such as Logistic regression, SVM, Random Forest and neural network, Nave Bayes and elastic Net. The experimental results show the elastic Net classifier (80.04%) out- performs to logistic regression (70.35%), Nave Bayes (70%), SVM (70.65%), neural network (70.65%), and random forest (70.51%) classifiers. Wu et al. [55] have proposed a framework to predict emoji in English tweets. They have used a residual CNN-LSTM to evaluate the performance of the approach. The English tweets are converted to word embedding. This embedding is then fed into deep learning classifiers, which has resulted in an accuracy of 82.9%.

Basile et al. [15] have proposed an approach to detect polarity in Italian tweets, and they have developed the first Italian tweet corpus. In order to collect Italian corpus, they have used language detection to collect only Italian tweets. The SVM has used to evaluate the performance of the approach which has resulted an accuracy of 66.4%. Kaya et al. [37] have proposed a framework to detect polarity in Turkish political news. The proposed technique has been evaluated by using four different machine learning algorithms such as Naive Bayes, Maximum entropy, SVM and character-based N-gram language model. Experimental results show SVM classifier (76.31%) outperforms Naive Bayes (72.05%), character N- gram (73.93%), and Maximum entropy classifiers (75.85%). Vural et al. [54] have presented a framework to detect polarity in Turkish text documents. The Turkish movie reviews are used to evaluate the performance of the approach, and the SentiStregth (English lexicon) is translated into Turkish to assign polarity into the dataset. Experimental results show the proposed framework has achieved an accuracy of 78.4%.

Ghorbel et al. [23] have proposed an approach to detect polarity in French movie reviews. The French movie reviews are collected online, which consist of 1000 positive and 1000 negative labels. Unigram and parts-of-speech features such as adjective, adverb, verb and noun are extracted to evaluate the performance of the approach. Experimental results show the unigram has achieved accuracy of 93%. Kap et al. [36] have proposed an approach to detect polarity Lithuanian text. The Internet comments in Lithuanian are collected, converted into word embedding using FastText, and classifiers such as Nave Bayes, SVM, CNN and LSTM are used to evaluate the performance of the approach. Experimental results show CNN classifier (87%) has achieved better accuracy as compared to SVM (61%), Nave Bayes (54%) and LSTM classifiers (73%). Rezaeinia et al. [48] have proposed an approach to detect polarity in movie reviews. These reviews are converted into word and parts-of-speech embedding. Afterwards, deep learning classifier such as CNN has been used to evaluate the performance of the approach which resulted an accuracy of 87%.

Wu et al. [55] have proposed a semi-supervised approach using variational autoencoder to assign polarity to unlabeled datasets. The multi-domain datasets (car, book, laptop, news) are used to evaluate the performance of this approach. Experimental results show the proposed approach work effectively to detect polarity in unlabeled datasets.

However, none of studies have explored a sentiment analysis approach for general Arabic and Jordanian dialect datasets. Therefore, in this paper, we propose a novel approach to detect sentiment polarity in general Arabic and Jordanian dialects.

Table I summarizes some of the best approaches, which have been used for different languages and their reported accuracies.

3. METHODOLOGY

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In this section, the proposed approach for Arabic tweets sentiment analysis framework is discussed in detail. Fig 1 depicts the proposed framework and its details are presented in subsequent sections. The dataset is preprocessed using tokenization and normalization techniques. Afterwards, ngram features are extracted from the sentences, and finally machine learning classifiers are used to evaluate the performance of the approach.

Pre-processing: The Arabic tweets dataset is used in order to evaluate the performance of the proposed framework. Dataset is distributed into training set (60%), testing set (30%), and validation set (10%) to apply machine learning classifiers. The dataset consists of positive and negative labels. The tweets are pre-processed using tokenization and normalization techniques. The NLTK tokenizer is used to tokenize Arabic tweets. For example, a text ”Its a good mobile” is converted into words such as ”Its”, ”a”, ”good”, and ”mobile”. Afterwards, normalization techniques are used to normalize the tweets. For example, a tweet ”It is gr8 time” is converted into ”It is great time”. [41] [56] [38].

N-gram: N-gram is a sequence of consecutive words in a text [22] [33]. For example, الفيلم كان جيدا , the unigram for sentence is الفيلم, كان, جيدا. In the proposed framework, we have used unigram, bigram and trigram.

Ensemble Classifier Averaging: In this method, we take an average prediction of all the classifiers which is used to make the final prediction for accuracy, precision, recall and F-measure.

$$Ensemble(AVG) = \sum \frac{1}{n}(a1 + a2 + a3 + ... + an) \quad (1)$$

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of our proposed approach, the tweets are converted into Bag Of Words (BOW). These BOW are fed into different machine learning algorithms such as SVM, Naive Bayes, Linear SVM, and RBF, to evaluate the performance of the approach. We have used different features such as unigram, bigram, trigram, mixture of unigram and bigram, mixture of bigram and trigram, and mixture of unigram and trigram to evaluate the performance of the dataset. Experimental results show the unigram achieved better accuracy as compared to other features such as Bigram, and Trigram.
In addition, the Naive Bayes classifier has received lower accuracy as compared to other algorithms.

ASTD: The Arabic tweets dataset consists of 1000 positive and 1000 negative tweets. These tweets are manually annotated into positive and negative labels by applying the Mechanical Turk approach. [43].

ATA: The dataset consists of 1000 positive and 1000 negative tweets, which are of various topics such as politics and arts. These tweets are in MSA and Jordanian dialects [6].

Results: The results for ASTD dataset are summarized into Table II and Table III. Among n-gram features, classifier trained on unigram has achieved better performance in terms of accuracy.

The experimental results show the effectiveness of the unigram feature using MLP classifier. In order to evaluate the performance of the approach, the following evaluation metrics are used to calculate precision, recall, f-measure and accuracy.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F\_measure = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where TP represents true positive, TN represents true negative, FP represents false positive, and FN represents false negative respectively.

The unigram has performed better in comparison with bigram and trigram, because the bigram and trigram features contain lots of noise, which affects the performance of the classifier. Fig 2 summarizes

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the accuracy of the experiments for ASTD dataset and Fig 3 summarizes the accuracy of the experiments for Arabic tweets dataset.

Table IV and Table V presents the Ensemble classifier and their results achieved on ASTD dataset and ATA dataset.

Table VI reports the comparison of the proposed framework for Arabic sentiment analysis with four recently proposed framework in the literature. In order to compare the results, we have used their dataset to evaluate the performance of our approach. Nabil et al. [43] have extracted unigram from ASTD dataset. TF-IDF has been applied into unigram and SVM is used to evaluate the performance of their approach. Rabab'ah et al. [46] have developed a lexicon and ASTD dataset is used to evaluate the performance of the proposed lexicon. Al-Thubaity et al. [14] have developed a lexicon for Saudi dialects, and they have used ASTD dataset to evaluate the performance of their lexicon. Table VII reports the comparison of proposed framework for Arabic sentiment analysis (ATA dataset) with recently proposed framework in the literature. In order to compare the results, we have used their dataset to evaluate the performance of our approach. Abooraig et al. [8] have developed a lexicon. Arabic tweets are used to evaluate the performance of the approach. Al Shboul et al. [12] have applied pre-processing technique known as stemming and used Naive Bayes classifier to evaluate the performance of the approach. Scikit learn (Python package) is used to train machine learning classifiers. The following experimental setup is used for our experiments:

Table 1: State-Of-The-Art Approaches And Achieved Accuracy

Ref	Purpose	Language	Accuracy
Hassan et al. [29]	Develop Arabic lexicon	Arabic	74%
Mostafa et al. [42]	Detecting polarity	Arabic	72.30%
Donia et al. [21]	Detect polarity in Arabic tweets	Arabic	72.83%
Zheng et al. [57]	Detect polarity in Chinese online reviews	Chinese	74.42%
Liu et al. [40]	Detect polarity in customer reviews	Chinese	87.50%
Jhanwar et al. [35]	Detect polarity in Facebook reviews	Hindi English	70.80%
Van et al. [53]	Irony detection in English	English	65.90%
Rauh et al. [47]	Detect polarity in German politic comments	German	72.30%
Nakayama et al. [44]	Detect polarity in Japanese reviews	Japanese	76%

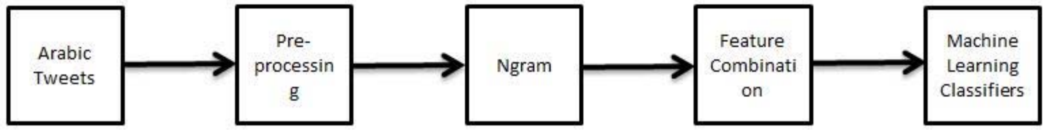

Figure 1: Proposed Framework

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: results on astd dataset

Classifier	Feature	Precision	Recall	F-measure	Accuracy (%)	Feature	Precision	Recall	F-measure	Accuracy
MLP	Unigram	0.74	0.75	0.74	75.47	Unigram + Bigram	0.75	0.75	0.72	75
LR	Unigram	0.74	0.75	0.72	74.76	Unigram + Bigram	0.74	0.74	0.7	74.04
Linear SVM	Unigram	0.77	0.69	0.57	72.43	Unigram + Bigram	0.75	0.7	0.59	69.68
RBF SVM	Unigram	0.78	0.68	0.55	68.01	Unigram + Bigram	0.78	0.68	0.55	68.01
NB	Unigram	0.47	0.69	0.56	68.05	Unigram + Bigram	0.63	0.59	0.6	66.92
MLP	Bigram	0.68	0.69	0.59	68.96	Unigram + Trigram	0.74	0.74	0.69	73.73
LR	Bigram	0.72	0.69	0.58	68.88	Unigram + Trigram	0.74	0.73	0.68	73.41
Linear SVM	Bigram	0.46	0.68	0.55	68.14	Unigram + Trigram	0.74	0.7	0.6	70
RBF SVM	Bigram	0.62	0.68	0.56	67.93	Unigram + Trigram	0.78	0.68	0.55	68.01
NB	Bigram	0.78	0.68	0.55	68.01	Unigram + Trigram	0.68	0.53	0.59	67.32
MLP	Trigram	0.69	0.68	0.57	68.41	Bigram + Trigram	0.69	0.69	0.58	68.8
LR	Trigram	0.73	0.68	0.57	68.49	Bigram + Trigram	0.73	0.69	0.57	68.73
Linear SVM	Trigram	0.46	0.68	0.55	67.93	Bigram + Trigram	0.78	0.68	0.55	68.01
RBF SVM	Trigram	0.65	0.68	0.58	68.33	Bigram + Trigram	0.62	0.68	0.56	67.93
NB	Trigram	0.46	0.68	0.55	67.93	Bigram + Trigram	0.64	0.51	0.56	67.54

Table 3: result on ata dataset

Classifier	Feature	Precision	Recall	F-measure	Accuracy (%)	Feature	Precision	Recall	F-measure	Accuracy (%)
MLP	Unigram	0.82	0.82	0.82	82.17	Unigram + Bigram	0.82	0.81	0.81	81.16
LR	Unigram	0.81	0.81	0.81	80.66	Unigram + Bigram	0.81	0.81	0.81	80.56
Linear SVM	Unigram	0.75	0.72	0.72	72.4	Unigram + Bigram	0.75	0.75	0.72	72.8
RBF SVM	Unigram	0.68	0.54	0.42	53.67	Unigram + Bigram	0.65	0.53	0.4	52.87
NB	Unigram	0.82	0.81	0.81	81.47	Unigram + Bigram	0.25	0.5	0.34	50.35
MLP	Bigram	0.74	0.66	0.63	65.86	Unigram + Trigram	0.81	0.8	0.8	80.46
LR	Bigram	0.75	0.68	0.66	68.47	Unigram + Trigram	0.8	0.8	0.8	79.65
Linear SVM	Bigram	0.75	0.52	0.37	55.24	Unigram + Trigram	0.75	0.74	0.73	73.51
RBF SVM	Bigram	0.59	0.54	0.47	54.27	Unigram + Trigram	0.65	0.53	0.4	52.87
NB	Bigram	0.75	0.52	0.37	51.56	Unigram + Trigram	0.75	0.5	0.34	49.94
MLP	Trigram	0.73	0.55	0.44	54.88	Bigram + Trigram	0.75	0.65	0.61	64.75
LR	Trigram	0.74	0.53	0.39	52.97	Bigram + Trigram	0.74	0.67	0.64	66.76
Linear SVM	Trigram	0.75	0.51	0.36	51.46	Bigram + Trigram	0.75	0.52	0.37	51.76
RBF SVM	Trigram	0.58	0.56	0.53	55.79	Bigram + Trigram	0.58	0.54	0.47	54.17
NB	Trigram	0.75	0.5	0.33	49.74	Bigram + Trigram	0.75	0.51	0.35	50.55

Journal of Theoretical and Applied Information Technology
15th September 2019. Vol.97. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

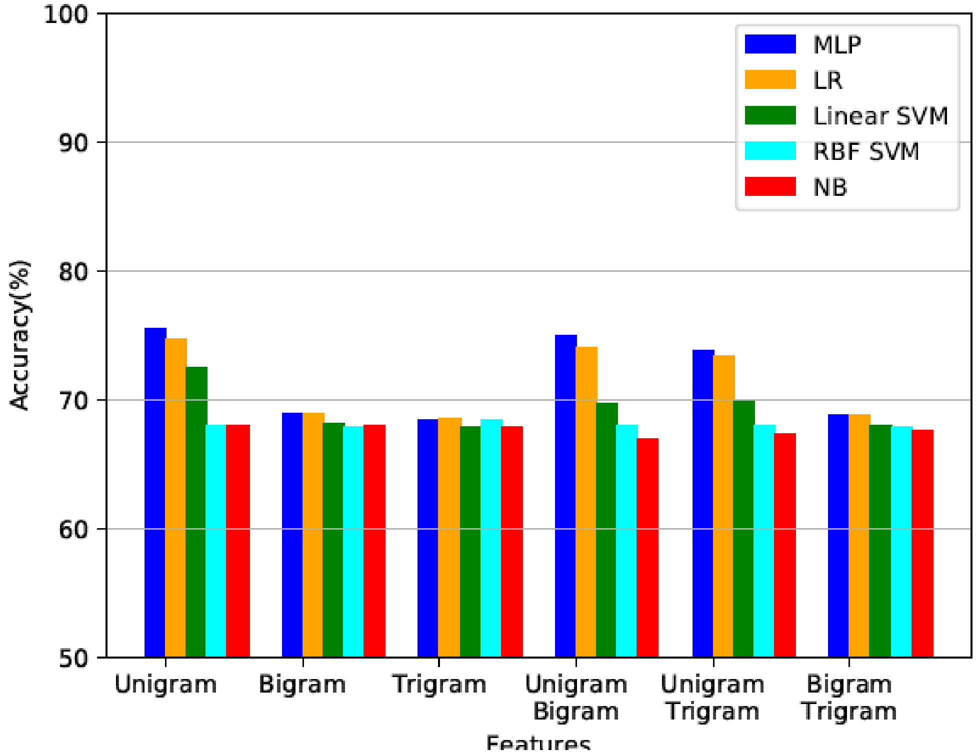

Figure 2: Feature Combination ASTD Dataset

Table 4: Ensemble Classifier Averaging Results On Astd Dataset

Feature	Precision	Recall	F-measure	Accuracy
Unigram	0.7	0.71	0.62	71.74
Bigram	0.65	0.68	0.56	68.38
Trigram	0.59	0.68	0.56	68.21
Unigram + Bigram	0.73	0.69	0.63	70.73
Unigram + Trigram	0.73	0.67	0.62	70.49
Bigram + Trigram	0.69	0.65	0.56	68.2

Table 5: Ensemble Classifier Averaging Results On Ata Dataset

Feature	Precision	Recall	F-measure	Accuracy
Unigram	0.77	0.74	0.71	74.07
Bigram	0.71	0.58	0.5	59.08
Trigram	0.71	0.53	0.41	52.96
Unigram + Bigram	0.65	0.68	0.61	67.54
Unigram + Trigram	0.75	0.67	0.61	67.28
Bigram + Trigram	0.71	0.57	0.48	57.59

Table 6: Accuracy Performance And Comparison With State Of The Art Models - Astd Dataset

References	Accuracy
Our proposed approach (ASTD Dataset)	75.47%
Nabil et al. [43]	68.07
Rabab'ah et al. [46]	62%
Al-Thubaity et al. [14]	65%

5. DISCUSSION

The main finding is that the sentiment analysis of Arabic tweets is important and it can improve the accuracy for detection of polarity in short sentences such as tweets.
The main advantage of the proposed approach is useful to detect polarity in Arabic tweets. As mentioned earlier, Arabic language consists of different dialects such as Arabizi or Sudanese [31]. However, the proposed approach is able detect polarity in modern standard Arabic and Jordanian dialect. In addition, the proposed approach uses traditional machine learning classifiers such as MLP, SVM, and NB. However, deep learning classifiers such as Convolutional neural network (CNN) and Long short-term memory (LSTM) have not been used. The extraction of features such as unigram, bigram and trigram can be improved. Furthermore, feature engineering such as part-of-speech tag features such as adjective, adverb, verb and noun can be used in the proposed approach. Finally, we have only evaluated the proposed framework for Arabic tweets. In the future, product and movie reviews in Arabic Language can also be used for comparative evaluation.

6. CONCLUSION

In this paper, a sentiment analysis framework has been proposed for Arabic twitter analysis. We have presented a comparative study analysis based on ngram features combination. The performance of the proposed framework has been evaluated by combining different features using two benchmark Arabic tweets datasets, and different machine learning classifiers. Comparative results show that the unigram features and MLP have achieved the best performance.

In future, we intend to apply deep learning classifiers to Arabic tweets and other benchmark datasets, and compare the results with other state of the art approaches..

REFERENCES:

[1] A dynamic partitioning algorithm for sip detection using a bottle-attachable imu sensor. International Journal of Advanced Computer Science and Applications.
[2] Learning deep transferability for several agricultural classification problems. International Journal of Advanced Computer Science and Applications.
[3] Linking context to data warehouse design. International Journal of Advanced Computer Science and Applications.
[4] Novel abcd formula to diagnose and feature ranking of melanoma. International Journal of Advanced Computer Science and Applications.
[5] Muhammad Abdul-Mageed, Mona Diab, and Sandra Kubler¨. Samar: Subjectivity and sentiment analysis for arabic social media. Computer Speech & Language, 28(1):20–37, 2014.
[6] Nawaf Abdulla, N Mahyoub, M Shehab, and M Al-Ayyoub. Arabic sentiment analysis: Corpus-based and lexicon-based. In Proceedings of The IEEE conference on Applied Electrical Engineering and Computing Technologies (AEECT), 2013.
[7] Nawaf A Abdulla, Mahmoud Al-Ayyoub, and Mohammed Naji Al-Kabi. An extended analytical study of arabic sentiments. International Journal of Big Data Intelligence 1, 1(1-2):103–113, 2014.
[8] Raddad Abooraig, Shadi Al-Zu'bi, Tarek Kanan, Bilal Hawashin, Mah-moud Al Ayoub, and Ismail Hmeidi. Automatic categorization of arabic articles based on their political orientation. Digital Investigation, 2018.
[9] Ahsan Adeel, Mandar Gogate, Saadullah Farooq, Cosimo Ieracitano, Kia Dashtipour, Hadi Larijani, and Amir Hussain. A survey on the role of wireless sensor networks and iot in disaster management. In Geological Disaster Monitoring Based on Sensor Networks, pages 57– 66. Springer, 2019.

[10] Khurshid Ahmad, David Cheng, and Yousif Almas. Multi-lingual sentiment analysis of financial news streams. In 1st International Workshop on Grid Technology for Financial Modeling and Simulation, volume 26, page 001. SISSA Medialab, 2007.

[11] Tareq Al-Moslmi, Mohammed Albared, Adel Al-Shabi, Nazlia Omar, and Salwani Abdullah. Arabic senti-lexicon: Constructing publicly available language resources for arabic sentiment analysis. Journal of Information Science, 44(3):345–362, 2018.

[12] Bashar Al Shboul, Mahmoud Al-Ayyoub, and Yaser Jararweh. Multi-way sentiment classification of arabic reviews. In 2015 6th International Conference on Information and Communication Systems (ICICS), pages 206–211. IEEE, 2015.

[13] Mohammad Al-Smadi, Bashar Talafha, Mahmoud Al-Ayyoub, and Yaser Jararweh. Using long short-term memory deep neural networks for aspect-based sentiment analysis of arabic reviews. International Journal of Machine Learning and Cybernetics, pages 1–13, 2018.

[14] Abdulmohsen Al-Thubaity, Mohammed Alharbi, Saif Alqahtani, and Abdulrahman Aljandal. A saudi dialect twitter corpus for sentiment and emotion analysis. In 2018 21st Saudi Computer Society National Computer Conference (NCC), pages 1–6. IEEE, 2018.

[15] Valerio Basile et al. Sentiment analysis on italian tweets. In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 100–107, 2013.

[16] Kia Dashtipour, Mandar Gogate, Ahsan Adeel, Amir Hussain, Ab-dulrahman Alqarafi, and Tariq Durrani. A comparative study of persian sentiment analysis based on different feature combinations. In International Conference in Communications, Signal Processing, and Systems, pages 2288–2294. Springer, 2017.

[17] Kia Dashtipour, Mandar Gogate, Ahsan Adeel, Cosimo Ieracitano, Hadi Larijani, and Amir Hussain. Exploiting deep learning for persian sentiment analysis. In International Conference on Brain Inspired Cognitive Systems, pages 597–604. Springer, 2018.

[18] Kia Dashtipour, Amir Hussain, and Alexander Gelbukh. Adaptation of sentiment analysis techniques to persian language. In International Conference on Computational Linguistics and Intelligent Text Process-ing, pages 129–140. Springer, 2017.

[19] Kia Dashtipour, Amir Hussain, Qiang Zhou, Alexander Gelbukh, Ah-mad YA Hawalah, and Erik Cambria. Persent: a freely available Persian sentiment lexicon. In International Conference on Brain Inspired Cognitive Systems, pages 310–320. Springer, 2016.

[20] Kia Dashtipour, Soujanya Poria, Amir Hussain, Erik Cambria, Ah-mad YA Hawalah, Alexander Gelbukh, and Qiang Zhou. Multilingual sentiment analysis: state of the art and independent comparison of techniques. Cognitive computation, 8(4):757–771, 2016.

[21] Marco Alfonse Donia Gamal, El-Sayed M El-Horbaty, and Abdel-Badeeh M Salem. Twitter benchmark dataset for arabic sentiment analysis. 2019.

[22] Amol S Gaikwad. Twitter sentiment analysis approaches: A survey. 2019.

[23] Hatem Ghorbel and David Jacot. Sentiment analysis of french movie reviews. In Advances in Distributed Agent-Based Retrieval Tools, pages 97–108. Springer, 2011.

[24] Mandar Gogate, Ahsan Adeel, and Amir Hussain. Deep learning driven multimodal fusion for automated deception detection. In Computational Intelligence (SSCI), 2017 IEEE Symposium Series on, pages 1–6. IEEE, 2017.

[25] Mandar Gogate, Ahsan Adeel, Ricard Marxer, Jon Barker, and Amir Hussain. Dnn driven speaker independent audio-visual mask estimation for speech separation. arXiv preprint arXiv:1808.00060, 2018.

[26] Dietmar Grabner,¨ Markus Zanker, Gunther Fliedl, Matthias Fuchs, et al. Classification of customer reviews based on sentiment analysis. Citeseer, 2012.

[27] Imane GUELLIL, Ahsan Adeel, Faical AZOUAOU, Ala-eddine Hachani, Amir Hussain, et al. Arabizi sentiment analysis based on transliteration and automatic corpus annotation. In Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 335–341, 2018.

[28] Imane Guellil, Faical Azouaou, Houda Saadane,ˆ and Nasredine Semmar. Une approche fondee´ sur les lexiques danalyse de sentiments du dialecte algerien´.

[29] Hend Hassan, Hitham M Abo Bakr, and Ibrahim E Ziedan. A framework for arabic

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

concept-level sentiment analysis using senticnet. International Journal of Electrical and Computer Engineering (IJECE), 8(6), 2018.

[30] Ahmad Hawalah and Maria Fasli. Utilizing contextual ontological user profiles for personalized recommendations. Expert Systems with Applications, 41(10):4777–4797, 2014.

[31] Intisar O Hussien, Kia Dashtipour, and Amir Hussain. Comparison of sentiment analysis approaches using modern arabic and sudanese dialect. In International Conference on Brain Inspired Cognitive Systems, pages 615–624. Springer, 2018.

[32] Hossam S Ibrahim, Sherif M Abdou, and Mervat Gheith. Sentiment analysis for modern standard arabic and colloquial. arXiv preprint arXiv:1505.03105, 2015.

[33] Cosimo Ieracitano, Ahsan Adeel, Mandar Gogate, Kia Dashtipour, Francesco Carlo Morabito, Hadi Larijani, Ali Raza, and Amir Hussain. Statistical analysis driven optimized deep learning system for intrusion detection. In International Conference on Brain Inspired Cognitive Systems, pages 759–769. Springer, 2018.

[34] Cosimo Ieracitano, Nadia Mammone, Alessia Bramanti, Amir Hussain, and Francesco C Morabito. A convolutional neural network approach for classification of dementia stages based on 2d-spectral representation of eeg recordings. Neurocomputing, 323:96–107, 2019.

[35] Madan Gopal Jhanwar and Arpita Das. An ensemble model for sentiment analysis of hindi-english code-mixed data. arXiv preprint arXiv:1806.04450, 2018.

[36] Jurgita Kapociˇut¯e˙-Dzikiene,˙ Robertas Damasevi˘cius,ˇ and Marcin Wozniak´. Sentiment analysis of lithuanian texts using traditional and deep learning approaches. Computers, 8(1):4, 2019.

[37] Mesut Kaya, Guven Fidan, and Ismail H Toroslu. Sentiment analysis of turkish political news. In Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01, pages 174–180. IEEE Computer Society, 2012.

[38] Aparup Khatua, Apalak Khatua, and Erik Cambria. A tale of two epidemics: Contextual word2vec for classifying twitter streams during outbreaks. Information Processing & Management, 56(1):247–257, 2019.

[39] Yujiao Li and Hasan Fleyeh. Twitter sentiment analysis of new ikea stores using machine learning. In 2018 International Conference on Computer and Applications (ICCA), pages 4–11. IEEE, 2018.

[40] Pengfei Liu, Ji Zhang, Cane Wing-Ki Leung, Chao He, and Thomas L Griffiths. Exploiting effective representations for chinese sentiment analysis using a multi-channel convolutional neural network. arXiv preprint arXiv:1808.02961, 2018.

[41] Navonil Majumder, Soujanya Poria, Haiyun Peng, Niyati Chhaya, Erik Cambria, and Alexander Gelbukh. Sentiment and sarcasm classification with multitask learning. arXiv preprint arXiv:1901.08014, 2019.

[42] Mohamed M Mostafa. Clustering halal food consumers: A twitter sentiment analysis. International Journal of Market Research, page 1470785318771451, 2018.

[43] Mahmoud Nabil, Mohamed Aly, and Amir Atiya. Astd: Arabic sentiment tweets dataset. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 2515–2519, 2015.

[44] Makoto Nakayama and Yun Wan. Is culture of origin associated with more expressions? an analysis of yelp reviews on japanese restaurants.
Tourism Management, 66:329–338, 2018.

[45] Jonathan Ortigosa-Hernandez,´ Juan Diego Rodr´ıguez, Leandro Alzate, Manuel Lucania, Inaki˜ Inza, and Jose A Lozano. Approaching senti-ment analysis by using semi-supervised learning of multi-dimensional classifiers. Neurocomputing, 92:98–115, 2012.

[46] Abdullateef M Rabab'ah, Mahmoud Al-Ayyoub, Yaser Jararweh, and Mohammed N Al-Kabi. Evaluating sentistrength for arabic sentiment analysis. In Computer Science and Information Technology (CSIT), 2016 7th International Conference on, pages 1–6. IEEE, 2016.

[47] Christian Rauh. Validating a sentiment dictionary for german political languagea workbench note. Journal of Information Technology & Politics, 15(4):319–343, 2018.

[48] Seyed Mahdi Rezaeinia, Rouhollah Rahmani, Ali Ghodsi, and Hadi Veisi. Sentiment analysis based on improved pre-trained word embed-dings. Expert Systems with Applications, 117:139–147, 2019.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[49] Hassan Saif, Yulan He, Miriam Fernandez, and Harith Alani. Contextual semantics for sentiment analysis of twitter. Information Processing & Management, 52(1):5–19, 2016.

[50] Nitesh Sharma, Rachit Pabreja, Ussama Yaqub, Vijayalakshmi Atluri, Soon Chun, and Jaideep Vaidya. Web-based application for sentiment analysis of live tweets. In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, page 120. ACM, 2018.

[51] Amira Shoukry and Ahmed Rafea. Sentence-level arabic sentiment analysis. In Collaboration Technologies and Systems (CTS), 2012 International Conference on, pages 546–550. IEEE, 2012.

[52] Xiao Sun, Chengcheng Li, and Fuji Ren. Sentiment analysis for chinese microblog based on deep neural networks with convolutional extension features. Neurocomputing, 210:227–236, 2016.

[53] Cynthia Van Hee, Els Lefever, and Veronique´ Hoste. Semeval-2018 task 3: Irony detection in english tweets. In Proceedings of The 12th International Workshop on Semantic Evaluation, pages 39–50, 2018.

[54] A Gural Vural, B Barla Cambazoglu, Pinar Senkul, and Z Ozge Tokgoz. A framework for sentiment analysis in turkish: Application to polarity detection of movie reviews in turkish. In Computer and Information Sciences III, pages 437–445. Springer, 2013.

[55] Chuhan Wu, Fangzhao Wu, Sixing Wu, Zhigang Yuan, Junxin Liu, and Yongfeng Huang. Semi-supervised dimensional sentiment analysis with variational autoencoder. Knowledge-Based Systems, 165:30–39, 2019.

[56] Frank Z Xing, Erik Cambria, and Roy E Welsch. Growing semantic vines for robust asset allocation. Knowledge-Based Systems, 165:297– 305, 2019.

[57] Lijuan Zheng, Hongwei Wang, and Song Gao. Sentimental feature selection for sentiment analysis of chinese online reviews. International journal of machine learning and cybernetics, 9(1):75–84, 2018.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

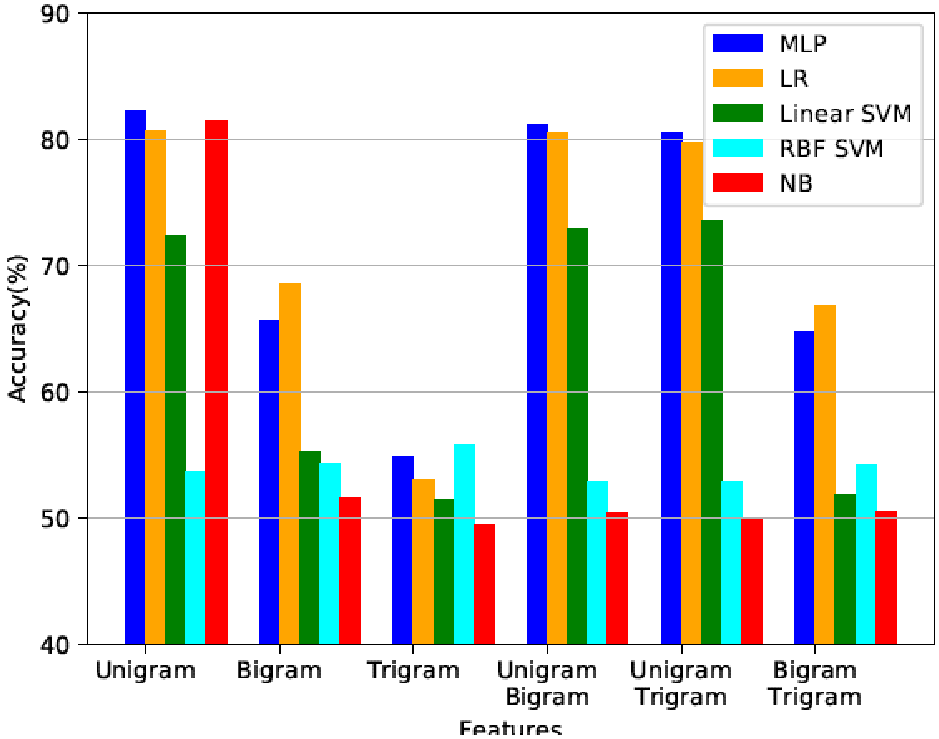

Figure 3: Feature Combination ATA Dataset

Table 7: Accuracy Performance And Comparison With State Of The Art Models - Ata Dataset

References	Accuracy
Our proposed approach	82.17%
(ATA Dataset)	
Al Shboul et al. [12]	64%
Abooraig et al. [8]	76.07%