

## Prediction of Daily Happiness Using Supervised Learning of Multimodal Lifelog Data

Predição da Felicidade Diária Usando Aprendizado Supervisionado de Registro de Dados Multimodais de Vida

Predicción de la Felicidad Diaria Mediante el Aprendizaje Supervisado de Datos Lifelog Multimodales

*Tetsuya Yamamoto*<sup>1</sup>

*Tokushima University*

*Junichiro Yoshimoto*

*Nara Institute of Science and Technology (NAIST)*

*Eric Murillo-Rodriguez*

*Universidad Anáhuac Mayab*

*Sergio Machado*

*Salgado de Oliveira University*

---

### Abstract

Developing an approach to predict happiness based on individual conditions and actions could enable us to select daily behaviors for enhancing well-being in life. Therefore, we propose a novel approach of applying machine learning, a branch of the field of artificial intelligence, to a variety of information concerning people's lives (i.e., a lifelog). We asked a participant (a healthy young man) to record 55 lifelog items (e.g., positive mood, negative events, sleep time etc.) in his daily life for about eight months using smartphone apps and a smartwatch. We then constructed a predictor to estimate the degree of happiness from the multimodal lifelog data using a support vector machine, which achieved 82.6% prediction accuracy. This suggests that our approach can predict the behaviors that increase individuals' happiness in their daily lives, thereby contributing to improvement in their happiness. Future studies examining the usability and clinical applicability of this approach would benefit from a larger and more diverse sample size.

*Keywords:* machine learning, lifelog, artificial intelligence, happiness, behavior

### Resumo

Desenvolver uma abordagem para prever a felicidade com base em condições e ações individuais pode nos permitir selecionar comportamentos diários para melhorar o bem-estar na vida. Portanto, propomos uma nova abordagem de aplicação da aprendizagem de máquina, um ramo do campo da inteligência artificial, para uma variedade de informações sobre a vida das pessoas (ou seja, um lifelog). Pedimos a um participante (um jovem saudável) que registrasse 55 itens de vida útil (por exemplo, humor positivo, eventos negativos, tempo de sono etc.) em sua vida diária por cerca de oito meses usando aplicativos de smartphones e um relógio inteligente. Em seguida, construímos um preditor para estimar o grau de felicidade dos dados de vida multimodal usando uma máquina de vetores de suporte, que atingiu 82,6% de precisão de previsão. Isso sugere que nossa abordagem pode prever os comportamentos que aumentam a felicidade dos indivíduos em suas vidas diárias, contribuindo para uma melhoria em sua felicidade. Estudos futuros examinando a usabilidade e a aplicabilidade clínica dessa abordagem se beneficiariam de um tamanho de amostra maior e mais diversificado.

*Palavras-chave:* aprendizado de máquina, log de vida, inteligência artificial, felicidade

### Resumen

El desarrollar un enfoque para predecir la felicidad, basado en las condiciones y acciones individuales, nos permitiría seleccionar comportamientos habituales para mejorar el bienestar en la vida. Por lo

---

<sup>1</sup> **Contact address:** Graduate School of Technology, Industrial and Social Sciences, Tokushima University1-1, Minamijosanjima-cho, Tokushima 770-8502, Japan, Contact: +81 88 656 761. E-mail: [t.yamamoto@tokushima-u.ac.jp](mailto:t.yamamoto@tokushima-u.ac.jp)

tanto, proponemos un novedoso enfoque de aplicación del aprendizaje automático, una rama del campo de la Inteligencia Artificial, a una variedad de información de la vida de las personas (es decir, un lifelog). Se le pidió a un participante (un sujeto joven sano) que registrara 55 elementos de lifelog (por ejemplo, humor positivo, eventos negativos, tiempo de sueño etc.) en su vida diaria, durante aproximadamente ocho meses, usando aplicaciones de teléfonos inteligentes, y un reloj inteligente. Posteriormente, construimos un predictor para estimar el grado de felicidad, a partir de los datos lifelog multimodales, utilizando un equipo de vectores de soporte, que logró una precisión de predicción del 82.6%. Estos datos sugieren que nuestro enfoque, puede predecir los comportamientos que incrementan la felicidad de las personas en su vida diaria, contribuyendo así, a una mejora en su felicidad. Los futuros estudios que examinen la usabilidad, y la aplicabilidad clínica de este enfoque, se beneficiarían al analizar un tamaño de muestra más grande, y más diversa.

*Palabras-clave:* aprendizaje automático, log de vida, inteligencia artificial, felicidad

## Introduction

Happiness has been considered an important factor for healthy human lives. Although numerous factors (e.g., social relationships) influencing happiness have been investigated (Bruni & Porta, 2005; Diener & Seligman, 2004), no study has examined an individualized approach that would help us know what we should do in the present to improve our happiness. Therefore, developing an approach for predicting happiness based on individual conditions and actions could enable us to select behaviors for enhancing well-being in life.

Toward this end, we propose a novel approach of applying machine learning, a branch of the field of artificial intelligence, to a variety of information concerning people's lives (i.e., a lifelog). Recording several lifelogs via various technologies such as smartphones and wearable devices enables us to obtain real-time behavioral, physiological, and psychosocial data (Marzano et al., 2015; Yamamoto & Yoshimoto, 2018). Additionally, machine learning automatically seeks informative patterns in those complex data and uses them to make predictions about and/or decisions in the future. These approaches could enable us to analyze our lifelog and thus greatly contribute to improving our daily happiness.

To examine the usefulness of this approach, we constructed a predictor to estimate the degree of happiness from lifelog data using a support vector machine (SVM) (Vapnik, 1995), a specific algorithm of supervised machine learning.

## Methods

We asked a participant (a man in his 20s with no physical and mental issues) to record the lifelog items below in his daily life for about eight months using two iPhone apps, Count Log Lite and MoodTools, and one smartwatch, the Fitbit Alta HR (Fitbit, Inc.), which allowed easy recording. We selected 55 features (variables) comprising: a) mood experiences (positive and negative mood, accomplishment, fatigue etc.); b) cognitive and behavioral experiences (positive and negative interpretation, rumination, gratitude etc.); c) psychiatric and physical symptoms (depression, headache, abdominal symptoms etc.); d) lifestyle characteristics (amount of sleep, physical activity, relaxation etc.); and e) life events (positive and negative). For the assessment of happiness, we asked the participant to rate the degree of happiness on that day, ranging from not at all (0) to extremely (10), and categorized the happiness score into three classes: low, middle, or high happiness. We have depicted all features in Table 1.

Table 1

*Lifelog data list*

Category	Item (feature)	
Mood experience	Positive mood	Happiness
	Negative mood	Daytime fatigue
	Accomplishment	Fatigue on awakening
Cognitive and behavioral experience	Positive thought	Reflection
	Negative thought	Rumination
	Mindfulness activity	Regret
	Breathing	Impulsiveness
	Exercise	Idleness
	Gratitude	
Psychiatric and physical symptom	Abdominal symptom	Low back pain on awakening
	Headache	Depression
	Daytime low back pain	
Lifestyle characteristic	Workload	Minutes awake
	Study	Number of awakenings
	Favorite activity	Time in bed
	Other activity	Minutes of REM sleep
	Reward for effort	Minutes of light sleep
	Seeing family pictures	Minutes of deep sleep
	Steps	Daytime nap
	Distance	Early-morning awakening
	Minutes of light activity	Going back to sleep
	Minutes of fair activity	Staying up late the previous night
	Minutes of high activity	Keeping late hours
	Activity calories	Rest day
	Minutes asleep	
Life event	Positive event	Meeting last night
	Negative event	Being away overnight last night
	Family time	Business
	Interpersonal exchange	Personal interview

We developed a predictor to estimate the degree of happiness from the above-mentioned features using a nonlinear SVM with a Gaussian kernel, which was implemented with the “e1071” package (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2014) in R (version 3.1.1; R Core Team, 2014). The detailed procedures were as follows.

Among the entire lifelog dataset for 237 days, the subset for the first 190 days was assigned to train the SVM, and the subset for the last 47 days was assigned to the test dataset for evaluating the generalization performance of the trained SVM. The data for 18 days in the training dataset and for one day in the test dataset were removed owing to the inclusion of extraordinary events such as overseas trips and missing values. Accordingly, we trained the SVM using the lifelog for 172 days and evaluated the generalization performance using the lifelog for 46 days (Figure 1).

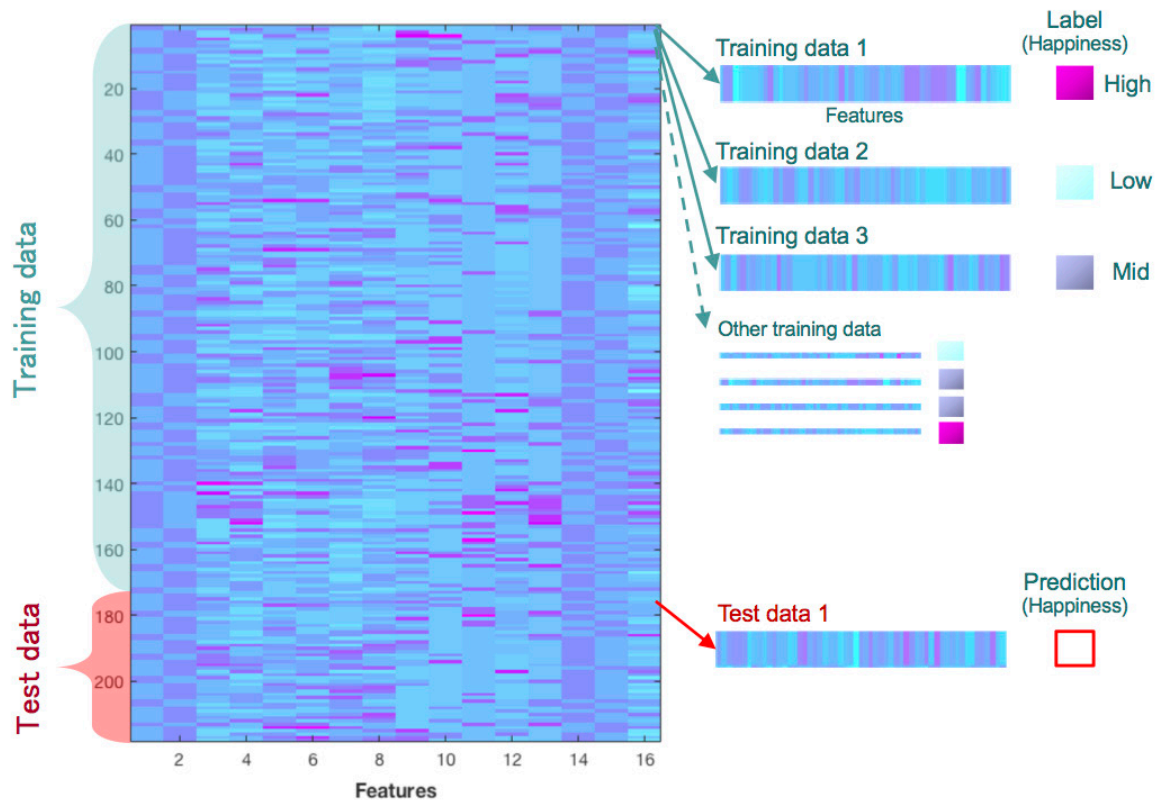


Figure 1. Schematic of our approach for happiness prediction based on the support vector machine.

In the preprocessing phase, we centered and normalized each feature by z-scores. To select the variables, we also computed the Pearson's correlation coefficient between the degree of happiness and each feature, and discarded features with non-significant correlation coefficients (significance level  $\alpha = 0.05$ ) or small effect size ( $-.20 < r < .20$ ; Cohen, 1988). Consequently, we selected 14 features (positive mood, positive event, mindfulness, reward for effort, family time, rest or work day, negative mood, negative event, daytime fatigue, workload, idleness, regret, staying up late the previous night, and minutes of being fairly active) as SVM inputs. The output of the SVM was a three-class categorical variable indicating one of three degrees of happiness: high, middle, or low.

In the training phase, we used the radial basis function kernel to consider a nonlinear mapping from the input to the output variables. The hyperparameters of the SVM ( $C, \gamma$ ) were tuned using a grid search in the range of  $C \in [10^{-2}, 10^2]$  and  $\gamma \in [10^{-5}, 10^5]$ , and the optimal values ( $C = 6.31$  and  $\gamma = 7.94 \cdot 10^{-3}$ ) were determined in order that the classification accuracy was maximized in the 10-fold cross-validation within the training dataset. In the test phase, we evaluated the prediction accuracy of the trained SVM using the lifelog for 46 days in the test dataset.

## Results

The participant recorded the lifelog items and wore the smartwatch with ease almost every day. Further, he reported increased awareness about his lifestyles through the investigation period.

The classification results are shown in Figure 2. The SVM model showed a prediction accuracy of 80% in the high happiness class and 100% in the middle happiness class, but 0% in the low happiness class.

		Predicted Class		
		High	Medium	Low
Actual Class	High	8 80%	2 20%	
	Medium		30 100%	
	Low		6 100%	0 0%

**Overall accuracy: 82.6%**

Figure 2. Confusion matrix for predicted and actual happiness. The rows correspond to the actual class and the columns correspond to the predicted class. The diagonal cells correspond to observations that are correctly classified. The off-diagonal cells correspond to incorrectly classified observations. Both the number of observations and the percentage of the total number of observations are shown in each cell.

### Discussion

Our results showed 82.6% overall accuracy of identifying middle and high happiness using the combination of lifelog data. This suggests that our approach can predict the behaviors that increase individuals' happiness in their daily lives. A concrete application example of our approach is shown in Figure 3. By entering features representing the condition of the day (e.g., negative mood, negative event, workload, etc.; step 1) and features representing an activity that can be performed in the day (e.g., reward for effort, mindfulness activities, active time, etc.; step 2) into the constructed predictor, the predicted happiness corresponding to the input value is outputted (step 3). Because the output of happiness differs according to the input value of the coping strategies (step 4), by referring to such output results, it is possible to use the output as useful information for selecting coping strategies for enhancing happiness.

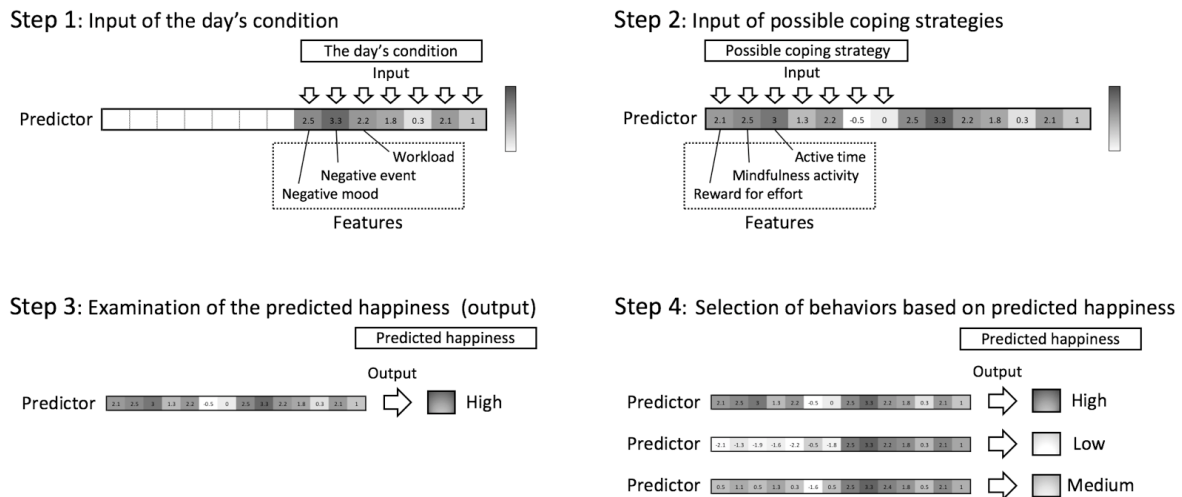


Figure 3. An application example of the support vector machine to the lifelog. Step 1-4 indicate the procedure for happiness prediction.

Furthermore, this approach enables the integration and visualization of multimodal data, such as behavior and physiological data, from multiple sources such as smartphone apps and wearable devices. It can also help increase the awareness of individual lifestyles. This technique could thus facilitate behavioral change to effectively lead to better health, potentially leading to clinical applications for various problems such as depression, stress, and lifestyle diseases.

Contrary to the middle and high happiness classes, the performance in the low happiness class was poor. This could be explained by the limited data for the reported low happiness days: 7% of all training data compared to high happiness (27.9%) and middle happiness (65.1%). In order to construct a more accurate predictor, future studies should include a more sensitive index to measure perceived happiness, examine the conditions in which this technique can be useful (e.g., people with clinical problems), and make performance comparisons by using other kernel functions and learning algorithms.

### Conclusion

Our findings indicate that the application of machine learning to lifelog data can predict daily happiness with high accuracy. Therefore, ours may be a useful behavioral approach to help people effectively improve their well-being. Future studies examining the usability and clinical applicability of this approach would benefit from a larger and more diverse sample size.

### Acknowledgements

We wish to thank C. Uchiumi for her contribution to the project. We also thank the participant for participating in the study.

### Funding

This work was supported by JSPS KAKENHI (grant number JP16H07011 and 18K13323).

## References

- Bruni, L., & Porta, P. L. (2005). *Economics and Happiness: Framing the Analysis*. Oxford: Oxford University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2a ed.). New York: Lawrence Earlbaum Associates.
- Diener, E., & Seligman, M. E. P. (2004). Beyond Money: Toward an Economy of Well-Being. *Psychological Science in the Public Interest*, 5(1), 1-31.
- Marzano, L., Bardill, A., Fields, B., Herd, K., Veale, D., Grey, N., & Moran, P. (2015). The application of mHealth to mental health: Opportunities and challenges. *The Lancet Psychiatry*, 2(10), 942-948.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2014). *e1071: Misc Functions of the Department of Statistics (e1071)*, TU Wien. R package version 1.6-4. Disponível em <http://CRAN.R-project.org/package=e1071>
- R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available in <http://www.R-project.org/>
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. Berlin: Springer.
- Yamamoto, T., & Yoshimoto, J. (2018, Nov). Artificial Intelligence-Based Approaches for Health Behavior Change. *Association for Behavioral and Cognitive Therapies 52nd Annual Convention*, Washington, DC.

Received: 21/09/2018

Revision received: 21/02/2019

Accepted: 27/02/2019

### About the authors:

**Tetsuya Yamamoto** – Ph.D. degree in Human Science from the Waseda University. Completed his post-doctoral training at the University of Pittsburgh and the Hiroshima University. Clinical Psychologist and a Director of Clinical Psychoinformatics Laboratory. His academic research field is linked to vulnerability assessment and intervention for depression. Associate Professor at Graduate School of Technology, Industrial and Social Sciences, Tokushima University. E-mail: [t.yamamoto@tokushima-u.ac.jp](mailto:t.yamamoto@tokushima-u.ac.jp), Orcid <http://orcid.org/0000-0003-4241-532X>

**Junichiro Yoshimoto** – Is an associate professor at the Division of Information Science, Graduate School of Science and Technology, Nara Institute of Science and Technology (NAIST). He received his Ph.D. in engineering from NAIST in 2002 and worked as a researcher at the Okinawa Institute of Science and Technology until 2015. His major is machine learning and its application to data science. Recently, he is engaged in data-driven studies for depression diagnosis and subtyping using multimodal physiological signals. E-mail: [juniti-y@is.naist.jp](mailto:juniti-y@is.naist.jp), Orcid: <http://orcid.org/0000-0001-7995-0321>

**Eric Murillo-Rodríguez** – Is Psychologist and holds a PhD. in Biomedical Sciences, both degrees obtained by the National Autonomous University of Mexico (UNAM). Upon graduation, he completed his postdoctoral training at Harvard Medical School. Currently, he is a Full Professor at School of Medicine, Universidad Anahuac Mayab (Merida, Yucatan, Mexico). He is an active member of several scientific societies as well as editorial board member of various journals. Dr. Murillo-Rodríguez has received outstanding academic achievements such as “Young Investigator Award” given by the World

Federation of Sleep Research Societies in 2003, and subsequently the “Young Investigator Honorable Mention” awarded by the Sleep Research Society in 2004. E-mail: [eric.murillo@anahuac.mx](mailto:eric.murillo@anahuac.mx), Orcid: <http://orcid.org/0000-0001-9307-3783>

**Sergio Machado** – Postdoctorate in Neurophilosophy pela Universidade Federal de Uberlândia (IFILO/UFU) and in Neuroscience of Physical Activity by National Institute of Science and Technology in Translational Medicine (INCT-TM). Postdoctorate, PhD and Master in Mental Health by Instituto de Psiquiatria (IPUB) from Universidade Federal do Rio de Janeiro (UFRJ). PhD in progress in Sports Science at Universidade da Beira Interior (UBI), Portugal. Graduated in Physical Education and in Psychology by Universidade Estácio de Sá (UNESA/RJ). Researcher from Laboratory of Panic and Respiration at IPUB/UFRJ and Permanent Professor of the Postgraduate Program in Sciences of Physical Activity of UNIVERSO, coordinating the Laboratory of Physical Activity Neuroscience (LABNAF). E-mail: [secm80@gmail.com](mailto:secm80@gmail.com), Orcid: <http://orcid.org/0000-0001-8946-8467>