Sentiment classification of user's reviews on drugs based on global vectors for word representation and bidirectional long short-term memory recurrent neural network

Hadab Khalid Obayes[1], Firas Sabah Al-Turaihi[2], Khaldoon H. Alhussayni[3]
[1]College of education for humanities studies, University of Babylon, Babylon, Iraq
[2]College of information technology, University of Babylon, Babylon, Iraq
[3]Computer Center, University of Babylon, Babylon, Iraq

Article Info

ABSTRACT

The process of product development in the health sector, especially pharmaceuticals, goes through a series of precise procedures due to its direct relevance to human life. The opinion of patients on a particular drug can be relied upon in this development process, as the patients convey their experience with the drugs through their opinion. The social media field provides many datasets related to drugs through knowing the user's rating and opinion on a drug after using it. In this work, a dataset is used that includes the user's rating and review on the drug, for the purpose of classifying the user's opinions (reviews) whether they are positive or negative. The proposed method includes two phases. The first phase is to use the global vectors for word representation model for converting texts into the vector of embedded words. As for the second stage, the deep neural network (bidirectional long short-term memory) is employed in the classification of reviews. The user's rating is used as a ground truth in evaluating the classification results. The proposed method presents encouraging results, as the classification results are evaluated through three criteria, namely precision, recall and F-score, whose obtained values equal (0.9543, 0.9597 and 0.9558), respectively.

Corresponding Author:

Hadab Khalid Obayes
Department of geography
College of education for humanities studies
University of Babylon, Babylon, Iraq
Email: hedhabsa@gmail.com, hedhab@uobabylon.edu.iq

1. INTRODUCTION

Since the turn of the century, sentiment analysis developed in such a way that it has become among the most actively researched areas of natural language processing (NLP), particularly due to the coinciding growth in social media communication such as reviews, forum discussion, and (micro) blogging services, as well as the enormous amount of opinionated data available in digital form. Other fields that overlap with this discipline are data mining, web mining, text mining, and information retrieval [1]. NLP can be defined as a number of computational techniques used to analyze natural language texts so as to enable computers of understanding human language, and the sentiment classification of online review data is considered to be an essential process and of high importance.

The discipline of opinion mining is found at the overlap between each of information retrieval and computational linguistics [2], further extending to fields of management and social sciences, including

finances, politics, communication, and the health sector, as it tends to be important not only to business but to the society as a whole. Its rapid growth and development can be traced back to the notion that opinions are the centre around which almost all human activities revolve, as these essentially influence our behavior. The way in which others view the world and create an evaluation of it are, to a certain extent determined by our beliefs and perspectives upon reality. Therefore, we often tend to look for other's opinions before making a certain decision, and the latter is true for both individuals as well as organizations [3].

A considerable number of tools have been developed for opinion mining to be used in analyzing consumer opinions in most major business environments, such as travelling, accommodation, consumables, materials, products, and services [4]. However, the health sector has witnessed relatively less development. Drug surveillance, for example, is considered to be an essential factor in terms of drug safely soon after the release of a certain drug to public use. Since drug trials are usually conducted with a restricted number of subjects, the chance of detecting any uncommon adverse effect tends to be minimal [2].

At present time, the safety of pharmaceutical products is determined by certain clinical trials and specified testing protocols [5]. These researches are often conducted under standard conditions, restricted by time and number of test subjects. This could potentially lead to cases where discrepancies in selecting patients, as well as treatment condition, could significantly affect the drug's efficiency and the possible risks of adverse drug reactions (ADRs) [6].

Deep learning methods could be described as being methods of several representation levels, which are obtained by means of the composition of simple yet non-linear modules, each of which has the ability of transforming a representation from a particular level (starting with the raw input) to a higher representation at a level of a little more abstractness. Composing a sufficient number of such transformations enables the learning of functions that are of considerable complexity [7]. For the analysis of patient's opinion on the effectiveness of drugs, this study proposes the integration of GloVe, an efficient NLP transfer learning model, with BiLSTM, being a typical state-of-the-art bi-directional model. The vector representation of a word is of great use in text classifying, clustering, and retrieving information. There are a number of benefits observed when comparing word embedding methods with bag-of-words representations. The BiLSTM takes into consideration efficient amounts of contexts both before and after a word, thereby eliminating the issues of context limitation that is found in feed-forward models.

2.   RESEARCH METHOD
2.1.   Global vectors for word representation
In this work, we used the pre-trained embedding of GloVe [8], [9] with a 300-dimensional, which is paired with an n-gram embedding. For the embedding of words within a vector space, the GloVe model is used in the examination of the Xij presence matrix in a large text block representing the user's reviews on drugs. Numeric vectors that represent user reviews are obtained. GloVe can be described as an unsupervised learning algorithm used to obtain vector representations for words [10]. The training of the main idea is stated as follows:

$$w_i^T w_j + b_i + b_j = \log(X_{ij}) \tag{1}$$

where $w_k \ and \ w_j$ represent the trained vectors, and $b_i \ and \ b_j$ indicate the scalar bias terms related to the words i and j, respectively. The essential aspects of training processes in GloVe include:
a) A weighting function f to eliminate very commonly occurring words (like stop words), as these tend to add noise and are not over weighted;
b) Rare words are not over weighted;
c) The co-occurrence strength, whenever modeled as a distance, needs to undergo smoothing by means of a log function.

The final loss function for GloVe [11] model will be as follows:

$$J = \sum_{i,j \in V} f(X_{ij})(w_i^T w_j + b_i + b_j - \log(X_{ij}))^2 \tag{2}$$

where V is a complete vocabulary, and $f(x) = (x/x_{\max})^\propto \ if \ x < x_{max}$ and f(x)=1 otherwise.

2.2.   Long-short-term-memory deep neural network (LSTM-DNN)
LSTM is a distinct case of RNNs. It is designed to cope with the problem of gradients that explode or vanish [12]. The basic idea behind the LSTM is the existence of a memory cell and a number of gates.

These memory cells and gates are added to each neuron in the network [13]. The principle of LSTM work is transmitting the important information in a reliable way over a number of time steps onto the next time step. LSTM cell whereby the gates in the LSTM model are used with additional parameters can be applied to a memory circuit used to memorize or store the information for a long term from the recurrent layer [14]. The input of RNN model is a sequence $\{x_1, x_2, ..., x_n\}$ that uses the following recurrence.

$$h_t = f(h_{t-1}, x_t) \tag{3}$$

where:
x_t: input at the time t.
h_t is the hidden state.
h_{t-1} is the previous state

The gates arrive to the recurrent function f to cope with vanishing or exploding problems. The LSTM cells are applied as follows:

$$i_t = \sigma(w_i.[h_{t-1}, x_t] + b_i) \tag{4}$$

$$f_t = \sigma(w_f.[h_{t-1}, x_t] + b_f) \tag{5}$$

$$o_t = \sigma(w_o.[h_{t-1}, x_t] + b_c) \tag{6}$$

$$\tilde{C} = \tanh(h_{t-1}, x_t] + b_c) \tag{7}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{8}$$

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

where:
i_t is the input gate.
f_t is the forgete gate.
o_t_t is the output gate

While W and b represent the parameters of LSTM, and C_t and $\tilde{C}$ the present cell state and the new candidate values for cell state respectively. The sigmoid function is used three times in (i_t, f_t and o_t), respectively. The output of these gates is between 0 and 1 as in (4) to (6). The decision of the three gates depends on the x_t current input and h_{t-1} the previous output. The forget gate f_t is responsible for allowing the amount of previous state to pass. The input gate i_t decides whether the recent data of the input demands an update or append to the cell state [15]. The output gate o_t determines the data to be output depending on the cell state. These two gates function side by side to learn and store information related to the long and short-term series [13]. The cell state is the core of the LSTM work, as it is regarded to be a type of conveyor belt [16].

2.3. Bidirectional LSTMs

Bidirectional LSTMs extend from traditional LSTMs [17], and they could enhance the model performance in terms of sequence classification issues. In situations wherein all time steps of the input sequence have been provided, the Bidirectional LSTMs trains two LSTMs [18] rather than just one on the input sequence: the first one on the input sequence will be as-is, whereas the other is on a reversed copy of the input sequence. Given that these present an extra context to the network, they will lead to a relatively faster and more comprehensive learning on the problem in case [19].

Bi-LSTMs have been found to be particular of use in cases that require a context for the input, including sentiment classification. In unidirectional LSTM, the information seems to flow backwards, thus from back to front. In contrast, in Bi-directional LSTM the information flows both backwards and forwards, by means of two hidden states, hence the Bi-LSTM's wider comprehension on the context [20]. Bi-LSTMs have been applied in escalating chunks of input information that are of use to the network. There are structures of RNN with LSTM and RNN with Bi-LSTM [21]. Essentially, the BRNN undergoes a procedure whereby the neurons of a normal RNN are broken to bidirectional ways. The first is for the backward states (negative time direction), whereas the second is for the forward states (positive time direction). There is no connection between the input of the reverse direction states and the two former states' results. The BiLSTM is structured as illustrated in Figure 1. As for the directions, the input data regarding the past and future of the current time frame could be applied, meanwhile the standard RNN needs the delays to include future data.
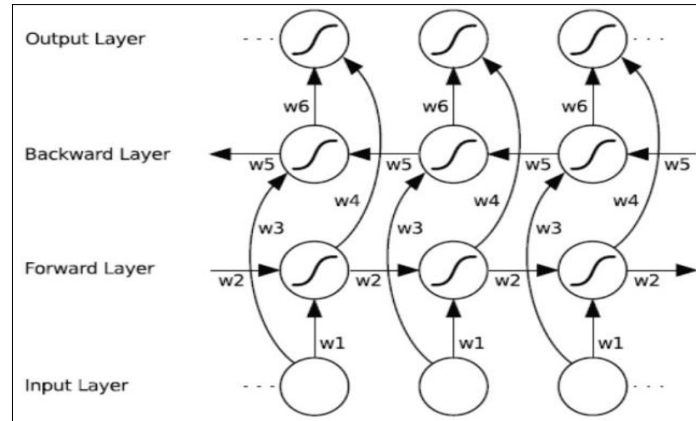
Figure 1. BiLSTM

3. RELEATED WORK

Qu *et. al.* introduced a bag-of-opinions review text representation model that differs from alternative models of this type, through which the sentiment phrases have been obtained. Viewpoints are composed of a triple aspect, namely a sentiment word, a modifier, and a negative word. To exemplify, in the phrase "not quite well", "well" is a sentiment word, whereas "quite" is a modifier, and "not" is negative word. As for sentimental classifications which only identify the tendency of a phrase, being positively or negatively, modifiers are not very significant, meanwhile for RRP, each of the modifier and the negative word are of importance. With a labeled domain-independent corpus (from multiple domains), a constrained ridge regression model will be first applied for learning every viewpoint feature (sentiment intensity and score), after which the traditional unigrams feature is extended by means of viewpoint features for achieving RRP [22].

There are a number of researches that make use of opinion mining within the medical field, particularly by means of machine learning techniques, greaves, ramirez-cano, millett, darzi, and donaldson (2013) conducted an analysis of patient's opinions on various factors of hospital performances in the United Kingdom. The authors applied four different methods, to see which gave the quickest and most accurate results: naïve bayes multinomials (NBM) decision trees, bagging, and support vector machines. There was 81%, 84%, and 89% agreement between quantitative ratings of care and those derived from free-text comments using sentiment analysis for cleanliness, being treated with dignity, and overall recommendation of hospital respectively (kappa scores: .40–.74, P<.001 for all) [23].

Nikfarjam *et al.* used association rule mining so as to identify patterns, i.e. combination of terms, or conditional random fields (CRFs), for the extraction of mentions of ADRs. Since it is implicitly assumed that a patient's post on ADRs typically expresses negatively tending sentiments, examined the impact of using sentiment analysis features to enrich a lexicon-based ADR identifying method [24]:

a) Gopalakrishnan *et al.* analyzed the extent to which patients were satisfied with a particular drug, through the use of a supervised learning sentiment analysis approach. There are three levels of polarity classified in this study, drawing a comparison between SVM and neural network based methods [2].

b) A. Navindgi and *et al.* (2016) made use of a subjectivity lexicon as well as machine learning algorithms to analyze the sentiment in posts on forums that discuss the topic of hearing loss. We report experiments on a sentiment-labeled corpus of posts taken from a medical support forum. they argue that not only is a more fine-grained approach to text analysis important but simultaneously recognizing the social function behind affective expressions enable a more accurate and valuable level of understanding [25].

4. THE PROPOSED SYSTEM

The current section presents a detailed explanation of the proposed system. Figure 2 shows the structure of the proposed system in this work. The dataset used in this work is obtained from the Kaggle website for retrieving user reviews and ratings on drug experience, namely the UCI ML Drug Review dataset. It involves user reviews on specific drugs, as well as any related conditions and a 10-star user rating that reflects a comprehensive view on the user's satisfaction. The emotion dataset consists of 215,063 instances, and there are 6 attributes in the dataset. The total number of individual drugs within the dataset is about 6345 drugs.

The dataset attributes are:
a)   Drug name (categorical): name of drug
b)   Condition (categorical): name of condition
c)   Review (text): patient review
d)   Rating (numerical): 10 star patient rating
e)   Date (date): date of review entry
f)   Useful count (numerical): number of users who found review useful

The dataset was explored to find any important statistics that help in understanding and analyzing the content of the data set. The dataset contains 215,063 instances of reviewers and 6345 drugs. Supervised machine learning approaches require labeled data to classify opinions in reviews. The text reviews (patient review) are assigned a label according to the rating (10 star patient rating). The text review consists of a written text that explains the rating score of the review. As the reviews are conducted by different persons with varying backgrounds, the intended meaning of a 10 star review will differ from one user to another. The patient review is then assigned an opinion class according to the rating in the data format. The assigned review class is positive whenever the rating is >=5, otherwise, if the rating is <5 then the class label will be negative. Figure 3 shows the count of the rating value.
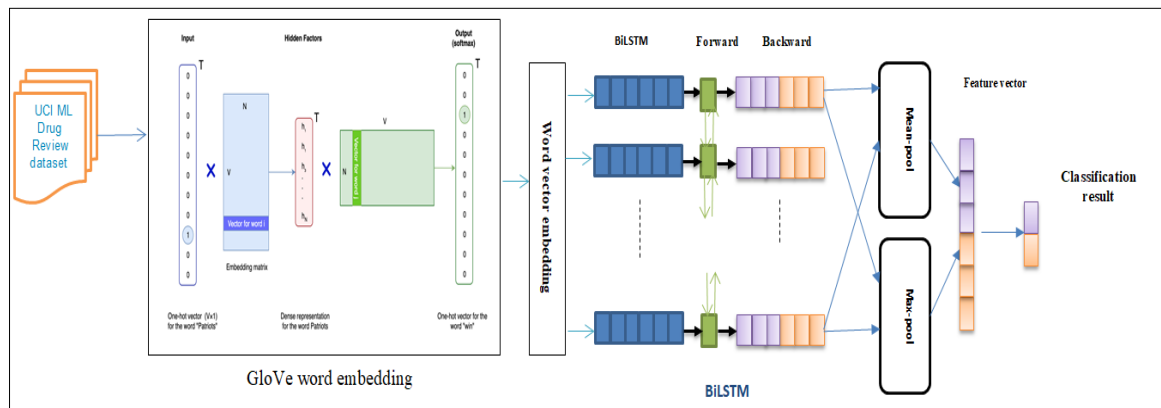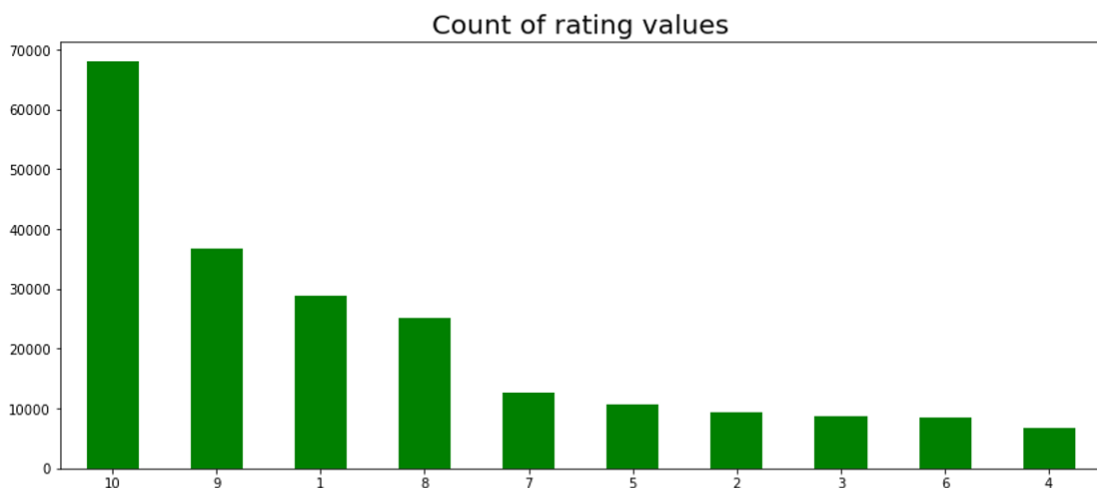

Figure 2. Proposes system

Figure 3. The count of rating value in the dataset

a)   The two rows of the dataset dealt with are the user reviews and rating.
b)   To get the words vector embedding, we employed the GloVe model, with the user reviews as its entry, and after its application this model the words vector embedding is obtained. The

pre-trained embedding of GLoVe word is used, which is paired with n-gram embedding characteristics, and is retained throughout the training.

c)   To perform the process of classifying user reviewers into two classes, which are positive reviewers and negative, the BiLSTM Deep Network is employed. The emotion dataset consists of 215,063 instances, split to 161,297 as training data and 53,766 as testing data. The rating per review labeled as integer numbers between 1 and 10.

d)   The units of the BI-LSTMs are described in 128 hidden dimensions, and the initial learning rate is trained using an ADAM optimizer. For the embedding layer, we defined the drop-out rate. The batch training was performed with a batch size of 128.

e)   RNNs will be used due of secret states, as they recall prior knowledge and link it to the current mission. Long short-term memory networks (LSTM) are an RNN subset, specialized in collecting information for longer periods of time. In addition, a bidirectional LSTM holds contextual details in both directions, which is very useful when classifying text.

5.   EXPERIMENTAL RESULT

For the evaluation of the efficiency of this classification model, the class must be compared with the ground line. There are multiple criteria to realize classification effectiveness. Experimental results are conducted on the ground-truth, based on the used dataset. As mentioned earlier, the user's reviews were sorted on the dataset of the corresponding rating value, and this enabled us to adopt rating as a category on the basis of which we classify the user's reviews. In this work, three criteria are used namely precision, recall and f-score.

Precision or confidence (as called in data mining) indicates the amount of predicted positive cases representing correct real positives.

$$precision = \frac{TP}{PP} = \frac{A}{A+B}$$ (10)

where TP is true positive and PP is predicted positives and TP is real positive.

Recall or sensitivity can be defined as the rate between the number of correct detections for anomalous measurements to the total sum of anomalous measurements.

$$Recall = \frac{TP}{RP} = \frac{A}{A+C}$$ (11)

F-Measure provides a single score that balances both the concerns of precision and recall in one number.

$$F_{Measure} = \frac{2*precision*Recall}{Precision+Recall}$$ (12)

when training the Bi-LSTM deep network, we obtained a precision value of 0.9543, while Recall and F-Measure were 0.9597 and 0.6558, respectively. Figure 4 shows the confusion matrix to the classification result.

For the purpose of examining the relative performances of classifying methods in the context of multiple accuracy evaluation measures, the result of Bi-LSTM has been compared with a number of classifiers that are well known in the machine learning field. Table 1 lists the result of comparison. Table 1 presents a summary of the value for each accuracy measure that has been obtained through five classification techniques applied to the user reviewer dataset. It indicates that the results of the Bi-LSTM are the best results. Figure 5 shows the criteria for evaluating the efficiency of the proposed work used in comparison with four other classifiers, as the results show the superiority of the proposed system over the rest of the classifiers.

Table 1. The classifiers results

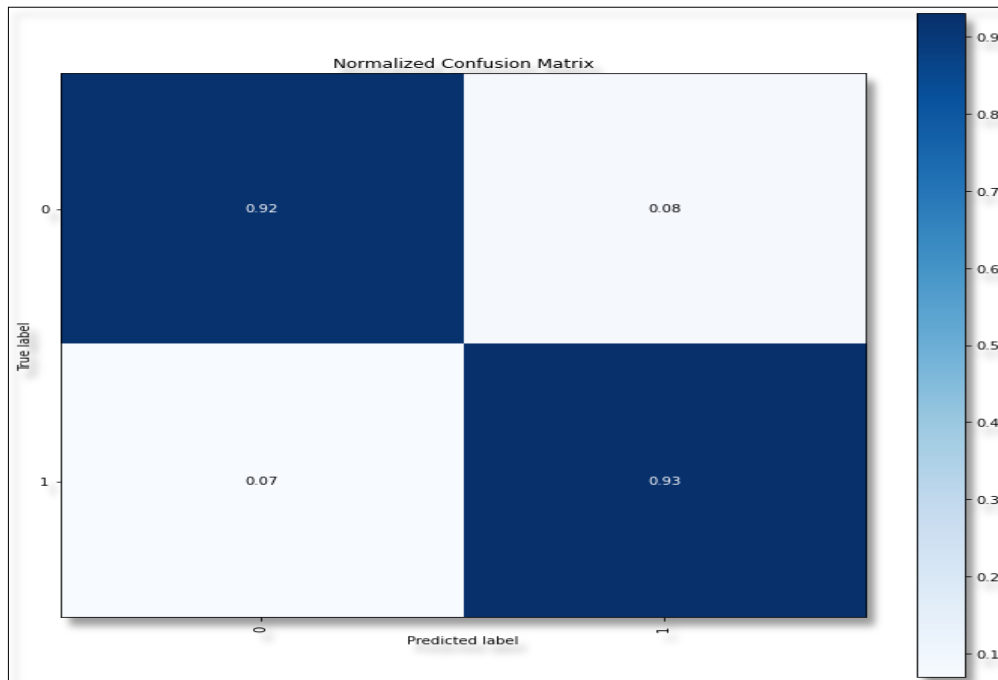| Classifier	Precision	Recall	F-Measure
Bernolli	0.8829	0.8386	0.8602
Random forest	0.8869	0.9451	0.9377
Decision tree	0.8863	0.9351	0.9376
Kneighbors	0.8586	0.9390	0.9235
BiLSTM	**0.9543**	**0.9597**	**0.9558**

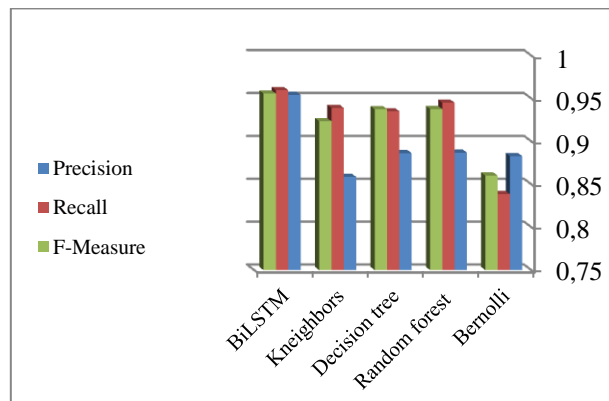Figure 4. The confusion matrix of classification result

Figure 5. The result of 5 classifiers

6. CONCLUSION

According to the experiment, the integrated GloVe and BiLSTM models have achieved good scores in terms of F1-score, Precision and Recall. This shows a prominent inspiration for new ideas to be applied in other future works. The model achieved a sufficient performance considering the cost of massive computing power. In this study, the user's reviewers are used as a label for the two classes, the positive and the negative class, thereby dividing the ten-scores into two-parts only. In future works, it is possible to use the model after implying several improvements in order to predict the user's rating through their reviewers.

REFERENCES

[1]   B. Wang, S. Xiong, Y. Huang, and X. Li, "Review rating prediction based on user context and product context," *Appl. Sci.*, vol. 8, no. 10, pp. 1-13, 2018, doi: 10.3390/app8101849.
[2]   V. Gopalakrishnan and C. Ramaswamy, "Patient opinion mining to analyze drugs satisfaction using supervised learning," *J. Appl. Res. Technol.*, vol. 15, no. 4, pp. 311-319, 2017, doi: 10.1016/j.jart.2017.02.005.
[3]   L. Zhang, S. Wang, B. Liu, "Deep Learning for Sentiment Analysis: A Survey," *WIRE's data mining and knowledge discovery.*, vol. 8, no. 4, e1253, 2018, doi: 10.1002/widm.1253.

[4]   P. Hajek, A. Barushka, and M. Munk, "Opinion mining of consumer reviews using deep neural networks with word-sentiment associations," vol. 583, pp. 419-429, IFIP. Springer International Publishing, 2020. In: Maglogiannis I., Iliadis L., Pimenidis E. (eds) Artificial Intelligence Applications and Innovations. AIAI 2020, doi: 10.1007/978-3-030-49161-1_35

[5]   S. Vijayaraghavan and D. Basu, "Sentiment Analysis in Drug Reviews using Supervised Machine Learning Algorithms," *arXiv*, 2020.

[6]   F. Gräßer, H. Malberg, S. Kallumadi, and S. Zaunseder, "Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning," *DH '18: Proceedings of the 2018 International Conference on Digital Health*, vol. 2018-April, pp. 121–125, 2018.

[7]   N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electron.*, vol. 9, no. 3, pp. 483, 2020, doi: 10.3390/electronics9030483.

[8]   Patel, V., Mishra, P., & Patni, J. C. (2018, June). PsyHeal: An Approach to Remote Mental Health Monitoring System. In 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE) (pp. 384-393). IEEE.

[9]   M. Kaneko and D. Bollegala, "Autoencoding Improves Pre-trained Word Embeddings," in *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, 2020, pp. 1699-1713.

[10] C. D. Pennington, Jeffrey; Socher, Richard; Manning, "GloVe: Global Vectors for Word Representation," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, vol. 19, no. 5, pp. 1532-1543.

[11] V. Makarenkov, B. Shapira, and L. Rokach, "Language Models with Pre-Trained (GloVe) Word Embeddings," *arXiv Comput. Lang.*, 2016.

[12] H. K. Obayes, N. Al A, and E. Al-shamery, "AIDS's Drugs Quantification and Surveillance Using Deep Learning," vol. 59, no. 6s, pp. 282-290, 2019.

[13] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," *Energies*, vol. 11, no. 7, pp. 1-20, 2018, doi: 10.3390/en11071636.

[14] C. Zhang and P. C. Woodland, "High Order Recurrent Neural Networks for Acoustic Modelling," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 5849-5853, doi: 10.1109/ICASSP.2018.8461608.

[15] H. K. Obayes, N. Al-A'araji, and E. Al-Shamery, "Examination and forecasting of drug consumption based on recurrent deep learning," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2 Special Issue 10, pp. 414-420, 2019, doi: 10.35940/ijrte.B1069.0982S1019.

[16] J. Kumar, R. Goomer, and A. K. Singh, "Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenterst," *Procedia Comput. Sci.*, vol. 125, no. January, pp. 676-682, 2018, doi: 10.1016/j.procs.2017.12.087.

[17] A. Pogiatzis and G. Samakovitis, "Using bilstm networks for context-aware deep sensitivity labelling on conversational data," *Appl. Sci.*, vol. 10, no. 24, pp. 1–17, 2020, doi: 10.3390/app10248924.

[18] B. Y. Lin, F. Xu, Z. Luo, K. Zhu, "Multi-channel BiLSTM-CRF Model for Emerging Named Entity Recognition in Social Media," *Proceedings of the 3rd Workshop on Noisy User-generated Text*, Copenhagen, 2017, pp. 160-165, doi: 10.18653/v1/W17-4421.

[19] A. Aziz Sharfuddin, M. Nafis Tihami and M. Saiful Islam, "A Deep Recurrent Neural Network with BiLSTM model for Sentiment Classification," *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 2018, pp. 1-4, doi: 10.1109/ICBSLP.2018.8554396.

[20] G. Guan and M. Zhu, "New Research on Transfer Learning Model of Named Entity Recognition," *J. Phys. Conf. Ser.*, vol. 1267, no. 1, 2019.

[21] T. Thireou and M. Reczko, "Bidirectional Long Short-Term Memory Networks for Predicting the Subcellular Localization of Eukaryotic Proteins," in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 4, no. 3, pp. 441-446, July-Sept. 2007, doi: 10.1109/tcbb.2007.1015.

[22] L. Qu, G. Ifrim, and G. Weikum, "The bag-of-opinions method for review rating prediction from sparse text patterns.," *Coling 2010 - 23rd Int. Conf. Comput. Linguist. Proc. Conf.*, Beijing, vol. 2, no. January, pp. 913–921, 2010.

[23] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, and L. Donaldson, "Use of sentiment analysis for capturing patient experience from free-text comments posted online," *J. Med. Internet Res.*, vol. 15, no. 11, 2013, pp. 1-9.

[24] A. Nikfarjam, A. Sarker, K. O'Connor, R. Ginn, and G. Gonzalez, "Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features," *J. Am. Med. Informatics Assoc.*, vol. 22, no. 3, pp. 671–681, 2015, doi: 10.1093/jamia/ocu041.

[25] A. Navindgi, C. Brun, C. Boulard Masson, and S. Nowson, "Steps Toward Automatic Understanding of the Function of Affective Language in Support Groups," *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*, Austin, 2016, pp. 26–33, doi: 10.18653/v1/W16-6205.

BIOGRAPHIES OF AUTHORS

Hadab Khalid Obayes is assistant professor at University of Babylon, Babel, Iraq. She received her PhD degree from University of Babylon, Iraq, in 2020. She published a number of papers in international journals and referred conferences. Her current research areas are, AI, Machine learning, neural networks, deep learning, natural language processing and Image Processing

Firas Sabah Al-Turaihi is a lecturer at University of Babylon, College of Information Technology, Networks Departments. He received his PhD degree from Brunel University London, United Kingdom. He received his BSc and MSc degrees in Computer Science. His research interests cover communication networks.
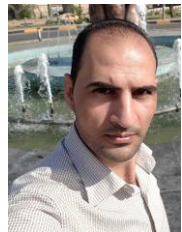
Khaldoon Hasan Alhussayni received the BSc degrees in Computer Science from the University of Babylon, Iraq, in 2008. After completing his BSc, he worked as a programmer at the Department of Computer Center, the University of Babylon in 2009. He received the MSc degrees in Computer Science and Computer Engineering from the Tula State University, Russian, in 2014. Secince Nov 2016. He entered the Tomsk state university (TSU), Department of computer science theory as a Ph.D. student in specialty Mathematical and software of computers, complexes and computer networks. His main research interests are machine learning, neural networks, deep learning, dialogue systems and natural language processing.