Adaptation of IDPT system based on patient-authored text data using NLP

Suresh Kumar Mukhiya*, Usman Ahmed*, Fazle Rabbi*,‡, Ka I Pun*,†, Yngve Lamo*
*Western Norway University of Applied Sciences, Norway
{suresh.kumar.mukhiya, usman.ahmed, fazle.rabbi, ka.i.pun, yngve.lamo}@hvl.no
†University of Oslo, Norway
‡University of Bergen, Norway

***Abstract*—*Background:* Internet-Delivered Psychological Treatment (IDPT) systems have the potential to provide evidence-based mental health treatments for a far-reaching population at a lower cost. However, most of the current IDPT systems follow a tunnel-based treatment process and do not adapt to the needs of different patients'. In this paper, we explore the possibility of applying Natural Language Processing (NLP) for personalizing mental health interventions. *Objective:* The primary objective of this study is to present an adaptive strategy based on NLP techniques that analyses patient-authored text data and extract depression symptoms based on a clinically established assessment questionnaire, PHQ-9. *Method:* We propose a novel word-embedding (Depression2Vec) to extract depression symptoms from patient-authored text data and compare it with three state-of-the-art NLP techniques. We also present an adaptive IDPT system that personalizes treatments for mental health patients based on the proposed depression symptoms detection technique. *Result:* Our results indicate that the performance of proposed embedding *Depression2Vec* is comparable to *WordNet*, but in some cases, the former outperforms the latter with respect to extracting depression symptoms from the patient-authored text. *Conclusion:* Although extraction of symptoms from text is challenging, our proposed method can effectively extract depression symptoms from text data, which can be used to deliver the personalized intervention.**

***Index Terms*—Internet-delivered interventions, NLP, tailored intervention, personalization, adaptive treatments, adaptive strategies, adaptive iCBT**

I. Introduction

Internet-Delivered Psychological Treatments (IDPT) has the potential to offer evidence-based mental health treatments for a larger population using less resources. However, despite extensive evidence that Internet Interventions can be an effective means in the treatment of mental health morbidities, many current IDPT systems are tunnel-based, inflexible, and non-interoperable. Lack of adaptability results in more dropouts and lower user adherence [1]. Hence, it is relevant to focus on the factors associated with enhancing user adaptation towards such interventions. One way to enhance user adaptation is to make IDPT systems adaptive such that they change behavior according to several factors (user preferences, user needs, user-health symptoms, user contexts, etc.). In this study, we aim to build an adaptive IDPT system by extracting depression symptoms from patient-authored text using Natural Language Processing (NLP) techniques.

Our hypothesis is based on the assumption that patients' depression symptoms are reflected in their writing when they communicate about their feelings. Based on this hypothesis, we consider that extracting depression-related symptoms from the patient-authored text should allow us to provide tailored intervention. We discuss an adaptive approach of intervention in detail in Section V. The proposed method to extract symptoms from the patient-authored text should help people be aware of the significance of their depression and realize if they should seek medical help. As outlined in [2], shyness, stigma, and embarrassment are key factors preventing people from getting treatments. In such a scenario, people tend to seek help using Internet technologies such as forums and blogs where they can talk about their feelings anonymously. However, due to such stigma related issues, people often do not recognize the severity of their depression. In such use cases and in the IDPT scenario, the extraction of depression symptoms could help patients know about the seriousness of their illness and the urgency to look for professional help.

In this study, we address the following questions: 1) How can patient-authored text data be harnessed to capture patients' depression symptoms? 2) How can we use the results (extracted depression score) to adapt interventions? 3) What are the relevant NLP techniques available to extract semantic similarity between authored text and actual symptoms?

To answer to these questions we perform the following research and development activities:

1) We propose *Depression2Vec*, an open set of depression word vectors/embedding that combines subword information from unlabelled online forums/websites.
2) We outline how the extracted symptoms from patient-authored texts can be used to adapt interventions in online intervention.
3) Finally, an evaluation of how the state-of-the-art NLP encoding libraries *Universal Sentence Encoder (USE)* [3], *Global Vector Representation (GloVe)* [4], *WordNet* [5] is extracting symptoms from patient-authored text.

We discuss the methods, results, adaptive strategies, and related work for adaptive IDPT systems in the rest of the paper.

II. Related work

Funk et al. [6] present a conceptual framework to apply NLP in digital health intervention to support automated analysis

2372-9198/20/$31.00 ©2020 IEEE
DOI 10.1109/CBMS49503.2020.00050

of texts authored by patients as well as messages exchanged between therapists and the patients. The study reports applying the framework to predict binge eating disorder and obtaining a result in an area under a curve between 0.57 and 0.72. However, the framework does not show how can we achieve adaptation in an IDPT environment. The feature engineering process used in the study considers the inclusion of several features, including metadata, word usage, topic models, word embedding, parts of speech, sentiment analysis, and others. In contrast, we use several word-embedding techniques and propose our word-embedding *Depression2Vec*. Yazdavar et al. [7] present a method to detect depressive symptoms, based on the PHQ-9 questionnaire, from Twitter. The study uses a semi-supervised statistical model to evaluate how the duration of these symptoms and their expression on Twitter (in terms of word usage patterns and topic preferences) align with the medical findings reported via the PHQ-9 questionnaire. The work uses two different methods, Latent Dirichlet Allocation (LDA) [8] and a proposed semi-supervised topic modeling over time (ssToT). Several studies have highlighted that the topics learned by LDA are not concrete enough to capture depressive symptom [8], [9]. To empower LDA shortcomings, the authors add supervision to the LDA method by using the terms that are strongly related to the PHQ-9 symptoms as the seeds to the topic clusters and guide the model to aggregate semantically-related terms into the same cluster. Similar to this technique, we use a seed term generation method. The main difference between our seeding model and theirs is that we do not use any dictionary to retrieve synonyms. Instead, we use WordNet to extract not just synonyms but also hypernyms, hyponyms, antonyms. Besides, we apply a different threshold for selecting the words for different methods. Karmen et al. [10] (2015) used a NLP method to detect symptoms of depression from forum texts. While this work focused on keyword density to extract depressive symptoms, we concentrate on finding a more effective approach to obtain depression symptom score from patient-authored text.

III. Methods

As depicted in Fig. 1, we generate embeddings for PHQ-9 questionnaire [11] and patient-authored text using a) three state-of-the-art embeddings and b) the proposed *Depression2Vec* embedding. Once we get the embedding, we use cosine similarity to get the PHQ-9 symptoms score. In this section, we explain these four methods we used to extract the depression symptoms from the patient-authored text data. An example of patient-authored text from an anonymous user is given below:

> *I was feeling slightly better and staying busy over the last week. My mood was a little better, and I wasn't as anxious feeling. Then today, I wake up earlier, not able to sleep much, and feeling anxious again.*

The diagnosis of psychological illness such as depression is more complicated than it appears according to the classification of ICD 10 [12]. One of the main reasons for such complications is the dynamic of the symptoms between patients. Hence, during the diagnosis of any patient, psychiatrists listen for all the hints of symptoms as the patient outlines one's condition and queries to extract additional information. The psychiatrists often use standardized questionnaires such as PHQ-9 and other psychometric tests as the aid to elicit symptoms relatedness. These questionnaires have a similar schema of detecting symptoms, determining the frequency of each symptom, calculating the score by summing the frequencies and using a predefined threshold to classify the intensity of the illness. For example, PHQ-9 has nine distinct symptoms and based on the frequency of the symptoms, it is possible to classify depression to as none, mild, moderate, or severe. We refer to this approach as the *clinical symptom elicitation process (CSEP)*. We aim to automate the CSEP process by extracting depression symptoms from patient-authored text using NLP techniques. Similar to the clinical process, we retrieve the symptoms and frequencies from the text and calculate the overall depression score. The underlying assumption is that the more symptom-related terms are detected in the patient-authored text, the more likely that the patient is suffering from depression. Fig. 1 shows the overall process of extracting PHQ-9 symptoms from the patient-authored text data.

In the following Section, we outline the steps involved in the symptoms extraction process:

A. Psychometric questionnaires (PQ)

We chose the PHQ-9 questionnaire as the standard to measure depression symptoms as it is one of the mostly used psychometric questionnaire for depression [11]. One may use other types of questionnaires for other cases. The PHQ-9 is a nine item depression scale (see Table 1 in supplement material VII in the appendix), which incorporates DSM-V[1]. Each symptom aims at eliciting the patient's condition in one dimension (such as sleeping issue, concentration issue, speaking issue and others.). In a standard CSEP, the psychiatrist extracts each symptom frequency by asking a question followed by frequency of observation of that symptom. Each question has the same options as follows: a) score 0: not at all, b) score 1: several days, c) score 2: more than half the days, and d) score 3: nearly every day. The psychiatrist calculates the overall score based on the patient's answers. Finally, the overall score determines the depression level of the patient.

B. Seed term generation

The seed term generation technique is based on initially fed keywords extracted from the PHQ-9 questionnaire. The initially handpicked terms associated with each symptom and the seed generation algorithm is presented in the appendix. For each symptom in the PHQ-9 questionnaire, we handpick the most relevant keywords. We used WordNet to find synonyms for each word in each symptom. Then, for each synonym, we find the associated hypernyms, hyponyms, and antonyms. For the *Depression2Vec* method, we chose the top five words for

[1]https://www.psychiatry.org/psychiatrists/practice/dsm

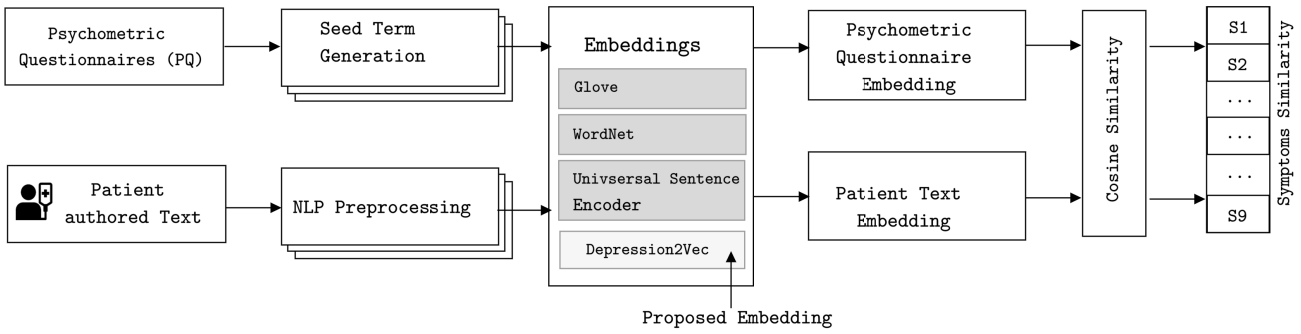

Fig. 1. A workflow architecture for estimating PHQ-9 symptoms from the patient-authored texts. Method 1: Universal Sentence Encoder, Method 2: WordNet, Method 3: GloVe, and Method 4: *Depression2Vec*. We first get the embeddings for both PHQ-9 questionnaire and patient-authored text, and then we use cosine similarity to compute the similarity between each symptom.

each symptom as our empirical results showed the top five words were highly correlated with the original symptom term. Similarly, for Method 3 (*GloVe*), we used only those words with a higher than 80% similarity. This threshold of 80% was empirically chosen so that the results are relevant to the initial handpicked symptoms.

C. Preprocessing

All symptom extraction methods discussed in this paper compare test sentences (patient-authored text) with PHQ-9 symptoms (S1-S9) and find similarity scores against these symptoms. To minimize outliers in the test sentences, we preprocess them using the algorithm provided as the supplement resource in the appendix. As shown in the algorithm, each test sentence goes through the following preprocessing: 1) convert the test sentence into UTF-8 format to maintain encoding consistency, 2) transform all letters into lower case, 3) remove any extra spaces around the sentences, 4) remove symbols (#, +, -, *, =, http, https) present in the sentence, 5) replace a short form of words with full formation, e.g., *won't* is replaced by *will not*, *can't* by cannot and so on.

D. Embeddings

In this section, we outline four different embedding methods used to encode both PHQ-9 questionnaire to generate the seed terms and the test sentences.

1) Method 1: Universal Sentence Encoder (USE) [3], publicly distributed by Tensorflow-hub, is a pre-trained model that helps to encode text into high dimensional vectors. The USE has two variations: a) one with a Transformer encoder, and b) one trained with the deep averaging network (DAN). These two variations have a trade-off of accuracy and computational resource requirement. While the one with a Transformer encoder has higher accuracy, it is computationally more intensive. The one with DAN encoding is computationally less expensive and with little lower accuracy [13]. In our approach, we chose to use transformer encoder in order to save computational resources.

2) Method 2: WordNet [5] is a machine-readable lexical database for English developed at Princeton University. The database consists of three separate databases respectively for nouns, verbs, and adjectives and adverbs. These databases are grouped into sets of cognitive synonymous called synsets, each expressing a unique concept. Synsets are interlinked using conceptual-semantic and lexicon relations. The words that are in the same synset are synonymous.

3) Method 3: Global Vector Representation (GloVe) [4] is an unsupervised learning algorithm for getting vector representation for words. It takes a corpus of text as input and transforms each word in that corpus of text into a position in a high-dimensional space. This technique essentially places semantically similar words together in a high-dimensional space. The pre-trained (see Table I) vectors are available for public use from the Standford official website[2].

4) Method 4: One novelty of our work is the proposed technique *Depression2Vec* (*emotionally aware word sense embedding*) to predict the appropriate symptoms from the text authored by depressed person. There exists an extensive literature in the NLP community available for emotion recognition. However, much less attention has been given to the representation of hybrid emotional knowledge (word embedding, sentence embedding, and word sense lexicon) and contextually-diverse embedding. Hence, in this method, we propose to use an embedding that incorporates contextually-diverse embedding that, in turn, combines depression lexicons as well as emotional knowledge by using online forums texts.

In emotion analysis, the text consists of ordered words that represent different contexts. A vector space is used to project the text document *(set of words)* to different contexts. The context is used to identify how a learning model will process the words and sentences. The bag of words *(BoW)* model is used to learn about the sentence structure in the original text document. It is a method to model text into a numerical representation. The BoW model processes the text without

[2]http://nlp.stanford.edu/data/glove.6B.zip

considering the word order, semantic structure, and contexts. Therefore, the BoW model is not able to capture the complex semantic structure of the text in its processing. Another issue with BoW model is its high dimensions *(large feature space)* and sparsity *(a large number of distinct features)*.

Depression word vector embedding using emotional lexicon: Word embedding has been introduced to overcome the issues with the BoW model (see Section III-D4). The word embedding uses unsupervised methods to embed any text (large corpus) into a vector space that represents the context (semantic meaning to a word). The state of art word embedding methods includes the *C and W model*, the *Skip-Gram Word2Vec model* [5], the *continuous bag-of-words (CBOW) model* [14] and the *GloVe model* [4]. The principle behind the word embedding method is the distributional vector hypothesis that *"You shall know a word by the company it keeps"* [15]. The statistical information *(i.e., co-occurrence frequencies)* of the words is used to embed many linguistics patterns and regularities in the vector space. The learned model produces word vector space in which unique word in a corpus of texts is assigned to a corresponding fixed-length vector, and similar-context words in the training corpus are located in close proximity. The word distribution in the vector space and its feasibility has been confirmed in many experiments [15]. However, most existing word embedding algorithms produce domain-specific contextual preserved vector space [4], [15].

The learned vector representation from algorithms *(pre-trained model)* is not applicable for emotional analysis. The contextual knowledge (large corpus pre-trained word2vec models) are built using public sentiment lexicons *(Wikipedia texts)*, and sentiment knowledge *(Twitter data)* is derived by using the word sense in the context and co-occurrence frequency. For example, ***"happy"*** and ***"sad"*** can represent a similar emotional context, i.e., ***"feelings"***, but both words are representing different emotions and mental state. Therefore, it is incorrect to infer emotional information by using the large corpus pre-trained word embedding that uses word co-occurrence method.

The lexicon-based method shows promising results due to the existence of a high-quality emotional lexicon. A fine-grained emotional lexicon embedding can improve the accuracy in the classification of various symptoms. In order to overcome aforementioned issue with word embedding, we proposed the contextually *(depression)* aware word embedding *(Depression2Vec)* that uses the traditional word embedding model, i.e., skip-gram, and uses the domain-specific contextual knowledge to learn the complex meaning of a word. To generate the word embeddings, a set of documents and emotional lexicons is given to the model.

Preprocessing: The objective of this method is to find the semantically related words for each word in the corpus. However, most of the words in the text corpus do not infer opinion, so part of speech (*noun*, *verb*, *adverb* and *adjective*) based reduction method is adopted. A corpus D, consisting of a set of texts, $D = \{d_1,\ d_2, \ldots, d_n\}$. As mentioned in Algorithm 1, for each text document in the corpus, we perform text processing (explained in Section III-C) and extracted part of speech. The *WordNet* knowledge graph is used to extract synonyms, antonyms, hypernyms, and physical meaning for each extracted part of speech. Then, the obtained emotion set, $E = \{E_1, E_2, \ldots, E_K\}$ is added into the original process text to make a domain-specific contextual corpus where E_i represents the extracted emotion words for each document. A vocabulary $V_E = \{v_1,\ v_2, \ldots, v_m\}$ is a set of unique terms extracted from E. For each V_i, word representation is mapped from the trained $Depression2Vec$ model, and then a set of word vectors is derived from the set V, i.e., $T = \{T_1,\ T_2, \ldots, T_m\} \in \mathbb{R}^{m \times \delta}$ where δ is the word vector dimension and m is the size of the vocabulary. In this study, the skip-gram model is used for word embedding (Depression2Vec) training. The final word list for each text document in the corpus is used by the skip-gram model to learn word vector representation which gives the vector representations, $T_V = \{T_{V_1},\ T_{V_2}, \ldots, T_{V_K}\}$.

Construction of sentence embedding: The final text embedding is extracted by averaging over all the word vector in the text. The resultant text vectorization (Depression2Vec) is a joint learning of word sense and emotional knowledge in the same latent space. By using the trained word embedding (Depression2Vec), the input text is converted into the word vector representation, and all the nine symptoms from PHQ-9 questionnaire lexicons are also converted into the corresponding word vector representation.

E. Similarity

Given two vectors **t**, **e**, where **t** representing vectors encoded from test sentence (patient-authored text) and **e** any particular PHQ-9 symptom, we use cosine similarity to compare these two vectors. The similarity measure is defined as follows:

$$\cos(\mathbf{t}, \mathbf{e}) = \frac{\mathbf{te}}{\|\mathbf{t}\|\|\mathbf{e}\|} = \frac{\sum_{i=1}^{n} \mathbf{t}_i \mathbf{e}_i}{\sqrt{\sum_{i=1}^{n} (\mathbf{t}_i)^2} \sqrt{\sum_{i=1}^{n} (\mathbf{e}_i)^2}} \quad (1)$$

A high cosine value indicates that the patient-authored text is closely related to the topic model associated with any particular PHQ-9 symptom.

IV. Results

We used three different algorithms to compare with the proposed embedding. Our embedding produced a significantly higher agreement with annotated data. Out of 15044 posts collected from online forums, websites, and social media, 100 posts were annotated by two humans and used as the testing set. The remaining 14944 posts were used for creating the *Depression2Vec* embedding (see Table II). We used the Amazon Mechanical Turk to hire two human annotators who were familiar with PHQ-9 and clinical text assessment. We asked human annotators to rate the posts according to the PHQ-9 rating method, such that 0 indicates not depressed, 1 mildly depressed, 2 moderately depressed, and 3 severely depressed. For each symptom, we converted this annotation into binary class such that 0 indicates the absence of symptom, and

Algorithm 1: Algorithm to train Depression2Vec

input : Corpus
output: Trained embedding

foreach *doc* $d \in corpus$ **do**
  d $\leftarrow algorithm_1(d)$;
  foreach *term* $t \in d$ **do**
    synonyms $\leftarrow$ wordnet.synonyms(t);
    foreach *synonym* $w \in$ *synonyms* **do**
      terms $\leftarrow$ wordnet.hyperonym(w);
      terms $\leftarrow$ wordnet.hyponym(w);
      terms $\leftarrow$ wordnet.antonyms(w);
    end foreach
  end foreach
  vocabulary $\leftarrow$ terms;
end foreach
$Embedding \leftarrow word2vec_{dep}(vocabulary, corpus, window = 2)$;
return seed_terms

TABLE I
MODEL EMBEDDING DETAILS FOR ALL FOUR METHODS USED IN THIS STUDY

Model	Embedding Corpus	Embedding Size
Depression2Vec	15043 (training set)	300
Universal sentence encoder	Pre_trained	512
Glove	Pre_trained	300
Wordnet	-	-

TABLE II
METADATA ABOUT THE DATASET USED FOR TRAINING AND TESTING

Type	Statistics
Corpus size (Number of posts collected)	15044
Number of sentences	133524
Average sentences per post	8.87
Number of words	3502245
Average words per post	232
Training set size (Number of posts)	14944
Testing set size (Number of posts)	100

TABLE III
ANNOTATION DONE BY TWO HUMANS A AND B SHOWING SYMPTOMS AND THEIR FREQUENCY OF PRESENCE IN THE TEST SENTENCES.

Symptom	Human A	Human B
1	37	13
2	82	92
3	17	13
4	39	2
5	6	3
6	30	25
7	11	3
8	3	0
9	24	23

the other indicates the presence of symptom. The depression annotation for test data is mentioned in Table III. Our primary hypothesis, as observed in this model, was *Depression2Vec* would generate more precise embedding and results compared to the state-of-the-art embeddings. We used the blind annotated test data and computed evaluation metrics (true positive rate, false-negative rate, accuracy, and f-measure) using all four methods. The results, as shown in Table IV, indicate WordNet and Depression2Vec perform better compared to the other methods. However, the results show a drop in detection for the symptom five using WordNet. The proposed algorithm, *Depression2Vec*, significantly performs better for depression detection. In comparison with other symptoms, the WordNet and *Depression2Vec* showed similar accuracy. However, the results show no improvement with existing pre-trained models.

TABLE IV
DEPRESSION2VEC COMPARISON WITH DIFFERENT MODEL: *UE - Universal encoding, Glove - Global vector representation, WN - WordNET, Dep2VEC - Depression vector, TPR - True positive rate, FNR - False negative rate, ACC - Accuracy, Fscore - Fmeasure. The bold text indicates models with better accuracies.*

Symptoms	Methods	TPR	FNR	ACC	Fscore
1	UE	0.00	1.00	0.59	0.44
	Glove	0.00	1.00	0.59	0.44
	WN	1.00	0.00	1.00	**1.00**
	DP2Vec	0.98	0.02	0.99	**0.99**
2	UE	0.00	1.00	0.04	0.00
	Glove	0.00	1.00	0.04	0.00
	WN	1.00	0.00	1.00	**1.00**
	DP2Vec	0.97	0.03	0.97	**0.97**
3	UE	0.00	1.00	0.76	0.66
	Glove	0.00	1.00	0.76	0.66
	WN	1.00	0.00	1.00	**1.00**
	DP2Vec	0.92	0.08	0.98	**0.98**
4	UE	0.00	1.00	0.61	0.46
	Glove	0.00	1.00	0.61	0.46
	WN	1.00	0.00	1.00	**1.00**
	DP2Vec	0.95	0.05	0.98	**0.98**
5	UE	0.00	1.00	0.93	0.90
	Glove	0.00	1.00	0.93	0.90
	WN	0.43	0.57	0.96	**0.95**
	DP2Vec	0.86	0.14	0.99	**0.99**
6	UE	0.00	1.00	0.57	0.41
	Glove	0.00	1.00	0.57	0.41
	WN	1.00	0.00	1.00	**1.00**
	DP2Vec	0.98	0.02	0.99	**0.99**
7	UE	0.00	1.00	0.87	0.81
	Glove	0.00	1.00	0.87	0.81
	WN	1.00	0.00	1.00	**1.00**
	DP2Vec	0.92	0.08	0.99	**0.99**
8	UE	0.00	1.00	0.97	0.96
	Glove	0.00	1.00	0.97	0.96
	WN	1.00	0.00	1.00	**1.00**
	DP2Vec	1.00	0.00	1.00	**1.00**
9	UE	0.00	1.00	0.69	0.56
	Glove	0.00	1.00	0.69	0.56
	WN	1.00	0.00	1.00	**1.00**
	DP2Vec	0.94	0.06	0.98	**0.98**

V. ADAPTATION STRATEGY

The study [16] discusses several adaptive strategies, including rule-based adaptation, policy-based adaptation, goal-driven adaptation, and predictive-algorithm in IDPT systems. In this section, we illustrate the amalgamation of the predictive algorithm and rule-based techniques to adapt to psychological treatments. We use the same scenario for Internet-Delivered Psychological Treatment (IDPT), as outlined in [16]. As discussed in the paper, any psychological treatment system has a case (such as depression, social anxiety, and others).

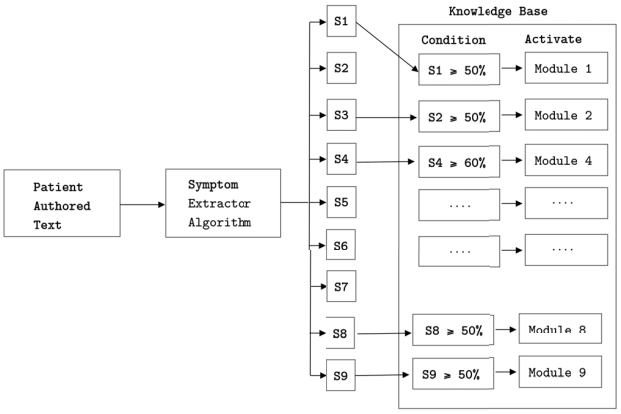

Fig. 2. The figure shows the adaptation of intervention modules based on symptom extractions techniques. It takes patient-authored text as input and gives the nine similarity scores as output. Based on the depression symptom similarity score, the system activates or deactivates related modules.

Each case has one or more modules (eating disorder, concentration, sleeping disorder, and others). Each module has one or more tasks, educational materials, and exercise. The exercise can be computerized exercise (e.g., question answers, quiz and feedback) and physical activity (e.g., breathing exercise and walk). To illustrate adaptive strategies, we assume a case for depression with nine modules, each module corresponding to symptoms used in the PHQ-9 questionnaire, $\forall S_i \in \{S_1, S_2, \ldots, S_9\}$, $\exists M_i \in \{M_1, M_2, \ldots, M_9\}$ where each M_i is a module providing psycho-education about the corresponding symptom S_i in the PHQ-9 questionnaire.

Adapting the intervention modules and feedbacks

To adapt and personalize module for each individual, we compute symptom similarity from the patient-authored text. The symptom extractor algorithm takes a patient-authored text as input and provides a similarity score for the nine symptoms as output (see figure 2). The symptom similarity score is used to recommend appropriate treatment modules. Once we get the symptom similarity score, we can check against the *knowledge base*. We assume that psychiatrists use their domain expertise to maintain a rule-based engine that shows which modules are relevant for what types of patients. For example, if a patient is showing a symptom of higher concentration problem, the patient is offered to go through the concentration module. A study by Bewick et al. reports feedbacks have thought to increase user adherence [17] for sixty-five percent of intervention participants. Personalizing such feedback based on their interaction with the IDPT system is required to improve user adherence [18]. Based on the proposed technique in this paper, we provide an adaptive IDPT system capable of providing automated feedback and remainders. For instance, using the proposed method, one can develop a system that can generate an automated alert to notify a therapist when a specific problem is identified in a patient-authored text.

VI. Discussion

The application of NLP to clinical text to extract relevant information has been thriving in recent years. A recent systematic literature review by Dreisbacha [19] et al. in 2019, synthesize the literature on the usage of NLP and text mining to extract symptoms from patient-authored text data. We can see two crucial observations in the review. First, researchers attempt to obtain symptoms related to several healthcare conditions, including pain, fatigue, cognitive function, affective mood, digestive symptoms, and others. According to the review, a limited number of studies considered the extraction of mental health symptoms from patient-authored text. Second, based on the review, none of the studies attempted to adapt the IDPT system based on symptom extraction. In this paper, we aim to fulfil this gap and use the knowledge of symptoms that the patients exhibit in the text authored by them, in the adaptation of IDPT.

A. Limitations

In this section, we point out some of the limitations of our study. The accuracy of the proposed algorithm is highly dependent on the datasets used for creating embeddings. Therefore, the text used for creating the embeddings must be rich in vocabulary. The absence of particular words in the patient-authored text would decrease the symptom similarity score. Again, the increase in the training dataset would increase the computation cost for training. We evaluated the performance of embedding based on a human-annotated dataset, which might have been biased. Involving domain experts such as linguists and psychiatrists for evaluating the performance of the embedding would significantly increase the quality of evaluation. Furthermore, our approach assumes that patient-authored texts are plain, simple, and do not contain complex natural language constructs:

- complex negated sentences (*e.g., The doctor says he does not have any explanation for my mild depression.*)
- conditional sentences (*e.g., If I experience sleeping issue, I will consult a psychiatrist.*),
- uncertain sentences (*e.g., I'm not sure if I have a speaking issue or if I interrupt others too much.*), and
- statements about history or family history (*e.g., I do not exhibit any symptoms, particularly; however, my mother has a severe headache and concentration issues.*)

B. Implications

Our objective is to use the symptoms extraction technique to tailor Internet interventions based on the patients' current context. To achieve such adaptability, a substantial accuracy of the symptoms extraction technique can work. Moreover, the IDPT is generally targeted for people with low or mild mental health issues. Hence, a false positive recommendation of a module or any other psychoeducation material will not have any impact on the patients' health. Instead, the recommendation of treatment modules based on the symptoms captured from the text authored by the patient can give a personalized and engaging treatment experience rather than

a tunnel-based general intervention. Moreover, the extraction of symptoms keywords from text authored by patients can provide an overview of the patients' mental health status over time. Any symptoms for severe depression or similar patterns can be detected early and reported to the therapists.

C. Future work

In this study, we incorporated text crawled from several forums/websites related to mental health intervention. While prescreening of the documents was done manually by the authors, it was not validated with the domain experts. Hence, one of the immediate future work is to verify the initial corpus with domain experts such as psychiatrists and linguistics. Another improvement in the embedding would be to incorporate Internet slang, correct spellings, and abbreviations before creating embedding. We did not attempt to validate and study the effects of complex negation within our initial study's time limits. Hence, one of the enhancements of this study is to verify and detect the presence of negation. We expect to improve the performance of the proposed embedding by identifying conditional sentences, uncertain sentences, and NLP dependency in the next phase of our study.

VII. Conclusion

We pursued an NLP approach to perform adaptation in IDPT systems for two main reasons: a) IDPT deals with a significant amount of texts in the form of computerized exercises for psycho-education, b) NLP method can provide an elegant way to adapt the intervention on the one hand and provide personalized feedback on the exercises. Several studies [1] have reported that lack of personalized feedback on their interventions was one of the leading causes of high dropouts. Hence, in this paper, we present an NLP based adaptive strategies to adapt intervention based on the symptoms exhibited by the patients' through text data. To extract symptoms from such patient-authored text data, we evaluate three different state-of-the-art NLP techniques and propose a novel technique. The results show that both WordNet and the proposed embedding *Depression2Vec* captures depressive symptoms better than other methods. However, there are several challenges, as presented by the study [20] associated with the detection of symptoms from text data. Outlining all complexities and challenges is beyond this paper's scope and is kept as one of the immediate future work.

Acknowledgment

This publication is a part of the INTROducing Mental health through Adaptive Technology (INTROMAT) project, partially funded by Norwegian Research Council (259293/o70).

Appendix

Supplement resources can be obtained from the following link *https://github.com/sureshHARDIYA/phd-resources/blob/master/Papers/AdaptationNLP/supplement.pdf*

References

[1] A. e. a. Konrad, "Finding the adaptive sweet spot: Balancing compliance and achievement in automated stress reduction," in *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2015-April, (New York, New York, USA), pp. 3829–3838, Association for Computing Machinery, 4 2015.

[2] S. K. Mukhiya, F. Rabbi, K. I. Pun, and Y. Lamo, "An architectural design for self-reporting e-health systems," in *Proceedings - 2019 IEEE/ACM 1st International Workshop on Software Engineering for Healthcare, SEH 2019*, pp. 1–8, Institute of Electrical and Electronics Engineers Inc., 5 2019.

[3] D. e. a. Cer, "Universal Sentence Encoder," *EMNLP 2018 - Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Proceedings*, pp. 169–174, 3 2018.

[4] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2014.

[5] G. Miller, C. Fellbaum, J. Kegl, and K. Miller, "WordNet: An Electronic Lexical Reference System Based on Theories of Lexical Memory," *Revue québécoise de linguistique*, vol. 17, pp. 181–212, 5 2009.

[6] B. e. a. Funk, "A Framework for Applying Natural Language Processing in Digital Health Interventions.," *Journal of medical Internet research*, vol. 22, p. e13855, 2 2020.

[7] A. H. Yazdavar, H. S. Al-Olimat, M. Ebrahimi, G. Bajaj, T. Banerjee, K. Thirunarayan, J. Pathak, and A. Sheth, "Semi-Supervised Approach to Monitoring Clinical Depressive Symptoms in Social Media," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2017*, 2017.

[8] D. Andrzejewski and X. Zhu, "Latent Dirichlet Allocation with Topic-in-Set Knowledge *," 2009.

[9] D. Ramage, C. D. Manning, and S. Dumais, *Partially Labeled Topic Models for Interpretable Text Mining*. 2011.

[10] C. Karmen, R. C. Hsiung, and T. Wetter, "Screening internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods," *Computer Methods and Programs in Biomedicine*, vol. 120, pp. 27–36, 6 2015.

[11] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The PHQ-9: Validity of a brief depression severity measure," *Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, 2001.

[12] N. e. a. Americans, "The ICD-10 Classification of Mental and Behavioural Disorders," *IACAPAP e-Textbook of child and adolescent Mental health*, vol. 55, no. 1993, pp. 135–139, 2013.

[13] X. Liu, J. Meehan, W. Tong, L. Wu, X. Xu, and J. Xu, "DLI-IT: A deep learning approach to drug label identification through image and text embedding," *BMC Medical Informatics and Decision Making*, 2020.

[14] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, International Conference on Learning Representations, ICLR, 2013.

[15] W. G. Charles, "Contextual correlates of meaning," *Applied Psycholinguistics*, vol. 21, no. 4, pp. 505–524, 2000.

[16] S. K. Mukhiya, J. Wake, Y. Inal, and Y. Lamo, "Adaptive Systems for Internet-Delivered Psychological Treatments (In review).," *IEEE Access*, 2020.

[17] B. M. Bewick, K. Trusler, B. Mulhern, M. Barkham, and A. J. Hill, "The feasibility and effectiveness of a web-based personalised feedback and social norms alcohol intervention in UK university students: A randomised control trial," *Addictive Behaviors*, vol. 33, pp. 1192–1198, 9 2008.

[18] Z. Hilvert-Bruce, P. J. Rossouw, N. Wong, M. Sunderland, and G. Andrews, "Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders," *Behaviour Research and Therapy*, vol. 50, pp. 463–468, 8 2012.

[19] C. Dreisbach, T. A. Koleck, P. E. Bourne, and S. Bakken, "A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data," 5 2019.

[20] S. Perera, A. Sheth, K. Thirunarayan, S. Nair, and N. Shah, "Challenges in understanding clinical notes: Why NLP engines fall short and where background knowledge can help," in *International Conference on Information and Knowledge Management, Proceedings*, (New York, New York, USA), pp. 21–26, ACM Press, 2013.