Future Generation Computer Systems 130 (2022) 106–113

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

EANDC: An explainable attention network based deep adaptive clustering model for mental health treatment

Usman Ahmed [a], Gautam Srivastava [b,d], Unil Yun [c], Jerry Chun-Wei Lin [a,*]

[a] Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, 5063, Bergen, Norway
[b] Department of Mathematics & Computer Science, Brandon University, Brandon, Canada
[c] Department of Computer Engineering, College of Electronics and Information Engineering, Sejong University, South Korea
[d] Research Centre for Interneural Computing, China Medical University, Taichung, Taiwan

ARTICLE INFO

Article history:
Received 2 July 2021
Received in revised form 19 October 2021
Accepted 12 December 2021
Available online 17 December 2021

Keywords:
Adaptive treatments
Internet-delivered interventions
NLP
Text clustering
Word sense identification
Explainable

ABSTRACT

Internet-delivered Psychological Treatment (IDPT) has been shown to be an effective method for improving psychological disorders. Natural language processing (NLP) requires an appropriate set of linguistic features for word representation and emotion segmentation. For psychological applications, models must be trained on extensive and diverse datasets to achieve expert-level performance. Labeling psychological texts authorized by patients is challenging because emotional biases can lead to incorrect segmentation of emotions and labeling emotional data is time consuming. In this paper, we propose an assistance tool for psychologists to explore the emotional aspects of mentally ill individuals. We first use an NLP-based method to create emotional lexicon embeddings, and then apply attention-based deep clustering. The learned representation is then used to visualize the emotional aspect of the text authorized by patients. We expand the patient authored text using synonymous semantic expansion. A latent semantic representation based on context is clustered using EANDC, which is a Explainable Attention Network-based Deep adaptive Clustering model. We use similarity metrics to select a subset of the text and then improve the explainability of learning using a curriculum-based optimization method. The experimental results show that synonym expansion based on the emotion lexicon increases accuracy without affecting the results. The attention method with bidirectional LSTM architecture achieved 0.81 ROC in a blind test. The self-learning based embedding then visualizes the weighted attention words and helps the psychiatrist to improve his explanatory power of the qualitative match for clinical notes and the remedy. The method helps in labeling text and improves the recognition rate of symptoms of mental disorders.

1. Introduction

As the use of technology grows [1–4], not only does it assist in a new frontier in mental health support and data collection (i.e., mobile phone, tablets, sensors and IoTs) but also creates serious mental illness for young adults [5]. The World Health Organization (WHO)[1] reports that impaired mental health is a worldwide problem. Increases and decreases in lockdown due to pandemics affect mental rather than physical health. Social isolation, low interaction at work and in education exacerbate mental health conditions. Frontline health workers also suffer from anxiety and depression due to family isolation, lack of protective equipment, stressful work environments, and social isolation. Depression and anxiety are emotions that describe a person's mental state. The diverse information comes from the extensive and growing literature. However, due to conflicting reports, it remains difficult to find out useful and up-to-date information on the treatments of these conditions [6].

A mental disorder is the result of a series of events that has occurred in the recent past of an individual and that may be triggered by a single immediate subject or event [7]. Recognizing the cause of the triggered emotion is still a challenging task [8]. The person may recognize vital signs and seek assistance from medical professionals. With the growing data hub of social media and Internet forums, individuals can connect and share their suffering anonymously online [9]. Sharing personal opinions often leads to an exposed situation [7,10]. Preventing and recognizing

* Corresponding author.
E-mail addresses: usman.ahmed@hvl.no (U. Ahmed), SRIVASTAVAG@brandonu.ca (G. Srivastava), yunei@sejong.ac.kr (U. Yun), jerrylin@ieee.org (J.C.-W. Lin).

1 https://www.who.int/teams/mental-health-and-substance-use/mental-health-and-covid-19.

https://doi.org/10.1016/j.future.2021.12.008
0167-739X/

such incidents through the texts can help in timely mediation and community welfare [11].

Depression is one of the most prevalent mental disorders worldwide, affecting approximately 264 million people [12]. Untreated depression has more severe consequences [13]. With over 800,000 people dying by suicide, depression is the second leading cause of death among 15 to 29-year-olds. In middle- and low-income countries, mental disorders that lead to social or personal problems are not treated regularly. Other problems include inaccurate assessment, lack of trained health care providers, and emotional bias associated with mental disorders [12]. Due to fear and shame of physiological treatment, people try to avoid medical treatment and medical advice [10]. For this reason, many patients avoid or do not admit to having mental illness.

Because of the resulting isolation, such as a pandemic quarantine, the health care system faces global challenges in preventing and treating mental health problems. The health care system is under financial and technical pressure to develop an adaptive system that enables online interventions. With fewer resources, Internet-delivered Psychological Treatment (IDPT) is helping to improve mental and physical problems of large populations [14]. The tunnel-based method is one of the existing solutions. However, it is inflexible and cannot be explained [7].

Interventions based on user adoption can lead to positive outcomes. Individual user behavior analysis can help the intervention system make personalized decisions. Behavioral analysis includes user's preferences, interests, social environment, and mental health symptoms [7]. This study aims to extract mental health symptoms from the text authorized by a patient based on the depression vital signs. In this work, we use the latent representation method to create an emotional embedding and then use this embedding to cluster the emptied texts based on deep attention-based clustering. Communication is the primary source for addressing mental state and emotional well-being. The patient composed a text that contained many important emotional signs that drew his attention to the triggering vital signs.

This paper addresses data labeling tasks in the extraction of depression-based emotional vital signs for mental health interventions. In particular, we have the following contributions:

1. We focus on these vital emotional signs to extract meaningful and contextual information, and then visualize it so psychologists can suggest preventive measures.
2. We used the concept lexicon construction, which expands the synonym based on the context in which it appears, and then uses this lexicon as a latent representation.
3. The learned representation is used to train and cluster the input text based on the attention-based method. The final embedding is used to visualize the emotional vital signs. The method works in batch and each batch is controlled using the curriculum learning method.

The remainder of the paper is organized as follows. Section 2 discusses related work. Section 3 describes the methodology, data collection tasks, and model training. Section 4 reports the results. Future work is discussed in Section 5. Finally, Section 6 concludes the paper with some closing remarks.

2. Related work

There are several attempts to detect depression using online intervention. This section addresses the existing method in this regard.

The Item Response Theory (IRT) method aims to extract depression symptoms and improve precision [15]. User intervention based on behavioral adaptation is achieved by using an adaptive questionnaire. The linked questionnaire selects the following most appropriate questions based on the previously answered questions. The adaptive questions help to understand the characteristics of the users and improve the precision in applying the remedial measure.

Dinakar et al. [16] analyzed the young community and proposed a stacked generalization modeling approach to analyze the online community. The method trained the ensemble classifier using the support vector machine model with SVM and stochastic gradient boosted decision trees (GBDT). The text is then categorized into 23 topics. Unigrams, bigrams, part-of-speech bigrams, and TF-IDF filters were used to extract the features. The stacking base classifier is then used to predict the output. The model evaluated the 7,147 personal stories on a popular teenage help website [16]. However, this data-driven deep learning (D-DL) is considered to be the technique that offers a clear advantage over previous techniques such as feature engineering because it generalizes the data and better accounts for sequence-to-sequence association, which is lacking in unigram and bigram models. Feature engineering makes modeling complex and depends on the domain. D-DL provides solutions without ambiguity to the problems.

Choudhury et al. [17] examined adolescent behavior in social forums. The model recognizes the depression potential of people. They obtained data for the labeling tasks from Amazon Mechanical Turk[2]. They used the questionnaire-based method to extract the user behavior responses and then analyzed the social media feed to stduy the relationship between the responses. Learning features were then modeled based on the extracted features. The learning method was tested on the geographically located Twitter data from the United States and found to have a strong correlation.

Another method based on the Twitter feed is also presented and discussed [18]. The text feed is called depression, bipolar disorder, and seasonal affective disorder. The Linguistic Inquiry Word Count (LIWC) detects a deviation in each disorder and groups them accordingly. The method analyzes the character occurrences based on their order to control the response. Then, the classifier can distinguish the groups. The Twitter feed of the group is analyzed to find the correlation between the class distribution of each group [19]. Lin et al. [20] used the four-layer network to compare the analysis of microblogs with traditional statistical-based classifiers such as Random Forest, SVM, Naïve Bayes. They analyze the pooling methods such as max-pooling, mean-over-instance, and mean-over-time. The best method for the DNN was mean-over-time pooling, among others.

A high-dimensional feature representation helps the neural network learn the unique representation [7]. The learning network uses the latent space to perform each task. The proposed network uses the averaging method to compute the inputs of the previous layer and the weights. The last layer of the network represents a non-linear activation function [21]. The hidden feature representation helps in predictive analysis. Proper selection of the hidden layer and hyperparameters also helps in storing complex problems [22]. In NLP domains [23–26], many algorithms were presented and a recurrent neural network (RNN) architecture was proposed by incorporating the word embedding method [27]. The encoder and decoder models encode the word sequence and decode it into a fixed-length vector representation. The gated method helps to reduce the loss function. The fixed-length vector loses information through vector compression based on the input features [22].

In [28], the authors used the Latent Dirichlet Allocation model to extract the themes. Based on the extracted themes, a comprehensive understanding of the emotion is then created. In another

2 https://www.mturk.com/

work, Glove Embedding is used to detect negative psychological emotions such as addiction, anxiety, depression, insomnia, stress, and obsessive cleansing disorder (OCD) [29]. Long-term text dependence is examined at the document level. A Bayesian network with fuzzy logic can also be used for dimensionality reduction and trajectory analysis [30].

The attention mechanism uses the alignment technique to improve the RNN encoder and decoder model. The values of neighbor features affect the sequence compared to multi-hop sequences [22]. The attention method applies a certain weight to selected inputs. The selection of inputs and their position helps the decoder to obtain the appropriate weights of the context vector for the higher feature representation. Then, the model is optimized using the RNN weight optimization method. The learned attention vector contributes to better context and feature representation [31]. The soft attention, hard attention, and global attention methods are proposed for the attention mechanism. The soft attention reduces the contextual information [9]. It also applies the averaging technique to learn the hidden feature representations. In hard attention, Xu et al. used the sampling method to calculate the context vector [32]. For local and global attention, the intermediate version of soft and hard [33] was used. Each attention point is selected from the stack of inputs and then the prediction function for the position of the attention point is used. The method helps to improve the domain-specific data.

Ahmed et al. used symptom extraction by attention-based models. The semi-supervised learning method improves by using the active learning model [9]. The method was trained with the active learning model and then used the visualized model to improve the accuracy over the baseline method. The method automatically learns from the texts authorized by the patients for psychoeducational exercises.

3. Explainable attention network based deep adaptive clustering (EANDC)

We used the emotion lexicon-based embedding in the depression symptom recognition model. The architecture is shown in Fig. 1, where we used the cosine similarity distance measure to evaluate the PHQ-9 symptom score. The lexicon extends its knowledge by using the word vector model. We explain the details of the Explainable attention network based deep adaptive clustering (EANDC) model as follows. An anonymous user text is mentioned below [7].

> *I had a Psychotic Break and have been super depressed ever since... one of my biggest worries is if I will ever be able to get back to work ... it been 6 months already, and I find the prospect of getting back to work terrifying... I'm afraid I'm so fragile now that I will not be able to take the long hours of work 9-6 again... has anyone been like this? Please help.*

3.1. Psychometric questionnaires (PQ)

The psychiatrist's method uses ICD10[3] classification. The answering based on standard questions helps the psychiatrist to suggest the supported diagnostic reliability according to patient's mental problems. The questionnaire includes the frequency of selected vital signs and then the rating is based on the frequency value using the predefined threshold. Each symptom is classified into mild, moderate, or severe conditions, which is called Clinical Symptom Elicitation Process (CSEP)[4]. We automate the process by first enriching the cadaver's knowledge with new words and then using the embedding when training the Deep Attention-based clustering method. The category of the symptom and its frequency are calculated to assign the score for the patient ratings.

The PHQ-9 method extracts the behavior types used in the DSM-V[5]. All nine symptoms are categorized as disorders, i.e., sleep, interest, eating, or social problems, as mentioned in Table 1 and sample document[6]. The scoring of the questions helps to classify the patient. During preprocessing, each term is lowercased, the non-meaningful full symbols (#, +, -, *, =, HTTP, HTTPS) are removed, slang or text-based words are converted to total words.

3.2. Word embedding using emotional lexicon

Knowledge-based emotional systems have not been well studied in the literature. In this work, we used the word sense lexicon to guide contextual embedding. The word sense lexicon is based on emotional knowledge extracted from an online forum. The word token in the patient text consists of a 300-dimensional pre-trained model called the global vector for word representation (Glove) [34]. The structural embedding of the sentence is trained using the concept of semantic composition and the hypothesis [35]. The frequency of co-occurrence of the vectorized words is calculated based on linguistic patterns. The unique word with fixed vector representation is the output of the model. In general, the pre-trained model performs better on the proposed task. However, they lack emotional analysis. Therefore, we extend the emotion semantics of words by extracting words using transfer learning. Since most of the models are based on the open-source data, i.e. *(Wikipedia texts)*, and sentiment knowledge *(Twitter data)*, the *happy* and *sad* showed the emotions of *feelings*. However, both words represent different mental orders. Therefore, its word lexicon is expanded based on the patient's authored texts. For the classification of documents, some preliminary definitions are discussed next.

Definition 3.1 (*Corpus*)**.** A corpus D consists of the set of texts, $D = \{t_1, t_2, \ldots, t_n\}$.

Definition 3.2 (*Emotion Set*)**.** It is used to extract synonyms, antonyms, hypernyms, and the physical meaning for each extracted part of speech. A sentiment word consists of the set $W = \{w_1, w_2, \ldots, w_K\}$ for each document.

Definition 3.3 (*Vocabulary Set*)**.** The vocabulary is built using the W set used to train the model. The resulting embedding is the learned vector V, i.e., $V = \{v_1, v_2, \ldots, v_m\} \in \mathbb{R}^{m\times\delta}$ where δ is the dimension of the word vector.

Referring to our earlier work [9], in this designed model, we first extract the part of speech, i.e., the (*noun*, *verb*, *adverb* and *adjective*). Then, we extract the synonyms using the *WordNet*. The synonyms include antonyms, hypernyms, and the physical meaning for each extracted part of speech after we obtain the domain-specific context corpus *(Algorithm 1, Lines 8 to 10)*. We then build the vocabulary using the word set. The latent representation is then built using the attention-based bidirectional LSTM. Using the vector averaging technique, we obtain the final sentence embedding. The vector representation is the emotional word representation. Using the trained embedding, we cluster the symptom-based on its similarity with the PHQ-9 questionnaire lexicons. The embedding of the questionnaire and input text are used to calculate the similarity, and then the label is assigned.

[3] https://www.who.int/classifications/icd/en/GRNBOOK.pdf
[4] https://www.who.int/classifications/icd/en/GRNBOOK.pdf
[5] https://www.psychiatry.org/psychiatrists/practice/dsm
[6] https://www.uspreventiveservicestaskforce.org/Home/GetFileByID/218

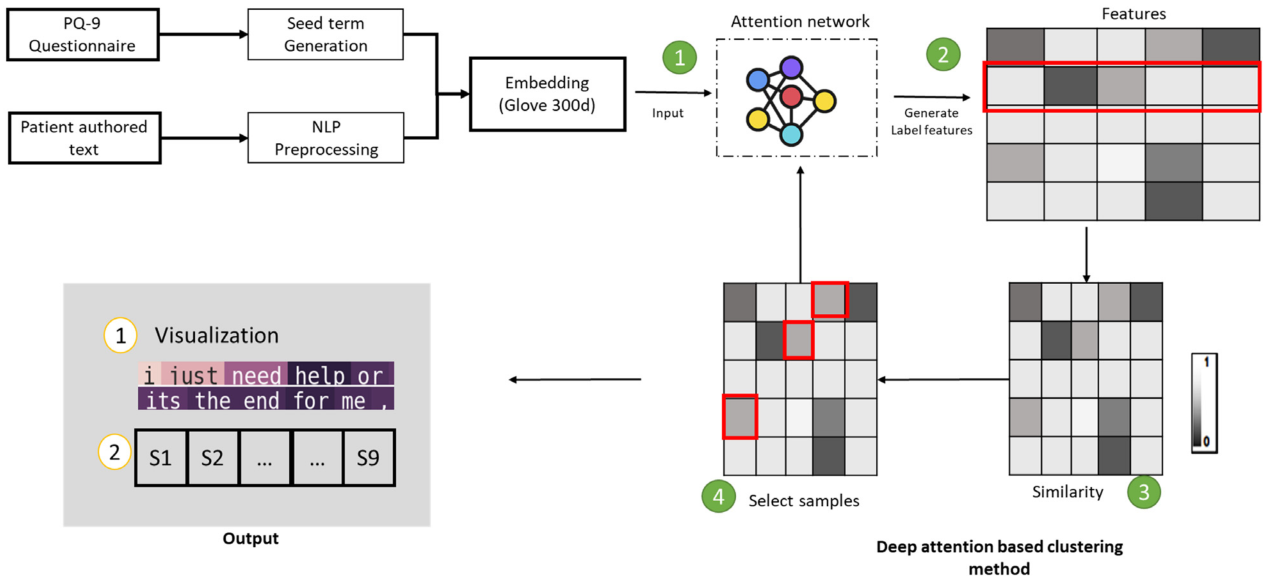

Fig. 1. A flowchart of training and adaptation using attention-based domain adoption. The similarity of visualization and symptoms is suggested to the psychiatrist to take notes and suggest remedies for specific symptoms to the patient.

Table 1
PHQ-9 questionnaire and seed terms for each symptoms.

Symptoms	PHQ-9	Seed terms
S1	Little interest or pleasure in doing things	Interest
S2	Felling down depressed or hopeless	Feeling, depressed, hopeless
S3	Trouble falling or staying asleep or sleeping too much	Sleep, asleep
S4	Feeling tired or having little energy	Tired, energy
S5	Poor appetite or over eating	Appetite, overeating
S6	Feeling bad about yourself or that you are a failure or have let yourself or your family down	Failure, family
S7	Trouble concentrating on things such as reading the newspaper or watching television	Concentration, reading, watching
S8	Moving or speaking so slowly that other people could have noticed or the opposite being or restless that you have been moving around a lot more than usual	Moving, speaking, restless
S9	Thoughts that you would be better off dead or of hurt yourself	Dead, hurt, suicide

The most similar ones are selected from the training samples and labels are assigned to them. Then RNN progresses in finding the optimized labels for the travel features by gradually taking a set of difficult samples. During the clustering process, λ is incrementally increased. The Algorithm 1 explains each step in detail, where the corpus represents the patient author's text with emotional embedding and training embedding, and the loss is calculated based on the gradient values *(Algorithm 1, input)*. We can control the frequency of the samples m *(Algorithm 1, input)*. First, we start with m samples and select the small batch with the average embedding from *(Algorithm 1, Lines 1 to 4)*. Then, we compute the similarity to obtain the label of the corpse with questionnaire *(Algorithm 1, Line 5)*. Then we update the gradient method *(Algorithm 1, Line 6)*. In the designed model, we used the `argmax` method to obtain the cluster classes for the final prediction of the output deviation points for the test samples *(Algorithm 1, Lines 8 to 10)*.

3.3. Dataset

The dataset used is accessed through an online forum, website, and social media site [7]. The Amazon Mechanical Turk[7] method is used to label the text. The other labeling is done using the EANDC. The labeling indicates the degree of depression in nine symptoms, i.e., 0 means not depressed, 1 is mildly depressed, 2 is moderately depressed, and 3 is severely depressed [7]. However, we convert to multi-label assignment, where the presence of each symptom means 1 and the absence means 0. The data collected are shown in Table 2.

Table 2
The statistical summary of the training and testing set.

Type	Statistics
Corpus size (Number of posts collected)	15044
Number of sentences	133524
Average sentences per post	8.87
Average words per post	232
Training set size (Number of posts)	14944
Testing set size (Number of posts)	100

3.4. Deep learning model

We used the feed-forward neural network as the basis for the experiments. The EANDC embedding is used as the feature extractor in all models, while the averaging method is used for the sentence vectors. The hidden layer consists of a (30, 20, 10) structure with a ReLU activation function [9]. We used the cross-entropy based loss function. For evaluation, we used the true

[7] https://www.mturk.com/.

Algorithm 1 Deep learning based clustering method.

Input: Corpus, m instances per batch; n, total instances; d, documents; C, set of instances in the batch; and T, set of training samples.
Output: Cluster label c_i.

for $d \in corpus$ **do**
  $d \leftarrow preprocessing(d)$;
  for term $t \in d$ **do**
    synonyms $\leftarrow Extract_{synonyms}(t)$;
    $z \leftarrow Extract_{hyperonym,hyponym,antonyms}(synonyms)$;
    for $w \in z$ **do**
      terms $\leftarrow$ wordnet (synonym);
    end for
    vocabulary $\leftarrow$ terms;
  end for
end for
Embed $\leftarrow attention_{BiLSTM}(vocabulary, corpus)$;
while $K \leq \{1, 2, ; \frac{n}{m}\}$ **do**
  Sample batch C from $corpus$;
  $Embed(C)$;
  Select training samples from T;
  Calculate similarity using the Cosine similarity;
  Update the gradient descent algorithm;
end while
while $C_i \in corpus$ **do**
  $\{c_i\} = argmax_h(l_{ih})$;
end while
Return Cluster label c_i.

Table 3
The mean ROC-AUC values of training and testing set.

Architectures	Train	Test
Baseline	0.76	0.73
Bidirectional LSTM	0.91	0.80
Bidirectional_LSTM_Attention	**0.96**	**0.88**

positive rate $TPR = \frac{TP}{TP+FN}$ and the false positive rate $FPR = \frac{FP}{FP+TN}$ as performance measures. We used an RNN gated architecture suitable for sequential tasks and employed it in the developed model. The memory method of bidirectional LSTM architecture helps in learning the representation and labeling of classes. We used the element-wise averaging method in the hidden layer. The architecture uses the two inputs specified for each token, which are combined and concatenated for output. We added a dropout ratio of 0 : 5 to avoid an overfitting problem that helps in the generations. Also, we used the attention method to incorporate the importance of emotional words in the prediction [9]. The added layer helps in learning the class and vector representation. A more extensive network with diverse data can achieve generalization. The cosine and casual learning method first selects the simple example in each group and removes the problematic example from each group. The model continues to train and converge to the position of the important word by using vector similarity based on the questionnaire and input feature labeling. The ability to fit the data distribution increases performance as the number of labeled instances increases. In addition, the learning expands over time.

Table 4
Precision recall results.

Architectures	Precision	Recall
Baseline	0.82	0.80
LSTM	0.62	0.58
Bidirectional LSTM	0.89	0.88
Bidirectional_LSTM_Attention	**0.90**	**0.89**

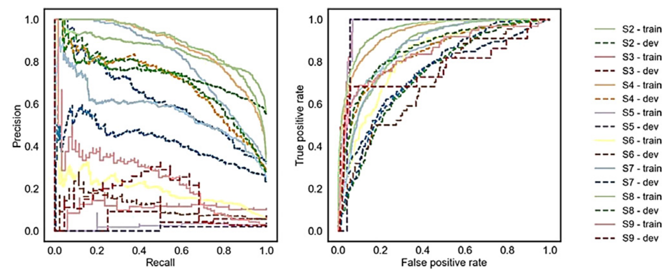

Fig. 2. The feed-forward neural network baseline model.

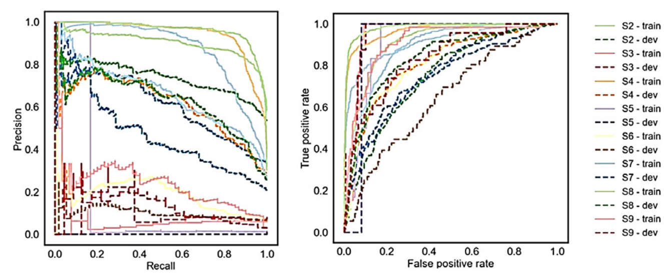

Fig. 3. The performance of bi-directional LSTM model.

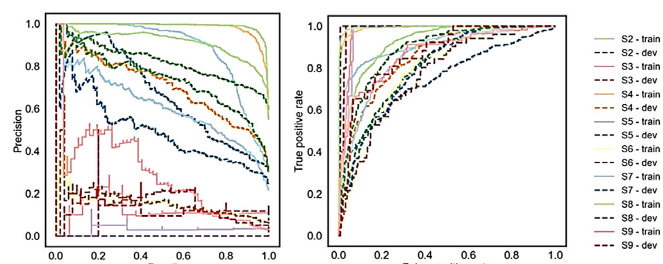

Fig. 4. The performance of bi-direction LSTM with attention.

4. Experimental result and analysis

We used the patient's authorized text to extract the emotional lexicon and then trained the attention network to create our embedding. Next, we trained and labeled the input text based on the PHQ-9 questionnaire using the embedding. For the vectorization method of transfer learning, we used the 300-dimensional glove vector. All nine symptoms represent different vectors as they represent different embeddings. Cosine similarity is used to label each instance. We evaluated the architectures, i.e., feed-forward network, bidirectional LSTM, and attention-based bidirectional LSTM. We evaluated the architecture based on the ROC curve and precision–recall. Adam optimizer is used and hyper-tuning is performed by keeping the learning rate static at 0.0005. Table 3 and Table 4 show the performance of the architecture and attention that performed best with the highest ROC test set. We ran the model for longer epochs and used early stopping to gradually recover the model. We also used the gradient pruning method to avoid gradient problems [36].

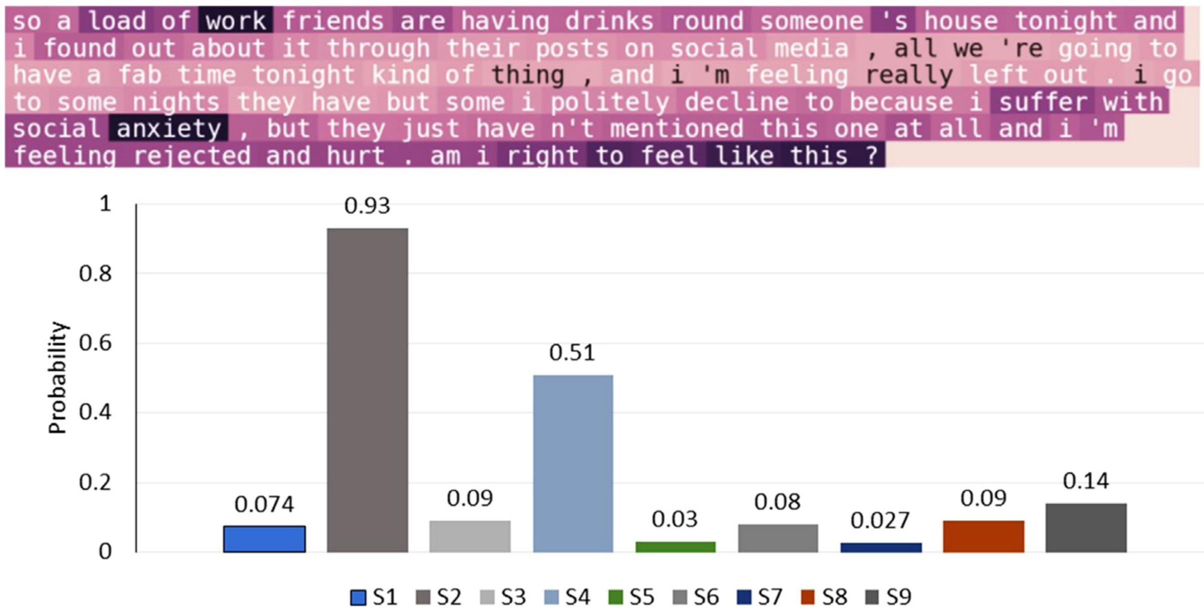

Fig. 5. An example patient-authored text and visualization of depression symptoms extracted by our approach.

The performance of the baseline model is then shown in Fig. 2. The model tends to overfit the values near the upper left corner of the precision–recall curve. The feed-forward architecture of the neural network goes through more epochs but does not find the pattern because it did not preserve the word sequence as relational features. The current data represents substantial emotional and positional importance that was not related by the simpler network.

In the bidirectional model, high accuracy is achieved, as mentioned in Fig. 3. The model runs from past to future and then vice versa. The two hidden states preserve the relational meaning. Forward and backward passage is possible to achieve high performance with low false positives. The BiLSTM model assumes each hidden state based on the previous state. As a result, it has achieved the importance of emotional embedding and performed better than the baseline model. However, the long-range dependencies affect the regularization of the model. Our model reduces the stack size by increasing complex patterns or low similarity values. Therefore, the performance of the model is not optimal.

In Fig. 4, the normalized attention network performs better. Due to the selected weight and position of attention, the effective sentence segmentation has strong words. As a result, the attention network captures the high-quality words related to the symptoms. These results clearly show that the attention model performs better than the LSTM model, indicating that the attention model can effectively capture the temporal variations in event sequences. The training set has a lower error rate, and the model also performs better in the testing set. The high positive rate and low false-negative rate indicate that the model learns local and global attention. The prediction of local word position helps to learn the importance and role of emotional segments in symptom classification. The performance of the model can also be further enhanced by expanding the vocabulary and grammatical permutations. Figs. 2, 3, and 4 represent three approaches. However, the attention network performed better than the baseline model and the bidirectional model. The positional encoder helps to achieve a ROC-AUC of about 0.88.

The attention method helps visualize the weighting and position of words. This helps psychiatrists to examine the trigger for the emotional feeling and its relationship to each type of symptom. In Fig. 5, the model shows that the intense relationship symptoms for two and four are related to low energy and depression, i.e., S1 (feeling down or hopeless) and S4 (feeling tired or low energy). The highlight region can serve as a computerized method for the intervention conducted over the Internet and helps in the assessment of the notes by the psychiatrist.

We used the clustering method for clinical text analysis. A deep neural network used large amounts of training data and automatically identifies features that distinguish between classes using a set of interconnected nodes called a neural network. However, models based on deep neural networks require large amounts of training data and do not have the size required for clinical performance. In addition, models built with a single data source are prone to bias or "overfitting" and do not generalize well to new data, limiting their applicability in new clinical settings.

5. Future work

In the future, we will use intense regulation and improve the emotional lexicon by linguistics and grammatical features. We will try to embed a character level classification with a visualizations method to tune neural architecture based on the embedding size. In [28], the authors used the Latent Dirichlet Allocation model to extract the topics. Then based on the extracted topics, a comprehensive understanding of the emotion is performed. In another work, Glove embedding is used for the detection of negative mental-health emotions like addiction, anxiety, depression, insomnia, stress, and obsessive cleaning disorder (OCD) [29]. The text long term dependency is investigated at the document level.

6. Conclusion

In this study, EANDC is the emotional text expansion method to expand expressive words in text authorized by the patients. We first expanded and trained the word embedding and then used the attention method to cluster the embedding based on cosine similarity. The model starts with simple patterns and the difficult patterns are introduced at each batch size. We visualized the attention word to support and understand the symptom classification and provide better explainability. The bidirectional LSTM with weighted word layers helps to identify the probability distribution of symptoms. The clustering method helps to generate a label and learn the representation. The output results are used as adaptive intervention and generated by using the personalized feedback from the patient.

CRediT authorship contribution statement

Usman Ahmed: Developed the main methodology, Wrote the draft of the manuscript. **Gautam Srivastava:** Validated the developed model, Verified the effectiveness, Proofread English writing. **Unil Yun:** Worked on the formal analysis of the developed model. **Jerry Chun-Wei Lin:** Investigated the main idea, Worked on the conceptualization of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W. Fang, X. Yao, X. Zhao, J. Yin, N. Xiong, A stochastic control approach to maximize profit on service provisioning for mobile cloudlet platforms, IEEE Trans. Syst. Man Cybern. Syst. 48 (4) (2018) 522–534.

[2] Y. Qu, N. Xiong, Rfh: A resilient, fault-tolerant and high-efficient replication algorithm for distributed cloud storage, in: The International Conference on Parallel Processing, 2012, pp. 520–529.

[3] M. Wu, L. Tan, N. Xiong, A structure fidelity approach for big data collection in wireless sensor networks, Sensors 15 (1) (2015) 248–273.

[4] H. Li, J. Liu, R.W. Liu, N. Xiong, K. Wu, T. h. Kim, A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis, Sensors 17 (8) (2017) 1792.

[5] J.A. Naslund, K.A. Aschbrenner, Technology use and interest in digital apps for mental health promotion and lifestyle intervention among young adults with serious mental illness, J. Affect. Disord. Rep. 6 (2021) 100227.

[6] A. Ebadi, P. Xi, S. Tremblay, B. Spencer, R. Pall, A. Wong, Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing, 2020, CoRR.

[7] S.K. Mukhiya, U. Ahmed, F. Rabbi, K.I. Pun, Y. Lamo, Adaptation of idpt system based on patient-authored text data using nlp, in: International Symposium on Computer-Based Medical Systems, CBMS, IEEE, 2020, pp. 226–232.

[8] D.E. Losada, P. Gamallo, Evaluating and improving lexical resources for detecting signs of depression in text, Lang. Resour. Eval. 54 (1) (2018) 1–24.

[9] U. Ahmed, S.K. Mukhiya, G. Srivastava, Y. Lamo, J.C.-W. Lin, Attention-based deep entropy active learning using lexical algorithm for mental health treatment, Front. Psychol. 12 (2021) 471.

[10] S.K. Mukhiya, J.D. Wake, Y. Inal, K.I. Pun, Y. Lamo, Adaptive elements in internet-delivered psychological treatment systems: Systematic review, J. Med. Internet Res. 22 (11) (2020) e21066.

[11] A. Neuraz, I. Lerner, W. Digan, N. Paris, R. Tsopra, A. Rogier, D. Baudoin, K.B. Cohen, A. Burgun, N. Garcelon, et al., Natural language processing for rapid response to emergent diseases: Case study of calcium channel blockers and hypertension in the COVID-19 pandemic, J. Med. Internet Res. 22 (8) (2020) e20773.

[12] S.L. James, D. Abate, K.H. Abate, S.M. Abay, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, J. Abdela, A. Abdelalim, et al., Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017, Lancet 392 (10159) (2018) 1789–1858.

[13] M.G. Mazza, R.D. Lorenzo, C. Conte, S. Poletti, B. Vai, I. Bollettini, E.M.T. Melloni, R. Furlan, F. Ciceri, P. Rovere-Querini, F. Benedetti, Anxiety and depression in COVID-19 survivors: Role of inflammatory and clinical predictors, Brain Behav. Immun. 89 (2020) 594–600.

[14] S.K. Mukhiya, J.D. Wake, Y. Inal, Y. Lamo, Adaptive systems for internet-delivered psychological treatments, IEEE Access 8 (2020) 112220–112236.

[15] T.M. Li, M. Chau, P.W. Wong, P.S. Yip, A hybrid system for online detection of emotional distress, in: Pacific-Asia Workshop on Intelligence and Security Informatics, Springer, 2012, pp. 73–80.

[16] K. Dinakar, E. Weinstein, H. Lieberman, R. Selman, Stacked generalization learning to analyze teenage distress, in: Proceedings of the International AAAI Conference on Web and Social Media, 2014, pp. 1–8.

[17] M.D. Choudhury, M. Gamon, S. Counts, E. Horvitz, Predicting depression via social media, in: Proceedings of the Seventh International Conference on Weblogs and Social Media, 2013, pp. 10–15.

[18] E. Chen, K. Lerman, E. Ferrara, Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus twitter data set, JMIR Public Health Surv. 6 (2) (2020) e19273.

[19] M. McDonnell, J.E. Owen, E.O. Bantum, Identification of emotional expression with cancer survivors: Validation of linguistic inquiry and word count, JMIR Form. Res. 4 (10) (2020) e18246.

[20] H. Lin, J. Jia, Q. Guo, Y. Xue, Q. Li, J. Huang, L. Cai, L. Feng, User-level psychological stress detection from social media using deep neural network, in: Proceedings of the 22nd ACM international conference on Multimedia, 2014, pp. 507–516.

[21] G. Nguyen, S. Dlugolinsky, M. Bobák, V.D. Tran, Á.L. García, I. Heredia, P. Malík, L. Hluchý, Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey, Artif. Intell. Rev. 52 (1) (2019) 77–124.

[22] K. Cho, B. van Merrienboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder–decoder for statistical machine translation, in: A. Moschitti, B. Pang, W. Daelemans (Eds.), The Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1724–1734.

[23] Y. Shao, J.C.W. Lin, G. Srivastava, A. Jolfaei, D. Guo, Y. Hua, Self-attention-based conditional random fields latent variables model for sequence labeling, Pattern Recognit. Lett. 145 (2021) 157–164.

[24] J.C.W. Lin, Y. Shao, Y. Djenouri, U. Yun, Asrnn: a recurrent neural network with an attention model for sequence labeling, Knowl.-Based Syst. 212 (2021) 106548.

[25] J.C.W. Lin, Y. Shao, J. Zhang, U. Yun, Enhanced sequence labeling based on latent variable conditional random fields, Neurocomputing 403 (2020) 431–440.

[26] J.C.W. Lin, Y. Shao, Y. Zhou, M. Pirouz, H.C. Chen, A bi-lstm mention hypergraph model with encoding schema for mention extraction, Eng. Appl. Artif. Intell. 85 (2019) 175–181.

[27] N.P. Jouppi, et al., In-datacenter performance analysis of a tensor processing unit, in: The Annual International Symposium on Computer Architecture, 2017, pp. 1–12.

[28] K. Hou, T. Hou, L. Cai, Public attention about covid-19 on social media: An investigation based on data mining and text analysis, Pers. Individ. Differ. 175 (2021) 110701.

[29] K. Dheeraj, T. Ramakrishnudu, Negative emotions detection on online mental-health related patients texts using the deep learning with mha-bcnn model, Expert Syst. Appl. (2021) 115265.

[30] Q. Zhang, C. Zhou, Y.-C. Tian, N. Xiong, Y. Qin, B. Hu, A fuzzy probability bayesian network approach for dynamic cybersecurity risk assessment in industrial control systems, IEEE Trans. Ind. Inf. 14 (6) (2017) 2497–2506.

[31] J. Lu, J. Yang, D. Batra, D. Parikh, Hierarchical question-image co-attention for visual question answering, in: D.D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems, 2016, pp. 289–297.

[32] K. Xu, J. Ba, R. Kiros, K. Cho, A.C. Courville, R. Salakhutdinov, R.S. Zemel, Y. Bengio, Show, attend and tell: neural image caption generation with visual attention, in: F.R. Bach, D.M. Blei (Eds.), The International Conference on Machine Learning, 37 of JMLR Workshop and Conference Proceedings, 2015, pp. 2048–2057.

[33] T. Luong, H. Pham, C.D. Manning, Effective approaches to attention-based neural machine translation, in: L. Màrquez, C. Callison-Burch, J. Su, D. Pighin, Y. Marton (Eds.), The Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1412–1421.

[34] J. Pennington, R. Socher, C.D. Manning, Glove: Global vectors for word representation, in: The Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1532–1543.
[35] W.G. Charles, Contextual correlates of meaning, Appl. Psycholinguist. 21 (4) (2000) 505–524.
[36] X. Chen, S.Z. Wu, M. Hong, Understanding gradient clipping in private sgd: A geometric perspective, in: Advances in Neural Information Processing Systems, Vol. 33, 2020.

Usman Ahmed is a Ph.D. candidate at the Western Norway University of Applied Sciences (HVL). He has rich experience in building and scaling high-performance systems based on data mining, natural language processing, and machine learning. His research interests are sequential data mining, heterogeneous computing, natural language processing, recommendation systems, and machine learning.

Gautam Srivastava was awarded his B.Sc. degree from Briar Cliff University in the U.S.A. in the year 2004, followed by his M.Sc. and Ph.D. degrees from the University of Victoria in Victoria, British Columbia, Canada in the years 2006 and 2012, respectively. Dr. G, as he is popularly known, is active in research in the field of Cryptography, Data Mining, Security and Privacy, and Blockchain Technology. In his 5 years as a research academic, he has published a total of 200papers in high-impact conferences in many countries and in high-status journals (SCI, SCIE). He is an IEEE Senior Member.

Unil Yun received the M.S. degree in computer science and engineering from Korea University, Seoul, South Korea in 1997, and the Ph.D. degree in computer science from Texas A&M University, College Station, TX, USA, in 2005. He was with the Multimedia Laboratory, Korea Telecom, from 1997 to 2002. After receiving the Ph.D. degree, he was a Postdoctoral Associate for almost one year in the Computer Science Department, Texas A&M University. After that, he was a Senior Researcher with the Electronics and Telecommunications Research Institute. In March 2007, he jointed the School of Electronic and Computer Engineering, Chungbuk National University, South Korea. Since August 2013, he has been with the Department of Computer Engineering, Sejong University, Seoul, South Korea. His research includes data mining, information retrieval, database systems, artificial intelligence, and digital libraries.

Jerry Chun-Wei Lin received his Ph.D. from the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan in 2010. He is currently a full Professor with the Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway. He has published more than 450 research articles in refereed journals (IEEE TKDE, IEEE TCYB, IEEE TII, IEEE TITS, IEEE TNSE, IEEE TETCI, IEEE SysJ, IEEE SensJ, IEEE IOTJ, ACM TKDD, ACM TDS, ACM TMIS, ACM TOIT, ACM TIST) and international conferences (IEEE ICDE, IEEE ICDM, PKDD, PAKDD), 12 edited books, as well as 33 patents (held and filed, 3 US patents). His research interests include data mining, soft computing, artificial intelligence/machine learning, and privacy preserving and security technologies. He is the Editor-in-Chief of the International Journal of Data Science and Pattern Recognition, the Guest Editor/Associate Editor for several IEEE/ACM journals such as IEEE TFS, IEEE TII, ACM TMIS, ACM TOIT, and IEEE Access. He has recognized as the most cited Chinese Researcher respectively in 2018, 2019, and 2020 by Scopus/Elsevier. He is the Fellow of IET (FIET), ACM Distinguished Scientist and IEEE Senior Member.