Cluster-based Approach to Improve Affect Recognition from Passively Sensed Data

Mawulolo K. Ameko[1], Lihua Cai[1], Mehdi Boukhechba[1], Alexander Daros[2],
Philip I. Chow[3], Bethany A. Teachman[2], Matthew S. Gerber[1], Laura E. Barnes[1]

Abstract—Negative affect is a proxy for mental health in adults. By being able to predict participants' negative affect states unobtrusively, researchers and clinicians will be better positioned to deliver targeted, just-in-time mental health interventions via mobile applications. This work attempts to personalize the passive recognition of negative affect states via group-based modeling of user behavior patterns captured from mobility, communication, and activity patterns. Results show that group models outperform generalized models in a dataset based on two weeks of users' daily lives.

I. INTRODUCTION

The extent to which individuals experience positive and negative affect on a daily basis is associated with mental health outcomes [1]. Higher levels of negative affect are associated with increased vulnerability to many mental disorders, including depression and anxiety disorders, two of the most common types of mental disorders in U.S. adults [2]. Mental health research typically relies on self-report questionnaires that assess negative affect at a moment in time. Repeated administration of these measures, such as in an ecological momentary assessment (EMA) framework, is resource intensive and susceptible to retrospective bias when participants are asked to recall their mood over a previous duration [3]. Ideally, negative affect would be recognized without asking participants, thereby reducing burden, improving compliance among participants, and allowing for continuous modeling of affect change. To aid recognition of negative affect, unobtrusive mobile sensing of location, texts and calls, and activity levels could also be used to enrich the information provided by participants' responses to questionnaires assessing negative affect and measures of mental health (e.g., social anxiety, depression).

Current affect recognition approaches are based primarily on generalized or individualized approaches [4]. In generalized approaches, the recognition model learns global patterns that the majority of participants followed during the experiment. These patterns are then used for prediction. Since user behaviors vary substantially, generalized models may fail to predict variations in affect for an individual person. In contrast, individualized models are designed to learn participants' patterns on a case-by-case basis, thus they are expected to be more accurate. However, individualized models require a

certain number of observations for each individual to obtain robust prediction performance. In short-term studies involving human subjects (e.g., two weeks), individual models may fail to adequately capture individual affective patterns because of a small pool of observations [5].

In our work, we propose a new group-based approach that integrates generalized and personalized models. We first propose a method for clustering multimodal behavioral profiles that groups participants based on their mental states, activity levels, communications, and mobility patterns. We then apply several prediction algorithms to investigate whether group models using multimodal user profiles outperform the generalized or population-based model.

II. RELATED WORK

Smartphone usage can be used as an indirect marker of mood. Passively sensed location information has been used to predict depressive symptoms [6]. Individuals with higher social anxiety levels were more likely to report negative affect during the day, which in turn was predictive of spending more time at home at subsequent measurements [7]. Self-reported stress and mental health indices were also successfully predicted in a 10-week long study design in college students with both passively and actively sensed data [8].

Prediction of affect from mobile sensing appears to be more difficult to replicate. In a feasibility study, LiKamWa et al. [5] explored a personalized feature selection approach to predict changes in mood from unobtrusively sensed indices of social activity (e.g., calls/texts, emails), physical activity (e.g., GPS), and general mobile phone use (e.g., application use, web browsing). The study relied on two months of data collected from 32 participants. Results indicated high levels of accuracy in predicting mood using personalized models. The personalized modeling also produced better accuracy compared to a generalized model using data from all users.

A follow-up study in which a personalized feature selection approach was used to predict affect ratings from 27 participants over 42 days found no clear benefits of using this approach [9]. However, these studies did differ in length, participant variability (e.g., depressive symptoms), and unobtrusive features assessed. It remains possible that personalized feature selection requires an intensive level of data collection that participants may perceive as burdensome. Given these findings, we use an intermediate approach between generalized and personalized models to recognize affect in a given situation.

[1]Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA 22904, USA

[2]Department of Psychology, University of Virginia, Charlottesville, VA 22904, USA

[3]Department of Psychiatry and Neurobehavioral Sciences, University of Virginia, Charlottesville, VA 22904, USA

III. STUDY DESIGN

Sixty-five undergraduate students were recruited for a two-week study period to understand dynamics of emotional, cognitive, and interpersonal processes associated with depression and social anxiety. University students provide a relatively homogeneous sample in terms of life phase and common psychological stressors, thereby mitigating the impact of a wide variety of potential nuance factors. Pre-study surveys were given to the students at enrollment, and one of these surveys measured students' social anxiety (SIAS) [10]. The study contained an ecological momentary assessment (EMA) phase that requests self-report data on psychological affect throughout the day. A customized mobile app (Sensus) [11] was installed on participants' personal Android smartphones and was programmed to deliver 6 EMAs throughout the day (each survey contained 12 questions), randomly scheduled in each 2-hour block from 9 a.m. to 9 p.m. (e.g., once between 9-11 a.m., once between 11 a.m.-1 p.m., etc.). Sensus was also configured to deliver an end-of-day survey at 10 p.m. each day. Prompts concerning affect first asked participants to rate how positive they were feeling from 1 (not at all) to 100 (very positive). The second question asked participants to rate how negative they were feeling from 1 (not at all) to 100 (very positive). In addition to these active assessments, Sensus also passively collected GPS coordinates every 150 seconds and accelerometer data at 1 Hz, in addition to call and text logs. All data were transmitted wirelessly to a secure Amazon Web Services server, where data were stored for further analysis (see Figure 1).
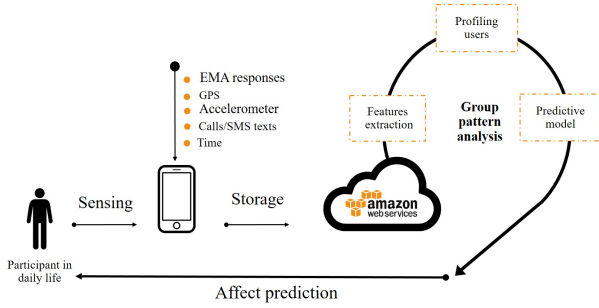

Fig. 1. Passive and affect data collection using smartphones.

IV. EXPERIMENTS

A. Data Preprocessing

We first processed participants' raw GPS data into semantic locations (e.g., leisure, education, and home) by combining a spatiotemporal clustering algorithm [12] and OpenStreetMap (OSM) geodatabase [13]. Our label taxonomy includes the following types: Education (e.g., university and libraries), Leisure (e.g., restaurants and cinemas), Out of town, In transition (e.g., going from one place to another), Home, and Other houses. Our algorithm has been trained to recognize Home as the place having a house OSM-tag (e.g., apartment, dormitory,

house, etc. See [13] for more details about OSM tags) where a subject stayed the most between 10 p.m. and 9 a.m.

For accelerometer data, we used statistical measures (mean, minimum, maximum, standard deviation, median and variance) on the 1-minute sliding window to extract several features of phones' motion around affect assessment moments. These features aim to represent the physical activity levels of the participants, and we used them to predict momentary negative affect. Note that our accelerometer features are extracted from the magnitude of acceleration $\sqrt{\frac{x^2+y^2+z^2}{3}}$ to make them orientation free, since the phones were used in participants' natural environments.

Individuals' affect may be associated with the degree to which they interact with others. Thus, we included communication events in our models. For each EMA we collected the number of text messages and phone calls as long as their duration overlapped with epochs prior to the EMA prompt. Here we chose 1 hour prior to the EMA prompt as the time window to record the number of text messages and phone calls.

B. Profiling Users

After preprocessing the data, we clustered the participants based on their behavioral profiles. There are different ways to cluster participants. For instance, a clustering strategy can be based on time spent at home to cluster people having depressive symptoms, drawing on the hypothesized correlation between home staying and affect fluctuation patterns. The following four passively sensed profile features were used to drive the clustering process.

1) Location: For location data, we considered five common point-of-interest classes consisting of {'out of town', 'education', 'friends' houses', 'home', 'leisure'}. Then we calculated the proportion of time spent in each of these locations over the study period for each participant.

2) Activity: From the accelerometer data, we chose thresholds of 0.2 and 0.3 between the minimum and maximum to define three levels of activity (e.g., {Low, Medium, High} in acceleration). We chose these cutoffs based on the observed distribution of the acceleration values. Then for each participant, we calculated the proportion of time being in these activity levels (e.g., proportion of time being in the high level).

3) Short-Message Service (SMS): From the SMS data, we aggregated the number of text messages sent and received within each 1-hour window during the study period. From this, we defined 5 text messaging levels based on text message frequencies (e.g., 'Very-Low','Low','Medium','High','VeryHigh') with intermediary cutoffs at 1, 10, 20 and 30 messages per hour based on their observed distribution.

4) Phone Calls: Similarly, we computed the proportion of calls occurring at each level of call activity defined as 'Low','Medium','High','VeryHigh' using thresholds of 1, 3 and 6 calls per 2-hour window. We used a 2-hour window to accommodate the lower hourly frequency of phone calls compared with text messages.

Formally, for the design matrix $X \in \mathbf{R}^{N \times d}$ with $X = \{\mathbf{x_i}\}_i^N$, the feature vector for each participant is $\mathbf{x_i} = \{\underbrace{x_{i1}, x_{i2}, \cdots, x_{ip_1}}_{M_1}, \underbrace{x_{i1}, x_{i2}, \cdots, x_{ip_2}}_{M_2}, \cdots, \underbrace{x_{i1}, x_{i2}, \cdots, x_{ip_n}}_{M_n}\}$.
Note that M_i ($i \in [1, n]$) represents the ith modality and p_i the number of levels in the ith modality.

With the above, we determine different clusters based on various combinations of these four passively sensed modalities in addition to SIAS using the G-means (Gaussian Means) [14] algorithm. The G-means algorithm is an extension of K-means where number of clusters is automatically determined by iteratively selecting k such that the data assigned to each cluster follows a Gaussian distribution.

C. Predictive Models

We used 4 algorithms to test the predictability of negative affect: Gaussian process, SVM, linear lasso, and random forest. Each of these models has merit with respect to the issues that may ensue from constraints of data availability for model training, which is the case in this study. Although random forest, SVM, and Lasso regression are well-studied, Gaussian processes have demonstrated promising performance in e-health applications [15] mostly because they enable experts to encode their beliefs about smoothness or periodicity using covariance functions. In addition, the complexity of the model is inherently regulated (see chapter 5 of [16]) and provides uncertainty over predicted values. In our case, we used the squared-exponential covariance function [16]:

$$K(x, x') = \theta_s^2 exp \left\{ \frac{-||x - x'||^2}{2\theta_\ell^2} \right\} \quad (1)$$

where $\theta_t = \{\theta_s, \theta_\ell\}$, with θ_s and θ_ℓ being the hyperparameters of the covariance function regulating the y-scale and x-scale, respectively.

V. RESULTS

Figure 2 presents the performance of various clustering strategies compared with generalized models using the predictive algorithms presented earlier. Before analyzing performance, we will present a brief interpretation of each grouping strategy. Using data from SMS, four groups were discovered as presented in Table I. The group labeled *freq* are most actively engaged with text messaging on their phones, while *reg1* and *reg2* fall in the middle with *reg2* being more frequent than *reg1*. The most inactive group is labeled by *infreq*. In the profiles learned using the phone call logs, two groups were discovered: an active group and an inactive group in terms of their phone call level distributions. Notice that for the majority of time prior to EMAs, phone calls were rarely made by our study participants, and thus we see high percentages in the 'low' level. Using acceleration as a proxy to characterize participants' activity level, we found two: one *active* group and one *inactive* group. Again notice that the differences in the acceleration level distribution between the two learned groups are minor and only relative between them. With respect to locations, in the first group, the participants split most of

	Gp	Label	#Part	Group Profile (%) Low+	Low	Med	High	High+
SMS	1	reg1	22	80.5	16.8	2.0	0.5	0.2
	2	reg2	12	68.6	25.9	4.3	0.7	0.4
	3	infreq	9	93.7	5.9	0.3	0.1	0.0
	4	freq	19	49.1	36.1	8.6	3.5	2.7
Call	1	inactive	54		89.5	9.1	1.3	0.1
	2	active	8		65.7	29.2	4.4	0.7
Acc	0	active	25		83.1	4.6	12.3	
	1	inactive	37		91.2	2.7	6.1	

	Gp	Label	#Part	Out	Edu	Friend	Home	Leisure
Loc	1	school-home	34	2.0	49.2	4.4	38.7	5.8
	2	school	18	3.0	83.0	2.7	4.9	6.5
	3	out	10	20.9	43.7	8.3	9.3	17.8

their time between school and home; in the second group, the participants spent over 80% of their time at school at the expense of other places; and in the third group, the participants spent the majority of their time away from home (e.g., traveling out of town, visiting friends, and at leisure place of interests).

We also used cutoffs of 34 and 43 in SIAS scores to divide participants into low, medium, and high social anxiety groups [17]. In total, we experimented with 10 grouping approaches based on location, activity level, communications (SMS and phone calls), and SIAS scores as shown in Figure 2. Specifically, *DailyActivity* applies a combination of location, activity level, communications profiles; communication is based on the combination of phone calls and SMS (re-grouped into active and inactive) producing three groups (active in both SMS and calls, only active in either SMS or calls, inactive in both SMS and calls).

From Figure 2, using most of the grouping strategies, we were able to obtain better overall performance in lower weighted RMSE in our group models when compared to the generalized model. Specifically, our generalized models using four different algorithms achieved a RMSE of 21.58 (random forest), 22.05 (Gaussian processes), 21.87 (linear lasso), and 22.31 (SVM), respectively. For each grouping strategy on Gaussian processes model, we were able to obtain average reductions of RMSE 0.8722 (Location), 0.6310 (activity level), 0.045 (SMS), 1.2330 (calls), 1.4505 (SIAS), 1.9268 (DailyActivity), 0.4264 (communication), 1.2675 (SIAS+communication), 2.1326 (All features - communication), 1.9231 (All features - SIAS), respectively.

Note from Figure 2 that the DailyActivity grouping strategy consistently performed better than most other grouping strategies, and this strategy is also closest to the individual model approach (65 individual models for 65 participants) because it resulted in the most (25) subgroups among all these strategies, thus we used it to further investigate whether there are any
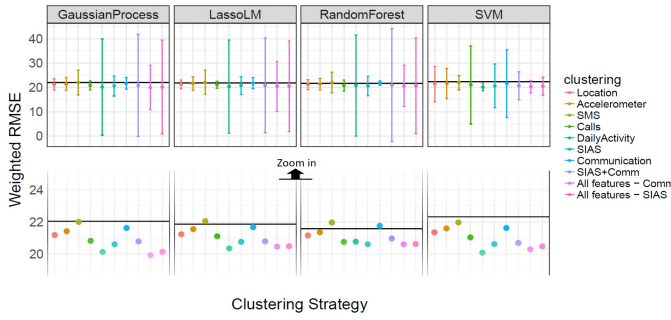
Fig. 2. The performance of each grouping strategy compared with the generalized model's performance (black horizontal line). The y-axis is the weighted root mean square error (WRMSE). The error bars represent 2 standard deviations of each grouping strategy.

specific patterns with respect to sample size to guide future design of group-level modeling approaches.

From Figure 3, we can see that there is a nonlinear relationship between sample size of groups and their performances. Groups with small sample size tend to perform either extremely poorly or extremely well. This signals potential weak generalizability of profiling strategies that forms many small groups. So the ideal situation will be to form groups with profiling strategies that evenly distribute the samples across different subgroups.
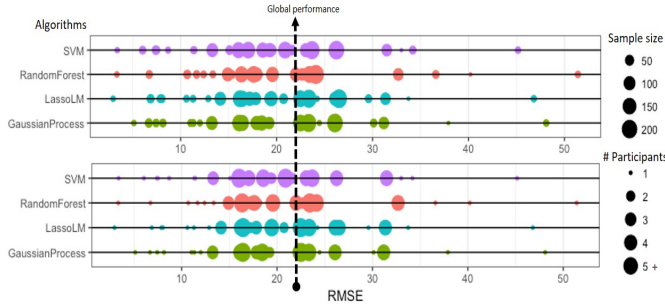

Fig. 3. The impact of sample size on the performance of groups formed by DailyActivity strategy.

VI. CONCLUSION

The focus of the present investigation was to provide a framework for accurately predicting negative affect from passively sensed data concurrent with individuals' affect ratings. Given that two weeks may be too short for algorithms to learn personalized models, we developed a method for predicting negative affect using a group-level approach. We first clustered participants using multimodal behavioral profiling, then we predicted negative affect from passively sensed data. The results indicate that profiling users based on their behavior improves the performance of the predictive model compared to generalized models. Future work will study the predictability levels among the different groups using validated questionnaire measures of personality and depression. The present study contributes to a body of research that aims to use passively sensed data to recognize user affect and launch interventions when and where they are most needed.

REFERENCES

[1] L. A. Clark, D. Watson, and S. Mineka, "Temperament, personality, and the mood and anxiety disorders." *Journal of Abnormal Psychology*, vol. 103, no. 1, p. 103, 1994.

[2] R. C. Kessler, P. Berglund, O. Demler, R. Jin, K. R. Merikangas, and E. E. Walters, "Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication," *Archives of General Psychiatry*, vol. 62, no. 6, pp. 593–602, 2005.

[3] A. Gentzler and K. Kerns, "Adult attachment and memory of emotional reactions to negative and positive events," *Cognition & Emotion*, vol. 20, no. 1, pp. 20–42, 2006.

[4] S. Yonekura, S. Okamura, Y. Kajiwara, and H. Shimakawa, "Mood prediction reflecting emotion state to improve mental health," vol. 3, no. 2, 2016, pp. 404–407.

[5] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "Moodscope: Building a mood sensor from smartphone usage patterns," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 389–402.

[6] S. Saeb, M. Zhang, C. J. Karr, S. M. Schueller, M. E. Corden, K. P. Kording, and D. C. Mohr, "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study," *Journal of Medical Internet Research*, vol. 17, no. 7, p. e175, jul 2015.

[7] P. I. Chow, K. Fua, Y. Huang, W. Bonelli, H. Xiong, L. E. Barnes, and B. A. Teachman, "Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students," *Journal of medical Internet research*, vol. 19, no. 3, 2017.

[8] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "StudentLife: Using smartphones to assess mental health and academic performance of college students," in *Mobile Health*. Springer International Publishing, 2017, pp. 7–33.

[9] J. Asselbergs, J. Ruwaard, M. Ejdys, N. Schrader, M. Sijbrandij, and H. Riper, "Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study," *Journal of medical Internet research*, vol. 18, no. 3, 2016.

[10] R. P. Mattick and J. C. Clarke, "Development and validation of measures of social phobia scrutiny fear and social interaction anxiety," *Behaviour research and therapy*, vol. 36, no. 4, pp. 455–470, 1998.

[11] H. Xiong, Y. Huang, L. E. Barnes, and M. S. Gerber, "Sensus: A cross-platform, general-purpose system for mobile crowdsensing in human-subject studies," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16*. ACM Press, 2016.

[12] J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello, "Extracting places from traces of locations," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 9, no. 3, p. 58, July 2005. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1094549.1094558

[13] C. Keßler, "OpenStreetMap," in *Encyclopedia of GIS*, S. Shekhar, H. Xiong, and X. Zhou, Eds. Cham: Springer International Publishing, 2015, pp. 1–5, dOI: 10.1007/978-3-319-23519-6_1654-1.

[14] G. Hamerly and C. Elkan, "Learning the k in k-means," in *Advances in neural information processing systems*, 2004, pp. 281–288.

[15] L. Clifton, D. A. Clifton, M. A. Pimentel, P. J. Watkinson, and L. Tarassenko, "Gaussian processes for personalized e-health monitoring with wearable sensors," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 193–197, 2013.

[16] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*, vol. 1.

[17] R. G. Heimberg, G. P. Mueller, C. S. Holt, D. A. Hope, and M. R. Liebowitz, "Assessment of anxiety in social interaction and being observed by others: The social interaction anxiety scale and the social phobia scale," *Behavior therapy*, vol. 23, no. 1, pp. 53–73, 1992.