Towards Developing Medical Sentiment Analysis Model

Moayyad Al-Bohnayyah
Department of Arabic Language
College of Arts
King Faisal University
Al-Ahsa, 31982, Saudi Arabia
malbohnayyah@kfu.edu.sa

Heider A. M. Wahsheh
Department of Information Systems
College of Computer Science and Information Technology
King Faisal University
Al-Ahsa, 31982, Saudi Arabia
Correspondence should be addressed to
hwahsheh@kfu.edu.sa;
heiderwahsheh@gmail.com

ABSTRACT

With the continued growth of curiosity in investigating public opinions on news, products, events, etc., many issues appear when executing the automatic procedure. Moreover, every language can have special characteristics that make executing techniques in other languages inaccurate, where the accuracy of final results is uncertain. Sentiment Analysis in Natural Language Processing (NLP) means utilizing computational linguistics methods to automatically extract, predict, and label the content polarity. It analyzes the reviews on social networks, interviews, phone calls, and blogs for products or services. Sentiment Analysis can be employed in several fields, such as marketing, decision-making, training, and medical systems. However, there needs to be more research recognizing these fields for the Arabic language. This study aims to develop a medical context sentiment extraction Artificial Intelligence model, whether they are written in Modern Standard Arabic (MSA), one of the Arabic Dialects, and/or Emoticons. A sizeable data collection of Arabic content was gathered from several social network websites (i.e., Facebook and Twitter), where the features were extracted and weighted. The adoption of both lexicon-based and support vector machine algorithms was employed to conduct sentimental analysis. SVM validates the results with more than 97% within Receiver Operating Characteristic (ROC) prediction quality.

Keywords: Big data, Social network, Sentiment Analysis, computational linguistics, Artificial Intelligence, opinion mining.
DOI Number: 10.48047/NQ.2022.20.17.NQ88032 **NeuroQuantology2022;20(17): 238-244**

1. INTRODUCTION AND BACKGROUND

The possibility to capture the general public's opinions has expanded, boosting interest within the scientific society, showing many exhilarating open challenges, and in the enterprise globe due to the incredible range of advantages envisaged, including commerce, business intelligence, and financial forecast [1]. Natural language processing (NLP) has been utilized to handle a wide range of issues, such as enhance for search engines and summarizing and classifying text. NLP is employed to improve the utility and administration of applications. It makes user information more comfortable and transforms text into more functional structures [2].

Recent Saudi Arabia social media statistics show that Saudi Arabia users generate billions of daily interactions and opinionated data about various topics. This wealth of narrow Arabic data includes textual content among user accounts to express their sentiments and thoughts about different issues [3, 4].

Sentiment analysis (SA) or opinion mining (OM) automatically analyzes data concerning different aspects of life expressed by reviews and comments posted by social media users [5]. Nowadays, several organizations investigate these posts to improve the provided quality of their services and products [6]. Consequently, there is no longer demanded to perform surveys or opinion polls by those parties. SA is a vital technique for making appropriate decisions for enterprises, governments, educational institutes, etc. For example, customers usually read other consumers' comments regarding a particular item before determining whether to buy it or search for alternatives [7].SA is not a direct task since the meaning of textual content is not permanently specified, but it relies on the discipline it is related to it. For example, I got a negative test result in the medical domain, which means good health (positive) status, while in education, it is considered a negative result [8].

Patients have powerful sentiments regarding the healthcare services they receive. Each interaction with healthcare centers will begin a positive or negative response. So sentiment analysis in healthcare is precious since the understandings extracted from a study of patient sentiments authorize healthcare providers to facilitate communication between healthcare centers and patients [9, 10].Multiple advanced Arabic sentiment analysis prototypes and models are available in the literature [11-13], but it receives little attention since only some are available as APIs services [14].The medical content is not a homogeneous domain; medical records are documented differently from scientific papers, series annotations, or public health procedures. Moreover, local dialects are not unique. For example, healthcare centers and laboratories use various terms with the same purpose. A client health information system must be capable of understanding both expert and non-experts medical terminology and mapping between them [1].In medical societies, health-related subjects and medical topics are compulsory for healthcare centers, doctors, and patients. The data availability via various media can revise the study behavior of decision-making. Several studies have emphasized the medical sentiments of clinical tests and social media texts to highlight patient's reviews in medical environments [15-18]. The study of [19] proposed a prediction technique that analyzes health points over social networks, where the results obtained an accuracy of 98%.

A typical model in the analysis of medical user sentiment relies on social networks was produced by [20]. The model utilized Latent Dirichlet Allocation (LDA) within a weighted scheme and it was capable of handling various interference with a patient medical disease, prescription, and treatment.

The study of [21] adopted Twitter data collection, marked user opinions as positive, negative, and neutral, and obtained 82% precision. The correlation between symptoms and diagnosis results among imbalanced datasets yielded 96%. The Naïve Bayes validates sentiment analysis with an accuracy of 80%.

The study of [22] used multiple text mining and machine learning methods to conduct sentiment analysis on a dispersed framework and created a model for easier and faster investigation. The adopted technique used a long short-term memory (LSTM) neural network as an alternative to conventional sentiment analysis for investigating considerable datasets' flow. The significance of the adopted procedures was assessed on several data collection. The results demonstrated that the adopted technique achieved better than the other methods in accuracy and implementation time metrics.

This study aims to build medical context sentiment extraction Artificial Intelligence model for Arabic language, whether they are written in Modern Standard Arabic (MSA), one of the Arabic Dialects, and/or Emoticons. Data collection of Arabic content will be gathered from several social network websites (i.e., Facebook and Twitter), where the features will be extracted and weighted. The adoption of both lexicon-based and support vector machine algorithms was employed to conduct sentimental analysis. The SVM Receiver Operating Characteristic (ROC) prediction quality metrics are utilized to assess the effectiveness of the proposed model.

The rest of this paper is organized as it follows: Section 2 explores the adopted methodology to accomplish this study; including the data collection and feature assigned weights and extractions.Section 3 shows the experiments and results. Section 4 presents the conclusion and future work.

2. METHODOLOGY

The proposed framework adopted in this study is to medical context sentiment extraction Artificial Intelligence model for the Modern Standard Arabic (MSA), Dialects, and/or Emoticons. It consists of the following main steps:

1. Compose a data collection of Arabic textual reviews that uses Modern Standard Arabic (MSA) and Dialects. These reviews were gathered automatically by the crawler of [14] from the following social networks (i.e., Facebook, Blogs, YouTube, and Twitter).
2. Build Arabic polarity (positive, negative, and neutral) lexicons for the Arabic text and emoticons.
3. Set various weights for the words/phrases/terms in the lexicons depending on their semantic meaning. This phase is performed by the authors, where the weights are decided based on their judgments.
4. Evaluate the adopted approach using Receiver Operating Characteristic (ROC) prediction quality measurements.

• DATA COLLECTION

The data collection utilized in sentiment analysis is either gathered manually or automatically [23], where there is a significant lack of availability of large tested datasets that can be used as a benchmark for the Arabic language. The collected data employed in this study is collected automatically by a specific crawler [14] from four main social network resources (i.e., Facebook, Blogs, YouTube, and Twitter).

The crawler employed specific keywords related to the medical domain.The total number of positive, negative, and neutral reviews was distributed equally to have a balanced dataset, with 10000 reviews for each polarity. The preprocessing phase differs significantly between Arabic NLP and other languages [14].

We conduct various preprocessing procedures as follows:

- Remove non-Arabic text, symbols, and punctuations.
- Normalize similar characters (i.e. (Alif, "أ، إ، ى، آ") to (Bare Alif, "ا"), (Taa', haa', "ة، ه") to (Haa', "ه"), (Yaa', "ي, ي ء") to (Yaa', "ي").
- Remove Arabic stop words.
- Tokenize Arabic text.

The percentage of MSA reviews is around 57%. Table 1 shows some examples of the medicals reviews.

Table 1. *A Sample from medical reviews.*

Arabic review	**English Translation**	**Labelled**
على الأقل لطمأنة الناس والثقة بأننا نسير في الاتجاه الصحيح وأن مستشفياتنا ، والحمد لله ، مؤهلة بشكل كاف لمواجهة حالات كوفيد.	At least to reassure people and trust that we are moving in the right direction and that our hospitals, praise be to God, are sufficiently qualified to face Covid cases.	Positive
نتيجة فحص كورونا ايجابية	The result of the Corona test (PCR) is positive.	Negative
الخدمة في المستشفى كانت سيئة لكن الطبيب كان متعاون	The service in the hospital was poor, but the doctor was helpful.	Neutral

• LEXICONS FEATURES AND WEIGHTS

This study built polarity lexicons that consist of emoticons and text. Table 2 presents the details for these lexicons.

Table 2. *Text and emotions lexicons information.*

Text-based lexicon includes textual emotions		
Positive	Negative	Neutral
5000	4000	1000

We assigned weights to distinct features based on the semantic meaning of the lexicon's characteristics. Table 3 shows a sample of these weights for the Arabic words/terms with their corresponding English translations.

Table 3. *Words and phrases* weights according to their semantic meaning

Lexicons features	English Translation	Assigned Weight	Polarity
أجسام مضادة	antibodies	1	Positive
جرعةثالثة	Third Dose	0.90	Positive

خطة رفع الحظر	Unlocking plan	0.75	Positive
ارتفاع الاصابات	High infections	1	Negative
اغلاق	Lockdown	0.85	Negative
آثار جانبية	Side effects	0.95	Negative
مسحة	Swab	1	Neutral

• FEATURE EXTRACTION METHODS

We developed an algorithm to determine the polarity for each review, which depends on the Term Frequency (*TF*) and the weights for the lexicons attributes. Figure 1 shows the proposed polarity algorithm.

Figure 1. The proposed medical polarity algorithm

Input:
MR: Medical review.
PL: Set of Positive lexicon.
NL: Set of Negative lexicon.
PW: Positive weights.
NW: Negative weights.
NeutW: Neutral weights..
Output:
MP: Medical polarity.
Initialization:
P_TF_W = 0, where *P_TF _*W is the *TF* for positive review and the assigned weights.
N_TF_W = 0, where *Neg_TF* is the *TF* for negative review and the assigned weights.
Neut_TF_W = 0, where *Neut_TF_W* is the *TF* for neutral review and the assigned weights.
Begin

```
Read MR
For every MR:
Remove punctuations from Arabic characters.
Remove stop words.
Normalize similar characters.
Divide MR into w/p word/phrase tokens.
For eachw/p, Search for similar w in PL, NL.
For each w/p, check PW, NW. NeutW
If w in PL then
                P_TF_W = P_TF + 1+ PW
                FP= Positive
        Else If w in NL then
                N_TF_W = N_TF + 1 +NW
                FP = Negative+NW
            Else
                Neut_TF_W = Neut_TF +1+ Neut_W
                FP= Neutral
            End If
End If
End For
If (P_TF_W> N_TF_W) then
        FP=Positive
Else If (N_TF_W> P_TF_W) then
        FP = Negative
Else
        FP = Neutral
End If
WriteMP to the final result file.
End For
```

End

For each textual review, we calculate TF plus the weights for the lexicons attributes, so if the TF_W of positive words/phrases is more significant than the TF_W of negative words/phrases in the review, then the medical polarity is recognized as a positive class labeled. While if the TF_W of negative words/phrases is more than the TF_W of positive words/phrases in the review, then the class label is determined as a negative. The neutral determination is considered if the TF_W of neutral words/phrases in the review is more than TF_W of positive or negative words/phrases, or when the TF_W of positive is equal to TF_W of negative words/phrases. In this paper, we considered every emoticon as a single feature; our lexicons convert the polarity for the words or phrases if the negation keywords [14] such as: (no, “لا (“and (not, "لم ") appeared in the text before them.

3. EXPERIMENTS AND RESULTS

Support Vector Machine (*SVM*) and Naïve Bayes (*NB*) machine learning classifiers are employed in this study to assess the adopted medical review approach.Support vector machine (SVM) is a supervised machine learning method that can be utilized for classification and regression. SVM is widely used in natural language processing fields such as sentiment analysis, spam prediction, and text categorization [24]. The Naïve Bayes (*NB*)model is a probabilistic supervised learning method that depends on the Bayes theorem and is employed for solving classification concerns. It is mostly utilized in text classification, spam filtration, and sentiment analysis. Naïve Bayes is one of the most straightforward and most useful classification techniques, which assists in creating fast machine learning models that can make quick forecasts [25].

We used the following measurement: True Positive (*TP*), True Negative (*TN*), False Positive (*FP*), False Negative (*FN*). Precision, Recall, Matthews Correlation Coefficient (*MCC*), precision recall curve (*PRC*) and F-Measure (*F-M*) are computed accordin to the following formulas.

$$Accuracy_i = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

$$Recall_i = \frac{TP}{TP+FN} \tag{2}$$

$$Precision_i = \frac{TP}{TP+FP} \tag{3}$$

$$F-measure = \frac{2TP}{2TP+FP+FN} \tag{4}$$

Matthews correlation coefficient (MCC) formula is presented by (Matthews, 1975), with coefficient values range from -1 to +1. MCC formula yields 0 for any two independent variables, -1 indicates total disagreement, and +1 indicates total agreement.

$$MCC = \frac{((TP)(TN))-((FP)(FN))}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{5}$$

The experiments for the reviewsused the tenfold cross-validation process using *SVM* showed an accuracy of 97%, Naïve Bayes (*NB*) showed an accuracy of 77.9%. Table 4explores the detailed results for the two classifiers used in this paper.

Table 4. Detailed results for NB and SVM.

Class	TP	FP	Precision	Recall	F-Measure	MCC	PRC	ROC
				NB classifier				
Positive	0.775	0.216	0.705	0.775	0.738	0.551	0.871	0.821
Neutral	0.980	0.000	1.000	0.980	0.990	0.987	0.992	0.989
Negative	0.683	0.152	0.750	0.683	0.714	0.540	0.874	0.801

Weighted AVG	0.779	0.147	0.782	0.779	0.779	0.634	0.896	0.847
SVM classifier								
Positive	0.980	0.040	0.962	0.980	0.941	0.941	0.959	0.961
Negative	0.960	0.020	0.980	0.960	0.970	0.941	0.959	0.926
Neutral	0.970	0.030	0.970	0.970	0.970	0.941	0.959	0.944
Weighted AVG	0.97	0.030	0.970	0.97	0.960	0.941	0.959	0.943

We examine the highest overall TP values, precision, recall, and F-measure when comparing the models.On the other hand, the FP rate should be minimized. Table 5 shows that the NB model successfully predicts the neutral class label with high-accuracy results. NB yielded accepted results for predicting positive class instances but failed to obtain high accuracy detection results for negative class with only 68%. The SVM validates the proposed medical sentiment analysis approach and Table 9 results show that Text-based opinions gain the best results while other types need more enhancements to yield better results.

8. CONCLUSION AND FUTURE WORK.

This study emphasized the necessity for resolvable sentiment analysis to extract a human-reliable understanding of sentiment and emotions from the user-generated medical text information. This work develops a medical context sentiment extraction Artificial Intelligence model for the Modern Standard Arabic (MSA), Dialects, and/or Emoticons. Data collection of Arabic content was collected from several social network websites, where the features were extracted and weighted. The Receiver Operating Characteristic (ROC) prediction quality metrics are utilized to evaluate the effectiveness of the proposed model with promising results. Future research should expend more awareness of various multimedia elements and assess each review in its whole context and not only the textual element. The adoption of both lexicon-based and support vector machine algorithms was employed to conduct sentimental analysis. SVM validates the results with more than 90% within Receiver Operating Characteristic (ROC) prediction quality.

ACKNOWLEDGMENT

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies andScientific Research, King Faisal University, Saudi Arabia [Grant No. GRANT989]'.

REFERENCES

1. Satapathy, R., Cambria, E., & Hussain, A. (2017). Sentiment Analysis in the Bio-Medical Domain.
2. Varol, A. E., Kocaman, V., Haq, H. U., & Talby, D. (2022). Understanding COVID-19 news coverage using medical NLP. arXiv preprint arXiv:2203.10338.
3. Alshalawi, A. S. (2022). Social Media Usage Intensity and Academic Performance among Undergraduate Students in Saudi Arabia. Contemporary Educational Technology, 14(2), ep361.
4. Aljefree, N. M., & Alhothali, G. T. (2022). Exposure to Food Marketing via Social Media and Obesity among University Students in Saudi Arabia. International Journal of Environmental Research and Public Health, 19(10), 5851.
5. Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E., & Hussain, A. (2022). Multimodal Sentiment Analysis: A Systematic review of History, Datasets, Multimodal Fusion Methods, Applications, Challenges and Future Directions. Information Fusion.
6. Verma, S. (2022). Sentiment analysis of public services for smart society: Literature review and future research directions. Government Information Quarterly, 101708.

7. Al-Kabi, M. N., Wahsheh, H. A., & Alsmadi, I. M. (2016). Polarity classification of Arabic sentiments. International Journal of Information Technology and Web Engineering (IJITWE), 11(3), 32-49.

8. Aljamaan, F., Alenezi, S., Alhasan, K., Saddik, B., Alhaboob, A., Altawil, E. S., ... & Temsah, M. H. (2022). Healthcare workers' worries and Monkeypox vaccine advocacy during the first month of the WHO Monkeypox alert: Cross-sectional survey in Saudi Arabia. Vaccines, 10(9), 1408.

9. Lai, S. T., & Mafas, R. (2022, April). Sentiment Analysis in Healthcare: Motives, Challenges & Opportunities pertaining to Machine Learning. In 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 1-4). IEEE.

10. Ghosh, A., Umer, S., Khan, M. K., Rout, R. K., & Dhara, B. C. (2022). Smart sentiment analysis system for pain detection using cutting edge techniques in a smart healthcare framework. Cluster Computing, 1-17.

11. Al-Ayyoub, M., Khamaiseh, A. A., Jararweh, Y., & Al-Kabi, M. N. (2019). A comprehensive survey of arabic sentiment analysis. Information processing & management, 56(2), 320-342.

12. Al-Hashedi, A., Al-Fuhaidi, B., Mohsen, A. M., Ali, Y., Gamal Al-Kaf, H. A., Al-Sorori, W., & Maqtary, N. (2022). Ensemble classifiers for Arabic sentiment analysis of social network (Twitter data) towards COVID-19-related conspiracy theories. Applied Computational Intelligence and Soft Computing, 2022.

13. Bensoltane, R., & Zaki, T. (2022). Aspect-based sentiment analysis: an overview in the use of Arabic language. Artificial Intelligence Review, 1-39.

14. Wahsheh, H. A., & Albarrak, A. S. (2023). Sentiment Analysis Online Tools: An Evaluation Study. In International Conference on Information Systems and Intelligent Applications (pp. 103-112). Springer, Cham.

15. Denecke, K., & Deng, Y. (2015). Sentiment analysis in medical settings: New opportunities and challenges. Artificial intelligence in medicine, 64(1), 17-27.

16. Grissette, H. (2022). Semisupervised neural biomedical sense disambiguation approach for aspect-based sentiment analysis on social networks. Journal of Biomedical Informatics, 104229.

17. Chintalapudi, N., Angeloni, U., Battineni, G., di Canio, M., Marotta, C., Rezza, G., ... & Amenta, F. (2022). LASSO Regression Modeling on Prediction of Medical Terms among Seafarers' Health Documents Using Tidy Text Mining. Bioengineering, 9(3), 124.

18. Zunic, A. (2022). Deep learning approach to sentiment analysis in health and well-being (Doctoral dissertation, Cardiff University).

19. Mittal, M., Kaur, I., Pandey, S. C., Verma, A., & Goyal, L. M. (2018). Opinion mining for the tweets in healthcare sector using fuzzy association rule. EAI Endorsed Transactions on Pervasive Health and Technology, 4(16), e2-e2.

20. Yang, F. C., Lee, A. J., & Kuo, S. C. (2016). Mining health social media with sentiment analysis. Journal of medical systems, 40(11), 1-8.

21. Salas-Zárate, M. D. P., Medina-Moreira, J., Lagos-Ortiz, K., Luna-Aveiga, H., Rodriguez-Garcia, M. A., & Valencia-Garcia, R. (2017). Sentiment analysis on tweets about diabetes: an aspect-level approach. Computational and mathematical methods in medicine, 2017.

22. Edara, D. C., Vanukuri, L. P., Sistla, V., & Kolli, V. K. K. (2019). Sentiment analysis and text categorization of cancer medical records with LSTM. Journal of Ambient Intelligence and Humanized Computing, 1-17.

23. Al-Kabi, M., Al-Ayyoub, M., Alsmadi, I., & Wahsheh, H. (2016). A prototype for a standard arabic sentiment analysis corpus. Int. Arab J. Inf. Technol., 13(1A), 163-170.

24. Ahmad, M., Aftab, S., Bashir, M. S., & Hameed, N. (2018). Sentiment analysis using SVM: a systematic literature review. International Journal of Advanced Computer Science and Applications, 9(2).

25. Kewsuwun, N., & Kajornkasirat, S. (2022). A sentiment analysis model of agritech startup on Facebook comments using naive Bayes classifier. International Journal of Electrical & Computer Engineering (2088-8708), 12(3).