A Deep Learning Approach to Mining the Relationship of Depression Symptoms and Treatments for Prediction and Recommendation

Juntian Lin, Guan Luo, Zhu Zhan and Xiaoyao Guan

Abstract Background: Behavior regulation and clinical intervention have a significant effect on depression treatments. This study aims to make a comparison between behavior regulation and clinical intervention for depression based on a large-scale dataset. Methods: We collect user-reported data from an online survey tool including depression symptoms, treatments and effectiveness of treatments (n = 91873). A deep learning approach is used to build an effective model to evaluate the effects on treatment methods for depression. The Skip-gram model is chosen to generate meaningful vector representations of symptoms and methods. Precision, recall and F1 score are calculated to evaluate the model performance. Results: Unidirectional model achieves higher F1 score than non-unidirectional model (0.71 vs. 0.63). The behavior regulation is better than the clinical intervention for mild depression symptoms. However, the clinical intervention for moderate or severe depression symptoms has obvious advantages. Conclusions: These experiments prove that the symptoms have unidirectional influence on the choice of regulatory methods. The behavior regulation and clinical treatment have different advantages for depression. These findings could help clinicians to choose better depression treatments.

Keywords Depression · Deep learning · Behavior regulation

J. Lin · G. Luo (✉) · Z. Zhan · X. Guan
National Laboratory of Pattern Recognition, Institute of Automation, CAS, Beijing, China
e-mail: gluo@nlpr.ia.ac.cn

J. Lin
e-mail: juntian.lin@ia.ac.cn

K. J. Kim and N. Baek (eds.), *Information Science and Applications 2018*,
Lecture Notes in Electrical Engineering 514,
https://doi.org/10.1007/978-981-13-1056-0_42

1 Introduction

Depression (Major depressive disorder) is a common but serious mental disorder. The World Health Organization (WHO) shows that more than 300 million people in the world are affected by depression [1]. Persistent feelings of sadness and lack of interest or pleasure in activities are defined to be the most important features of depression in the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V). Besides the two above symptoms, DSM-V lists the other seven typical depression symptoms (Table 1).

The depression treatment can be divided into antidepressant and psychology. Antidepressant is the most common choice for depression treatment. However, it cannot treat depression effectively because of its poor compliance and persistence. Sansone and Sansone [2] suggest that approximately half of the depressive patients cannot persist in antidepressant treatment. To enhance the efficacy of depression treatment, psychological treatment is usually used as an important complement to antidepressant. Psychology can also improve the quality of life assessment [3].

In addition to the above clinical treatments, the use of behavior regulation, which contains energy regulation, emotion regulation and so on, is also regarded as an important aspect of recovery for depression [4]. Patients with higher level of depressive symptoms prefer using rumination and suppression, rather than reappraisal [5]. Many studies focus on the mechanism by which behavior regulation affects depression. However, the public data on the efficacy of self-regulating is small.

The aim of this study is to examine the effect of behavior regulation and clinical intervention for depression on a large-scale dataset. We build a model to predict whether the symptoms of the participants can be alleviated, so that we can use this model to evaluate the effect of each method in any condition.

Table 1 Typical depression symptoms (DSM-V)

Symptom	Description
A	Depressed mood or irritable nearly every day
B	Diminished interest or pleasure in activities
C	Significant weight loss (5%) or change in appetite
D	Insomnia or hypersomnia
E	Psychomotor agitation or retardation
F	Fatigue or lack of energy
G	Feelings of inappropriate guilt or worthlessness
H	Decreased ability to concentrate, or indecisiveness
I	Suicidality

2 Related Work

Many studies have shown that behavior regulation and clinical intervention have a significant effect for depression treatment [6–8]. These studies form the basis of making comparisons on different methods through a large-scale data. Martin and Dahlen [9] show that emotion regulation strategies, especially positive reappraisal and rumination, is valuable factors to predict negative emotions like depression. Nevertheless, depressed people cannot decrease negative feelings in consequence of using emotion regulation strategies [10]. These studies imply that methods are affected by depression symptoms. We use a deep learning approach to enrich the information about methods and to compare behavior regulation with clinical intervention.

3 Method

3.1 Dataset Description

We use an online survey tool to collect data from voluntary participants. The dataset eventually contains 91873 valid data. Three aspects of data are recorded for every participant, i.e. depressive symptoms, methods and effect. The Patient Health Questionnaire (PHQ-9) is a self-report inventory designed for depression screening [11]. We use PHQ-9 to obtain nine depressive symptoms, denoted as letter A-I. Each depression symptom is divided into four levels: “Not at all”, “Several Days”, “More than half the days”, “Nearly every day”, and is recorded as 0–3, like A1 means “Depressed mood in several days”. At the same time, we list eleven commonly used depression adjustment methods, denoted as a-k, for participants to follow, and evaluate the effect of the chosen methods (0-good, 1-bad) (Table 2).

Table 2 Depression adjustment methods

Method	Description
a	Hot showers
b	Exercising
c	Talking
d	Suppression
e	Journey
f	Eating
g	Reading
h	Sex life
i	Shopping
j	Reappraisal
k	Seeking medical advice

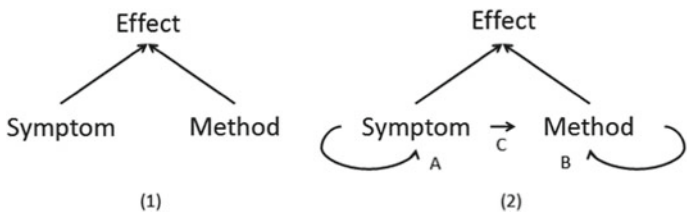

Fig. 1 Research approaches to depression

All eligible participant should answer all of questions independently. Considering that the purpose of this study is to predict the result of a combination of methods for a patient with some depression symptoms, we drop out those data with no symptoms or no methods. At last, we have 91873 valid records. Participants are cognitively healthy men and women aged in 18–70. The depression mean score for males (n = 27018) is 9.90 ± 6.54 and the mean score for female (n = 64855) is 9.92 ± 6.36. There is no significant difference in PHQ-9 score between males and females ($t = -0.58$, $df = 49281$, $p = 0.57$). More than half of participants (52.11%) feel that those methods can decrease their depressive symptoms.

3.2 Study Design

To predict the patient's condition is improved or not after he/she uses the clinical or self-regulating methods, a common approach in Fig. 11 is to treat each depression symptom and method as an independent feature and then use a classification algorithm to solve the problem. This structure is simple but it does not consider the interaction between features. Thus, we design a more reasonable structure to describe the relationship among symptoms, methods and effect as shown in Fig. 12.

First, different symptoms and methods are correlated in their respective sets. Second, the symptoms have influence on the choice of regulatory methods [12]. The first problem can be solved by using the traditional interactive terms in Fig. 11. However, the second one cannot do this because the interaction terms cannot describe the unidirectional relationship. Since depression symptoms have effects on the chosen of methods, the opposite does not hold true. In order to model the unidirectional relationship between symptoms and methods, we propose a word embedding approach to model the relationship.

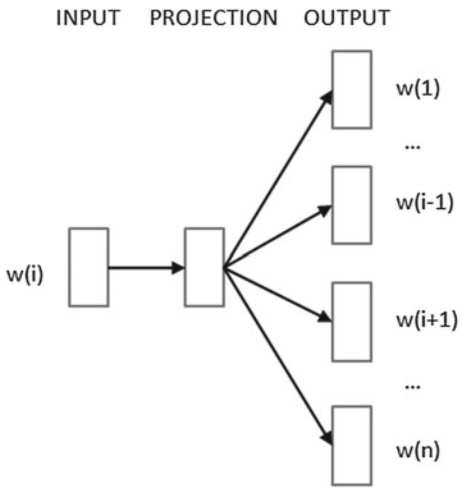

Fig. 2 Skip-gram model structure

3.3 Learning Symptom and Method Vectors

We treat the symptoms and methods in PHQ-9 as words, and the combination of individual symptoms and methods as a sentence. The efficacy of the methods is the learning target. Thus, by training a neural network using the word embedding algorithm (Word2Vec), we learn the interrelationships between symptoms and methods. In particular, we make an improvement on the Word2Vec's algorithmic to learn unidirectional information between symptoms and methods (Fig. 2).

Mikolov [13] shown that Skip-gram model in Word2Vec can capture more semantic information than Continuous Bag-of-Words(CBOW). Hence, in our approach we use Skip-gram model to learn symptom and method vectors. In symptom vector training process, we select one sample at each epoch, then we take turns choosing depression symptom w(i) as input and the rest of the symptoms as the model output w(1),…, w(i − 1), w(i + 1),…, w(n), where n is the number of symptoms in this sample. The method vector training process is the same.

The difference between our approach and the Skip-gram prototype lies in the fact that our vector embedding models do not have the concept of a sequence. To reduce the sensitivity of the model to these orderings, we shuffle the symptoms and methods each time in training process. Since there are not many symptoms or methods to choose from, the native loss is calculated directly.

3.4 Model Building and Evaluation

For classification, we use the same linear softmax classifiers as fastText to classify the samples [14]. For a training sample, we summed up the symptom vectors as features

Table 3 Algorithm performance for different learning methods

	Precision	Recall	F1 score
Logistic regression	0.61	0.60	0.60
Non-unidirectional	0.69	0.58	0.63
Unidirectional	0.75	0.67	0.71

of the symptom part. Similarly, the method vectors were summed up as the features of the method part, and then the two vectors were combined as the final sample features. The dataset is randomly divided into training set and test set according to 7:3.

We use precision, recall, and F1 scores to evaluate the performance. In a classification task, the test set could be divided into four types as follows: true positives (TP), false positives (FP), true negatives (TN), false negatives (FN). If the effect of treatment is proven present in a participant, the given model also indicates the effect of treatment in this participant, the result of the model is considered true positive. The other three types are similar to the definition of TP. Precision is the fraction of TP in the sum of TP and FP. Recall is the fraction of TP in the sum of TP and FN. F1 score is the harmonic mean of precision and recall.

4 Result

4.1 Learning Unidirectional Relationship

We choose logistic regression method as the baseline model to compare the performance. To learn the symptom and method vectors, the dimensionality of symptom and method vectors is set to 20. The results are shown in Table 3.

We see that the F1 score is 0.60 for baseline model, and 0.63 for non-unidirectional model. These two models are closer in the F1 score. Method vectors added symptoms information during the training and the F1 score is 0.71 for unidirectional model. The result indicates that the method is affected by the symptoms.

4.2 Symptom Vector Analysis

By computing the cosine similarity between different symptoms, the symptoms of depression show a stratification characteristic, mild ("Several Days"), moderate ("More than half the days") and severe ("Nearly every day") symptoms are clustered based on the severity of symptoms. For example, Table 4 shows A1/A2/A3 and its nearest nine symptoms.

Table 4 The nearest symptoms (context) to symptom A (center)

Center	Context
A1	B1, F1, G1, H1, C1, E1, D1, I1
A2	C2, G2, F2, B2, D2, E2, H2, I2
A3	D3, F3, E3, C3, G3, B3, H3, I3

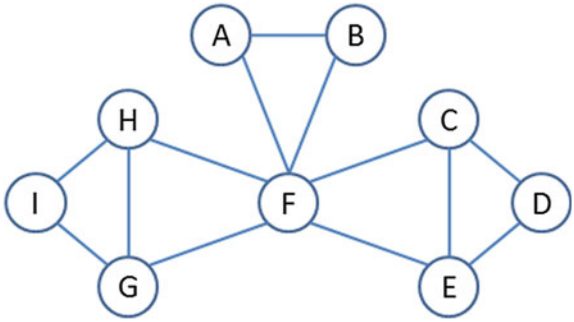

Fig. 3 The mild symptom network

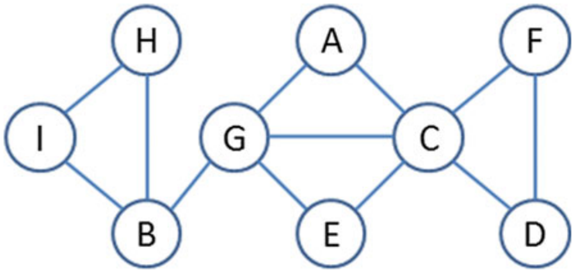

Fig. 4 The moderate symptom network

Based on the fact of stratification, we construct the symptom networks for the symptoms of grade1, grade2, and grade3 respectively. The symptom network has nine vertices corresponding to nine symptoms, the edge between vertices means the symptoms relationship. The symptom network is an undirected graph because there is no direction for similarity.

The symptom network is established as follows: the network vertices are nine depressive symptoms and for every symptom 'X', the other two symptoms 'Y' and 'Z' which have the two highest cosine similarity scores symptoms are selected to create edges. Finally, the symptom networks are shown in Figs. 3, 4 and 5, mild (Fig. 3), moderate (Fig. 4) and severe (Fig. 5).

The mild symptom network consists of one core symptom 'F' and three groups 'AB', 'CDE', 'GHI'. The moderate symptom network has no core symptom, one core group 'ACEG' and two subgroups 'DF' and 'BHI' are made up of this network. The severe symptom network consists of two completely independent networks.

First of all, we calculate the average degree of three symptom networks. They are 2.78, 2.67, 2.22, respectively from mild to severe of the symptom score. The link

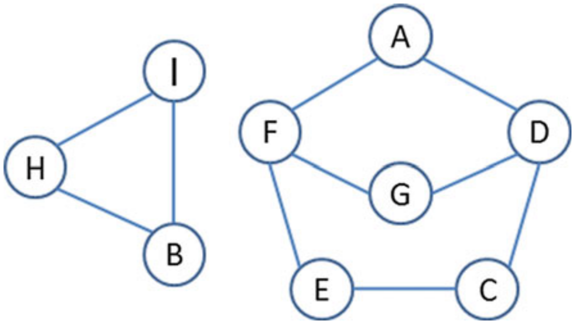

Fig. 5 The severe symptom network

density of the symptom network is reducing. This phenomenon indicates that the relationship between the symptoms become more concentrated with the increasing of symptom scores. At the same time, the maximum degree of the three symptom networks are 6, 5, 3, which is reducing with the increasing of severity of symptoms. This phenomenon reflects the disappearance of the central symptoms in depression network. In other word, the more severe the symptoms, the greater discrepancy between participants.

Second, the development of symptoms 'B', 'G', 'H', and 'I' is interesting. In the mild symptoms network, symptom 'B' is mainly related to symptoms 'A' and 'F'. The symptoms 'A' and 'B' are the prerequisite for depression diagnosis. 'CDE' and 'GHI' are superficial and deep symptom groups respectively. In the moderate symptoms network, Symptom 'B' dives to establish connect with 'GHI' and to replace Symptom 'G' in 'GHI' groups. In the severe symptoms network, 'BHI' forms an independent network, which shows that the participant may have symptom 'H' and 'I' if he has severe symptom 'B'. This result implies that serious lack of interest (symptom 'A') and severe depression mood (symptom 'B') represent two types of severe depression, of which type B is likely to endanger the participant's health.

4.3 *Method Effect Analysis*

Based on the fact of stratification, we can assume that there is only the same grade of symptom combinations between the depression symptoms, so the number of depression symptoms combinations for each grade is 2 ^ 9 – 1 = 511 species. We combine these 511 combinations with 11 methods and use the model to calculate the average effect of each method, of which the first 10 were behavior regulation and the last was clinical intervention. The results are shown in Table 5.

The results show that with the exacerbation of symptoms, both the behavior regulation and the clinical intervention can relieve the symptoms of depression. And the behavior regulation is more effective than clinical intervention for relieving mild

Table 5 The effectiveness of different methods on depression treatment

	Grade mild (%)	Grade moderate (%)	Grade severe (%)
Behavior regulation	18.3	59.7	91.7
Clinical intervention	1.0	80.0	98.0

depression symptoms. The clinical intervention for lessening moderate or severe depression symptoms has obvious advantages.

5 Conclusion

In this paper, we built an effective model to evaluate the effect of treatment methods for depression patients. The experiments show that the symptoms have unidirectional influence on the choice of regulatory methods. Our analysis shows that behavior regulation and clinical treatment have different advantages. This approach could be applied to more mental illnesses.

Acknowledgements This work was supported by Beijing Haola Technology Co., Ltd and the Mental Health Center of West China Hospital, Sichuan University. The online survey tool was designed by Beijing Haola Technology Co., Ltd. The authors acknowledge Prof. Zhang Jun at the Mental Health Center of West China Hospital, Sichuan University for his research assistance and insightful comments to the manuscript.

References

1. Anderson IM, Tomenson BM (1995) Treatment discontinuation with selective serotonin reuptake inhibitors compared with tricyclic antidepressants: a meta-analysis. BMJ 310:1433
2. Sansone RA, Sansone LA (2012) Antidepressant adherence: are patients taking their medications? Innov Clin Neurosci 9:41–46
3. Pampallona S, Bollini P, Tibaldi G, Kupelnick B, Munizza C (2004) Combined pharmacotherapy and psychological treatment for depression. Arch Gen Psychiat 61:714
4. Cuijpers P, Cristea IA, Ebert DD, Koot HM, Auerbach RP, Bruffaerts R, Kessler RC (2016) Psychological treatment of depression in college students: a metanalysis. Depress Anxiety 33:400–414
5. Joormann J, Gotlib IH (2017) Emotion regulation in depression: relation to cognitive inhibition. Cogn Emot 24:281–298
6. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT (2008) Initial severity and antidepressant benefits: a meta-analysis of data submitted to the food and drug administration. PLoS Med. 5:0260–0268
7. Carl JR, Gallagher MW, Barlow, DH (2017) Development and preliminary evaluation of a positive emotion regulation augmentation module for anxiety and depression. Behav Therapy
8. Schuch FB, Vasconcelos-Moreno MP, Borowsky C, Zimmermann AB, Rocha NS, Fleck MP (2015) Exercise and severe major depression: effect on symptom severity and quality of life at discharge in an inpatient cohort. J Psychiat Res 61:25–32

9. Martin R, Dahlen ER. Cognitive emotion regulation in the prediction of depression, anxiety, stress, and anger
10. Millgram Y, Joormann J, Huppert JD, Tamir M (2015) Sad as a matter of choice? emotion-regulation goals in depression. Psychol Sci 26:1216–1228
11. Kroenke K, Spitzer RL (2002) The PHQ-9: a new depression diagnostic and severity measure. Psychiat Ann 32:509–515
12. Pearlstein TB, Zlotnick C, Battle CL, Stuart S, O'Hara MW, Price AB, Grause MA, Howard M (2006) Patient choice of treatment for postpartum depression: a pilot study. Arch Womens Ment Health 9:303–308
13. Mikolov T, Chen K, Corrado G, Dean J (2015) Efficient estimation of word representations in vector space. In: IJCAI international joint conference on artificial intelligence, Jan, pp 4069–4076 (2015)
14. Joulin A, Grave E, Bojanowski P, Mikolov T (2016) Bag of tricks for efficient text classification