Applying Deep Learning in Depression Detection

Research-in-Progress

Wenwen Li
Faculty of Business and Economics
The University of Hong Kong
Hong Kong, China
liwwen@connect.hku.hk

Michael Chau
Faculty of Business and Economics
The University of Hong Kong
Hong Kong, China
mchau@business.hku.hk

Abstract

According to the World Health Organization, one in twenty people in the world have suffered from depression and emotional distress in the previous twelve months. How to manage and provide appropriate treatment to people suffering from depression and emotional distress is a highly pressing issue. However, many people with depression and emotional distress are not sufficiently recognized and treated and do not actively seek help. It is therefore highly desirable to devise a method to effectively and proactively identify these people. Following the design science approach, we propose DK-LSTM, a novel design based on deep learning to identify people with depression and emotional distress. Based on Long Short-Term Memory (LSTM), a type of deep learning networks, our model incorporates both general knowledge and domain knowledge in the learning process through word embedding and parallel LSTM units.

Keywords: deep learning, depression, social media, classification

Introduction

Depression is a common but serious mental health issue in modern societies. According to the World Health Organization, about 5% of people worldwide have suffered depression in the previous year, and depression is contributing significantly to the global burden of diseases around the world (WHO 2012). It has also been suggested that depression and emotional distress are comorbid with other chronic diseases, such as angina, arthritis, asthma, and diabetes, and can worsen the associated health outcomes (Moussavi et al. 2007). Therefore, providing professional help and appropriate treatment to these patients is helpful to not only the mental health of these patients but also the conditions of their chronic diseases.

Unfortunately, many people suffering from recurring depression and emotional distress are not sufficiently recognized and treated. Many of them do not actively seek for help and their symptoms are often difficult to observe. It would be impractical to provide help to people with depression and emotional distress if they could not be identified. Therefore, having an effective way to identify these people in need is an important issue in the management of people with depression and emotional distress.

Written language has been shown to be an independent and meaningful way of exploring personality (Pennebaker et al. 1999). Previous research has shown that traits of depression and emotional distress can be identified in one's writing (Rude et al. 2004). With the popularity of social media in recent years, it is very common to see people expressing their emotions and writing about their daily lives on social media such as social networking websites and blogs. Some recent studies have shown that by analyzing these social media postings, it is possible to find people showing depression symptoms through manual evaluation or machine classification approaches (Huang et al. 2007; Li et al. 2012). However, it has also been reported that traditional classification methods may not perform the classification task very well because of the difficulty in extracting the subtle meaning hidden in social media writing (Cheng et al. 2017). Some more advanced methods, such as deep learning which has achieved great success in various domains in recent years, may be needed to identify people with depression through online channels such that professional assistance may be provided in a timely and effective manner.

In this study, we propose a design called DK-LSTM (which stands for Domain-Knowledge-enhanced Long Short-Term Memory) based on deep learning. The proposed design encodes and processes semantic and domain knowledge in written language, and then use Long Short-Term Memory (LSTM) networks, a popular type of deep learning architecture (Hochreiter and Schimidhuber 1997), to detect depression and emotional distress in social media contents. we propose a unique way to incorporate domain knowledge into LSTM in our design. The performance of DK-LSTM is demonstrated through an experiment that compares the proposed approach with several machine learning and deep learning models on identifying people showing depression and emotional distress from a set of blogs written in Chinese.

Research Background

Depression and Emotional Distress and the Internet

Depression is a leading contributor to the global burden of disease (Whiteford et al. 2013). Depression is a chronic recurring disorder, but in most current practice depression is not treated as a chronic disease. It is undesirable to wait for patients to take the initiative to seek professional help when recurring depression episodes occur (Andrews, 2001). Thus, it is all the more important to take preventative precautions by identifying those with depressive symptoms and emotional distress, and then providing them with appropriate interventions and treatments.

It has been suggested that user-generated online contents, such as narratives and diaries written on blogs and web forums, have great potential for gathering useful data to understand individuals' emotional status such as depressive symptoms and distress (Cavazos-Rehg et al. 2016). Emotional distress and depressive symptoms can be found in user-generated online contents shown in different ways such as emotional responses, negative expressions, and even suicide notes. Meanwhile, Domain-specific lexicons have been used to improve the performance of these automation techniques. For example, the Linguistic Inquiry and Word Count (LIWC) lexicon (Pennebaker et al., 2007) has been used in many studies in the public health domain to identify people with depression and emotional distress.

Web mining and text analysis techniques have achieved satisfactory performance in identifying affects and sentiments in social media (Cambria 2016). However, when it comes to the detection of emotional distress and depression, the classification performance can be very low (Guan et al. 2015). There is a good research opportunity to take a more sophisticated approach using advanced techniques, such as deep learning models which have gained a lot of academic and industrial attention in recent years, to improve the detection of depression and emotional distress in social media.

Deep Learning

Deep learning is a kind of representation-learning methods. It discovers the representations from raw data (LeCun et al. 2015) and attempts to learn multiple levels of representation. Different from traditional machine learning approaches that need careful human selection of feature sets, deep learning approaches learn features from data via a general-purpose learning procedure (LeCun et al. 2015). Thus, deep learning methods can represent high-dimensional data at a more abstract level. For instance, language, images, and videos, which are abundantly available but not thoroughly analyzed by traditional machine learning research, have become hot topics in deep learning research. Many exciting findings and achievements have been reported (He et al. 2016; Hinton et al. 2012).

Deep learning has become one of the most powerful tools for natural language processing (NLP). Instead of using linear models, deep learning approaches apply non-linear neural network models to process language, such as Recurrent Neural Networks and Convolutional Neural Networks. Deep learning models have shown promising performance in various text analysis tasks, such as sentiment analysis (Socher et al. 2013) and machine translation (Choi et al. 2017).

Long Short-Term Memory networks

Long Short-Term Memory networks are a variation of Recurrent Neural Networks (RNNs) with Long Short-Term Memory units. LSTM was first introduced in 1997 (Hochreiter and Schmidhuber 1997), and has been improved and applied since then. One of the advantages of LSTM over other RNNs is its ability to deal with the error backflow problem. The performance of previous models, such as Back-Propagation Through Time and Real-Time Recurrent Learning, are heavily influenced by long periods of time because of the vanishing or exploding of the error signals (Hochreiter and Schmidhuber 1997). LSTM overcomes these gradient vanishing or exploding problems by having memory cells with gates (Wang et al. 2016).

As important structures of LSTM, the gates of a memory cell keep irrelevant information from being stored in the memory, decide information storage, and output filtered information. Essentially, the gates open and close access to constant error flow through the memory cell (Hochreiter and Schmidhuber 1997). During the training process, the memory cells learn when to block or pass on information by adjusting weights via gradient descent.

LSTM has shown superiority in many natural language processing (NLP) tasks. Text analysis can be regarded as sequential problems with varying lengths, and LSTM can effectively map sequences to sequences. Furthermore, LSTM is often combined with other deep learning approaches (e.g. convolutional neural networks) to tackle image question-answering problems (Ren et al. 2015). A common approach is to extract an image feature vector using a convolutional neural network and encode the corresponding question as a feature vector using LSTM.

Proposed Design

Overview

Given the potential of deep learning, we follow the design science approach (Hevner et al. 2004; Gregor & Hevner 2013) and propose an architecture based on deep learning that aims to automatically identify people who are showing emotional distress and depression symptoms in their online postings. We have observed that text can contain multi-level information, and humans need specialized knowledge to comprehend such information. For instance, it is difficult for untrained people to identify distressed emotion and depression tendency through written texts, due to the lack of psychological knowledge to differentiate depression and emotional distress from negative sentiment. In a similar way, typical word embedding methods capture general syntactic and semantic word relationships (Mikolov et al. 2013), but lack domain knowledge for specific tasks. To address this problem, I incorporate domain knowledge into deep neural networks through a novel design. The main idea of the architecture is to represent social media contents from both aspects and process them in parallel.

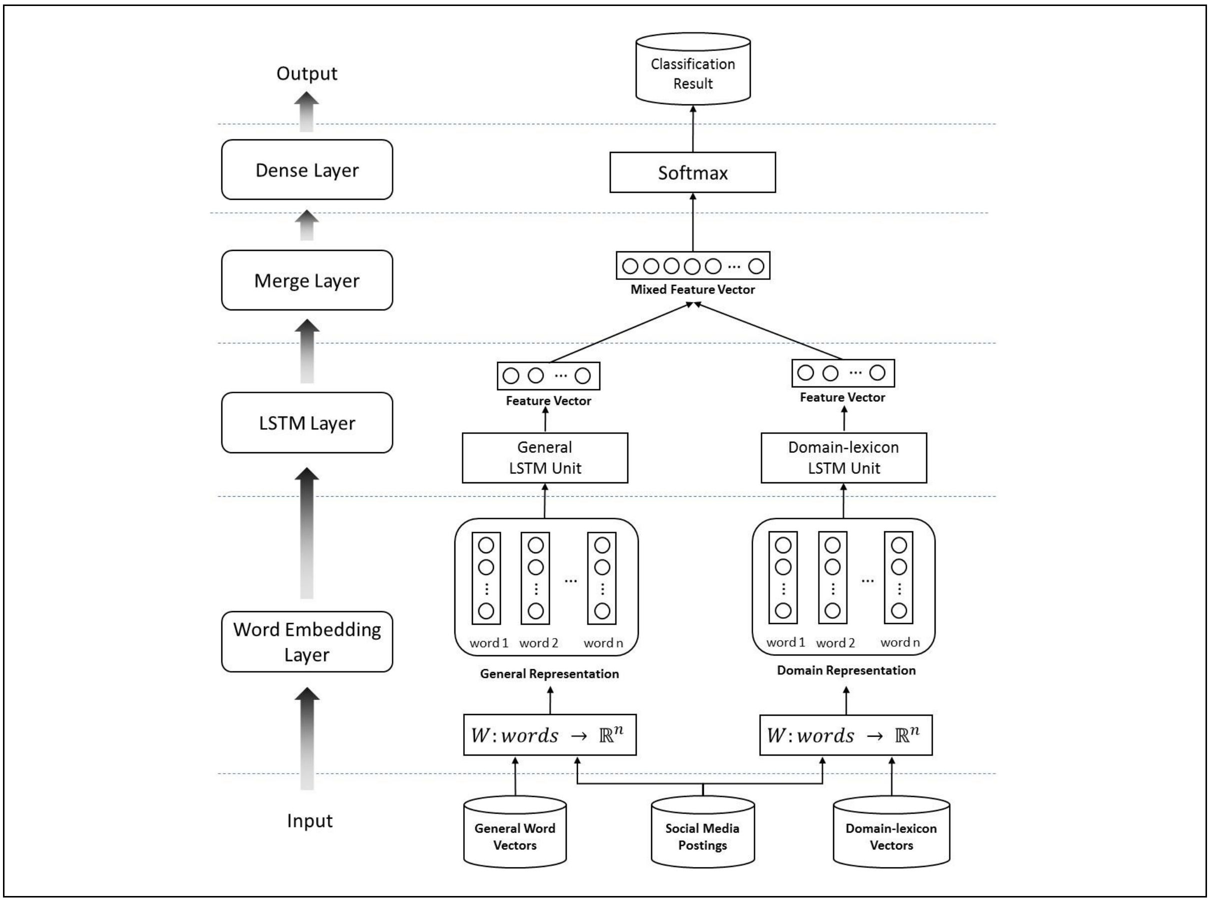

Figure 1. The Architecture of DK-LSTM

Word Embedding Layer

An important component of our framework is word embedding. Word embedding encodes meaningful syntactic and semantic relationships and have been widely used in deep learning research in the pre-training phase. To perform word embedding, we first need to build the corresponding word vectors, namely the general word vectors and domain-lexicon vectors. The general word vectors are generated by the Word2Vec method (Mikolov et al. 2013) and the domain-lexicon vectors are generated based on the LIWC lexicon (Pennebaker et al. 2007). More details are given in the following.

We first discuss how we build the general word vectors. We choose 100 as the size (i.e., number of dimensions) of the vectors according to previous research (Chen et al. 2013). The process is shown in Figure 2. A large corpus, such as the contents of the entire Wikipedia, is first obtained from the Internet. Then there are three steps of pre-processing the data. Firstly, we cleanse the data by removing formatting characters and non-UTF-8 characters, because these irrelevant characters may influence further analysis. Secondly, we segment sentences into words. Thirdly, we conduct stop-word filtering by removing words with high occurrence frequency but low information value. Deleting these words will improve the accuracy of the final result.

After data pre-processing, we use the Word2Vec approach (Mikolov et al. 2013) to train the general word vectors. Word2Vec is a tool enabling the training and use of word embedding. Word2Vec takes a large text corpus as input, and produces a numeric vector representation for each word as output. There are two popular models that can be used in the Word2Vec process, namely CBOW and Skip-gram. According to previous research (Mikolov et al. 2013), we apply the Skip-gram model to train the word vectors. After running the Word2Vec process, every word that appears in the input corpus is represented as a 100-dimensional vector, and the vectors for all words together constitute the general word vectors.

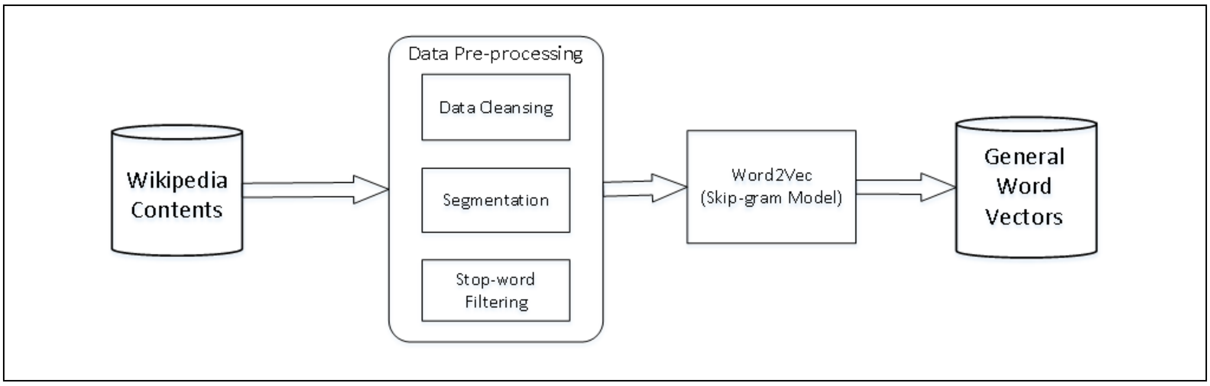

Figure 2. The Process of Building the General Word Vectors

The domain-lexicon vectors are generated based on the LIWC lexicon (Pennebaker et al. 2007). LIWC has been widely used in psychology research. It categorizes words into 71 groups, so that I build the LIWC vectors with 71 dimensions. Specifically, we use a 71-dimensional vector to represent each input sequence. If a word belongs to one or more categories in LIWC, we will change the value in the corresponding element(s) in the vector to one. Otherwise, all elements in the vector will have a value of zero.

After both the general word vectors and the domain-lexicon vectors are built, we start to encode each input document. Given an input social media document, each word in the document is represented by the corresponding general word vector and domain-lexicon vector. If a word has not been seen before and the corresponding vector is not found (i.e., the word does not appear in the Wikipedia contents or the LIWC lexicon), a vector with all elements equal to zero will be used. Thus, each input social media document is encoded into two sets of vectors according to the words contained in the document, based on the general word vectors and domain-lexicon vectors. These two set of vectors used to encode each document are called the general representation and domain representation of the document, respectively. In other words, each "representation" of a document is a set of vectors (as shown in Figure 1). These representations are then passed to the LSTM Layer for analysis.

LSTM Layer

The standard LSTM has limited ability to detect domain-based information. In order to address this issue, DK-LSTM aims to capture both general information and domain-specific information at the same time. There are two LSTM units in the LSTM Layer to process general representation and domain representation separately. The two units are called the General LSTM unit and the Domain-lexicon LSTM unit, and they take the general representation and domain-lexicon representation as input, respectively. Each unit is a standard LSTM unit, and readers may refer to Section 2.5 for more technical details of the LSTM units.

The General LSTM unit is used to analyze semantic and syntactic information. In other words, it helps to capture information related to the basic meaning of texts. The input of this unit is the general representation that contain general information of the input data.

The Domain-lexicon LSTM unit helps to capture information related to the specific domain based on existing knowledge. Usually, existing knowledge comes from established dictionaries, lexicons, and experts. In this design, we obtain domain knowledge from the LIWC lexicon through the domain representation.

Merge and Dense Layers

As discussed earlier, I divide the Word Embedding Layer and the LSTM Layer into two parts and these two parts are processed separately until the Merge Layer. The Merge Layer combines the feature vectors produced by the two LSTM units and generate a mixed feature vector. In our architecture, as we process texts using two representations (general-based and domain-based), we need to combine the outputs of these two LSTM units to get a final result. Thus, in this layer, outputs of the two LSTM units are concatenated. This is an intuitive way to encode and combine the general and domain information from the input texts efficiently.

The output of the Merge Layer is a mixed feature vector, which is highly abstract and contains general and domain information of input data. Then, we pass the mixed feature vector through the Dense Layer to get the final classification result. The Dense Layer uses the softmax function to provide a probability distribution over the possible outputs.

Conclusion

Potential contribution

There are three main contributions of the study. First, we propose a design based on deep learning that tackles the problem of depression and emotional distress by analyzing online contents in social media, and demonstrate the effectiveness of the proposed design. To our knowledge, we are the first to apply deep learning models in detecting depression and emotional distress from text. The improved performance over existing models is important to the management of this chronic and often recurring disease. Second, while most previous studies only used word vectors to represent knowledge learned from general corpus such as Wikipedia, we demonstrate how to represent domain knowledge through word vectors using the word embedding approach. Third, we propose a novel design that incorporates domain knowledge into an LSTM network using parallel LSTM units.

Future work

We will implement the DK-LSTM using Keras based on the TensorFlow backend. To demonstrate the performance of the proposed model, we will evaluate it against existing machine classification methods and standard deep learning methods.

Acknowledgements

The authors thank to reviewers and colleagues for their constructive comments on the paper.

References

Andrews, G. 2001. "Should depression be managed as a chronic disease?" BMJ: British Medical Journal (322:7283), pp. 419.

Cambria, E. 2016. "Affective computing and sentiment analysis," IEEE Intelligent Systems (31:2), pp. 102-107.

Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., and Bierut, L. J. 2016. "A content analysis of depression-related tweets," Computers in Human Behavior (54), pp. 351-357.

Chen, Y., Perozzi, B., Al-Rfou, R., and Skiena, S. 2013. "The expressive power of word embeddings," in Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013.

Cheng, Q., Li, T. M. H., Kwok, C. L., Zhu, T., and Yip, P. S. 2017. “Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study,” Journal of Medical Internet Research (19:7), e243.

Choi, H., Cho, K. and Bengio, Y. 2017. "Context-dependent word representation for neural machine translation," Computer Speech & Language (45) pp. 149-160.

Gregor, S. and Hevner, A. R. 2013. “Positioning and presenting design science research for maximum impact,” MIS Quarterly, 37(2), 337-356.

Guan, L., Hao, B., Cheng, Q., Yip, P. S., and Zhu, T. 2015. “Identifying Chinese Microblog Users With High Suicide Probability Using Internet-Based Profile and Linguistic Features: Classification Model,” JMIR Mental Health (2:2), e17.

He, K., Zhang, X., Ren, S. and Sun, J. 2016. "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778.

Hevner, A. R., March, S. T., Park, J., & Ram, S. 2004. “Design science in information systems research,” MIS Quarterly, 28(1), 75-105.

Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N. and Kingsbury, B. 2012. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," IEEE Signal Processing Magazine (29:6), pp. 82-97.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long short-term memory," Neural Computation (9:8), pp. 1735-1780.

Huang, Y., Goh, T., and Liew, C. L. 2007. “Hunting suicide notes in Web 2.0 – preliminary findings,” in Proceedings of the IEEE International Symposium on Multimedia – Workshops.

LeCun, Y., Bengio, Y. and Hinton, G. 2015. "Deep learning." Nature (521:7553), pp. 436-444.

Li, T. M. H., Chau, M., Wong, P. W. C., and Yip, P. S. F. 2012, May. “A hybrid system for online detection of emotional distress,” In M. Chau, G. A. Wang, W. T. Yue, & H. Chen (Eds.) Proceedings of the Pacific Asia Workshop on Intelligence and Security Informatics 2012: Kuala Lumpur, Malaysia (pp. 73-80), Heidelberg, Germany: Springer.

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

Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V., and Ustun, B. 2007. “Depression, chronic diseases, and decrements in health: results from the World Health Surveys,” The Lancet (370:9590), pp. 851-858.

Pennebaker, J. W., and King, L. A. 1999. "Linguistic styles: language use as an individual difference," Journal of Personality and Social Psychology (77:6), pp, 1296.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., and Booth, R. J. 2007. The Development and Psychometric Properties of LIWC 2007. LIWC.Net, Austin, TX, USA.

Ren, M., Kiros, R. and Zemel, R. 2015. "Exploring models and data for image question answering," in Advances in Neural Information Processing Systems, pp. 2953-2961.

Rude, S., Gortner, E., and Pennebaker, J. 2004. "Language use of depressed and depression-vulnerable college students," Cognition & Emotion (18:8), pp. 1121-1133.

Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y. and Potts, C. 2013. "Recursive deep models for semantic compositionality over a sentiment treebank," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), vol. 1631, p. 1642.

Wang, Y., Huang, M., Zhu, X., and Zhao, L. 2016. "Attention-based LSTM for Aspect-level Sentiment Classification," in Proceedings of the Conference on Empirical Methods on Natural Language Processing, pp. 606-615.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., ... and Burstein, R. 2013. “Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010” The Lancet (382:9904), pp. 1575-1586.

World Health Organization 2012. “Depression: A Global Public Health Concern,” http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf. Accessed 7.7.2017