Exploiting Temporal Information in a Two-Stage Classification Framework for Content-Based Depression Detection

Yu-Chun Shen[1], Tsung-Ting Kuo[1], I-Ning Yeh[2], Tzu-Ting Chen[3], and Shou-De Lin[1,*]

[1] Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
{r98922071,d97944007,sdlin}@csie.ntu.edu.tw
[2] St. Joseph Hospital, Yunlin, Taiwan
yehining@gmail.com
[3] National Taiwan University Hospital Yun-Lin Branch, Yunlin, Taiwan
tzuting.chen@gmail.com

Abstract. Depression has become a critical illness in human society as many people suffer from the condition without being aware of it. The goal of this paper is to design a system to identify potential depression candidates based on their write-ups. To solve this problem, we propose a two-stage supervised learning framework. The first stage determines whether the user possesses apparent negative emotion. Then the positive cases are passed to the second stage to further evaluate whether the condition is clinical depression or just ordinary sadness. Our training data are generated automatically from Bulletin Board Systems. The content and temporal features are designed to improve the classification accuracy. Finally we develop an online demo system that takes a piece of written text as input, and outputs the likelihood of the author currently suffering depression. We conduct cross-validation and human study to evaluate the effectiveness of this system.

Keywords: Depression Detection, Time Information, Text Classification.

1 Introduction

Depression has gradually become a common mental illness in the modern era. According to World Health Organization, 121 million people are affected by depression, but less than 25% of those people receive adequate treatment (Saraceno, B. 2002).

Depression is a type of mental disease without apparent symptoms, especially during the early stage (Feightner et al. 1990). Sometimes the patients do not understand their drastic mood swings are caused by depression. With the goal to combat this illness, this paper presents an early-detection mechanism that is capable of identifying potential depression cases given the written materials of the subjects.

Depression detection has been studied for a while but most of the existing depression detection systems are not fully automatic. Existing systems usually require the potential subjects to take online evaluation tests. Then a simple rule-based system

* Corresponding author.

J. Pei et al. (Eds.): PAKDD 2013, Part I, LNAI 7818, pp. 276–288, 2013.

or a more complicated learning-based system can be designed. Such process has significant limitations in real-world usage. First, the potential candidates need to be manually identified (usually by another human). Second, even if the potential candidates can be identified, the validity of the test results remain to be questioned as the test takers may not take the test seriously.

To address the above concerns, we design a supervised-learning system for text-based depression detection. Our system does not need the potential candidates to undergo an evaluation, as it scans through their write-ups on public platforms (e.g. on blogs or BBS) to make a decision. To design such a system, there are several issues to address:

Q1. What kind of classifier should be designed? A binary classification tool that separates depression from non-depression? A multi-class classifier that distinguishes multiple mental diseases including depression? Or something else?

Q2. Given the answer from Q1, how to obtain the labeled training data?

Q3. What kinds of features are useful for this task?

Q4. What kind of evaluation procedure is considered as a sound mechanism to assess the quality of the system?

In the following sections, we will discuss the above issues to design a text-based depression detection framework.

2 Methodology

We investigate how to design a learning-based system that is capable of determining whether the author of a piece of writing is suffering or on the verge of suffering depression. Furthermore, we investigate how the temporal information can be exploited in such task.

Our training data is collected from the most popular bulletin board system (BBS) in Taiwan, the PTT. Figure 1 is a screenshot of the user interface of PTT. Due to the anonymous property of the cyber world, users are more willing to express their true feelings on web platforms than they do in the real world. This makes such public-sharing platform a good source of data for depression detection, and consequently enables our study on depression detection through text mining.

In PTT, there are more than twenty thousand boards focusing on wide range of topics, such as politics, sports, life and game. Over 1 million registered users post tens of thousands of new posts every day. The main reason we use data from PTT for depression analysis is twofold. First, there is an existing depression board, the *Prozac* board, on PTT for depressed persons to express their feelings and thoughts. Second, besides this board, there are other boards that allow users to express their feelings that can be exploited as the negative training data or as the testing platform. For example, there is a *Sad* board for sad people to share their feelings, and there is a *Diary* board that allows users to write down what are on their minds.

2.1 Methodology Overview

To solve Q1 as described previously, one conventional solution for depression detection is to treat this task as a binary classification problem. One can collect some positive and negative samples to train a classifier. However, it is non-trivial to choose the source of negative data for learning in this case, as there are many different kinds of non-depression content to choose from (e.g. technical papers, news, announcements, etc.). If, say, a bunch of technical papers are chosen as negative samples, then it is very possible all non-technical manuscripts (e.g. poetry) will be classified as ‘depression articles’. Similar drawbacks happen to cases when news, novel, or other kinds of writings are used as negative examples. If a variety of different corpora are used as negative samples, then we will inevitably face a serious data-imbalance problem as there are significantly more negative instances than positive ones. Treating this task as a multi-class classification problem brings up similar concern as there are too many classes to be distinguished from depression.

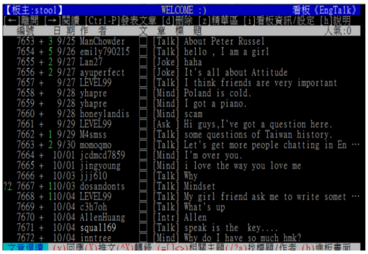


Fig. 1. PTT bulletin board

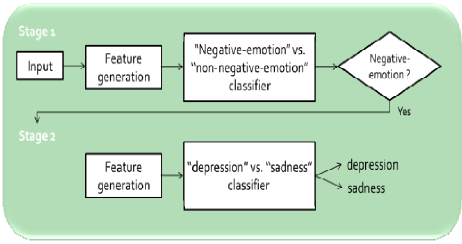


Fig. 2. Methodology overview

To further investigate the above issues, we first discuss two practical usage scenarios of our system. First, as we believe our system will be most useful in situation when somebody intents to know whether a target manuscript is written by a person suffering depression. For such manuscript to be identified in the first place, it should, to some extent, already carry certain negative information. However, not all write-ups with negative emotion are written by depression patients and it is generally hard for a non-expert to distinguish between subjects suffering ‘potential depression’ or ‘simply carry some negative thoughts’. Therefore we believe it is very important to design an automatic system for such task. That says, this usage scenario suggests that an effective depression detection engine should be able to distinguish writings that contain depressed emotion from those that contain quasi-depressed emotions.

The second usage scenario is to allow the detection system to scan through a large amount of writings (most likely crawled from the WWW) and highlight the potential candidates for early treatment. Combining with a Web Crawler, the depression detection system can act as an active detector that automatically identifies suspicious candidates for further treatment. In this case, the system needs to distinguish depression articles from many other types of writings.

To handle these two scenarios, we design a two-stage classification framework (as shown in Figure 2). In the first stage, we train an emotion classifier to classify the input manuscript into “negative-emotion” or “non-negative-emotion”. Inputs being classified as negative-emotion are then passed to the second stage, which contains a “depression” or “sadness” classifier to determine whether the author should be considered as a depression candidate or simply a user with sad emotion. The proposed framework satisfies both scenarios as we need a sadness/depression detector for the first scenario and an additional positive-negative emotion detection system can be used to filter non-negative posts in the second scenario. Intuitively, the first classifier deals with a simpler task than the second classifier, because depression and sadness are closely related concepts that cannot be easily distinguished.

2.2 Feature Generation

In the first stage, we exploit only content information for classification. We chose TF-IDF values of words as the feature representation (i.e. term features). Shown in Equation 1 and Equation 2, $n_{i,j}$ is the frequency of term t_i occurring in a person p_j's written articles. In equation 2, the denominator in idf equation represents the number of people whose articles contain the term t_i.

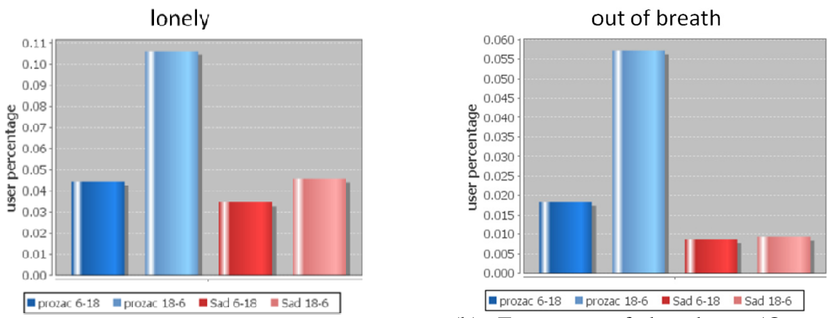
$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

$$idf_i = \log \frac{1}{|\{j : t_i \in p_j\}|} \quad (2)$$

In the second stage, we try to exploit not only content-based unigram features but also temporal information to handle a more difficult problem in distinguishing depression and sadness. The intuition is that people suffering from depression differ from normal people not only in the written content, but also on the timing of such content being produced. Below we will describe several of our observations in this aspect.

We first select two boards from PTT for observation: *Prozac* and *Sad* board, whose members consist of mostly people suffering clinical depression and people with ordinary sadness emotion (more detailed descriptions of these two boards will be stated in section 3.1). Figure 3 shows two of many examples we observed that contain time clues.

In Figure 3(a), we observe that people in both boards use the term "lonely". However, on the *Prozac* board, the term is used significantly more frequently during night time (from 18:00 to 6:00 next day), but on the *Sad* board, such difference is not significant. Similarly in Figure 3(b), the clinically depressed persons use "out of breath" much more frequently during the night time, while sad persons do not show such trend. Both observations motivate us to bring the temporal information into play to better recognize depression cases.



(a). Frequency of the word 'lonely' being used. In *Prozac* board, the frequency of usage at night is much higher than that of daytime. In *Sad* board, the difference is not significant.

(b). Frequency of the phrase 'Out of breath' being used. In *Prozac* board, the frequency of usage at night is higher than that of daytime. In *Sad* board, the difference is not significant.

Fig. 3. Examples of the data observation

The posting time of a post on BBS, blogs, or micro-blogs is usually fine-grained, accurate to the minutes or even seconds. To incorporate this information into our model, we define the temporal feature as a pair containing the term itself and the time of it being posted. We encode the posting time into different spans as shown in Table 1. Then we can apply TF-IDF in equation 1 and 2 to generate the value of temporal features. The only difference is that now $n_{i,j}$ stands for the frequency of a " timeslot-term" pair.

Table 1. Predefined temporal categories

Category	#	Description	Feature (tired)
12Hour(6-18)	2	6-18, 18-6	18-6 tired
12Hour(8-20)	2	8-20, 20-8	20-8 tired
12Hour(10-22)	2	10-22, 22-10	10-22 tired
6Hour(6-12)	4	6-12, 12-18, 18-24, 24-6	18-24 tired
6Hour(8-14)	4	8-14, 14-20, 20-2, 2-8	20-2 tired
6Hour(10-16)	4	10-16, 16-22, 22-4, 4-10	16-22 tired
24Hour	24	24 hours	20 tired
12Month	12	12 months	March tired
4Season	4	spring, summer, fall, winter	Spring tired
2Season	2	spring-summer, fall-winter	Spring-summer tired
Season change	2	April+May+September+October, others	Others tired
Workday-Weekend	2	workday, weekend	Workday tired

Each of these temporal categories has its own meaning. The first three categories divide a day into two sections, “daytime” vs. “night” or “working hours” vs. “non-working hours”. The next 3 categories further separate a day into four sections. Next, we use more fine-grained slots such as “24Hour” and “12Month”. Units such as “4Season”, “2Season” and “Season change” are motivated by the observation that weather may affect human beings’ emotions, in particular for people suffering depression. Finally, we define the category “Workday-Weekend” to reflect the observation that depressed people sometimes suffer occupational function impairment, which leads to different mental conditions or behaviors between workday and weekend.

There are 12 different categories in Table 1. Thus, one single term feature can be converted to 12 different temporal features. Assuming there is a message posted on *Mon Mar 21 20:16:11* with a term *tired*, the resulting temporal features are listed in the right-most column in Table 1.

3 Experiment

3.1 How the Training Data Can Be Obtained?

We conduct experiment on PTT data, and use 5-fold cross validation to obtain the accuracy. As in the first stage, we hope to train a classifier that distinguishes messages with negative emotions from those with non-negative emotions. We choose posts in *Gossiping* and *Happy* boards on PTT to represent messages with non-negative emotions and posts from *Prozac* and *Sad* boards to represent the ones with negative emotions. The Gossiping board is chosen to be the data source not only because it is the most popular board on PTT but also due to the variety of posts with different purposes and emotions that we believe are general enough to represent many different types of write-ups.

In the second stage, we train a *depression vs. sadness* classifier. We use the posts from *Prozac* and *Sad* boards as the positive and negative examples respectively, while the former contains mostly depression posts and the latter contains mostly sad (but not necessarily depressed) messages. *Prozac* board is a place for people with clinical depression or potential candidates to interact with each other, share their experiences and express their emotions. The word Prozac is the name of medicine for treating depressive disorder. The categories of posts in *Prozac* board are clearly specified, as listed in Table 2. Note that we use only the articles from the *cloudy day* category, which contain posts created by people with depression to express their thoughts and feelings. Other categories such as *information* or *sunny* day are excluded while generating the training data to ensure the quality of the training data. Table 3 shows the basic statistics of the training data.

Table 2. Categories of posts in Prozac board

Self-induction	For the new users to introduce themselves
Medical	For the discussion of hospital, doctor, counseling, medicine and symptoms
Experience	For the discussion of experience, treatment situation
Information	For news, information, research and books about depression
Question	For asking questions
Transcription	For transcription of posts from other boards
Cloudy day	For self-introduced users to vent bad emotions
Sunny day	For self-introduced users to share good emotions
Chat	For self-introduced users to use when other categories are not applicable

Table 3. Basic statistics of training data

Statistics	Gossiping	Happy	Prozac	Sad
Time period	08/01~10/12	10/01~10/12	08/01~10/12	10/01~10/12
Posts number	6505	11209	6015	4900
Users number	1699	2695	1027	1652

3.2 Stage 1 – Negative vs. Non-negative Classifier

Most of the posts in PTT are in Traditional Chinese. Therefore we focus on Chinese posts in this experiment, though the proposed technique is language universal. We first use Yahoo's segmentation API service to perform word-segmentation on the posts because of its stability and efficiency. We then filter out single-character terms in Chinese as they are more likely to be stop-words without apparent meaning. Besides, terms used by too few people (< 25) are removed to avoid over-fitting. In the end, 5622 unigram terms are chosen.

Here we collect all posts of a user to extract the unigram features. Note that the subject is persons not posts, because in practice it is more critical to know whether a person is a potential depression patient. We exploit linear SVM using Liblinear(Fan et al. 2008) for classification. Note that each feature is scaled to [0, 1]. The 5 fold cross validation accuracy reaches 96.17%, which means negative-posts and non-negative-posts can be easily separated based on content.

3.3 Stage 2 – Depression vs. Sadness Classifier

The feature generation process is the same as mentioned in previous section. Similar to what we did in stage 1, we first extract 2460 unigrams for learning. The results are shown in Table 4. The 5-fold cross validation accuracy is 81.86%, which signifies that depression and sadness posts are indeed separable, but doing so is much harder than separating negative vs. non-negative write-ups.

Table 4. Experiment results in stage 2 (term features only)

CV accuracy(5 fold)	81.86%
CV AUC(5 fold)	0.8863

To verify the quality of training data and learning process, we also examine the weight of each feature. Table 5 shows the top 40 depression features, and Table 6 shows the top 40 sad features. All features have been translated into English.

Table 5. Top 40 depression features and weights

Take medicine	2.17	Normal people	1.34	World	1.08	Day after tomorrow	0.98
Doctor	1.96	Horrific	1.27	Prozac	1.06	Crowd	0.98
Depression	1.79	Boyfriend	1.26	Yesterday	1.05	Dark night	0.98
School	1.78	Medicine	1.24	Squeeze out	1.02	Subsist	0.98
Suicide	1.62	Emotion	1.23	Clinic	1.02	Tight Chest	0.97
Counseling	1.50	Pain	1.20	Destroy	1.02	Agree	0.97
Sick	1.44	Psychologist	1.18	Patient	1.02	State	0.96
They	1.44	Afraid	1.17	Thought	1.00	Stable	0.95
Happy	1.42	Bedtime	1.14	Ward mate	0.99	Appetite	0.95
Trembling	1.39	What if	1.09	Can't hold out	0.98	Social	0.94

Table 6. Top 40 sad features and weights

Miserable	-1.70	Cant' let go	-1.24	Day	-1.01	Opportunity	-0.92
Cheer up	-1.37	Hesitate	-1.15	Can't see	-1.00	Immediately	-0.92
We	-1.37	Mood	-1.15	Put	-1.00	Misunderstand	-0.90
Hope	-1.31	Feel	-1.12	Memory	-0.99	Time	-0.90
Torture	-1.30	Quietly	-1.10	Really	-0.98	A little	-0.90
Sad	-1.27	Think	-1.08	And	-0.97	Relieved	-0.90
Broken-heart	-1.27	Actual	-1.03	Tear	-0.96	Sorry	-0.88
Such	-1.26	Future	-1.02	Slowly	-0.95	Popular feeling	-0.87
Obviously	-1.26	Always	-1.02	Children	-0.94	Try one's best	-0.86
Exaggerate	-1.24	Affair	-1.01	Future	-0.94	Outcome	-0.85

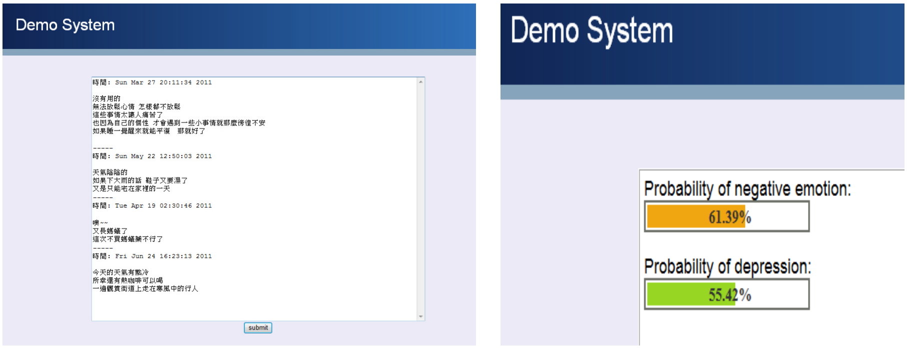
From the top 50 depression features, we can see that some of them are related, such as *depression, suicide* and *pain*; some are related to somatic symptoms like *chest tightness* and *appetite*; and some provide more insights into the mental world of people with depression such as terms including *normal people*, *squeeze out*, *social* and *crowd*. The term *knife* might indicate the most common tool they used or imagined to hurt themselves with. Furthermore, *dark night, destroy, unable to hold out* and *terror* show the darkness and fear inside their minds. On the other hand, the top keywords for *sad* posts are more general.

Table 7. Results of using temporal features only

Feature	#	Accuracy	AUC
Term	2460	81.86%	0.8863
12Hour(6-18)	4920	81.97%	0.8817
12Hour(8-20)	4920	80.55%	0.8817
12Hour(10-22)	4920	80.55%	0.8790
6Hour(6-12)	9833	81.34%	0.8897
6Hour(8-14)	9812	80.55%	0.8746
6Hour(10-16)	9812	80.55%	0.8746
24Hour	49644	76.15%	0.8109
12Month	28759	78.09%	0.8439
4Season	9839	80.03%	0.8711
2Season	4920	80.81%	0.8797
Season change	4920	80.93%	0.8825
Workday	4920	80.89%	0.8821
All	147237	84.47%	0.9167

Table 8. Results of using both term features and temporal features

Feature	#	Accuracy	AUC
Term	2460	81.86%	0.8863
12Hour(6-18)	7380	82.34%	0.8931
12Hour(8-20)	7380	82.34%	0.8940
12Hour(10-22)	7380	81.71%	0.8892
6Hour(6-12)	12290	82.79%	0.8987
6Hour(8-14)	12293	83.58%	0.9089
6Hour(10-16)	12272	82.79%	0.8970
24Hour	52104	82.38%	0.8980
12Month	31219	83.20%	0.9033
4Season	12299	82.98%	0.8941
2Season	7380	81.71%	0.8887
Season change	7380	81.49%	0.8915
Workday	7380	81.34%	0.8882
All	149697	84.51%	0.9176



(a). The input screen (b). System output

Fig. 4. Screenshots of the demo system

We then conduct experiments to observe the performance of using only temporal features. The results are shown in Table 7. Note that the accuracies are not significantly higher than that without temporal features. We believe it might due to the lack of sufficient data to train such fine-grained feature set, as the experiments also show that the more fine-grained a temporal category is (e.g. 24 hours, 12 month), the lower the accuracy it produces. However, when both term features and temporal features are exploited (see Table 8), the performance does show consistent improvement. In most of the temporal categories, we receive higher accuracy than using only term features, and the highest accuracy lies in "6Hour(8-14)" category. If we include all 12 temporal features, the accuracy can be boosted to 84.51%, which is significantly higher than using only term features.

4 System Demo and Manual Evaluation

To further confirm the effectiveness of our model, we construct a real-time system for demo and manual evaluation. The user interface of the system can be seen below. People can simply copy/paste user posts into the window and submit for evaluation. There is no need to fill in any personal information or questionnaires.

Figure 4 shows the screenshots of our demo system. Once the post content and timestamp are submitted, the system evaluates the posts and outputs the probabilities of the subject possessing negative emotion and depression.

4.1 Manual Evaluation

We extract posts from another board on PTT, the *Diary* board, to evaluate whether our system can identify some potential depression candidates based on their diary writings. *Diary* board is a place for PTT users to write things about their lives, feelings and thoughts. People have various reasons to post on the *Diary* board. Nevertheless, we believe that among the users who post on the *Diary* board, there might be a small fraction of them suffering from depression, and we want to test whether our system can identify some of them. We collect 15374 posts from the *Diary* board, total 2071 users. Figure 5 illustrates the flowchart of manual evaluation.

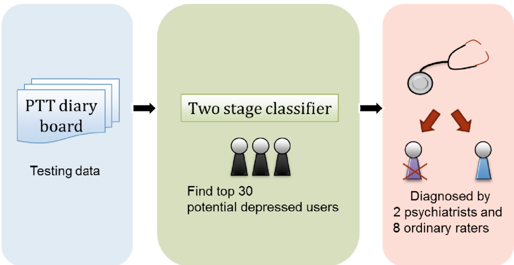


Fig. 5. Flowchart of manual evaluation

We feed the posts from the *Diary* board into our system, and rank the users using our two-stage classifier. We then asked 2 psychiatrists and 8 ordinary evaluators to diagnose (based on the writings) whether the top suspicious candidates our system identified are truly potential depression candidates. Since ordinary evaluators lack the knowledge to diagnose depression, before making the evaluation they are asked to read a brief document about the diagnosis principle released in the fourth edition of Diagnostic and Statistical Manual of Mental Disorders (DSM IV) published by American Psychiatric Association (2000). After reading the posts, evaluators have to choose from four options: non-diagnosable, major depression, moderate or minor depression, and no depression.

Table 9 shows the diagnosis results of two psychiatrists. Note that we regard a case as “major depression” if at least one of the doctors thinks so, or “moderate or minor depression” if at least one of the doctors thinks so. The results show that out of the 30 top suspects our system identified, 18 of them do possess some depression symptoms.

The agreement is even higher (21/30) for non-expert evaluators (see Table 10). The manual evaluation demonstrates that our system can effectively identify potential subjects who are suffering depression but are unaware of it from their writings.

Table 9. Diagnosis of 2 doctors

30 detected users	
major depression	9
moderate and minor depression	9
no depression	12
non-diagnosable	0

Table 10. Diagnosis of 8 ordinary raters

30 detected users	
major depression	7
moderate and minor depression	14
no depression	9
non-diagnosable	0

5 Related Work

5.1 Automated Depression Detection

Neuman and Kedma (2010) propose a system called Pedesis to automatically detect users with depression on the web, specifically on blogs. Their main idea is that depression can be hidden in some metaphors, which implies that obvious depressed terms are not the only indicators. By the help of a search engine, they extract sentences in the form of "depression is like *", and identify some popular metaphors for depression such as *black hole* or *dark cloud*. These metaphors are then used as keywords to identify users with depression. This knowledge-driven method is fundamentally different from our learning-based method.

There are researches applying machine learning methods for depression detection. We can classify them into four categories based on the sources used for detection: text, speech, facial expression and electroencephalogram (EEG).

Text. Jarrold et al. (2010) investigate whether language features can be used to diagnose depression. They use only binary classification and did not include temporal information in the features. Besides, the corpus they use takes lots of time and effort to create, while we use PTT which contains a gradually increasing resource whose data as well as labels can be extracted automatically with limit amount of efforts. Aamodt et al. (2010) develop a decision support system for depression diagnosis using case-based reasoning. However, to compare a new case with past cases, the system requires the patient to fill out a questionnaire first, thus the usage and coverage is limited.

Speech. Based on the theory that people with depression have slow and monotonous voice, Low et al. (2010) use speech characteristics as features for depression detection and compare performances of different feature combinations. Their experiments show MFCC and short time energy outperform other features. Sanchez et al. (2011) successfully exploit prosodic and spectral speech features to decide whether the speaker is depressed.

Facial expression. Cohn et al. (2009) consider not only verbal features but also facial actions to perform depression detection. They found both manual FACS coding and active appearance modeling (AAM) are correlated with depression. Maddage et al. (2009) also have similar idea of using facial features for depression detection.

EEG. Hosseinifard et al. (2011) use EEG signal as the features for depression classification and achieve high accuracy in the experiments. Unfortunately, a depression detection system like this can hardly be used as screening tool for large amount of candidates.

The related work shows that different sources of information can serve as the clue for depression detection. Our work is novel because we have not yet found any work that proposes similar two-stage online screening framework with empirical demonstration on the effectiveness of the designed temporal features.

5.2 Temporal Information in Depression

Early studies have suggested that depression might be a seasonal variation disorder. Eastwood and Stiasny (1978) conduct an analysis of hospital admission of neurotic and endogenous depression, and reported significant peaks in spring and fall. Morken et al. (2002) focus on monthly variation of depression, with gender information also considered. They find that depression admission for women reaches highest peak in November, and for men, in April. Both studies have confirmed a correlation between incidence of depression and time. Kerkhofs et al. (1991) investigate the 24 hour sleep patterns of 22 people, with 12 of them having major depressive disorder and 10 being normal people. The result shows a difference of sleeping behavior for the two groups. Normal people tend to take a nap in the early afternoon, and depressed people have no consistent sleeping period. If a person naps in the morning more often, indicating that the person lacks sleep at night, the probability of depression might be higher. These researches indicate the existence of time clues for detecting depression and non-depression candidates, which strengthen our proposal to introduce the temporal information as features.

6 Conclusion

The main contributions of this paper are listed below:

1. We propose two practical real-world usage scenarios for depression detection, and design a novel two-stage learning framework based on such scenarios.
2. We design a strategy to incorporate temporal information into the content feature and improve the detection accuracy significantly.
3. We identified an important resource, namely BBS, which allows us to automatically extract data and labels for training and testing. Furthermore, the data grows daily, which means one can obtain more training data to enhance the performance of the system. Such resources can potentially be used for other classification tasks because there are thousands of boards in this system, and one can easily obtain labeled data of specific purpose to facilitate training and testing.
4. We developed an online real-time depression detection engine which does not require the users to fill out any questionnaires or reveal their identities. For evaluation, we conduct cross-validation on the PTT dataset and find experts and normal people to judge the usability of the system. The results show that our system can indeed discover potential or hidden depression candidates that are otherwise hard to find.

The framework, data, and features we have proposed can easily be applied to design detection engines for other kinds of mental diseases such as delusional disorder, schizophrenia, anxiety disorder, etc. In the future, we plan to integrate other dimension of information into our system including the weather and geographical information, which we believe are also useful clues for depression detection.

Acknowledgements. This work was supported by National Science Council, National Taiwan University and Intel Corporation under Grants NSC101-2911-I-002-001, NSC101-2628-E-002-028-MY2 and NTU102R7501.

References

1. Aamodt, A., Gundersen, O.E., Loge, J.H., Wasteson, E., Szczepanski, T.: Case-Based Reasoning for Assessment and Diagnosis of Depression in Palliative Care. In: The International Symposium on Computer-Based Medical Systems, pp. 480–285 (2010)
2. American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 4th edn., Text Revision. American Psychiatric Association, Washington, DC (2000)
3. Cohn, J.F., Kruez, T.S., Matthews, I., Yang, Y., Nguyen, M.H., Padilla, M.T., Zhou, F., De la Torre, F.: Detecting Depression from Facial Actions and Vocal Prosody. In: International Conference on Affective Computing and Intelligent Interaction (2009)
4. Eastwood, M.R., Stiasny, S.: Psychiatric Disorder, Hospital Admission, and Season. Archives of General Psychiatry 35, 769–771 (1978)
5. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: LIBLINEAR: A Library for Large Linear Classification. Journal of Machine Learning Research 9, 1871–1874 (2008)
6. Feightner, J.W., Worrall, G.: Early Detection of Depression by Primary Care Physicians. Can. Med. Assoc. J. 142, 1215–1220 (1990)
7. Hosseinifard, B., Moradi, M.H., Rostami, R.: Classifying Depression Patients and Normal Subjects Using Machine Learning Techniques. In: Iranian Conference on Electrical Engineering, pp. 1–4 (2011)
8. Jarrold, W.L., Peintner, B., Yeh, E., Krasnow, R., Javitz, H.S., Swan, G.E.: Language Analytics for Assessing Brain Health: Cognitive Impairment, Depression and Pre-symptomatic Alzheimer's Disease. Brain Informatics, 299–307 (2010)
9. Kerkhofs, M., Linkowski, P., Lucas, F., Mendelwicz, J.: Twenty-Four-Hour Patterns of Sleep in Depression. Sleep 14, 501–506 (1991)
10. Low, L.A., Maddage, N.C., Lech, M., Sheeber, L., Allen, N.: Influence of Acoustic Low-Level Descriptors in the Detection of Clinical Depression in Adolescents. In: ICASSP, pp. 5154–5157 (2010)
11. Maddage, N.C., Senaratne, R., Low, L.A., Lech, M., Allen, N.: Video-based Detection of the Clinical Depression in Adolescents. In: International Conference on Engineering in Medicine and Biology Society, pp. 3723–3726 (2009)
12. Morken, G., Lilleeng, S., Linaker, L.M.: Seasonal Variation in Suicides and in Admissions to Hospital for Mania and Depression. Journal of Affective Disorders 69, 39–45 (2002)
13. Neuman, Y., Kedma, G., Cohen, Y., Nave, O.: Using Web-Intelligence for Excavating the Emerging Meaning of Target-Concepts. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 22–25 (2010)

14. PTT, http://www.ptt.cc/index.html (retrieved June 27, 2011)
15. Sanchez, M.H., Vergyri, D., Ferrer, L., Richey, C., Garcia, P., Knoth, B., Jarrold, W.: Using Prosodic and Spectral Features in Detecting Depression in Elderly Males. In: INTERSPEECH, pp. 3001–3004 (2011)
16. Saraceno, B.: The WHO World Health Report 2001 on mental health. Epidemiol. Psychiatr. Soc. 11, 83–87 (2002)