Tackling Mental Health by Integrating Unobtrusive Multimodal Sensing

Dawei Zhou, Jiebo Luo, Vincent Silenzio, Yun Zhou, Jile Hu, Glenn Currier, and **Henry Kautz**
University of Rochester
Rochester, NY 14627

Abstract

Mental illness is becoming a major plague in modern societies and poses challenges to the capacity of current public health systems worldwide. With the widespread adoption of social media and mobile devices, and rapid advances in artificial intelligence, a unique opportunity arises for tackling mental health problems. In this study, we investigate how users' online social activities and physiological signals detected through ubiquitous sensors can be utilized in realistic scenarios for monitoring their mental health states. First, we extract a suite of multimodal time-series signals using modern computer vision and signal processing techniques, from recruited participants while they are immersed in online social media that elicit emotions and emotion transitions. Next, we use machine learning techniques to build a model that establishes the connection between mental states and the extracted multimodal signals. Finally, we validate the effectiveness of our approach using two groups of recruited subjects.

Introduction

Mental health is a significant problem on the rise with reports of anxiety, stress, depression, suicide, and violence (Crown 2011). A large survey from BRFSS (Behavioral Risk Factor Surveillance System) (CDC, 2008, 2011, 2012) found that an astonishing 46% met criteria established by the American Psychiatric Association (APA) for having had at least one mental illness within four broad categories at some time in their lives. The categories were anxiety disorders, mood disorders (including major depression and bipolar disorders, impulse-control disorders, and substance use disorders (including alcohol and drug abuse)). Mental illness has been and remains a major cause of disability, dysfunction, and even violence and crime.

Tackling mental health on a large scale is a major challenge in terms of resources. Traditional methods of monitoring mental health are expensive, intrusive, and often geared toward serious mental disorders. More importantly, these methods do not scale to a large population of varying demographics, and are not particularly designed for those in the early stages of developing mental health problems. In the meantime, effective intervention for mental health has been severely limited by the availability of resources in healthcare professionals and treatments. Advances in computer vision and machine learning, coupled with the widespread use of the Internet and adoption of social media, are opening doors for a new and effective approach to tackling mental health using physically noninvasive, low-cost, and pervasive multimodal sensors already ubiquitous in people's daily lives.

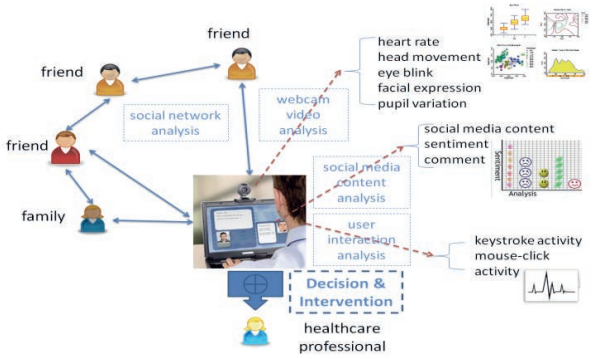

Figure 1. The proposed framework for multimodal monitoring.

We propose to combine the analysis of a subject's social media stream and image signals from a close-up video of the subject captured by today's mobile devices to construct a continuous, multifaceted view of the subject's mental health (Figure 1). We will employ both computer vision and data mining techniques to produce a real-time assessment of mental health based on the constructed fine-grained signals at the individual subject level. This study makes several contributions, including

1) Extracting fine-grained psycho-behavioral signals that reflect the mental state of the subject from imagery unobtrusively captured by the webcams built in most mobile devices (laptops, tablets, and smartphones). We develop robust computer vision algorithms to monitor real-time psycho-behavioral signals including the heart rate, eye blink rate, pupil variations, head movements, and facial expressions of the users.

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

2) Analyzing effects from personal social media stream data, which may reveal the mood and sentiment of its users. We measure the mood and emotion of the subject from the social media posted by the subject as a prelude to assessing the effects of social contacts and context within such media.
3) Establishing the connection between mental health and multimodal signals extracted unobtrusively from social media and webcams using machine learning methods.

We expect that the outcome of this research will be an effective tool for assessing the affective states of individuals on a large scale. It can be used as an enabling component for developing new mental health solutions, including identifying the onset and severity of mental health problems in individuals and may prove to be of use to clinicians, for self-awareness, and for support from family and friends. It is important to note that the proposed framework is intended to complement, not replace, existing assessment and treatment mechanisms.

Related Work

There is no reliable laboratory test for diagnosing most forms of mental illness; typically, the diagnosis is based on the patient's self-reported experiences, behaviors reported by relatives or friends, and a mental status examination.

The Internet and other technological advances offer a set of new and intriguing tools to assist in this challenge. We recognize the possibility of using computer vision, social media and machine learning technologies, as well as behavior science, to assist in identifying those individuals who may be most at risk, and doing so early enough in the process to alter their trajectory away from suicide (or to a lesser degree, depression) entirely.

Literature reviews (Tao & Tan 2005; Zeng et al. 2009) provide a comprehensive coverage of the state of the art in affective computing. Typically, physiological data are acquired using specialized wired or wireless invasive sensors that not only are cumbersome but may also cause the subject to become overly aware of being monitored that it inhibits natural behavior. It is desirable to monitor mental health in a non-intrusive fashion using low-cost, ubiquitous sensors naturally existing in people's lives. This constitutes the one of the major appreciable differences between this work and the prior art in affective computing. An early work (Cohen et al. 2009) compared clinical diagnosis of major depression with automatically measured facial actions and vocal prosody in patients undergoing treatment for depression. Both face and voice demonstrated moderate concurrent validity with depression.

Recently, McDuff et al. (2014) have shown that changes in physiological parameters during cognitive stress can be captured remotely (at a distance of 3m) of the face of the participant using a digital camera. They built a person-independent classifier to predict cognitive stress based on the remotely detected physiological parameters (heart rate, breathing rate and heart rate variability) and achieved a high accuracy of 85% (35% greater than chance).

Our methods for detecting emotional information using both online social media and passive sensors can potentially enhance the effectiveness and quality of new services delivered online or via mobile devices (Matthews et al. 2008) in current depression patient care.

Approaches

There is significant variability in the ways in which affect, emotion, and related concepts are articulated in the literature of various fields, including psychology, psychiatry, and the social sciences. For the purposes of this study, we adapt the following terminology. We use the term affective as an umbrella concept to refer broadly to moods, feelings, and attitudes. Emotion and sentiment refer to short-term affective states, which typically remain stable over a period of time on the order of minutes to hours. Mood refers to the predominant emotional pattern over periods lasting from days to weeks. This time scale for persistent affective states is chosen in order to remain compatible with those commonly used in the clinical classification of mood disorders.

It is also important to clarify what we mean by "unobtrusive sensing" here. First, similar mobile-phone apps (Lane et al. 2011) already exist for monitoring physical activities and the information we extract is primarily for self-awareness and mental health self-management, and is not shared with anyone else without the consent of the subject. Compared with wired or wireless sensors that people would otherwise need to carry or wear, our technologies are "unobtrusive" in the sense that the users would not feel being monitored as smartphones and social media are already embedded in their lives.

Multi-modal Signals

Our approach takes full advantage of online social media related activities as a source to extract signals related to different mental states and applies machine learning techniques to provide an instant and actionable reflection on an online user's mental health state for stress management and suicide prevention. To that end, we employ computer vision, user interaction logging, and social media content sentiment analysis to extract a suite of twelve signals from a user's online activities.

Webcam Video Tracking

Since most mobile devices (laptops, tablets and smartphones) are equipped with built-in video cameras, we

propose to use webcam video analysis to monitor head movement, heart rate, eye blink, pupillary response, and facial expression. These indicators, though noisy, may correlate with a user's mental states.

Head Movement Analysis. When people are reading, thinking or have emotional fluctuation, they tend to have various head motions (e.g., nod, shake, turn). To extract various head motions, we first apply the Cascade Classifier[1] in OpenCV to locate the face region, and then use the Tracking-Learning-Detection (TLD) algorithm (Kala et al. 2010) to locate feature points on the face and track them across consecutive frames. TLD is designed for real-time tracking in unconstrained environments. TLD can track and learn simultaneously, as well as resetting the tracker when it fails. It has been applied to a number of existing systems, especially for tracking applications on smartphones, tablets and AR glasses.[2] Although it may be more accurate to build a 3D head model (La Cascia et al. 2000) to track head motions. We chose to use 2D tracking for low power requirements and real time processing.

Different head motions are detected and counted as a time series of head movements. Figure 2 shows one person's head movements in three mental states (Positive, Neutral and Negative). Our data suggests that one tends to show a higher rate of head movements in positive and negative moods, but more periodic head moves in the neutral mood.

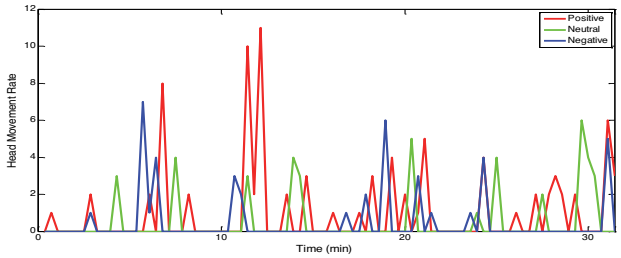

Figure 2. Detected head movements.

Heart Rate Analysis. There are a number of vital signs to measure human body's basic functions, including body temperature, pulse rate, respiration rate and blood pressure. These biological signals can be useful for determining emotional conditions. We are inspired by research on non-contact video-based analysis of human vital signs, where promising results were reported for estimating the heart rates of infants and adults (Wu et al. 2012). The foundation behind extracting the heart rate from a non-contact video of the face is that subtle color changes will be induced by blood flow. The Eulerian video magnification method by Wu et al. (2012) basically takes time series of color values for each pixel and amplifies the variation in a specific frequency band of interest. We first decompose the input video sequence into different spatial frequency bands with a Gaussian pyramid. Since the normal static heart rate for adults ranges from 45 to 120 beats per minute, or 0.75 to 2 Hz (Balakrishnan et al. 2013), we select the band between 0.6667 to 2Hz for the bandpass filter.

However, the heart rate signal can easily be overwhelmed by involuntary head movements. Therefore, we have improved the techniques by Wu et al. (2012) to account for head movement. From color amplified videos, we found that head motions generate either dark or bright parts in the images. We use four steps to exploit this effect to remove noisy estimates. First, we filter out those images with blurry faces due to motion. Next, we apply an eye pair detector (Castrillon et al. 2007) in order to locate the areas of forehead, left cheek and right cheek. Each of these three areas is then split into 10*10 patches, respectively. The average luminance trace for each small patch over time is accessible by using a sliding window strategy. The length of the window is set to ten seconds. We then explore the frequency of maximum power within the range of heart beat frequency in the spectrum to derive the heart rate estimate. Finally, a median filter is applied to smooth the estimates for each patch. The final heart rate estimate is an average of the estimates from these three areas.

Eye Blink Analysis. There are three types of eye blink: spontaneous blink, voluntary blink and reflex blink. Spontaneous blink is a physiological phenomenon, while the other two are triggered by the external environment. Therefore, the rate of eye blink would directly or indirectly reflect participant's emotion to some extent.

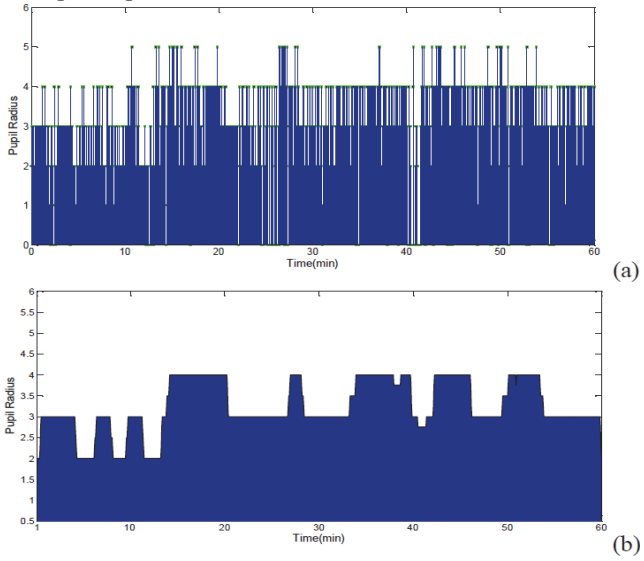

Figure 3. Pupil radius (a) and smoothed signal (b) over time.

Although a deformable model would provide higher accuracy for eye location, it incurs higher computational complexity. Therefore, we select circular Hough transform for good accuracy and low computation cost. Both eye

[1] http://docs.opencv.org/doc/tutorials/objdetect/cascade_classifier/cascade_classifier.html
[2]http://www.tldvision.com/

opening and closing result in variations in pixel intensity. The variation is calculated recursively from frame to frame. We consider it as one eye blink when the variation exceeds a threshold. Based on the time series of the pupil radius, we use a moving average filter to remove the noise and obtain the pupillary response signal as shown in Figure 3.

Pupillary Variation Analysis. Existing physiological studies suggest that the pupil will become wider when people are experiencing dramatic changes of emotions, such as happiness, anger and interest. In contrast, the pupil will be narrower due to the termination of iris sphincter muscle if people feel bored or disgusted, or lose interest.

Using a deformable template, we can determine the pupil radius. In our system, Deformable Template Matching (DTM) (Yuille et al. 1992) not only improves the accuracy but is also less time consuming for processing video. The pupil model is formulated as two concentric circles; the outer one fits the iris edge and the inner one matches the pupil. Based on the time series of the pupil radius, we use a moving average filter to remove the noise and obtain the pupillary response signal.

Facial Expression Analysis. Expression is the most apparent indication of human emotion. The consistency in one's expression will produce repeatable patterns that can be recognized and contribute to emotion evaluation.

In our system, we analyze the participants' facial expressions as they go through tweets. These tweets also include a wide variety of images that could trigger emotions. The face video stream was captured from a webcam (640 x 480 pixels) and the lighting condition was not purposely controlled. In the entire one-hour process, the participants were asked to behave naturally as they would when skimming their Twitter feeds.

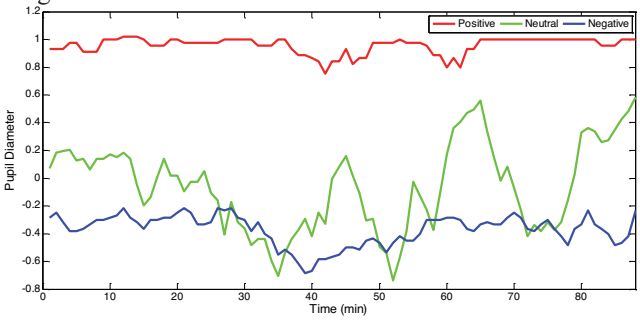

Figure 4. Facial sentiment analysis in three mental states.

To strike a balance between computational expense and reliable performance, we chose to use the Eigen Face method to map the test image set to a low-dimensional face space (Agarwal et al. 2010). Since the Chi square statistic (χ^2) is effective for measuring the difference in multimodal signal analysis, we use it as the dissimilarity measurement (Kaur et al. 2010) for facial expressions. Based on the Chi square distance, we separate the expressions into three categories. Figure 4 shows that the sentiments expressed in different expressions can be roughly separated from each other.

User Interaction Tracking

In our system, we also track the user's online interactions by detecting keystrokes and mouse operations (Kabi et al. 2013). User interaction activities have already applied to many fields (Rybnik et al. 2008), where they are treated as biometrics to protect computer systems or to verify one's identity. In our application of emotion recognition, they are more functional for providing strong evidence that reflects the changes of emotion. User's typing pace and mouse-click activity can change during emotion changes, as evident in previous studies (Kolakowska 2013).

We track the total number of keystrokes and mouse operations, which include mouse click, mouse moving distance and wheel slide, as indicators of different emotions. As our statistics (Figure 5) show, when the participants are in a negative mood, they may stay on the page for a long period and directly express their feelings in words. So they may have more key strokes but only a few mouse clicks. On the other hand, when the participants are in a positive mood, they may be more active in browsing the content through mouse operations. In this case, they may also show their emotion by more facial expressions instead of interactions with the system in words. In some ways, the interactions in a neutral mood are similar to the negative mood, when the amount of the interactions is less.

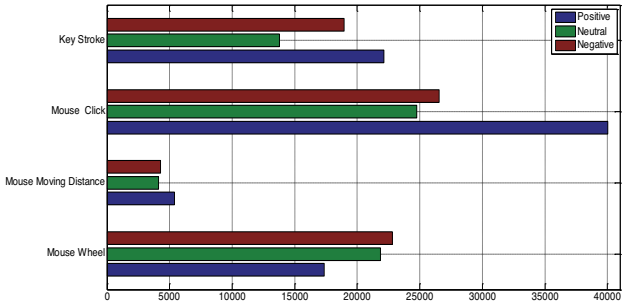

Figure 5. Interaction distribution of online users.

Content Tracking

As users engage themselves in online social media, the content will affect users' mental states to some extent. The content could even become a stimulus to their specific mental state which can be reflected by their replies to the tweets. Users' replies can be treated as a direct output signal of their mental state. Therefore, we record the contents of tweets and their responses at the same time and employ Sentiment 140[3] a Natural Language Processing (NLP) tool to discover the sentiment of tweets and obtain the sentiment polarity of the tweets. To exploit the images in tweets, we employ an image sentiment prediction algorithm based on middle-level attributes (Yuan et al. 2013;

[3]http://help.sentiment140.com/home

You et al. 2015). These three types of content-based signals are then incorporated to predict the mental states.

Pattern Classification and Mining

Data Preprocessing

Before combining all the information extracted from different signals and making a prediction of a users' mental health state, it is necessary to pre-process the signals since they are generated from multiple and heterogeneous sources which may introduce noisy, missing and inconsistent data. We take the following steps to preprocess the raw data:

Data cleaning. Signals, such as the pupil radius and heart rate, can be affected by user motion, illumination change and other factors. For example, when user look down on the keyboard or turn their heads around, the camera cannot detect their eyes. Furthermore, a user's subtle movement can also introduce noise. Some users may involuntarily shake their legs in a relaxed environment, causing the variation of illumination on the skin and adding a noise frequency into our heart rate analysis. All of these would result in noise and missing values in our signals. Therefore, we apply different methods (e.g. heart rate analysis using Nearest Neighbor Interpolation, and pupil radius analysis using Linear Interpolation) to smooth out the noise.

Data transformation. In order to consider temporal correlation and trends, we apply a moving average strategy. We use 1-min sliding window containing three instances of 20 seconds each. The sliding step is also 20s.

Time series normalization. We normalize the values of different signals from different subjects into a common range of [0, 1] to preserve the relationships among the original data values and enhance the contrast of features.

Feature Analysis

We use a set of acronyms to represent our 12 feature signals and prediction labels. Table 1 summarizes all the labeling conventions.

Table 1. A summary of the notations

Notations	Meaning
PR	Pupil radius
HM	Head movement rate
EB	Eye blinking rate
FE	Facial expression
HR	Heart rate
TS	Textual tweets sentiment
RTS	Textual tweets reply sentiment
IS	Image sentiment
MD	Mouse moving distance
KR	Keystroke rate
MR	Mouse click rate
WM	Mouse wheel slide
Label '1'	Positive prediction result
Label '0'	Neutral prediction result
Label '-1'	Negative prediction result

Table 2. Feature ranking

Label '-1'	Weight	Label '0'	Weight	Label '1'	Weight
HM	8.93	MD	2.68	KR	3.16
IS	6.5	HM	1.56	EB	0.43
EB	3.26	WM	0.96	HR	0.22
RTS	3.20	FE	0.70	FE	0.08
PR	3.07	HR	0.61	RTS	-0.18
TS	0.88	MR	-0.15	TS	-0.75
KR	0.39	RTS	-0.65	MR	-0.87
MD	0.29	PR	-0.68	WM	-0.94
FE	-0.29	KR	-0.71	PR	-2.61
MR	-1.29	TS	-0.82	IS	-2.69
HR	-1.43	EB	-0.95	MD	-5.87
WM	-5.75	IS	-2.25	HM	-16.2

In Table 2, we show the ranking of all the feature signals based on their contribution weights in our prediction model for each mental state, which ranges from the most positive to the most negative. We notice that head movement always plays an important role, either for positive contribution or negative contribution to our prediction results. However, this feature may not be effective when the user does not move much, e.g., reading tweets in bed. Furthermore, we look into our data and find how each feature signal affects the prediction. For example, it is shown that eye blink gives a big contribution to both 'Label1' and 'Label-1', as reading either positive or negative content will lead to eye blink reactions. Beyond these, each feature has its limitations. Expression only has a positive contribution to 'Label1' and 'Label0'. Although our facial expression module can recognize five different expressions, this feature fails for 'Label-1' because people tend to have less negative expression while reading social content alone in a relaxed setting. In addition to the feature ranking, an ablation analysis may provide further insight on how different components affect the integrated system.

Pattern Prediction Based on Time Series

As mentioned earlier, all the features extracted from the preprocessing stage are weighted and fed into a multi-class classifier (logistic regression), which learns different characteristics of the three different emotion states (negative, neutral and positive). Besides logistic regression, we have also tried an SVM classifier. Logistic regression yields a better performance, probably due to the small size of our data set as SVM may require much more data to train. The emotion inference is done in real-time based on a likelihood indication classification rule. With this rule, the classifier predicts the maximum possibility of the three states. For example, if the probability for the positive state is larger than the other two, the current emotion is positive.

Figure 2 shows the example prediction results for one participant's mental states based on the model trained by data of other participants. As mentioned earlier, we generate an instance every 20s and employ a 1-min sliding win-

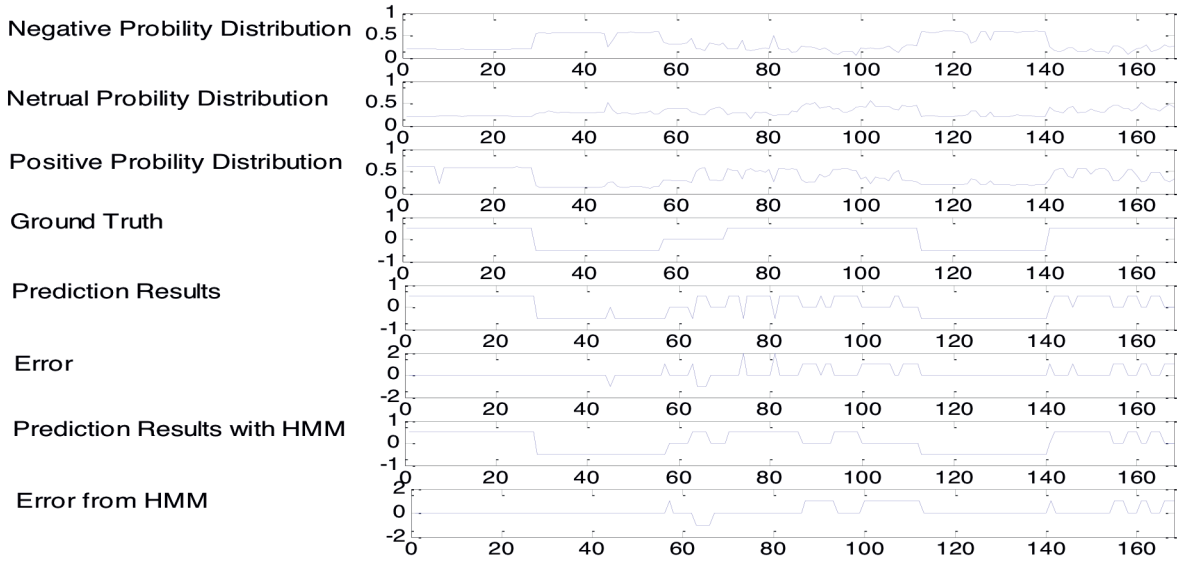

Figure 2. Example prediction results for one participant's mental states.

dow to calculate the moving average. Therefore, for each person, we have six 10-min long videos which are then split into 168 instances. The error is the difference between the prediction and ground truth. In Figure 9, from the ground truth we know this participant has high fluctuations in mood - there are only a few neutral instances and most of the time the participant is in an emotionally excited state. The trained classifier accurately predicts the mental states in the positive and negative stages. The error mainly emerges in the neutral stage and when the polarity of emotion changes rapidly. One's emotion could be easily affected by the context read before, so it is possible that between instances 60 to 80 some negative emotions are predicted. However, the same situation occurs between instances 90 to 110. The prediction results can be improved with a temporal model such as HMM (Eddy 1996), which also accounts for the different likelihoods of different emotion transitions, i.e. user's previous mental state can influence user's current mental state. The result after HMM is also included.

Experiments and Evaluation

Experiment 1: Normal Users

In order to ensure that our test and analysis are meaningful, we collect data in realistic, life-like environments. We conducted a series of mock-experiments in a living room equipped with sofas, desks, snacks, PCs, and our monitoring system.

All the participants were recruited through the University mail-list and social networks. In particular, they have social connections with each other so that they can form a sub-network in Twitter. This enables us to account for mutual influence through social connections in Twitter, although this is beyond the scope of this paper (Tang et al. 2009). For example, you may feel sorry when one of your friends is upset and you may become excited when you find your friends discussing something interesting.

To ensure our experiment can meet the diversity distribution requirements of Twitter users, we enrolled 27 participants (16 females and 11 males) in our experiments. They include undergraduate students, PhD students, and faculties, with different backgrounds in terms of education, income, and disciplines. The age of the participants ranges from 19 to 33, consistent with the age of the primary users of social networks.

Before the participants took the experiments, they were informed about the nature of the user study and incentives they would receive for their time (Amazon gift cards). They were also aware that we would record their activities during the experiments, but their personal background information and the experiment records would not be disclosed to the public and used in other ways.

During the experiments, we recorded the participants' activities with a webcam while they were reading the tweets. All the textural content and images we provided had been selected from real tweets. In this way, we make the participants feel more realistic as if they are browsing friends' tweets rather than being involved in a test. More specifically, we constructed 6 tweet sets in a predefined order of *positive-negative-positive-neutral-negative-neutral*, with every part lasting about 10 minutes. However, the participants were not aware of this design. We asked

for self-reporting of their feelings (every two minutes to be not too disruptive) as the ground truth. In addition, we recorded the number of mouse clicks and key strokes, respectively. Furthermore, every participant was required to reply to all the tweets and images in the tests and the information is considered reflective of their true mental states.

Our experiment is based on 162 videos from 27 people of both genders under different illumination conditions while reading various social media contents. This ensures a reasonably good anthropometric variation and also gives us the opportunity to look into the strengths and limitations of our system.

To evaluate our entire system, we use a leave-one-subject-out procedure where data from the testing participant is not used in the training phase. In Table 3, we compare 4 measurements: precision, recall, F-1 measure and area under the receiver operating characteristic (ROC) curve (AUC) for different classes. F-1 measure is the weighted harmonic mean of the precision and recall. Generally, the higher F1 is, the better the overall performance is. Similarly, the overall performance improves when AUC increases. All the 4 measurements fall into similar distributions, where negative and positive classes have higher accuracy than the neutral class. It seems that the signals extracted from negative and positive states are more distinctive.

Table 3. Leave-One-Subject-Out Test for Experiment 1.

	TP	FP	Prec.	Rec.	F-1	AUC
Negative	0.89	0.08	0.82	0.89	0.84	0.95
Neutral	0.56	0.13	0.67	0.56	0.59	0.79
Positive	0.78	0.17	0.76	0.78	0.75	0.91

Experiment 2: Depression Patients

If we constrain our system on the negative mood, our system can potentially make a prediction of depression levels or suicide possibility. Surveys (Goldsmith 2002) show there is an increasing number of suicides in the United States. Our system can offer another powerful instrument for psychologists to comprehend people's psychological states on a significantly larger scale.

Table 4. Leave-One-Subject-Out Test for Experiment 2.

	Patients vs. Control in *positive* mood	Patients vs. Control in *negative* mood	Patients vs. Control in neutral mood
precision	0.814	0.817	0.813
recall	0.674	0.738	0.717

We conducted a preliminary study using a group of 5 depression patients (2 severe/suicidal and 3 moderate) and a control group of five normal users. Given the small sample size, leave-one-subject-out test is performed. The ground truth was provided by a psychiatrist based on the video and audio content of Apple's *FaceTime*^{TM} chats with the patients. The results in Table 4 are quite promising, including differentiating between two levels of depression.

Discussions

We note that significant effort was involved in making the proposed system work, starting from building individual components to the system integration. The main challenges in the integration include:

1) acquiring adequate videos and social media streams from the recruited subjects to facilitate the investigation;
2) making individual weak signals strong and meaningful enough so each can contribute to, rather than hurt, the overall system performance; and
3) choosing the proper machine learning techniques and temporal models to handle such diverse, noisy multimodal data in order to obtain the promising accuracy we demonstrated.

With the help of our system, people can become more aware of their emotion fluctuations and pay more attention to their mental health. From this point of view, our system can assist users to monitor their emotions and allow users to understand the influence of social networks on themselves.

We envision making an Android app. Both the front cameras and processors on most of today's android devices can satisfy our computation needs. The high utilization rate of these mobile devices will provide the convenience for users to acquaint with their own emotion fluctuations and make adjustment themselves. With a big data set from more users, more interesting patterns can be mined and deep understanding about how different signals relate to different groups can be achieved. Although ethical concerns are real, we note that mobile applications have already been developed to support mental health interventions (Matthews et al. 2008).

Conclusion and Future Work

In this paper, a highly integrated multimodal signal system for mental health prediction is developed. Using non-contact multimodal signals, user interaction and content analysis, our system is able to infer the user's current mental state. In the future, we will also explicitly integrate the influence of social networks as shown by Tang et al. (2009). We are encouraged by this study to evaluate the effectiveness of the integrated system for mental health monitoring and intervention using a larger population of patients in an existing mental health program that we will have access to. We will further develop the system to recognize a more fine-grained depression scale using facial

expression dynamics, as shown in the recent work by Jan et al. (2014). This new multimodal approach to monitoring and managing mental health is an example of integrating computer vision, data mining, social media, and behavior science for social good. It has the potential to revolutionize mental health care in cost and effectiveness.

References

Crown, R. W. 2011. Anatomy of an Epidemic: Magic Bullets, Psychiatric Drugs, and the Astonishing Rise of Mental Illness in America. Broadway Books.

Centers for Disease Control and Prevention (CDC). Behavioral Risk Factor Surveillance System Survey Data. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2008, 2011, 2012.

Tao, J.; Tan, T. 2005. Affective computing: A review. In Affective computing and intelligent interaction (pp. 981-995). Springer Berlin Heidelberg.

Zeng, Z; Pantic, M; Roisman, G.I.; Huang, T.S. 2009. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE PAMI.

Kalal, Z.; Mikolajczyk, K.; Matas, J. 2010. Face-TLD: Tracking-learning-detection applied to faces. IEEE International Conference on Image Processing (ICIP).

La Cascia, M; Sclaroff, S.; Athitsos, V. 2000. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models. IEEE PAMI, 22(4): 322-336.

Wu, H. Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. Eulerian video magnification for revealing subtle changes in the world.ACM Transactions on Graphics (TOG), 31(4), 65, 2012.

Balakrishnan, G.; Durand, F.; Guttag, J. 2013. Detecting Pulse from Head Motions in Video. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Castrillón, M.; Déniz, O.; Guerra, C.; Hernández, M. 2007. ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. Journal of Visual Communication and Image Representation, 18(2).

Yuille, A.L.; Hallinan, P.W.; Cohen, D.S. 1992. Feature extraction from faces using deformable templates. International Journal of Computer Vision, Volume 8, Number 2, Page 99.

Agarwal, M.; Jain, N.; Kumar, M. M. 2010. Face Recognition Using Eigen Faces and Artificial Neural Network. International Journal of Computer Theory and Engineering, 2(4): 1793-8201.

Kaur, M.; Vashisht, R.; Neeru, N. 2010. Recognition of facial expressions with principal component analysis and singular value decomposition. International Journal of Computer Applications, 9(12): 36-40.

Kabi, B.; Samantaray, A.; Patnaik, P.; Routray, A. 2013. Voice Cues, Keyboard Entry and Mouse Click for Detection of Affective and Cognitive States: A Case for Use in Technology-Based Pedagogy. IEEE International Conference on Technology for Education (T4E).

Rybnik, M.; Tabedzki, M.; Saeed, K. 2008. A keystroke dynamics based system for user identification. IEEE Computer Information Systems and Industrial Management Applications.

Kolakowska, A. 2013. A review of emotion recognition methods based on keystroke dynamics and mouse movements. The 6th IEEE International Conference Human System Interaction (HSI).

Yuan, J.; You, Q.; McDonough, S.; Luo, J. 2013. Sentribute: Image Sentiment Analysis from a Mid-level Perspective. ACM SIGKDD Workshop on Issues of Sentiment Discovery and Opinion Mining.

You, Q.; Luo, J.; Jin, H.; Yang, J. 2015. Robust Image Sentiment Analysis using Progressively Trained and Domain Transferred Deep Networks. The Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI).

Eddy, Sean R. 1996. Hidden Markov Models. Current opinion in structural biology. 6.3: 361-365.

Mignault, A.; Chaudhuri, A. 2003. The many faces of a neutral face: Head tilt and perception of dominance and emotion. Journal of Nonverbal Behavior, 27(2), 111-132

Goldshmidt, O. T.; Weller, L. 2000. Talking emotions: Gender differences in a variety of conversational contexts. Symbolic Interaction, 23(2), 117-134.

Waters, A. M., Lipp, O. V., & Spence, S. H. The effects of affective picture stimuli on blink modulation in adults and children. Biological psychology, 68(3), 257-281, 2005.

Tran T.; Phung, D.; Luo W. 2013. An integrated framework for suicide risk prediction. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining.

Tang J.; Sun, J.; Wang, C. 2009. Social influence analysis in large-scale networks. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.

Matthews, M.; Doherty, G.; Coyle, D.; Sharry, J. 2008. Designing Mobile Applications to Support Mental Health Interventions. in Handbook of Research on User Interface Design and Evaluation for Mobile Technology, Lumsden Jo (Ed.), IGI, Information Science Reference, pp.635 - 656.

Cohn, J.F.; Simon, T.; Matthews, I.; Yang, Y.; Nguyen, M.H.; Tejera, M.; Zhou, F.; De la Torre, F. 2009. Detecting Depression from Facial Actions and Vocal Prosody, Affective Computing and Intelligent Interaction (ACII).

Goldsmith, S.K. 2002. Committee on Pathophysiology & Prevention of Adolescent & Adult Suicide. Reducing suicide. Washington, DC: The National Academies Press.

Jan, A.; Meng, H.; Gaus, Y.F.A.; Zhang, F.; Turabzadeh, S. 2014. Automatic depression scale prediction using facial expression dynamic and regression. The 4th International Audio/Visual Emotion Challenge and Workshop.

Lane, N.D.; Choudhury, T.; Campbell, A.; Mohammod, M.; Lin, M.; Yang, X.; Doryab, A.; Lu, H.; Ali, S.; Berke, E. 2011. BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing. Pervasive Health 2011-- 5th International ICST Conference on Pervasive Computing Technologies for Healthcare.

McDuff, D.; Gontarek, S.; Picard, R. W. 2014. Remote Measurement of Cognitive Stress via Heart Rate Variability. In the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).