How to Predict Mood? Delving into Features of Smartphone- Based Data

Conference Paper · August 2016

6 authors, including:

Dennis Becker
Leuphana University Lüneburg

Vincent Bremer
Leuphana University Lüneburg

Burkhardt Funk
Leuphana University Lüneburg

Joost Asselbergs
Vrije Universiteit Amsterdam

How to Predict Mood? Delving into Features of Smartphone-Based Data

Full papers

Dennis Becker
Leuphana University Lüneburg
dbecker@leuphana.de

Vincent Bremer
Leuphana University Lüneburg
vincent.bremer@leuphana.de

Burkhardt Funk
Leuphana University Lüneburg
funk@uni.leuphana.de

Joost Asselbergs
VU University Amsterdam
j.a.g.j.asselbergs@vu.nl

Heleen Riper
VU University Amsterdam
h.riper@vu.nl

Jeroen Ruwaard
VU University Amsterdam
j.j.ruwaard@vu.nl

Abstract

Smartphones are increasingly utilized in society and enable scientists to record a wide range of behavioral and environmental information. These information, referred to as Unobtrusive Ecological Momentary Assessment Data, might support prediction procedures regarding the mood level of users and simultaneously contribute to an enhancement of therapy strategies. In this paper, we analyze how the mood level of healthy clients is affected by unobtrusive measures and how this kind of data contributes to the prediction performance of various statistical models (Bayesian methods, Lasso procedures, etc.). We conduct analyses on a non-user and a user level. We then compare the models by utilizing introduced performance measures. Our findings indicate that the prediction performance increases when considering individual users. However, the implemented models only perform slightly better than the introduced mean model. Indicated by feature selection methods, we assume that more meaningful variables regarding the outcome can potentially increase prediction performance.

Keywords

Unobtrusive EMA, E-Mental-Health, Mood Prediction, Smartphone-Based Data, Bayesian Modeling.

Introduction

A good state of mental health is an important factor for every individual as it promotes a healthy social environment, provides general motivation in achieving life goals, and can even benefit the economy. Using today's modern resources, individuals are being more proactive about their mental health. The Internet is increasingly being used to search for medical information and thus individuals are becoming better informed about their health (Spil and Schuring, 2006) as the demand for medical knowledge in the population is steadily growing (Andreassen et al., 2007). Seeking ”new” ways to treat depression and other mental diseases, healthcare treatments supported by electronic processes such as e-mental-health have been established. Internet-based treatments have the potential to provide high quality treatment by sustaining deeper insight into the daily lives of the clients and thus enhancing the overall therapy success (Eysenbach, 2001). These treatments additionally allow for the creation of relatively new kinds of data.

Ecological Momentary Assessments (EMA) are data collection methods that provide researchers with data in regard to symptoms, behavior, and cognition close in time to the clients' experience in their natural environment (Iida et al., 2012; Stone and Shiffman, 1994). Due to technological development, instead of the traditional pen and paper diary or submission through a bulky at-home desktop, EMA data can now take on the form of accessible and convenient electronic devices that can lead to a greater ecological validity (Iida et al., 2012). Smartphones, in particular, provide advanced computing and storage capabilities that in turn lead to the ability to record a wide range of behavioral and environmental information (Gaggioli et al., 2013; Gimpel et al., 2015), can possibly be a beneficial factor for gaining deeper insight into clients' behavior, and simultaneously provide guidance for an improvement of future therapy strategies. Thus, smartphones are increasingly utilized as sensors for physical, social, and other activities humans are occupied with (Asselbergs et al., 2016).

The Smartphone sensing process, also referred to as unobtrusive EMA, silently accumulates the clients' data without prompting the user for additional information. Various unobtrusive measures can therefore be collected and utilized when predicting or inferring certain psychological concepts such as academic performance (Wang et al., 2014), sleep duration (Zhenyu et al., 2013), and depression (Saeb et al., 2015). Therefore, this data can be utilized as additional information besides subjectively reported measures by the clients (Gaggioli et al., 2013) or even as predictor to forecast traditional measurements. This collection method has certain advantages. It obviously reduces usability issues tremendously since no interaction between the client and software is required. Furthermore, almost everyone can potentially be monitored since nowadays a smartphone is a fundamental device which is wide spread in the society and part of every day life (Zhenyu et al., 2013; Abdelzaher et al., 2007).

Previous studies have already investigated the usage of smartphone-based data. Burns et al. (2011), for example, develop a mobile application that assists depressive clients in difficult situations with Ecological Momentary Interventions. This application infers the users' mental state using measures of location, activity, social environment, and smartphone usage. Based on the measurements, this model then determines if a supportive intervention is triggered. Furthermore, Saeb et al. (2015) demonstrate that daily movement patterns, estimated from a phones' GPS record, help in determining depressive symptoms. Features such as the variance in visited locations, frequency of visited locations, and time spend there are being generated from the GPS data. These movement patterns have then been shown to correlate with depressive symptoms. Moreover, Ma et al. (2012) develop a program called MoodMiner. This is a smartphone application that infers the owner's mood. LikamWa et al. (2013) also create a program for achieving a similar goal (MoodScope). However, even though many studies exist in this field and LikamWa et al. (2013) find that smartphone usage data similar to ours correlates well with the users' mood level, it is not assured that unobtrusive measures can generally contribute significantly towards predictions of any kind. Asselbergs et al. (2016), for example, demonstrate that unobtrusive measures may not contribute as much as the study of LiKamWa et al. (2013) suggests. Thus, in this paper we seek to analyze this aspect even further.

We make an attempt to predict the mood level of healthy Dutch students by using the aforementioned mobile phone usage data and simultaneously compare different statistical methods. Furthermore, we seek to reveal the importance of various unobtrusive measures and specify their contribution and influence on the prediction performance. The analysis of unobtrusive mobile phone measurements and their contribution to prediction performance is yet to be analyzed more intensively in research. Additionally, the fact that we utilize data of healthy clients might be an interesting factor because the gained insight into healthy individuals' behavior might provide guidance for improving actions of unhealthy clients. In the following chapters, we introduce the data used for our approach, illustrate the utilized models, conclude our results, and briefly outline future procedures.

Data & Methods

The Data

We utilize smartphone-based obtrusive and unobtrusive data from an explorative uncontrolled pilot study for our analyses (Asselbergs et al., 2016). The dataset consists of 27 healthy Dutch students who reported their mood level on their smartphones for six weeks at a frequency of 5 times per day and simultaneously provided unobtrusive measures by the usage of a smartphone application called iYouVU that silently collects data in the background (Asselbergs et al., 2016). Throughout this study, we will refer to these students as "subjects" or "clients" for clarity and for the sake of convenience. Usually, the clients provided traditional mood measures (one-dimensional mood measure – 10-point scale; two-dimensional measures valence and arousal – -2 to 2 scale) at specific times: 9am, 12pm, 3pm, 6pm, and 9pm; however, they were allowed to push back the measurement requests which results in differing measurement times amongst the subjects. Asselbergs et al. (2016) already utilized and aggregated the data on a daily basis in an explorative study. However, we only aggregate the unobtrusive data up to the point of each mood request. This procedure results in 1335 observations of the 27 clients. Because no treatment was provided, the data reflect the natural course of mood over time. Appendix A illustrates the attributes of the dataset in more detail.

The Approach

For analyzing the data, we conduct several analyses including different statistical methods. We first analyze the data on a non-user level. Afterwards, we conduct analyses on a user level in order to possibly reveal advantages of the hierarchical structure. In the following accompanying information, we illustrate the different approaches:

First, we perform analyses in the instances in which every data point is being considered as independent from the other touchpoints – individuals are not taken into account. We call this analysis the non-user level analysis. For this method, we split the dataset into a training (75%) and a test dataset (25%). As mentioned before, this splitting process does not consider any user specific touchpoints. Then, based on the specific model utilized and the used independent variables, we make an attempt to predict the mood level of the test dataset. Additionally, a feature ranking is used to reveal the contribution of each phone measure to the mood prediction. All phone measures are then sorted and listed according to their importance.

Second, we analyze the data on a personalized level. Particularly, we consider individuals in this case. We also introduce the weekday and time of the request as additional variables, which explains why the first week of the data points is required for model training. We then start the estimation of the daily mood based on the data points of the previous week. After the first mood prediction, the next data point is added to the training set, the model is retrained, and the next day is predicted based on all previous data. Specifically, when estimating the mood level on the 8th day, data from day one to seven are used for training. The records of the variables on the 8th day are then used to predict the mood level on the 8th day based on the trained model. We call this method the user level analysis because again, we do consider individual clients. In the following sub-chapters, we introduce the utilized models for the prediction of the subject's mood level, for the creation of feature importance rankings, and subsequently the used performance measures for model comparison.

The Mean Model

As method for comparison, we use a mean model in our analyses. Specifically, we utilize the outcome mean of the observations for predictions. For the non-user level analysis, we use the mean value over all mood values of the training set and utilize this mean for predictions. However, since we consider individuals in the user level analysis, we calculate the mean based on every subjects' mood observations upon the current measurement point. Particularly, when estimating the mood level on the 8th day, we utilize the mean of day one to seven for this individual as the mean model prediction.

Linear Regression

We use linear models and generalized linear models to predict the mood level of the subjects. The relationships between the outcome variable (mood level in this case) and the independent variables (all unobtrusive measures) are represented by the coefficients. The intercept illustrates the prediction the model creates if all independent variables were zero. The other coefficients (one for each attribute) represent the change in the predicted value of the outcome per unit of change in the corresponding independent variable (i.e. app usage) while all other x variables are fixed (Rencher and Christensen, 2012).

Support Vector Machine

We also utilize support vector machines (SVM). Support vector machines exhibit great classification performance and have been used in various fields (Burges, 1998). They are often applied to supervised learning problems (Vapnik, 1998). For predicting the mood level, we use the so called ε-Support Vector regression (Vapnik, 2000). Support vector regression seeks to find a linear function that does not penalize mood values that are smaller than a specific ε value. By not penalizing the training samples within this ε bound, it is ensured that most samples lie within this bound. Simultaneously, the estimated weights are supposed to be as small as possible to produce a flat solution (Smola and Schölkopf, 2004). During training, a balance between the flatness and a small ε parameter is found in order to produce a solution that sufficiently fits both requirements.

Lasso Regression

Furthermore, we use Lasso regression in our project (Least Absolute Shrinkage and Selection Operator). Lasso regression is a linear regression algorithm including an additional linear penalty term. The cost function of the regression, which is to be optimized, consists of the mean square error of the misclassified samples. By minimizing the classification error, over-fitting can possibly occur. In this case, the training error steadily decreases, essentially improving predictions. However, the error on a new test set increases because the algorithm generalizes poorly. An additional penalty term is introduced to the cost function to prevent over-fitting. Specifically, the Lasso regression penalizes the absolute value of the regression coefficients (Tibshirani, 1994). This linear penalty is the sum of the absolute values over all weights and enforces useless coefficients to shrink towards zero in order to produce a sparse solution. The optimization problem that arises is shown in Equation 1:

$$\underset{\beta}{minimize}\ \{||Y\text{-}x\beta||^2 - \lambda||\beta||_1\} \qquad (1)$$

The λ term influences the "strength" of the penalty. Specifically, the higher the value of λ, the higher the penalty. A higher penalty leads to sparser solutions (more coefficients equal zero). The optimal λ can be found by utilizing cross validation. In that case, the dataset is separated into a specific amount of chunks. One chunk is then being predicted by the rest of the separated parts. This procedure is repeated for all λ´s. The λ with the smallest cross validation error is then used for the final prediction.

Bayesian Hierarchical Linear Regression

To consider the hierarchical structure of the data and include the phone measures of all subjects in every estimation, we develop a Bayesian hierarchical linear regression model. In a clinical setting, it is expected that people would not be willing to constantly contribute mood measures for training purposes. In this case and similar to the study of LikamWa et al. (2013), our model can consider data of additional subjects in order to improve predictions for another client.

The implemented model is similar to the model used by MacKay (1996), Neal (1996), and Tipping et al. (2000). However, we additionally implement a hierarchical prior on the individual weights. Since the model is trained for each subject and each day in the user level analysis, we implement a variational approximation for speed benefits (Bishop, 2006). This approximation is possible because only conjugate priors are used. Figure 1 illustrates the plate notation of the model:

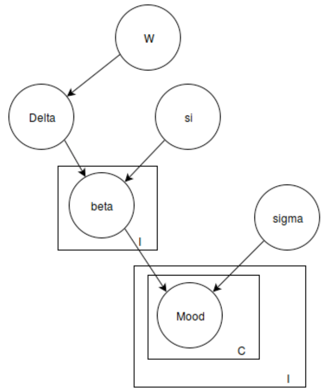

Figure 1: Plate Notation

Every user has its own set of beta coefficients. These coefficients are dependent on the hierarchical uninformative prior Delta that is sampled with variance w. These individual gamma priors w are used to estimate the influence of each weight and determine their importance. Specifically, w stands for the variance of each individual weight and simultaneously allows for the creation of a feature importance ranking. Additionally, one gamma prior si is used for the individual weights for all users. This prior defines how strongly the coefficients are allowed to differ from the hierarchical prior Delta. Moreover, sigma (the variance for mood) represents the additional noise in the mood level.

Performance Measure

We use the Root Mean Square Error (RMSE) as comparison and performance measure. The RMSE can be defined as the average distance between predicted and observed values. Based on the RMSE, a confidence interval can be derived. The interval limits for the 95% confidence interval lie approximately on +/- 2*RMSE. Thus, the actual value (the value that is attempted of being predicted) lies within this confidence interval with a probability of 95%. In addition to the RMSE, we utilize a specific performance measure that is illustrated by Equation 2:

$$Performance = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(residual_i^2 < .25) \quad (2)$$

This performance measure illustrates the percentage of correctly classified predictions. It classifies forecasts as correct when the prediction lies within a certain boundary around the true value, in this case .5 (LikamWa et al., 2013). In the next chapter, we present the results from our analyses. We subdivide the chapters into our non-user and user level analyses.

Results

Analysis – Non User Level

Table 1 illustrates the results of the non-user level analysis. We execute the methods already introduced in the previous chapter. Furthermore, we illustrate the feature importance ranking of the Lasso procedures in Appendix B.

Methods	RMSE	Performance
Mean Model	1.11	0.38
SVM	0.87	0.41
Regression	0.83	0.40
Lasso	0.86	0.39
Lasso (IE)	0.84	0.38

Table 1. Results Non-User Level

In our non-user level analyses, we fit linear models by using the glm function and apply support vector machine procedures by utilizing the e1071-package (Meyer et al., 2015) in R (R Core Team, 2015). In a first attempt, we use all variables as predictors (SVM & Regression). Afterwards, we perform an analysis that executes Lasso procedures in a linear regression and simultaneously selects features by using the glmnet- package (Friedman et al, 2010). Additionally, we multiply all columns with one another and center the results to consider potential interaction effects between the different concepts. For feature selection in this case, we repeatedly take advantage of the glmnet-package (Friedman et al, 2010). The RMSE values of our analysis are between .83 and 1.11. The mean value prediction model achieves a RMSE of 1.11 and a performance measure of .38, as shown in Table 1. The predictions of this model are solely based on the mean mood of the training data. The RMSE and performance measures of the other models, which use more features besides the mood measure, perform slightly better.

Their RMSE error is smaller than the RSME of the mean model. Consequently, the additional measurements might contain information about the current mood level. Therefore, we do correctly classify more values and perform slightly better by utilizing various models compared to the mean model. However, the results of the interaction effect analysis indicate that only the psychological concepts valence and arousal influence the mood significantly. Almost all other attributes either only slightly contribute or completely fail to contribute to the prediction at all. The importance ranking in Appendix B supports this result even further. To conclude, although the linear regression model that is inclusive of all variables demonstrates the best performance, we do not correctly predict significantly more values than the mean model - irrespective of which model is used. As these results are insufficient, we also examine the provided dataset on a user level in an attempt to potentially further reduce the RMSE by considering individuals and the hierarchical structure.

Analysis – User Level

For this analysis, we add further features to the dataset. Specifically, we add the weekday (extracted from the time stamp) and a feature that indicates the n-th measure of this particular day (measureOfDay). The results are illustrated in Table 2:

Methods	RMSE	Performance
Mean Model	0.90	0.49
SVM	0.85	0.51
Regression	0.61	0.57
Lasso	0.53	0.56
Bayesian Hierarchical	0.58	0.57

Table 2. Results User Level

In the user level analysis, the RMSE differs between .53 and .90. Therefore, Table 2 indicates that considering individual clients can in fact improve the results of analyses. The applied models perform slightly better than the mean model. As indicated in Table 2, the improvement is also higher compared to the non-user level analysis. The hierarchical model results in a smaller RSME than the regression and SVM model. Nevertheless, the Lasso regression results in an even lower RSME. This is unexpected because the hierarchical model uses the data of all users and Lasso regression utilizes only the current users' data. However, when inspecting both algorithms, it is noticeable that the Lasso procedures set the weights to 0 based on the lambda value. This procedure is not to be found in the hierarchical model. Even tough the weights are all individually penalized to enforce sparseness and provide a feature ranking, they never truly become zero. Therefore, the utilization of all the non-contributing variables might be responsible for the slightly higher RSME. Even though the models perform better than the models in the non-user level analysis and even slightly better than the mean model, we do not deem these results sufficient. Therefore, we seek to gain more insight as to which features can potentially be useful for mood prediction. As a result, we turn to a feature ranking using all data estimated from the Lasso regression and the hierarchical model.

Lasso Analysis			Bayesian Analysis	
Rank	Feature Name	Lambda	Feature Name	Prior
1	valence	.657	valence	1.06
2	arousal	.085	activity	98.14
3	activity	.058	arousal	116.54
4	measureOfDay	.049	weekday7	132.52
5	appCat.entertainment	.040	weekday2	141.29
6	appCat.weather	.037	weekday3	144.18
7	weekday7	.033	weekday5	146.86
8	call	.019	weekday4	152.81
9	appCat.builtin	.016	measureOfDay	155.61
10	weekday3	.013	weekday6	159.13
11	sms	.011	sms	160.27
12	appCat.social	.010	appCat.weather	160.52
13	appCat.office	.009	appCat.finance	161.16
14	appCat.finance	.008	appCat.travel	161.17
15	appCat.other	.008	call	161.35
16	appCat.unknown	.007	appCat.office	161.35
17	weekday5	.006	appCat.game	161.36
18	weekday2	.004	appCat.utilities	161.36
19	appCat.utilities	.003	appCat.other	161.36
20	screen	.003	appCat.social	161.36
21	appCat.game	.002	appCat.enterntainment	161.37
22	weekday4	.002	appCat.unknown	161.37
23	weekday6	.002	screen	161.37
24	appCat.travel	.002	appCat.builtin	161.37
25	appCat.communication	.002	appCat.communication	161.37

Table 3. Feature ranking

As indicated by both feature rankings in Table 3, the coefficient of valence, arousal, and activity appear useful when inferring the mood from mobile phone measurements. The influence of valence and arousal are not surprising because these measurements correlate with mood and are taken at the same time the mood level is reported. Both rankings also indicate that weekday and measureOfDay might influence the current mood. To analyze this phenomenon further, we implement a hierarchical Bayesian regression model in JAGS (Appendix C). Hence, we are enabled to reveal the significant and influencing factors regarding the mood level. Table 4 illustrates the results of this analysis:

Attributes	Mean	2.5% & 97.5%
arousal	0.09	(0.04; 0.14)
valence	1.00	(0.91;1.10)
activity	0.08	(0.04; 0.13)
Monday	-0.06	(-0.13;-0.001)
Saturday	0.07	(0.002; 0.13)
measureOfDay4	0.09	(0.02; 0.15)
measureOfDay5	0.14	(0.08; 0.20)

Table 4. Significant features for predicting mood - Bayesian hierarchical regression

According to Table 4, the activity level and Saturdays have a significant positive effect on the mood level. Furthermore, Mondays seems to affect the mood in a slightly negative manner. This can possibly be due to the fact that individuals often have a hard time returning to work after a weekend (Helliwell and Wang, 2015). Our results also show that the mood level of the subjects steadily increases in time. Specifically, the fourth measure of mood (measureOfDay4) is taken after work and the fifth measure (measureOfDay5) is taken around 9pm when clients are presumably at home or engaged in leisure time activities. Since an individuals' mood is increasing after work or on days off, the influences of time and weekday on mood can potentially be referred to work activities.

Conclusion

We perform a multitude of different statistical methods on a dataset that includes UEMA and EMA observations of healthy Dutch students. We try to reveal important variables and seek to predict the future mood level of the subjects. In conclusion, we find that individuals generally have an increased mood level on weekends and during leisure time. Additionally, the weekday Monday influences healthy individuals negatively. However, since the utilized dataset consists of only students, we might not be able to generalize the findings and assume that an older population would be influenced similarly. We further illustrate that analyzing this dataset on a user level, which means the consideration of the hierarchical structure and therefore individuals, can improve analyses. However, the implemented models only perform slightly better than the introduced mean model. We are still not satisfied with these results. Although the implementation of the more complex Bayesian hierarchical model and the Lasso procedures enhance the prediction performance in the user level analysis, only a slight improvement is achieved. The feature ranking analyses further raises the assumption that the variables do not tremendously contribute to the predictions. This circumstance would support the inability of creating better prediction performance. Since the UEMA method is considered a young field (Asselbergs et al., 2016), this kind of data might still be a powerful tool for future findings. In our case, with the provided dataset and the implemented models, we are not able to reach a significantly better prediction performance. In the future, we aim to build models that are able to predict the mood level of subjects more accurately and find ways to select important features more reliably. Implementing more individualized hierarchical models that account for heterogeneity amongst the clients might be one option to increase the prediction performance of the mood level. Moreover, we assume that obtaining more meaningful features that might contribute to the forecast more intensively can boost prediction performance as well.

REFERENCES

Abdelzaher, T., Anokwa, Y., Boda, P., Burke, J., Estrin, D., Guibas, L., Kansal, A., Madden, S., and Reich, J. (2007). Mobiscopes for Human Spaces. IEEE Pervasive Computing, 6(2): 20–29.

Andreassen, H. K., Bujnowska-Fedak, M. M., Chronaki, C. E., Dumitru, R. C., Pudule, I., Santana, S., Voss, H., and Wynn, R. (2007). European citizens' use of E-health services: a study of seven countries. BMC public health, 7(53).

Asselbergs, J., Ruwaard, J., Ejdys, M., Schrader, N., Sijbrandij, M., and Riper, H. (2016). Mobile Phone-based Unobtrusive Ecological Momentary Assessment of Day-to-Day Mood: An Explorative Study.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Secaucus, NJ, USA.

Burges, C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2(2):121–167.

Burns, M. N., Begale, M., Duffecy, J., Gergle, D., Karr, C. J., Giangrande, E., & Mohr, D. C. (2011). Harnessing context sensing to develop a mobile intervention for depression. Journal of Medical Internet Research, 13(3).

Eysenbach, G. (2001). What is e-health? Journal of Medical Internet Research, 3(2):e20.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1):1-22.

Gaggioli, A., Pioggia, G., Tartarisco, G., Baldus, G., Corda, D., Cipresso, P., and Riva, G. (2013). A mobile data collection platform for mental health research. Personal & Ubiquitous Computing, 17(2):241.

Gimpel, H., Regal, C., and Schmidt, M. (2015). myStress: unobtrusive smartphone-based stress detection. Proceedings of the 23th European Conference on Information Systems, Münster, Germany.

Helliwell, J. F. and Wang, S. (2015). How Was the Weekend? How the Social Context Underlies Weekend Effects in Happiness and Other Emotions for US Workers. PLoS ONE, 10(12): e0145123.

Iida, M., Shrout, P., Laurenceau, J., and Bolger, N. (2012). Using Diary Methods in Psychological Research. In Cooper, H., Camic, P. M., Long, D. L., Panter, A. T., Rindskopf, D., and Sher, K. J., editors, APA Handbook of Research Methods in Psychology: Vol. 1. Foundations, Planning, Measures and Psychometrics, pages 277– 305. US: American Psychological Association, Washington, DC.

LikamWa, R., Liu, Y., Lane, N. D., and Zhong, L. (2013). MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. In: Chu H, Huang P, Choudhury RR, Zhao F, editors. MobiSys '13; Taipei, Taiwan. New York: ACM; 2013. p. 389.

Ma, Y., Xu, B., Bai, Y., Sun, G., & Zhu, R. (2012). Daily mood assessment based on mobile phone sensing. In Proceedings - BSN 2012: 9th International Workshop on Wearable and Implantable Body Sensor Networks (pp. 142–147).

MacKay, D. J. C. (1996). Bayesian methods for back propagation networks. In Models of neural networks III, pages 211–254.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Friedrich Leisch (2015). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien.

Neal, R. M. (1996). Bayesian learning for neural networks. Number 118. Rencher, A. and Christensen, W. (2012). Multivariate regression. In Methods of Multivariate Analysis, chapter 10, page 19 ff. John Wiley & Sons, 3rd edition.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M.E., Kording, K. P., and Mohr, D. C. (2015). Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. Journal of medical Internet research; 17(7):e175.

Smola, A. J. and Schölkopf, B. (2004). A Tutorial on Support Vector Regression. Statistics and Computing, 14(3):199–222.

Smyth, J. M. and Stone, A. (2003). Ecological momentary assessment research in behavioral medicine. Journal of Happiness Studies, 4(1):35–52.

Spil, T. and Schuring, R. (2006). E-Health Systems Diffusion and Use: The Innovation, the User and the Use IT Model. Idea Group Pub.

Stone, A. A. and Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. Annal Behav Med 16:199–202.

Tibshirani, R. (1994). Regression Selection and Shrinkage via the Lasso. Journal of the Royal Statistical Society, Series B, 58:267-288.

Tipping, M. E. (2000). The Relevance Vector Machine. In S.A. Solla, T. K. Leen, and K.-R. Müller (Eds.), Advances in Neural Information Processing Systems 12, pp. 652-658.

Vapnik, V. N. (1998). Statistical Learning Theory (Adaptive and learning Systems for Signal Processing, Communications and Control). John Wiley & Sons, New York, NY, USA.

Vapnik, V. N. (2000). The Nature of Statistical Learning Theory (Information Science and Statistics). Springer-Verlag New York, New York, NY, USA.

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., and Campbell, A. T. (2014). Student Life: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students using Smartphones. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM,. p. 3–14.

Zhenyu, C., Mu, L., Fanglin, C., Lane, N. D., Cardone, G., Rui, W., Li, T., Chen, Y., Choudhury, T., and Campbell, A. T. (2013). Unobtrusive sleep monitoring using smartphones. Pervasive Health '13: 7th International Conference on Pervasive Computing Technologies for Healthcare, 2013, pp. 145-152.

Appendix

A

Attribute	Description
valence	Two dimensional construct (-2 - 2)
arousal	Two dimensional construct (-2 - 2)
activity	average percentage of high accelerometer data points
measureOfDay	which number of data record a day
appCat.entertainment	app use frequency, entertainment app category
appCat.office	app use frequency, office app category

appCat.finance	app use frequency, business tools category
appCat.other	app use frequency, other app category
appCat.unknown	app use frequency, unknown app category
appCat.weather	app use frequency, weather app category
appCat.builtin	app use frequency, builtin app category
appCat.social	app use frequency, social app category
appCat.utilities	app use frequency, utilities app category
appCat.game	app use frequency, games app category
appCat.communication	app use frequency, communication app category
appCat.travel	app use frequency, travel app category
call	number of calls made (5 most frequent contacts)
sms	number of SMS's sent (5 most frequent contacts)
screen	screen-on frequency
weekday2	weekday – Monday
weekday3	weekday – Tuesday
weekday4	weekday – Wednesday
weekday5	weekday – Thursday
weekday6	weekday – Friday
weekday7	weekday – Saturday

B

Rank	Feature	Lambda
1	valence	.63
2	arousal	.12
3	activity	.003
4	appCat.builtin	.003
5	appCat.weather	.003
6	call	.003
7	appCat.game	.002
8	appCat.social	.002
9	appCat.communication	.002
10	appCat.utilities	.001
11	appCat.unknown	.001
12	appCat.finance	.001
13	appCat.office	.001
14	appCat.travel	.001

C

```
model{
    for(j in 1:N){
      mu[j]<-inprod(X[j, ],beta[id[j], ])
      mood[j]~dnorm(mu[j],tau)
    }

    for(u in 1:USERS){
      for(i in 1:p){
        beta[u,i]~dnorm(delta[i],indiv_gamma[i])
      }
    }

    for(j in 1:p){
      indiv_gamma[j]~dgamma(0.001,0.001)
    }

    delta~dmnorm(mubar,sigma)
    tau~dgamma(0.001,0.001)
}
```