Identifying Cognitive Distortion by Convolutional Neural Network based Text Classification

Xuejiao Zhao[12], Chunyan Miao[12], and Zhenchang Xing[3]

[1]Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY), NTU, Singapore,
[2]School of Computer Science and Engineering, NTU, Singapore
[3]Research School of Computer Science, Australian National University, Australia

{xjzhao, ASCYMiao}@ntu.edu.sg zhenchang.xing@anu.edu.au

Abstract

Cognitive distortions have a way of playing havoc with our lives. The most important step to untwist the irrational thinking is identifying the forms of the cognitive distortion. The daily narration or diaries of the patients are always used by the cognitive-behavioral therapists as a clue to identify the cognitive distortion. But these natural language materials are always diverse and desultory which affect the efficiency and accuracy of identification. In this research, we propose a model called *ICODLE* (Identifying Cognitive Distortion by Deep Learning) which utilizes the daily narration or diaries of the patients to identify the forms of the cognitive distortion. *ICODLE* collect the daily narration and diaries from the authoritative books and webpages in *CBT* (Cognitive-Behavioral Therapy) domain. Then *ICODLE* creates the database of the 10 forms of cognitive distortion which were defined by David D. Burns. By utilizing the advanced deep learning techniques (e.g., Word Embedding, *CNN* (Convolutional Neural Network), etc.), *ICODLE* can identify the forms of the patients' cognitive distortions without the features extraction. *ICODLE* can effectively assist the patients and the cognitive-behavioral therapists to diagnose the cognitive distortions. *ICODLE* also benefit to build up the online persuasion system.

Keyword: *CBT*, Cognitive Distortion, Word Embedding, *CNN*

I. Introduction

Cognitive distortion is the systematic mistakes of in the perception and information processing, are individuals convinces themselves of something that isn't really true [1]. These inaccurate thoughts reinforce negative thinking and caused depression. The patients will tell themselves the things that sound rational but actually only serve to keep them feeling bad about themselves [1] [2].

Cognitive therapy is a treatment process using for the depression. It assists the patients to change their beliefs and behaviors that generate the certain mood states. Many prior works discovered that depressed patients continually had spontaneously arising negative awareness ("Automatic thoughts" that were verbal or imaginal in nature) about themselves, their worlds, and their future. But if the cognitive-behavioral therapists help them to solve their problems, analyze their dysfunctional behavior, and the distortions in their thinking, they can achieve sustained improvement in their emotion, symptoms, behaviors, and relationships [3]. The first therapeutic strategy of cognitive-behavioral therapists using to untwist the irrational thinking is to identify the forms of the cognitive distortion [4] [5].

TABLE I: THE 10 FORMS OF COGNITIVE DISTORTIONS [4]

No.	Forms of Distortion	Description
1	All-or-nothing thinking	You think things in absolute, black-and-white categories.
2	Overgeneralization	You look at a negative event as a never-ending pattern of defeat.
3	Mental filter	You dwell on the negatives.
4	Discounting the positives	You insist that your accomplishments or positive qualities don't count.
5	Jumping to conclusions:	
	Mind-reading	You regard the others' response as negative to you but there is no exact evidence;
	Fortune-telling	You arbitrarily predict the things will turn out badly.
6	Magnification or minimization	You blow things way out of proportion or you shrink their importance.
7	Emotional reasoning	You reason from how you feel: "I feel like an idiot, so I really must be one."
8	Should statement	You criticize yourself (or other people) with "should," "ought," "must" and "have to."
9	Labeling	Instead of saying "I made a mistake," you tell yourself, "I'm a jerk," or "a fool," or "a loser."
10	Personalization and blame	You blame yourself for something you weren't entirely responsible for, or you blame other

Tab. I shows the 10 forms of cognitive distortion defined by David D. Burns. [4]. The daily narrations or diaries of the patients are always used by the cognitive-behavioral therapists as a clue to identify the cognitive distortion.

David D. Burns encourages the patients to record their daily mood log as shown in **Tab**. II. The daily mood log is a diary to write the "automatic thoughts" which occur spontaneously when the patient experience some specific problems or upsetting events which bothering them. The form of distortion can identify by the cognitive-behavioral therapists or themselves. Then, depending on the forms of the distortion, the cognitive- behavioral therapists or themselves can generate the rational responses to instead of their "negative thoughts". However, the cognitive-behavioral therapists can't always follow the patients whenever their negative thoughts occur. For the patients, they are not professional enough to identify themselves' cognitive distortions and even they tend to believe their original thought. Therefore if there are some online therapists which can identify the forms of cognitive distortion of the "automatic thoughts" and give the rational responses in real time, that will be very helpful to the patients. Nonetheless the "automatic thoughts" are always diverse and desultory which affect the efficiency and accuracy of the identification.

TABLE II: THE DAILY MOOD LOG [4]

Automatic Thoughts	Distortions	Rational Responses
I just know that I'm going to have an awful day.	fortune telling	Today might have some obstacles, but I can overcome them and still have a good day.

Traditional natural language classifiers are almost constructed based on the human-designed features (e.g., Dictionaries, Knowledge Bases, Special Tree Kernels, etc.) [6] [7]. However recently, deep learning approaches are widely used to natural language classification tasks [6] [8] [9]. By utilizing the layers with convolving filters, *CNN* (Convolutional neural networks) can capture the local features of sentences automatically [10]. So *CNN* have been shown to be effective for natural language classification and other NLP (Natural Language Processing) tasks. On the word

representation aspect, the traditional NLP techniques always represent words as indices in a vocabulary, which not include the relationship between words.

Compared with *1 of V*, the word embedding learned by deep learning approaches targets to encode the semantic relationships, linguistic regularities and patterns into the embedding space explicitly. So word embedding technique will learn a dense, low-dimensional, high-level feature representation real-valued vector for each word in an unsupervised way [11] [12]. This makes it suitable to use as an input of natural language classification.

In this paper, we present a model called *ICODLE* based on deep learning approach to identify the forms of cognitive distortion. *ICODLE* collects the daily mood log and narration of the cognitive distortion patients from the authoritative books [1] [3] [4] [5] and webpages [13] [14] in CBT (Cognitive- Behavioral Therapy) domain. First, *ICODLE* creates the database of the 10 forms of cognitive distortion which defined by David D. Burns [4]. Then, *ICODLE* uses *1 of V* to represent the words into one hot vector. After that, *ICODLE* uses *word2vec* based on *CBOW* (Continuous Bag of Words Model) to encode the word vector to a dense, low-dimensional, high-level feature representation real-valued vector by 100 billion words of Google News. With the word vectors obtained from word2vec, *ICODLE* trains the *CNN* classifier with one layer of convolution to classify the daily narration or diaries of the patients to the 10 forms of cognitive distortion. An experiment on a large scale of real-world materials is ongoing to evaluate the efficiency of *ICODLE*. *ICODLE* also benefit to build up the online persuasion system due to identifying the forms of cognitive distortion is the precondition to form a reasonable persuasion.

II. Related Work

CBT (Cognitive Behavior Therapy) is a popular research topic about psychosocial intervention [1] [15], and it's the most popular evidence-based practice for mental disorders treating [16]. The cognitive-behavioral therapists who identify cognitive distortions make big contribution to the treatment of depression and anxiety [17-20]. For example, the treatment manual for depression of

Aaron T. Beck et al. [3] presents an explicit theory and practice of cognitive therapy for depression and provides the guidance to address suicidal ideation and possible relapse. Dawes, RM [21] proposes that the cognitive-behavioral therapists can use CBT techniques to assist the patients to challenge their beliefs and replace cognitive distortion such as “overgeneralizing”, “minimizing positives”, etc. to more rational responses, thus reducing the emotional distress and their self-defeating behavior. The most common forms classification of cognitive distortions is proposed by David D. Burns, M.D [4]. This book includes an explanation of the principles of CBT and describes in detail about how to improve individual’s mood and life by identifying and eliminating common cognitive distortions. There are some exercises to assist the reader in identifying the forms of cognitive distortion and replace them with more rational responses.

Machine learning technique is widely used to the CBT domain. James K et al. [22] propose the intelligent real-time therapy which uses machine learning to optimize the delivery of momentary cognitive-behavioral interventions. Masson, Kristoffer NT, et al. [23] predicts long-term outcome of internet- delivered cognitive behavior therapy for social anxiety disorder using fMRI (Functional Magnetic Resonance Imaging) and SVM. J Li et al. [24] uses the machine learning method to identify the sentiment of the sentences. IR Galatzer-Levy et al. [25] shows a machine learning application to forecast PTSD (Post Traumatic Stress Disorder) from early trauma responses quantitatively. Clark, Ian A et al. [26] predict intrusive memories of traumatic film footage using machine learning on fMRI data.

To our best knowledge, we do the first attempt for identify the forms of cognitive distortion by deep learning method.

III. The Approach

A. System Overview

Fig. 1 shows the structure of *ICODLE*. There are 2 main lines in *ICODLE*, one is the training phase another is prediction phrase. In the training phase, the input is the collected daily mood

log or narration of the word vector matrix patients from the authoritative books and webpages in the *CBT* domain. Then we send the labeled text to the word embedding part to transform the text to word vector matrix.

Next, we use the word vector matrixes of different forms of cognitive distortion to train the *CNN* model and get the trained *CNN* multiclass classifier. In the prediction phase, if a cognitive patient input an automatic thought, *ICODLE* will go through same wording embedding progress and use the trained *CNN* multiclass classifier to identify the form of cognitive distortion if current automatic thought.

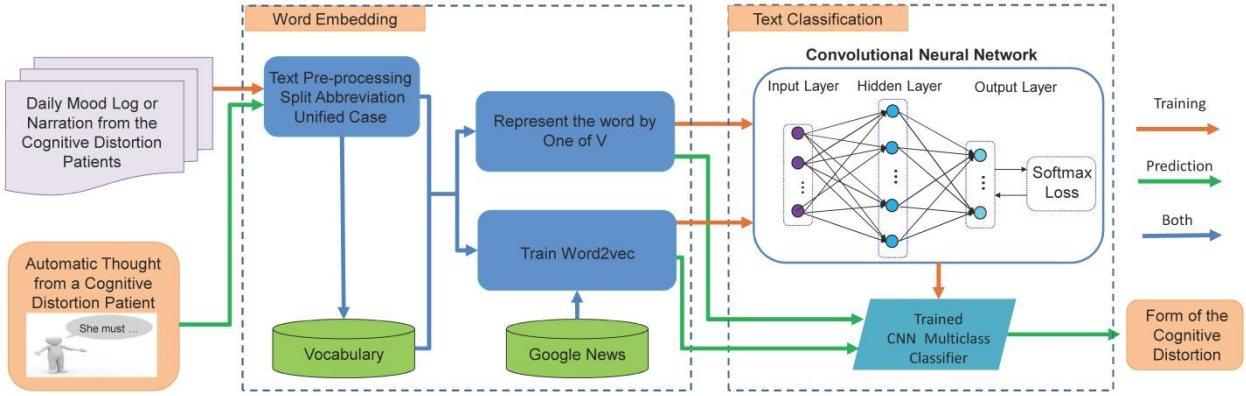

Fig. 1: The Structure of *ICODLE*

B. *Data Preprocessing*

In the data preprocessing of *ICODLE*, we separate symbol like "()", "?" to words. Then we split the abbreviation of the sentences like "that's" to "that is", "won't" to "will not". Last we unify all the word to lowercase.

C. *Word Embedding*

Word embedding is used to compute distributed representations of words, and transform the words into the form of continuous vectors. Word embedding assumes that words appear in similar context may have similar meanings [27]. Therefore in the embedding space, the semantically similar words are close to each other. The input of word embedding is the result of *1 of V*. *1 of V* generates a fixed-size vocabulary from the materials with *V* members in total.

So V is the dimension of *1 of V*. Each input word will be coded as a vector of size *V* with all zeros except for the element corresponding to the word's order in the vocabulary.

Word embedding is unsupervised word representations contained many different methods. Here we introduce the technique used in *ICODLE* call *word2vec*. The *word2vec* can use either of the two model architectures proposed by Mikolov et al. to generate a distributed representation of the words: one is called *CBOW* and another is called *Skip-gram* (Continuous Skip-gram Model) [28] [29].

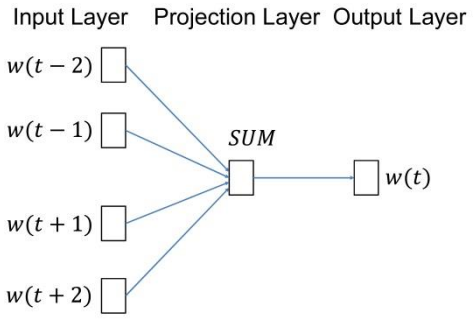

Fig. 2: The Structure of *CBOW* [28]

Fig. 2 shows the structure of the *CBOW* model [28]. There are 3 layers in *CBOW*, respectively are input layer, projection layer and the output layer. The *CBOW* model predicts the current word by utilizing the surrounding context words. And the CBOW model assumes that the words sequence of the context does not affect the result of prediction. Supposing $w(t)$ is any word in the corpus, the $\mathrm{Context}(w(t))$ is the n words before and after the target word $w(t)$. So we have a training sample $(Context(w(t)), w(t))$, and the window is $2n + 1$, here we assume the n = 2. The objective function of *CBOW* model is to maximize the sum of log probabilities of the surrounding context words conditioned on the center word:

$$\sum_{w \in C} logp(w(t)|Context(w(t))) \qquad (1)$$

where *C* is the vocabulary of all words.

For the projection layer, we accumulate the 2n vectors from input layer as following:

$$X_w = \sum_{i=1}^{2n} (Context(w(t))_i) \qquad (2)$$

From the projection layer to the output layer there is a Huffman Tree, using to compute the *p(w(t)|Context(w(t)))*.

$$p(w(t)\ Context(w(t))) = \frac{e^{y_{w(t),i_{w(t)}}}}{\sum_{i=1}^{2n} e^{y_{w(t),i}}} \qquad (3)$$

Using this equation, we can compute a score to each word according to the vocabulary.

D. ***Identify the Forms of Cognitive Distortion by CNN***

The input of the *CNN* is a sentences matrix generated by *word2vec*. The row of the matrix corresponds to each word. The number of columns is the dimension of the word embedding. For a 9 word sentence using a 300-dimensional embedding, we would have a 9 × 300 matrix as our input.

In the convolution layer, *ICODLE* performs convolutions over the embedding matrix using multiple filter sizes and generate feature maps. The width of the filters is 300 which are same as the width of the input matrix. The height, or sliding windows, is 2-5 words. The convolutional layer contains multiple filter widths and feature maps.

Next, *ICODLE* max-pool the result of the convolutional layer into a long feature vector, i.e., the largest number from each feature map is recorded and adds dropout regularization.

Last, we classify the result using a softmax layer.

IV. Conclusion and Future Work

Cognitive therapy is a treatment process using for the depression. It assists the patients to change their beliefs and behaviors that generate the certain mood states. The most important step to untwist the irrational thinking is identifying the forms of the cognitive distortion. But these natural language materials are always diverse and desultory which affect the efficiency and accuracy of identification. In this research, we propose a model called *ICODLE* which utilizes the daily mood log and narration of the cognitive distortion patients to identify the cognitive distortion. Based on the deep learning techniques like *word2vec, CNN, etc., ICODLE* can identify the forms of the patients' cognitive

distortions without the features engineering. *ICODLE* can assist patients and the cognitive-behavioral therapists to diagnose the cognitive distortions sufficiently. *ICODLE* is also significant to build up the online persuasion system.

In the future, we will conduct an experiment on a large scale of real-world materials to evaluate the efficiency of *ICODLE*.

V. Acknowledgements

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative.

References

[1] J. S. Beck, Cognitive behavior therapy: Basics and beyond. Guilford Press, 2011.

[2] J. M. GROHOL, "15 common cognitive distortions," Psych Central, 2009.

[3] A. T. Beck, Cognitive therapy of depression. Guilford press, 1979. [4] D. D. Burns, The feeling good handbook, Rev. Plume/Penguin Books,

[4] 1999.

[5] F. O. Henker, "Feeling good: The new mood therapy," 1982.

[6] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification." in AAAI, vol. 333, 2015, pp. 2267– 2273.

[7] X. Zhao, Z. Xing, M. A. Kabir, N. Sawada, J. Li, and S.-W. Lin, "Hdskg: Harvesting domain specific knowledge graph from content of webpages," in Software Analysis, Evolution and Reengineering (SANER), 2017 IEEE 24th International Conference on. IEEE, 2017, pp. 56–67.

[8] Y. Kim, "Convolutional neural networks for sentence classification," arXiv preprint arXiv:1408.5882, 2014.

[9] G. Kumaran and J. Allan, "Text classification and named entities for new event detection," in Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2004, pp. 297–304.

[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[11] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation." in EMNLP, vol. 14, 2014, pp. 1532–1543.

[12] Y. Goldberg and O. Levy, "word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method," arXiv preprint arXiv:1402.3722, 2014.

[13] S. McLeod, "Cognitive behavioral therapy," https://www. simplypsychology.org/cognitive-therapy.html, published 2008, updated 2015.

[14] M. David D. Burns, "Daily mood log," http://www.burdenbearersdv. com/documents/Daily_Mood_with_example.pdf, published 1980, Revised 2004.

[15] S. D. Hollon and A. T. Beck, "Cognitive and cognitive-behavioral therapies." 1994.

[16] T. A. Field, E. T. Beeson, and L. K. Jones, "The new abcs: A practitioner's guide to neuroscience-informed cognitive-behavior therapy," Journal of Mental Health Counseling, vol. 37, no. 3, pp. 206–220, 2015.

[17] D. Robertson, The philosophy of cognitive-behavioural therapy (CBT): Stoic philosophy as rational and cognitive psychotherapy. Karnac Books, 2010.

[18] G. Long et al., Enchiridion. Courier Corporation, 2004.

[19] A. Ellis and D. Ellis, "Rational emotive behavior therapy," Current psychotherapies, pp. 196–234, 2011.

[20] A. T. Beck and B. A. Alford, Depression: Causes and treatment. University of Pennsylvania Press, 2009.

[21] R. M. Dawes, "Cognitive distortion," Psychological Reports, vol. 14, no. 2, pp. 443–459, 1964.

[22] J. Kelly, P. Gooding, D. Pratt, J. Ainsworth, M. Welford, and N. Tarrier, “Intelligent real-time therapy: Harnessing the power of machine learning to optimise the delivery of momentary cognitive–behavioural interventions,” Journal of Mental Health, vol. 21, no. 4, pp. 404–414, 2012.

[23] K. N. Månsson, A. Frick, C.-J. Boraxbekk, A. Marquand, S. Williams, P. Carlbring, G. Andersson, and T. Furmark, “Predicting long-term outcome of internet-delivered cognitive behavior therapy for social anxiety disorder using fmri and support vector machine learning,” Translational psychiatry, vol. 5, no. 3, p. e530, 2015.

[24] J. Li and M. Sun, “Experimental study on sentiment classification of chinese review using machine learning techniques,” in Natural Language Processing and Knowledge Engineering, 2007. NLP-KE 2007. International Conference on. IEEE, 2007, pp. 393–400.

[25] I. R. Galatzer-Levy, K.-I. Karstoft, A. Statnikov, and A. Y. Shalev, “Quantitative forecasting of ptsd from early trauma responses: A machine learning application,” Journal of psychiatric research, vol. 59, pp. 68–76, 2014.

[26] I. A. Clark, K. E. Niehaus, E. P. Duff, M. C. Di Simplicio, G. D. Clifford, S. M. Smith, C. E. Mackay, M. W. Woolrich, and E. A. Holmes, “First steps in using machine learning on fmri data to predict intrusive memories of traumatic film footage,” Behaviour research and therapy, vol. 62, pp. 37–46, 2014.

[27] Z. S. Harris, “Distributional structure,” Word, vol. 10, no. 2-3, pp. 146–162, 1954.

[28] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distribut- ed representations of words and phrases and their compositionality,” in Advances in neural information processing systems, 2013, pp. 3111–3119.

[29] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.

Xuejiao Zhao is currently a Ph.D. student in the School of Computer Science and Engineering, Nanyang Technological University in Singapore. She obtained B.Eng. and M.Eng. degrees from China West Normal University and University of Electronic Science and Technology of China, in 2011 and 2014, respectively. Her research interest includes natural language processing, knowledge graph, explainable AI, smart home for elderly, post-operation rehabilitation systems and Artificial intelligence (AI) powered cognitive-behavioral therapy.

Dr. Chunyan Miao is a Professor with the School of Computer Science and Engineering (SCSE) at Nanyang Technological University (NTU), Singapore. She is the Director of the NTU-UBC Joint Research Centre of Excellence in Active Living for the Elderly (LILY). Prior to joining NTU, she was an Instructor and Post-Doctoral Fellow with the School of Computing, Simon Fraser University, Canada.

Her research focuses on studying the cognitive and social characteristics of intelligent agents in multi-agent and distributed AI/CI systems, such as trust, emotions, incentives, motivated learning, ecological and organizational behavior. She has worked on new disruptive Artificial intelligence (AI) approaches and theories that synergize human intelligence, artificial intelligence and behavior data analytics (AI powered by humans). Her current research interests include human-agent interaction, cognitive agents, human computation and serious games.

Dr. Zhenchang Xing is now a Senior Lecturer in the Research School of Computer Science, Australian National University. Previously, he was an Assistant Professor in the School of Computer Science and Engineering, Nanyang Technological University, Singapore, from 2012-2016. Dr. Xing's research interests include software engineering, data mining and human-computer interaction. His work combines software analytics, behavioral research methods, data mining techniques, and interaction design to understand how developers work, and then build recommendation or exploratory search systems for the timely or serendipitous discovery of the needed information.