Can Your Smartphone Infer Your Mood?

Robert LiKamWa[*], Yunxin Liu[+], Nicholas D. Lane[+], Lin Zhong[*]
Rice University[*], Microsoft Research Asia[+]

ABSTRACT
Our driving vision is a smartphone service, called *MoodSense*, that can infer its owner's mood based on information already available in today's smartphones. The service will fundamentally enhance context-awareness by providing clues about mobile users' mental states. We report early results from studying 25 iPhone users in the field and the correlation between their mood and phone usage. We show that user mood can be inferred into four major types with an average accuracy of 91%. This is achieved using only three weeks of training data and simple smartphone usage statistics. The results, though preliminary, strongly suggest the feasibility of mood inference without using the power-hungry and socially invasive microphone and camera.

Categories and Subject Descriptors
C.4 [**Performance of systems**]: Measurement techniques

General Terms
Human Factors

Keywords
Smartphone usage, mood

1. Introduction
Mood is a mental state that is induced by complicated causes. Compared to emotion, mood is a less intense state, but lasts much longer, e.g., days or hours instead of minutes [1]. Mood affects how we behave and make decisions and, more importantly, is an important social signal that others leverage to better interact with us. If our phone knows our mood, it will not only allow more personalized, situated services but also facilitate our social interaction, as suggested by affective computing research [10, 16]. Compared to the extensive literature in inferring the *physical* and *social* context of a mobile user, there has been little success in inferring the user's mental state. Existing work often focuses on emotions, a transient mental state that is intrinsically difficult to infer, and relies on invasive means such as microphone and camera, e.g., [9, 11, 17, 19], and even body-worn sensors, e.g., [7]. The resulting solutions incur considerable cost in computation, energy, and usability.

Our project, therefore, seeks to answer the following question: *can a smartphone infer its user's mood with the information it already has?* Our motivations are twofold. First, although we often fake our emotions, e.g., to be polite, we rarely can fake our mood because mood is a long-lasting mental state. This makes mood a more tractable target for machine learning and more useful for enriching context-awareness. Second, our smartphones already have rich user information: where we have been (location), what we have done (calendar), whom we communicate with and how (email, SMS, and phone call), what applications we use, which websites we visit, and more. While a smartphone initially lacks background knowledge of a user's behavior, it may achieve mood inference by learning about its user over time with its analytical power.

Toward answering the question, we carry out two user studies. The first study employs focus group discussions with smartphone users to learn about how mood affects their interaction with their devices and how they would like their mood to be exploited by their devices. We are currently in the middle of the second user study that monitors device usage and mood of 25 iPhone users in the field. The study employs a continuously running logger that captures almost every bit of the iPhone usage and a convenient iPhone application for manual mood input. Our hypothesis is that *by correlating the logged usage to the mood input by the user, we will be able to evaluate how accurate mood can be inferred and which predictors are most indicative of mood.*

In this paper, we report the following findings. (*i*) Smartphone usage does indicate user mood. Users use different applications and communicate with different people depending on their mood. In particular, most participants in our study indicated that they are more likely to communicate with others when they are happy. Using only six pieces of usage information, (SMS, email, phone call, application usage, web browsing, and location) from all participants, a simple clustering classifier can infer a participant's mood out of four categories with 61% accuracy on average. (*ii*) How usage indicates the mood is highly personal. The accuracy of mood classification is improved to 91% on average when inference is based on the same participant's data. This strongly suggests that mood inference should be personalized, using the same user's data to train the classifier.

Our experience so far suggests that usage can be used to predict mood. As our field study collects more weeks of data, we will extend our work to include more sophisticated predictors, infer more fine-grained moods, investigate more powerful machine learning algorithms, and study techniques for reducing the required mood input by users as the ground truth for machine learning. This will be discussed in Section 4. Results from the project will enable us to build *MoodSense*, a system service that will gather information already available in smartphones, occasionally request user mood input, and accurately infer the user mood over time.

2. Studying Mood in the Field
We have performed two user studies toward mood inference with 25 iPhone users. We recruited participants from the general public in Beijing, China by posting recruiting advertisements on online forums. All are existing iPhone

users, aged between 20 and 29. 8 of them are female and 17 are male. 17 of the users are students but the rest covers a diverse set of occupations including two software engineers, one web editor, one salesman and one teacher. To quantify mood, we adopt the Circumplex Model [13], a widely used and validated [3, 7, 16, 18] model that measures mood with two orthogonal dimensions: *pleasure* (happy-sad) and *activeness* (aroused-sleepy). For instance, mood of high pleasure and high activeness is *excited* while that of low pleasure and high activeness is *stressed*.

2.1 Focus Group

Before the field study, we conducted a two-part focus group study with the participants to gain motivation and intuition of automatic mood inference.

The first part of the focus group dealt with the impact of mood changes on phone usage habits. We asked whether the participants use different applications or communicate with different people depending on their mood. All but one of the participants answered "yes". They also reported other ways their phone usage changes depending on their mood, including changing ringtones and wallpapers.

The second part of the focus group asked for the participants' opinions on mood sharing. We asked how they would publish their mood, with whom they would share their mood, when they would like others to see their mood, whose moods they would be interested to see, how they would like their phone to automatically adapt to their mood, and how sharing mood would affect their life. All but one of the participants responded that they would like to publish their mood through microblogs, instant messenger or other applications. For the participants who would publish mood, they all wished to publish their "current mood". Some others also indicated they would like to publish their "mood over the past week." All of the participants indicated that they would like to share their mood within certain social circles (e.g., friends or co-workers) but they would not want people to see their current mood in some specific cases, such as if they were in extremely bad moods and did not want to talk about it. All participants were interested in seeing others' mood, particularly the mood of their friends. One participant even responded that he would like to know everyone's mood, including the mood of strangers. All but one of the participants indicated that they would like their smartphone to automatically change with their mood, e.g., changing the color scheme, reordering the applications and notifying them on their mood changes. For the open question of how sharing mood affects their life, most participants mentioned that sharing mood would have very positive impacts on their life, e.g., sharing happiness with friends or getting rid of bad moods with help from friends. The results of this focus group are highly motivational and encourage us to conduct the work in this paper.

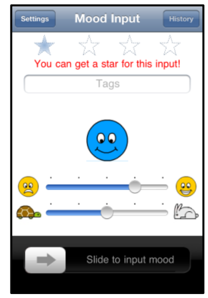

Figure 1: Mood Input Application

2.2 Field Study

We employ a longitudinal field study to collect real-world data in order to study the correlation between mood and smartphone interactions. The study involves two pieces of custom developed software: a mood input application and a background logger. Each participant went through a training session to ensure they knew how to use the mood input application to provide accurate mood data, how the logger works and what data we collect. Then we asked them to use the mood input application and run the logger for at least two months. The study is currently in progress; the users have completed one month of the study.

2.2.1 Mood Input Application

Figure 1 shows the primary GUI of the Mood Input Application. The application allows users to report their mood conveniently. Two slider bars allow the user to set their mood along the Circumplex pleasure and activeness dimensions previously described. For the sake of simplicity, each slider limits the users to one of five options, and to prevent confusion, we intentionally omit showing numbers. As the user inputs a mood, the mood is represented as a bouncing cartoon face in the center of the screen. As the pleasure slider is changed, the facial expression and color of the face changes. As the activeness slider is changed, the speed of the bouncing movement changes. If desired, users can also enter in text tags expressing why they feel a certain way. We allow users to see their previous inputs through a calendar mode and also through a chart mode. To minimize the influence of previous inputs on users' mood, previous inputs are shown in a different GUI and not provided to users when they input their mood.

Users are asked to input their mood at least four times a day, with at least three hours between each input. This is motivated with a star achievement system, which gives a user a star for an input if their last star was given at least three hours prior. The user can gather up to four stars per day and is notified when a star is available through smartphone alerts. This star system enables us to regularly capture a user's self-reported mood.

2.2.2 Smartphone Interaction Logger

We transparently collect a participant's smartphone interaction to link with the collected moods. We leverage the

LiveLab iPhone Logger [15] to capture user behavior using daemons operating in the background. The data is archived nightly to a server over a cell data or Wi-Fi connection.

Using this tool, we gather relevant information to form our mood models. Application usage, phone calls, email messages, SMSes, web browsing histories, calendar entries and location changes are collected as user behavior features. We ensure that user privacy is properly protected, hashing all private user data (e.g., contact identifiers used in phone calls and SMSes).

2.2.3 *Participation Incentives*

To provide incentives for our participants to actively use our mood input applications and provide accurate data, we offer RMB500 compensation for each participant and an iPad 2 raffle after completion of the study. The raffle is based on the star achievement system described above: each star a participant earns is an entry into the raffle.

3. Mood Inference

Our participants actively reported their mood using the Mood Input Application. On average each participant received 3.6 stars per day. Using all the data collected from the field study, we next investigate the correlation between mood and smartphone usage. MoodSense can be essentially built on a model that predicts mood using smartphone usage statistics. Our analysis of data collected so far suggests that such a model is indeed feasible.

3.1 Preliminary Model Design

A variety of everyday behaviors have been established as indicators of mood. In particular, patterns of how people socialize [6] and the daily activities they engage in [4] are strongly correlated with the mood of an individual. In our proof-of-concept mood model, we focus on these two categories of mood-sensitive behavior, social interaction and daily activities. Phone-collected data allows us to indirectly observe these patterns in a number of different ways.

We leverage six types of phone-collected data. Specifically, we use patterns in application usage, phone calls, SMSes, emails, web browsing history, and location. Collectively, phone calls, SMS and email signal changes in social interaction. We treat these three types of data identically, and count the number of exchanges the user has with different contacts. Similarly, we use patterns in browser history, phone application usage and location history as a coarse indicator of daily activities. Application and browser activities are monitored based on the frequency of application use and unique URLs visits. A time-series of location estimates for the user is clustered using the DBSCAN [5] clustering algorithm, which allows us to then count user visits to each location cluster.

Using this collection of measurements, we compute six separate histograms, one for each type of phone data. For features, our model uses the normalized frequency count of the ten most frequently occurring bins in each histogram. This produces a vector of 60 features, which we use to describe user behavior and correlate self-reported mood from our field study. We compute features using a non-overlapping three-day window of data; we find that smaller window sizes capture meaningless variations in user behavior (e.g., phone call patterns). Ground-truth mood is determined by averaging self-reported mood survey data within each window.

We discriminate between four discrete mood states defined by the two dimensions of the Circumplex model, the pleasure and activeness indices. Each dimension is split into high and low states based on the median user reported scores during the trial. As a result, user mood is represented by the following states: `PA` – above median pleasure and above median activeness; `pA` – below median pleasure and above median activeness; `Pa` – above median pleasure and below median activeness; and finally, `pa` – below median pleasure and below median activeness.

3.2 Data Analysis

We first perform an experiment where we use a general one-size-fits-all model of mood and apply it to every user in the field trial. We find overall accuracy of our model is 61% per user for four-fold cross validation (with each fold corresponding to one week of data). As seen in Table 1, where we present the confusion matrix, we see that even the extremes of our discrete mood states (`PA` and `pa`) are being confused. To study the influence of each different type of data we repeat this experiment and train a new model, except each time only using a single type of phone data. The accuracy of each of these models is reported in Table 2. Most interestingly we discover that some of the different types of features in isolation do not perform much worse than the whole model; combining features does not help the overall model significantly. Further, we find that the accuracies of these feature-specific models vary significantly from user to user, suggesting the mood of different users is more easily recognized based on certain features.

We examine the need to *personalize* our mood model by training and testing separate models based solely on the data collected by each user. We observe average per-user accuracy (applying four-fold cross validation) increases to 91%. Variance in accuracy between users is narrow, ranging between 77% and 98%. We repeat our experiment to identify which features have the strongest discriminative power and find our results vary from user to user. We illustrate this variation for two representative users in Figure 2. These figures show a single feature for each type of data under different mood states. We can see that these two individuals have noticeably different relationships between their behavior and mood – explaining why personalize mood models are more effective. However, this improved performance comes at a cost, as users must provide their own training data to personalize their models.

Figure 2: Normalized Frequency of Most Frequent Feature in each category for Two Users

T able 1: Confusion Matrix for Mood Model (Actual on left, Classified on top)

	pa	pA	Pa	PA
pa	557	126	180	95
pA	99	375	47	97
Pa	111	52	445	33
PA	121	140	181	503

Table 2: Discriminative Power of Feature Categories

Feature Category	Discriminative Power
Location	0.60
Web	0.54
SMS	0.51
Email	0.47
Call	0.39
Apps	0.37

4. Ongoing Work

Our results are promising; we find personalized mood models are 91% accurate with as little as three weeks of training data. However, our results remain preliminary until we can conduct larger field experiments with a more diverse set of participants. As our study lasted only one month, it is unknown if relationships between user behavior and mood are robust to changes in user lifestyle or external factors, such as weather and seasonal effects. As our study population is too homogenous in age, geography, ethnicity and culture, our results may not be directly generalized for the general population. Nevertheless, prior studies [14, 18] indicate observations in Chinese-based mood experiments do generalize to other countries. We need additional longitudinal field trials with diverse populations to refine our preliminary model and verify its effectiveness. We plan to stage additional studies not only in China but also in the United States.

4.1 Evaluating Mood Inference

In this work, we evaluated mood inference with four coarse categories. This evaluation can be improved in two ways. First, a significant portion of reported moods in our field study were close to neutral along either the pleasure or the activeness dimension. That is, many reported moods were simply neutral or close to neutral. For applications' perspective, it is more interesting to know a user's mood when it is far from neutral. This implies that mood inference should be evaluated as a detection task, instead of as a classification task. That is, it should be evaluated based on how many non-neutral moods it correctly detects and misses, respectively. Furthermore, since we measure mood with the Circumplex model, mood inference can be evaluated as an estimation task that estimates pleasure and activeness as two separate parameters. Because pleasure and activeness estimations can be directly translated into mood detection or classification, estimation of these parameters is fundamental to mood evaluation.

4.2 Improving Inference

We are building MoodSense as a system service based on the mood model to supply the user mood to interested applications. Our current focus is to improve the mood model in the following directions. First, we will consider a richer array of features associated with user social interaction and daily activity. Currently our features are based on the frequency of simple clearly recognized user actions, such as, how often a user calls particular contacts. This is an overly simplistic view of social interaction and daily activities; ignoring factors such as when and where interactions and activities take place, e.g., weekend/weekday or workplace/home.

Moreover, we plan to investigate a broader variety of relationships. For example, sentiment analysis applied to the content of SMS and email messages will likely prove effective indicators of mood. This work also neglects time-series patterns in user actions as we only model mood based on features extracted from a single window of time. We anticipate introducing temporal models to capture mood indicators that are visible at different time-scales. By incorporating new information in addition to extracting richer features we may be able to recognize finer classes of mood states than we use in this work. We found that our field trial lacked sufficient examples of various user mood types, constraining our model to only four coarse mood states.

Third, we are interested in determining if personalized mood models are necessary. Although our early results show that personalized mood models achieve higher accuracy, they require each user to supply training data, which may have negative usability implications. Based on previous studies of smartphone usage classes and community similarity [8, 12], we conjecture that a much smaller number of models might be sufficient, based on a few categories of user mood behavior, rather than a unique model being necessary for each user. Therefore, MoodSense can start from a generic model for the category the user belongs to and then personalize it as user-specific data is gathered.

Finally, we are interested in applying on-line learning techniques that improve inference based on user feedbacks. For example, by allowing the user to fix an incorrectly inferred mood, MoodSense can incrementally adjust its model.

4.3 Building Mood-based Applications

We are also interested in building smartphone applications using mood inference from *MoodSense*. When our smartphones know our mood, many appealing applications will be possible. For example, a smartphone can automatically share its user's mood through Facebook and Twitter to keep friends updated, which was shown to be desirable by a user study reported in [3]. Furthermore, because mood is an important social signal, knowing mood can transform smartphones into a social instrument that helps a user connect to unknown people. For example, a smartphone can make its owner's mood available to people in the physical vicinity through wireless broadcast to facilitate social interaction and friend-making. Yet another type of application could adapt device and environment according to user mood. A smartphone could automatically select suitable songs depending on a user's mood. Similarly, when a user goes back home, lights and background music could automatically adapt to his or her mood. User mood inference can drive these and many more potential applications.

5. Conclusions

Gradually, mobile phone sensing is reaching beyond the recognition of physically observable human behaviors or environmental context. Researchers are beginning to develop techniques that allow devices to infer the internal mental state of users [2, 16, 17, 19], opening exciting new opportunities for mobile applications [7]. Our work contributes to this new direction as the first demonstration of automatic mood inference using routinely collected phone data (e.g., browser, call, SMS, location history etc.).

In this paper we presented our initial findings based on the analysis of data collected from a longitudinal field study involving 25 iPhone users. In addition to refining and verifying our mood model, we plan to build MoodSense, a mood inference system service, on the mood model, and study potential mood-enhanced mobile applications. Initially, we will develop a number of mobile applications that benefit from user mood awareness. This will be followed by the deployment and evaluation of these applications within a user study designed to evaluate not only their benefit; but gain deeper insights into applications that utilize this new form of additional sensor-based user inference.

6. REFERENCES

[1] C. D. Batson, L. L. Shaw and K.C. Oleson, "Differentiating affect, mood and emotion: toward functionally-based conceptual distinctions", *Emotion*, Newbury Park, 1992.

[2] G. Chittaranjan, J. Blom and D. Gatica-Perez, "Who's who with Big-Five: Analyzing and Classifying Personality Traits with Smartphones", in *ISWC 2011*.

[3] K. Church, E. Hoggan and N. Oliver, "A study of mobile mood awareness and communication through MobiMood", in *NordiCHI*, 2010.

[4] L. A. Clark and D. Watson, "Mood and the mundane: Relations between daily life events and self-reported mood", *Journal of Personality and Social Psychology*, vol 54(2), 1988.

[5] M. Ester, H.-P. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", in *Second International Conference on Knowledge Discovery and Data Mining*. pp. 226–231, 1996.

[6] J. P. Forgas, G. H. Bower and S.E. Krantz, "The influence of mood on perceptions of social interactions", *Journal of Experimental Social Psychology*, vol 20(6), 1984.

[7] A. Gluhak, M. Presser, L. Zhu, S. Esfandiyari and S. Kupschick, "Towards mood based mobile services and applications", in *EuroSSC*, 2007.

[8] N. D. Lane, Y. Xu, H. Lu, S. Hu, T. Choudhury, A.T. Campbell, and F. Zhao, "Enabling large-scale human activity inference on smartphones using community similarity networks (csn)", in *Ubicomp*, 2011.

[9] A. Matic, A. Papliatseyeu, S. Gabrielli, V. Osmani and O. Mayora-Ibarra, "Happy or moody? Why so? Monitoring daily routines at work and inferring their influence on mood", in *UbiHealth*, 2010.

[10] R. W. Picard, "Affective computing", *MIT Press*, 1997.

[11] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth and A. Aucinas, "EmotionSense: a mobile phones based adaptive platform for experimental social psychology research", in *UbiComp*, 2010.

[12] A. Rahmati, C. Shepard, C. Tossell, M. Dong, Z. Wang, L. Zhong, P. Kortum, "Tales of 34 iPhone Users: How they change and why they are different", Tech. Rep., 2011, Arxiv preprint arXiv:1106.5100.

[13] J. A. Russell, "A circumplex model of affect", Journal of Personality and Social Psychology, vol. 39, no. 6, 1980.

[14] J. A. Russell and M. Yik, "Emotion among the Chinese", in The handbook of Chinese psychology, pp. 166-188. *Hong Kong: Oxford University Press*, 1996.

[15] C. Shepard, A. Rahmati, C. Tossell, L. Zhong and P. Lortum, "LiveLab: measuring wireless networks and smartphone users in the field", in *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 38, no. 3, 2010.

[16] J. Tao and T. Tan, "Affective computing: a review", Affective Computing and Intelligent Interaction, LNCS 3784, Springer, 2005.

[17] A. Vinciarelli, M. Pantic and H. Bourlard, "Social signal processing: survey of an emerging domain", *Image and Vision Computing*, 27, 2009.

[18] M. Yik, "Studying Affect Among the Chinese: The Circular Way", *Journal of Personality Assessment*, 91(5), 2009.

[19] Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, 2009.