Depression Detection via Harvesting Social Media:
A Multimodal Dictionary Learning Solution

Guangyao Shen[1], **Jia Jia**[1*], **Liqiang Nie**[2], **Fuli Feng**[3], **Cunjun Zhang**[1], **Tianrui Hu**[4], **Tat-Seng Chua**[3] **and Wenwu Zhu**[1]

[1]Department of Computer Science and Technology, Tsinghua University; TNList

[2] Department of Computer Science and Technology, Shandong University

[3] School of Computing, National University of Singapore

[4]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications

{ thusgy2012,nieliqiang,fulifeng93}@gmail.com, {jjia,wwzhu}@tsinghua.edu.cn, {18811351350,hutr96}@126.com, dcscts@nus.edu.sg

Abstract

Depression is a major contributor to the overall global burden of diseases. Traditionally, doctors diagnose depressed people face to face via referring to clinical depression criteria. However, more than 70% of the patients would not consult doctors at early stages of depression, which leads to further deterioration of their conditions. Meanwhile, people are increasingly relying on social media to disclose emotions and sharing their daily lives, thus social media have successfully been leveraged for helping detect physical and mental diseases. Inspired by these, our work aims to make timely depression detection via harvesting social media data. We construct well-labeled depression and non-depression dataset on Twitter, and extract six depression-related feature groups covering not only the clinical depression criteria, but also online behaviors on social media. With these feature groups, we propose a multimodal depressive dictionary learning model to detect the depressed users on Twitter. A series of experiments are conducted to validate this model, which outperforms (+3% to +10%) several baselines. Finally, we analyze a large-scale dataset on Twitter to reveal the underlying online behaviors between depressed and non-depressed users.

1 Introduction

Depression is a leading cause of disability worldwide. Globally, an estimated 350 million people of all ages suffer from depression[1]. Depressed people have various depression symptoms manifested by distinguishing behaviors. In clinical diagnosis, psychological doctors often make face-to-face interviews referring to the commonly used *Diagnostic and Statistical Manual of Mental Disorders* criteria. Nine classes of depression symptoms are defined in the criteria, describing the distinguishing behaviors on daily lives. Although this is the most effective method for depression diagnosis, people are somehow ashamed or unaware of depression. More than 70% of people in the early stages of depression would not consult the psychological doctors, deteriorating their conditions[2].

On the other hand, people are increasingly relying on social media platforms like Twitter and Facebook to disclose emotions and moods as well as share their personal statuses. The user generated contents (UGC) on social media instantly reflect not only the daily lives, but also the mental states of users. In the past decade, social media were widely used for physical and mental wellness researches, especially the mental wellness [Coppersmith *et al.*, 2014][Lin *et al.*, 2016][Akbari *et al.*, 2016]. Inspired by these, some efforts have been dedicated to depression studies. Some researchers asked users to fill questionnaires or participate interviews on social media. For example, [Park *et al.*, 2012] analyzed behaviors and the use of languages of depressed users on Twitter. These methods are effective but expensive, time-consuming and hard to get sufficient data to guarantee their findings are robust and generalizable. Besides, these questionnaires and interviews focus on the depression behaviors already defined in depression criteria. However, the symptoms of depression evolve as the world develops, especially the online behaviors, which may not be covered detailedly in the previous depression criteria. On the other hand, some work considered online behaviors on social media. [Choudhury *et al.*, 2013] extracted several feature groups like engagement and emotion features to detect depressed users on Twitter. However, these feature groups were not regarded as different modalities, so that the relation across different feature groups can hardly be captured without a systematic multimodal framework.

In this paper, we work towards timely depression detection via harvesting social media. This work is non-trival owning to the following challenges: 1) As far as we know, there is no public available large-scale benchmark datasets for depression research that are suitable to our study. 2) Users' behaviors on social media are multi-faceted. It is hard to characterize the users from discriminant perspectives and capture the relation across different modalities. 3) Although users' behaviors are rich and diverse, only a few are symptoms of depression, so the depressive-oriented features are sparse on social media and hard to be captured. Towards this end, we first construct well-labeled depression and non-

---

[1]http://www.who.int/mediacentre/factsheets/fs369/en/.
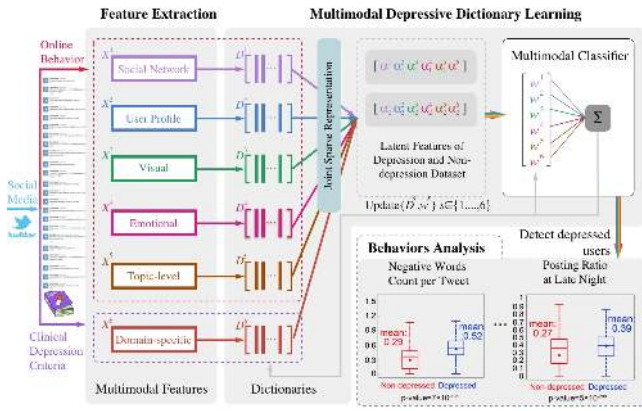
[2]http://tinyurl.com/zrnrw5j.

Figure 1: Illustration of our framework.

depression datasets on Twitter by rule-based heuristic methods. Second, to represent each user in our datasets, we extract depression-oriented features, inspired not only by the offline symptoms from the depression criteria, but also online behaviors on social media. These features are composed of six groups, namely, social network features, user profile features, visual features, emotional features, topic-level features, and domain-specific features. Third, taking each feature group as a single modality, we devise a multimodal depressive dictionary learning model (MDL) to learn the sparse user representations. In addition to the sparse representation, this model also considers the relation among the six modalities. We carry out extensive experiments on the labeled datasets to validate our model (+3% to +10% better than baselines in terms of F1-Measure) and analyze the contributions of features.

Finally, we conduct a series of case studies to reveal the underlying behavior discrepancy between depressed and non-depressed users on social media. On account of the limited users on well-labeled depression dataset, we construct a large-scale dataset on Twitter and uncover more depressed users by our proposed method. In particular, we get some statistical findings, such as: 1) Depressed users tend to leak off more emotions on social media, especially negative emotions. Their negative emotion words are 79% more than those of non-depressed users. 2) Depressed users tend to post 44% more tweets between 23:00 and 5:00 on average. More detailed results can be found in section 6.3. Our framework is presented in Figure 1.

We summarize the main contributions in three aspects:

- We construct benchmark datasets for online depression detection and analysis, including the well-labeled depression and non-depression datasets as well as a large-scale depression-candidate dataset. In addition, we release these datasets[3] with features to facilitate wellness study for computer science and psychology.
- We extract six groups of discriminant depression-oriented features to describe users from different aspects. As only few of the users' behaviors are symptoms of depression, we present a multimodal depressive dictionary learning model to learn the sparse representa-

---

[3] http://depressiondetection.droppages.com/.

tion of users. Our methods can be used to timely detect depression, take proactive care to prevent the depressed condition to be deteriorated.
- We analyze feature contributions and online behaviors of depression. We make our efforts to reveal the behaviors not covered in depression criteria, trying to provide more perspectives and insights for depression researches.

2 Related Work

2.1 Depression Criteria and Symptoms

Depression studies came much earlier than that of Internet. Based upon the user study or questionnaire survey, many widely-accepted scales and criteria have been developed. For example, *Beck's Depression Inventory* [Beck *et al.*, 1961] comprises of 21 questions about users' mental and physiological state. Another example, *CES-D Scale* [Radloff, 1977] contains 20 questions about the mental conditions like users' guilty feelings and sleep conditions. The questions either have several options aligned with different scores or require users to feedback the degree of their situations. The depression level is diagnosed according to the scale of the total score.

On the other hand, as a standard criterion for depression diagnosis, the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* [Whooley and Owen, 2014] introduces nine kinds of depressive indicators such as depressed mood and diminished interest. Clinicians usually check whether these symptoms have been presented during a period of time to make their final decision.

Beyond all doubts, these criteria are well validated and applied to real-world cases for many years. However, as the DSM took 12 years to be evolved from DSM-IV to DSM-V, it is relatively slow for these criteria to be updated strictly, so they may not comprehensively cover the new behaviors and symptoms such as those conveyed by the timely social media.

2.2 Online Depression Analysis and Detection

With the era changing, we, including the depressed users, almost cannot survive without social media. Researchers thereby started to analyze the online behaviors of depressed users. As a preliminary research, Park et al. [2012] explored the use of language in describing depressive moods utilizing real-time moods captured from Twitter users. In their follow-up work, Park et al. [2013] conducted face-to-face interviews with 14 active Twitter users to explore their depressive behaviors. Most recently, Xu et al. [2016] attempted to explain how web users discuss depression-related issues from the perspective of the social networks and linguistic patterns.

With the aforementioned work, depression detection via social media became possible. Choudhury et al. [2013] explored the potential of using social media to detect and diagnose major depressive disorders in individual. Resnik [2015] studied the topic models in the analysis of linguistic signals for detecting depression. These depression detection efforts demonstrated that it is possible to analyze massive depressed users on social media.

However, there are some limitations of the existing work: 1) Most of them conducted their experiments on a very small

set of samples, so it is difficult to justify the robustness and generality of their findings on a large-scale depression group. 2) Few of them characterized the potential depressed users from multiple modalities of social media. 3) They did not systematically investigate the depression behaviors on social media, and very little work looked into the newly-presented depression symptoms as the world develops.

3 Problem Formulation

To formulate our problem, we declare some notations in advance. In particular, we use bold capital letters (e.g., $\mathbf{X}$) and bold lowercase letters (e.g., $\mathbf{x}$) to denote matrices and vectors, respectively. We employ non-bold letters (e.g., N) to represent scalars, and denote parameters with Greek letters (e.g., λ). Besides, we use curlicue letters to represent sets (e.g., $\mathscr{V}$). If not clarified, all vectors are in column form.

Suppose that we have a set of N_L labeled depression or non-depression samples, denoted as $\mathscr{V}^L$, and an unlabeled depression-candidate set, denoted as $\mathscr{V}^U$ with N_U samples. Let $\mathscr{S} \triangleq \{1, ..., S\}$ be a finite set of available modalities for each sample. Supposing there are M features in total, we denote M^s as the feature dimension of the sth modality, $s \in \mathscr{S}$. For each sample $v_n \in (\mathscr{V}^L \cup \mathscr{V}^U)$, we denote $\mathbf{x}_n \in \mathbb{R}^M$ as the feature vector of v_n, and let $\mathbf{x}_n^s \in \mathbb{R}^{M^s}$ be the feature vector for the sth modality of v_n. Besides, $\mathbf{y} \in \mathbb{R}^{N_L}$ denotes the label vector of our labeled samples.

With the notations above, we can formally define our problem as: *Given a set of labeled samples $\mathscr{V}^L$, we aim to build a model for depression detection by learning the sparse representations of $\mathscr{V}^L$ with S modalities, and further detect more depressed users in a large-scale unlabeled set $\mathscr{V}^U$ to analyze their common behaviors in the social networks.*

4 Data and Features

4.1 Data Collection

To make depression detection via social media, we constructed two datasets of depression and non-depression users on Twitter, which has mature APIs and is prevalent around the world. Given a Twitter user, we collected the profile information of the user and an anchor tweet to infer the mental state. As people should be observed for a period of time according to clinical experience, all the other tweets published within one month from the anchor tweet were also obtained.

Depression Dataset D_1. Based on the tweets between 2009 and 2016, we constructed a depression dataset D_1. Inspired by [Coppersmith *et al.*, 2014], users were labeled as depressed if their anchor tweets satisfied the strict pattern "(I'm/ I was/ I am/ I've been) diagnosed depression". In this way we obtained 1,402 depressed users and 292,564 tweets within one month.

Non-Depression Dataset D_2. We constructed a non-depression dataset D_2, where users were labeled as non-depressed if they had never posted any tweet containing the character string "depress". As twitter has over 300 million active users and 10 billion tweets per month, we select the tweets on December 2016.

Although D_1 and D_2 are well-labeled, the depressed users on D_1 are too few so we constructed a larger dataset D_3 for depression behaviors discovery.

Depression-candidate Dataset D_3. Based on the tweets on December 2016, we constructed an unlabeled depression-candidate dataset D_3, where users were obtained if their anchor tweets loosely contained the character string "depress". Although the depression-candidate dataset contained much noise, it contained more depressed users than randomly sampling. Finally we obtained 36,993 depression-candidate users and over 35 million tweets within one month, which will be used for online behavior analysis.

The statistics of the datasets are summarized in Table 1.

Table 1: Datasets.

Dataset	D_1	D_2	D_3
Users	1,402	>300 million	36,993
Tweets	292,564	>10 billion	35,076,677

4.2 Data Preprocessing

Before feature extraction, we noticed that the words are flexible and variant in the raw data of social media, which causes great difficulties in word matching and semantic analysis. Therefore, we carried out the following data preprocessing procedures: 1) *Emoji processing*. Emoji are incompatible with many text processing algorithms. We thus removed the emoji in Tweets' texts with an emoji library collected from Twitter, then counted them separately. 2) *Stemming*. As the keyword matching strategy is widely used, words must have unified representations regardless of tense and voice. For example, "married" and "marrying" should be represented as "marri" uniformly. We here adopted the Porter Stemmer [Porter, 2001] as the stemming algorithm. 3) *Irregular words processing*. The words on social media may be irregular because of typographical mistakes or abbreviations of common words. We thus leveraged a word2vec model trained on 400 million tweets by [Baldwin *et al.*, 2015] to obtain the regular representations of irregular words. For each word encountered, we tried to find it in the WordNet [Miller, 1995]. Once failed, we used the word2vec model to find its most related five words with the NLTK toolbox [Bird, 2006]. Afterwards, we concatenated the preprocessed content of each user's recent tweets in one month to a single document for extraction of text related features.

4.3 Feature Extraction

We intended to detect and analyze depressed users with offline and online behaviors. Regarding offline behaviors, there are clear definitions in depression criteria, which have been widely used in depression diagnosis. On the other hand, we harvested the social media and found some common online behaviors. With references in computer science and psychology, we finally defined and extracted six depression-oriented feature groups to comprehensively describe each user, presented in our data-released website for more details.

Social Network Feature. It was found that depressed users are less active in social networks and depressed individuals

perceived Twitter more as a tool for social awareness and emotional interaction [Park *et al.*, 2013]. Therefore, the social network features were worth considering, such as: 1) *Number of tweets*. We extracted the number of tweets posted historically and recently by the given user to assess the user's activeness. 2) *Social interactions*. We considered the social interaction features such as the number of the user's followings and followers to describe users' online social behaviors. 3) *Posting behaviors*. We also extracted the different posting behaviors of users to reflect the state of their lives, such as the posting time distribution.

User Profile Feature. The user profile features refer to users' personal information in social networks. It was found that people having a college degree or a regular job are less likely to be depressed [Park *et al.*, 2012]. However, there are quite few personal informations returned by Twitter APIs. Therefore, we employed the bBridge [Farseev *et al.*, 2016], a big data platform for social multimedia analytics, to obtain the genders, ages, relationships, and education levels of users.

Visual Feature. Visual features were proved effective in cross-modal problems [Zhang *et al.*, 2016] and modeling sentiments and emotions in social networks [Wang *et al.*, 2015]. Compared with texts, images are more vivid, freer, and pass more complex message. In our work, we considered users' avatars in their accounts' home pages as the first impression of users in social networks, and extracted their five-color combinations, brightness, saturation, cool color ratio and clear color ratio as visual features.

Emotional Feature. Emotional status of depressed users differs from that of common users, so the emotional features are beneficial in depression detection. We studied: 1) *Emotion words*. We extracted positive and negative words count in recent tweets with LIWC [Pennebaker *et al.*, 2001]. 2) *Emoji*. We invited three annotators to vote for the sentiment of the emoji in our library mentioned before. Based on the majority voting, we obtained a sentimental emoji library, with which we extracted the sentimental emoji counts. 3) *VAD features*. We leveraged *Affective norms for English words* [Bradley and Lang, 1999] to extract the VAD features, that is, valence, arousal, and dominance, which is proved effective to explain human emotions.

Topic-level Feature. The topics concerned by depressed users and non-depressed users are likely to be significantly different, and topic models had been found to be effective in predicting depression on social media [Resnik *et al.*, 2015]. In our work, we employed the unsupervised Latent Dirichlet Allocation (LDA) model to extract the topic distribution of the documents. Based on perplexity metric frequently utilized to find the optimal number of hidden topics, we ultimately obtained 25 dimensional topic features.

Domain-specific Feature. From the depressed users, we gained some insightful depression-related observations. Inspired by this, we extracted: 1) *Antidepressant*. We established a lexicon of antidepressants from the Wikipedia page of "Antidepressant", with which we counted the average number of antidepressant names mentioned. 2) *Depression symptoms*. Referring to the nine groups of symptoms in DSM-IV criteria, we extracted the corresponding keywords respectively. However, different from the formal text in the criteria,

the linguistic style varies widely on Twitter. Therefore, we extended the keywords with the word2vec model mentioned before to construct a lexicon of the popular words of these symptoms on Twitter. In such way, we ultimately obtained the word counts of the nine symptom categories for each user.

5 Multimodal Depressive Dictionary Learning

Intuitively, give a sample v_n, the original multimodal feature representation $[\mathbf{x_n^1}, ..., \mathbf{x_n^S}]$ of v_n has some common patterns. Besides, depression representations are sparse with regard to the depression criteria. Thus, we present a *multimodal depressive dictionary learning model (MDL)* to detect depressed users, with the general idea of: 1) learn the latent and sparse representation of users by dictionary learning; 2) jointly model cross modality relatedness to capture the common patterns and learn the joint sparse representations; and 3) train a classifier to detect depressed users with the learned features specifically.

5.1 Uni-modal Dictionary Learning

Although we extract a rich set of features from each modality, not all of them are obviously related to depressed users. In addition, as the contents on social media are usually in freestyle, some noises were extracted as well, somehow impacting the detection accuracy. We thus turned to learn the latent and sparse representation of users by dictionary learning. Given the original feature representation $\mathbf{X} = [\mathbf{x_1}, ..., \mathbf{x_{N_L}}] \in \mathbb{R}^{M \times N_L}$, dictionary learning aims to learn a set of latent concepts or feature patterns, $\mathbf{D} = [\mathbf{d_1}, ..., \mathbf{d_D}] \in \mathbb{R}^{M \times D}$ and a latent sparse representation $\mathbf{A} = [\alpha_1, ..., \alpha_{N_L}] \in \mathbb{R}^{D \times N_L}$, with the following empirical cost,

$$\min_{\mathbf{D}} \frac{1}{N_L} \sum_{n=1}^{N_L} l(\mathbf{x_n}, \mathbf{D}), \ s.t., ||\mathbf{d_k}||_{l_2} \le 1, \forall k = 1, ..., D, \quad (1)$$

where the unsupervised loss $l(\mathbf{x_n}, \mathbf{D})$ is defined as,

$$\min_{\alpha_\mathbf{n}} \frac{1}{2} ||\mathbf{x_n} - \mathbf{D}\alpha_\mathbf{n}||_{l_2}^2 + \lambda_1 ||\alpha_\mathbf{n}||_{l_1} + \frac{\lambda_2}{2} ||\alpha_\mathbf{n}||_{l_2}^2, \quad (2)$$

where λ_1 and λ_2 are the regularizing parameters. The l_1-norm is applied to regulate the learned representation $\alpha_\mathbf{n}$ to be sparse.

5.2 Multimodal Joint Sparse Representation

In fact, different modalities are not independent of each other and share some common patterns which cannot be captured by the uni-modal dictionary learning. Therefore, dictionary learning was extended to multimodal to fuse features across modalities and learn the joint sparse representation to obtain the latent features. As our samples have S modalities, we denote $\mathbf{D^s} \in \mathbb{R}^{M^s \times D}$ as the corresponding dictionary of the sth modality and the latent feature $\mathbf{A_n} = [\alpha_\mathbf{n}^\mathbf{1}, \alpha_\mathbf{n}^\mathbf{2}, ..., \alpha_\mathbf{n}^\mathbf{S}] \in \mathbb{R}^{D \times S}$ as the sparse representation of the nth sample v_n. So the empirical cost $l(\mathbf{x_n}, \mathbf{D})$ would be,

$$\min_{\mathbf{A_n}} \frac{1}{2} \sum_{s=1}^{S} ||\mathbf{x_n^s} - \mathbf{D^s}\alpha_\mathbf{n}^\mathbf{s}||_{l_2}^2 + \lambda ||\mathbf{A_n}||_{l_{21}}, \quad (3)$$

where λ is a regularization parameter to balance the joint sparse and the reconstruction error. The l_{21}-norm of $\mathbf{A_n}$ is defined as,

$$||\mathbf{A_n}||_{l_{21}} = \sum_{d=1}^{D}(\sum_{s=1}^{S} A_{n_{ds}}^2)^{\frac{1}{2}}, \qquad (4)$$

which encourages row sparsity in $\mathbf{A_n}$. By regularizing the l_{21}-norm of $\mathbf{A_n}$, we encourage collaboration cross modalities so that the same dictionary atoms from different modalities present the same concept and the sparse representations from different modalities consist with each other.

The overall optimal dictionaries and joint sparse representation of each sample can be obtained by optimizing Eqn.(1) and Eqn.(3) with alternating direction method of multipliers [Parikh and Boyd, 2014] and projected stochastic gradient [Aharon and Elad, 2008], respectively.

5.3 Depression Classification

With the learned joint sparse representations $\mathbf{A_n^*}, n = 1, ...N_L$, we can train a binary classifier to detect depressed users based on cumulative loss as follows,

$$\min_{\mathbf{W}} \sum_{s=1}^{S} \sum_{n=1}^{N_L} l_{su}(y_n, \mathbf{w^s}, \alpha_{\mathbf{n}}^{\mathbf{s}*}) + \frac{p}{2} \sum_{s=1}^{S} ||\mathbf{w^s}||_{l_2}^2, \qquad (5)$$

where $\mathbf{W} = [\mathbf{w^1}, \mathbf{w^2}, ...\mathbf{w^S}] \in \mathbb{R}^{D \times S}$ is the coefficient matrix, p is a regularization parameter, and $l_{su}(y_u, \mathbf{w^s}, \alpha_{\mathbf{v}}^{\mathbf{s}*})$ is a loss function that measures how the classifier, parametrized by $\mathbf{w^s}$, can predict y_n by observing $\alpha_{\mathbf{n}}^{\mathbf{s}*}$. For our binary classification problem, l_{su} can be naturally chosen as the logistic regression loss,8

$$l_{su}(y_n, \mathbf{w^s}, \alpha_{\mathbf{n}}^{\mathbf{s}*}) = log(1 + e^{-y_n \mathbf{w^s}^T \alpha_n^*}). \qquad (6)$$

Then, Eqn.(5) can be directly solved via gradient descent.

6 Experiments

6.1 Experimental Setup

We validated the effectiveness of the MDL method on depression dataset D_1 and non-depression dataset D_2. We first tuned three key parameters of the proposed MDL and tried different scales of depressed users to obtain the best performance, then compared it with baseline methods. Next, we compared different modality combinations to validate the effectiveness of the extracted feature groups. Finally, in order to obtain large-scale depressed users, we made detection on depression-candidate D_3 and compared the results to D_2 to analyze online behaviors of depression.

Comparison Methods

We compared the following classification methods:

- **Naive Bayesian (NB)**. Naive Bayesian is a widely used classifier [Pedregosa et al., 2011]. All the features are directly inputted into the classifier and the classifier outputs whether this user is depressed.
- **Multiple Social Networking Learning (MSNL)**. MSNL [Song et al., 2015] is a novel model for multi-view learning by seamlessly analyzing information from multiple sources. It was proposed for a binary classification problem, volunteerism tendency prediction.

- **Wasserstein Dictionary Learning (WDL)**. WDL [Rolet et al., 2016] is a dictionary learning model using the Wasserstein distance as the fitting error between each original point and its reconstruction to leverage the similarity shared by the features.
- **Multimodal Depressive Dictionary Learning (MDL)**. The proposed method in this paper.

We trained and tested these methods under 5-fold cross validation, with over 10 randomized experimental runs.

Metrics

We evaluated the detection performance of our method and comparison methods in terms of Accuracy (Acc.), Macro-averaged Recall (Rec.), Macro-averaged Precision (Prec.), and Macro-averaged F1-Measure (F1).

6.2 Performance

Parameter Tuning

There are three key parameters in the MDL: two regularization parameters, λ in Eqn.(3) and p in Eqn.(5), as well as an implicit parameter D. The optimal values of these parameters were carefully tuned via grid search with small but adaptive step size. The search range for λ, p, and D are [0.001, 0.04], [$10^{-5}, 10^{-1}$], and [50, 200], respectively. The parameters corresponding to the best F1 measure were used to report the final results. For other competitors, the procedures to tune the parameters are the same to ensure fair comparison. We finally observed that MDL reached the optimal performance when $\lambda = 0.007$, $p = 10^{-2.5}$, and $D = 130$. Figures 2(a), 2(b) and 2(c) illustrate the performance of our method with respect to these three parameters, which were tuned by varying one and fixing the others with optimal values.

Scalability Analysis for Depressed Users

We justified the model performance on different scales of depressed users. With 1,402 depressed users in total, we fixed the capacity of our dataset to 1,500 and varied the scale of depressed users from 10% to 90% with increment of 10%. Figure 2(d) shows the trend of detection performance with different proportions of depressed users. It can be found that our method achieved an outstanding performance when depression users' scale laid between 40% to 60%, with the best performance in 50%. However, we retrieved a decent performance under imbalanced scales. Therefore, in the following experiments, we randomly select 1,402 non-depressed users on D_2 to make the scale of depressed users to be 50%.

Method Comparison

We compared the detection performance of MDL and baselines in terms of the four selected measures. The comparison results were summarized in Figure 2(e). From this figure, we have the following observations: 1) WDL achieved better performance than NB by 10% which shows that latent and sparse representation are effective in depression detection. 2) The MSNL and the presented MDL outperformed WDL by 5% to 8%, which verifies that modeling relation among modalities is beneficial for depressed user detection. 3) The MDL method surpassed the WDL method by 3%, which justifies the important role of modality collaboration during sparse representation learning. 4) Our MDL method achieved the best perfor-
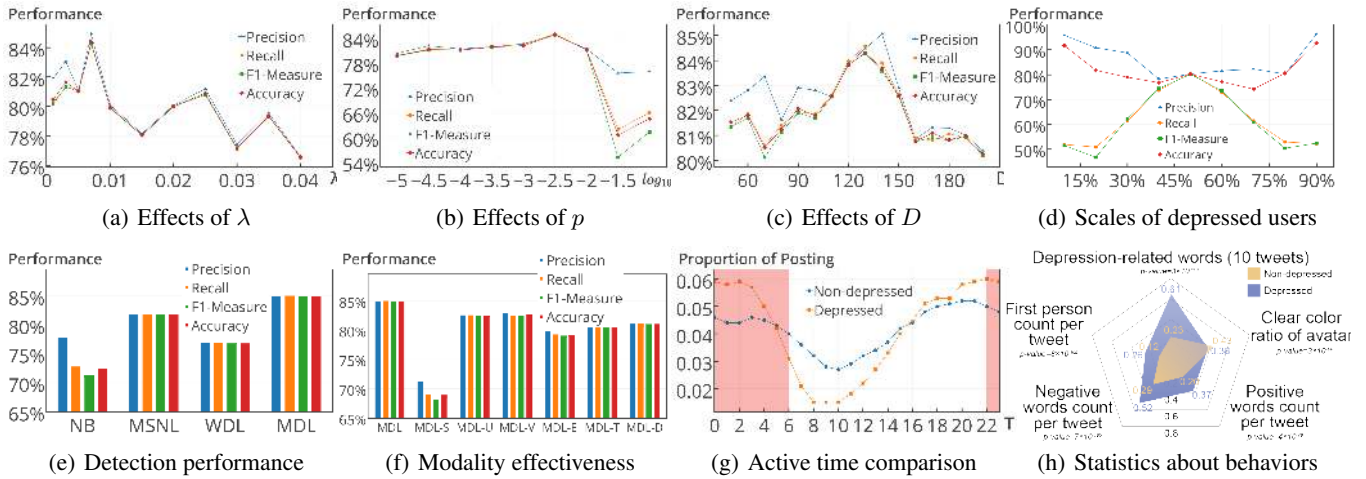
Figure 2: Experimental results of (a) effects of λ with $p = 10^{-2.5}$ and $D = 130$; (b) effects of p as $\lambda = 0.007$ and $D = 130$; (c) effects of D when $\lambda = 0.007$ and $p = 10^{-2.5}$; (d) performance over different scale of depressed users; (e) effectiveness comparison between MDL and baselines (%); (f) effectiveness comparison of the MDL using different modalities; (g) active time distribution comparison, and (h) difference between depressed and non-depressed users w.r.t. five representative behaviors.

mance with 85% in F1-Measure, indicating that combining multimodal strategy and dictionary learning strategy is effective in depression detection.

Modality Contribution Analysis

To study the effectiveness of different modalities, we further constructed an experiment to feed our model with one modal unselected each time. Specifically, we first used all modalities, denoted as MDL. We then removed the six modalities separately and denoted them as MDL-S, MDL-U, MDL-V, MDL-E, MDL-T, and MDL-D, respectively. From the results shown in Figure 2(f) we could see that the MDL outperforms the others by 2% at least, which indicates that each modality do contribute to depression detection. Besides, it can be seen that the performance suffers the most from removing the social network modality by 16%, indicating that users' posting behaviors are more discriminative than the contents they posted. In addition, the emotion modality also contributes much to the performance, which shows the depressed users usually have different emotion status over a long period.

6.3 Depression Behaviors Discovery

Besides analyzing the effectiveness of our MDL model, we further expected to compare some online behaviors between depressed and non-depressed users. Although the depressed users are well-labeled on D_1, its size is so small that we made detection on D_3 to find more depressed users. In this way, we obtained 19,233 depressed users, which were compared to the non-depressed users on D_2.

In this way, we have some interesting findings shown in Figure 2(g) and 2(h). 1) Posting time. The depressed users are more likely to post tweets (+44% on average) between 23:00 and 6:00, indicating that they are susceptible to insomnia. 2) Emotion catharsis. Depressed users have 0.37 positive words and 0.52 negative words per tweet, which surpasses those of non-depressed users by 0.17 and 0.23. It shows that although all users are likely to say more about their bad moods, depressed users express more emotions, especially negative emotions, on social media. Furthermore,

the clear color ratio of depressed users' avatars are 5% lower, which presented more repressed emotions to others. 3) Self-awareness. Compared to non-depressed users, nearly 200% more first personal pronouns (0.26 per tweet) are used in tweets of depressed users, which may reflect their suppressed monologues and strong senses of self-awareness. 4) Live sharing. Depressed users post antidepressant and depression symptom words 165% more (0.061 per tweet) than non-depressed users (0.023 per tweet) on average, indicating that they are willing to share what they encountered in the real life.

7 Conclusion

This paper aims to make timely depression detection via harvesting social media. With the benchmark depression and non-depression datasets as well as well-defined discriminative depression-oriented feature groups, we proposed a multimodal depressive dictionary learning method to detect depressed users in Twitter. We then analyzed the contribution of the feature modalities and detected depressed users on a large-scale depression-candidate dataset to reveal some underlying online behaviors discrepancy between depressed users and non-depressed users on social media. Since online behaviors cannot be ignored in modern life, we expect our findings to provide more perspectives and insights for depression researches in computer science and psychology.

References

[Aharon and Elad, 2008] Michal Aharon and Michael Elad. Sparse and redundant modeling of image content using an image-signature-dictionary. *Siam Journal on Imaging Sciences*, 1(3):228–247, 2008.

[Akbari *et al.*, 2016] Mohammad Akbari, Xia Hu, Liqiang Nie, and Tat-Seng Chua. From tweets to wellness: Wellness event detection from twitter streams. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 87–93, 2016.

[Baldwin *et al.*, 2015] Timothy Baldwin, Marie Catherine De Marneffe, Bo Han, Young Bum Kim, Alan Ritter, and Wei Xu. Shared tasks of the 2015 workshop on noisy user-generated text: Twitter lexical normalization and named entity recognition. In *Proceedings of the ACL 2015 Workshop on Noisy User-generated Text*, pages 126–135, 2015.

[Beck *et al.*, 1961] Aaron Beck, Martin Mendelson, J Mock, and J Erbaugh. An inventory for measuring depression. *Archives of General Psychiatry*, 4(6):561–571, 1961.

[Bird, 2006] Steven Bird. Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72, 2006.

[Bradley and Lang, 1999] Margaret M Bradley and Peter J Lang. Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, Technical report C-1, the center for research in psychophysiology, University of Florida, 1999.

[Choudhury *et al.*, 2013] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. In *Proceedings of the International Conference on Weblogs and Social Media*, pages 128–137, 2013.

[Coppersmith *et al.*, 2014] Glen Coppersmith, Mark Dredze, and Craig Harman. Quantifying mental health signals in twitter. In *The Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal To Clinical Reality*, pages 51–60, 2014.

[Farseev *et al.*, 2016] Aleksandr Farseev, Ivan Samborskii, and Tat-Seng Chua. bbridge: A big data platform for social multimedia analytics. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 759–761. ACM, 2016.

[Lin *et al.*, 2016] Huijie Lin, Jia Jia, Liqiang Nie, Guangyao Shen, and Tat-Seng Chua. What does social media say about your stress? In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 3775–3781, 2016.

[Miller, 1995] George A. Miller. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[Parikh and Boyd, 2014] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations & Trends in Optimization*, 1(3):127–239, 2014.

[Park *et al.*, 2012] Minsu Park, Chiyoung Cha, and Meeyoung Cha. Depressive moods of users portrayed in twitter. In *Proceedings of the ACM SIGKDD Workshop on healthcare informatics*, pages 1–8, 2012.

[Park *et al.*, 2013] Minsu Park, David W. McDonald, and Meeyoung Cha. Perception differences between the depressed and non-depressed users in twitter. In *Proceedings of the International Conference on Weblogs and Social Media*, pages 476–485, 2013.

[Pedregosa *et al.*, 2011] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[Pennebaker *et al.*, 2001] James W. Pennebaker, Martha E. Francis, and Roger J. Booth. Linguistic inquiry and word count. *Psychosomatic Medicine*, 63:517–522, 2001.

[Porter, 2001] Martin F Porter. Snowball: A language for stemming algorithms. *Retrieved March*, 1:2014, 2001.

[Radloff, 1977] Lenore Sawyer Radloff. The ces-d scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3):385–401, 1977.

[Resnik *et al.*, 2015] Philip Resnik, William Armstrong, Leonardo Claudino, Thang Nguyen, Viet An Nguyen, and Jordan Boyd-Graber. Beyond lda: Exploring supervised topic modeling for depression-related language in twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 99–107, 2015.

[Rolet *et al.*, 2016] Antoine Rolet, Marco Cuturi, and Gabriel Peyre. Fast Dictionary Learning with a Smoothed Wasserstein Loss. In *International Conference on Artificial Intelligence and Statistics*, pages 630–638, 2016.

[Song *et al.*, 2015] Xuemeng Song, Nie Liqiang, Luming Zhang, Mohammad Akbari, and Chua Tat-Seng. Multiple Social Network Learning and Its Application in Volunteerism Tendency Prediction. In *The International ACM SIGIR Conference*, pages 213–222, 2015.

[Wang *et al.*, 2015] Xiaohui Wang, Jia Jia, Jie Tang, Boya Wu, Lianhong Cai, and Lexing Xie. Modeling emotion influence in image social networks. *IEEE Trans. Affective Computing*, 6(3):286–297, 2015.

[Whooley and Owen, 2014] Whooley and Owen. *Diagnostic and Statistical Manual of Mental Disorders (DSM)*. John Wiley & Sons, Ltd, 2014.

[Xu and Zhang, 2016] Ronghua Xu and Qingpeng Zhang. Understanding online health groups for depression: Social network and linguistic perspectives. *Journal of Medical Internet Research*, 18(3):e63, 2016.

[Zhang *et al.*, 2016] Hanwang Zhang, Xindi Shang, Huan-Bo Luan, Meng Wang, and Tat-Seng Chua. Learning from collective intelligence: Feature learning using social images and tags. *TOMCCAP*, 13(1):1:1–1:23, 2016.