Self-reported Well-being Score Modelling and Prediction: Proof-of-Concept of an Approach based on Linear Dynamic Systems

Xinyang Li[1], Riccardo Poli[1], Gaetano Valenza[2], Enzo Pasquale Scilingo[2] and Luca Citi[1,✉]

***Abstract*— Assessment and recognition of perceived well-being has wide applications in the development of assistive healthcare systems for people with physical and mental disorders. In practical data collection, these systems need to be less intrusive, and respect users' autonomy and willingness as much as possible. As a result, self-reported data are not necessarily available at all times. Conventional classifiers, which usually require feature vectors of a prefixed dimension, are not well suited for this problem. To address the issue of non-uniformly sampled measurements, in this study we propose a method for the modelling and prediction of self-reported well-being scores based on a linear dynamic system. Within the model, we formulate different features as observations, making predictions even in the presence of inconsistent and irregular data. We evaluate the proposed method with synthetic data, as well as real data from two patients diagnosed with cancer. In the latter, self-reported scores from three well-being-related scales were collected over a period of approximately 60 days. Prompted each day, the patients had the choice whether to respond or not. Results show that the proposed model is able to track and predict the patients' perceived well-being dynamics despite the irregularly sampled data.**

I. Introduction

The detection or recognition of perceived well-being is fundamental for the development of assistive healthcare systems for people with physical and mental conditions [1]. Generally speaking, psycho-physilogical state detection can be performed through two types of measurements: physiological/behavioural observations and self-reported data [1]–[3].

Exemplarily, analysis of heart rate variability series from long-term electrocardiogram monitoring has been proven to be an important tool for mood detection and prediction (see, e.g., [3] and references therein). Furthermore, behaviour measurements, including locations, frequency of phone calls, SMSs and/or social interaction, as well as accelerometer data have been used for monitoring affective disorders [4], [5]. Although physiological measurements are crucial for a comprehensive understanding of psychophysiological state-related hormonal, immunologic, or autonomic nervous system responses, behaviour measurements that could be gathered via a smartphone without the need of additional sensors are less obtrusive and are more readily accepted by patients [3], [6], [7]. Importantly, it has been demonstrated that increases in the autocorrelations and cross-correlations of fluctuations of autorecorded emotions often predict stage transitions between a normal state and a state of depression [7]. Scores gathered from structured interviews and psychological questionnaires currently constitute the standard clinical assessment for care in mental health [8]. Although these scores may be biased by subjective evaluations, both from the patient's and clinician's sides, they are typically used as ground truth to build machine learning models for objective mood recognition [2]–[4], [7].

For these reasons, an ubiquitous, unobtrusive and comfortable psycho-physiological data collection strategy is adopted in the Horizon 2020 project "NEVERMIND", within which the research presented in this paper was carried out. Specifically, the NEVERMIND project aims at providing effective self-management tools to help patients suffering from potential depressive symthoms induced by a primary disease (e.g., cancer, miocardial infarction, amptutation, nephropaty). NEVERMIND plans to integrate physiologcial and psychological measurements for mood recognition and prediction in a single system, which will still respect users' autonomy as much as possible. This implies that measurements, particularly self-assessment scores and behavioural data, are gathered in an inconsistent manner, i.e. their frequency and timing may considerably vary according to patient's willingness. Consequently, more flexible machine learning approaches that are able to process irregular data gathered from different sources without prefixed dimensions for features are needed.

To address this major issue, in this study we propose a method for modelling and prediction of self-reported well-being scores, based on a Linear Dynamic System (LDS) [9]. By formulating different features as observations, this type of modeling allows for making predictions in the presence of inconsistent and irregular measurements.

The rest of the paper is organized as follows. Section II introduces mood modelling based on LDS. In Section III, the proposed model is evaluated with simulation and experimental studies. Concluding remarks are given in Section IV.

II. Linear Dynamic System for Self-reported Well-being Score Prediction

A. Problem Formulation

In this work, we assume that the well-being of the user is represented by a scalar value, and adopt a second-order LDS

[1]Xinyang Li, Riccardo Poli, and Luca Citi are with the School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom. Email: {x.li, rpoli, lciti}@essex.ac.uk.

[2]Gaetano Valenza and Enzo Pasquale Scilingo are with the Department of Information Engineering and the Research Centre "E. Piaggio", School of Engineering, University of Pisa, Pisa 56126, Italy. Email: g.valenza@ieee.org, e.scilingo@centropiaggio.unipi.it.

of the form:

$$\mathbf{x}(t) = A\mathbf{x}(t-1) + B\mathbf{u}(t) + \epsilon_{\mathbf{x}}(t) \quad (1)$$
$$\mathbf{y}(t) = C\mathbf{x}(t) + \mu_{\mathbf{y}} + \epsilon_{\mathbf{y}}(t) \quad (2)$$

where

$$\mathbf{x}(t) = \begin{bmatrix} x(t-1) \\ x(t) \end{bmatrix}, \ A = \begin{bmatrix} 0 & 1 \\ a_2 & a_1 \end{bmatrix}, \quad (3)$$

$\mathbf{x}(t) \in \mathbb{R}^2$ is the latent state, $\mathbf{y}(t) \in \mathbb{R}^{n_\mathbf{y}}$ the observation and $\mathbf{u}(t) \in \mathbb{R}^{n_\mathbf{u}}$ the input, while $B \in \mathbb{R}^{2\times n_\mathbf{u}}$ and $C \in \mathbb{R}^{n_\mathbf{y}\times 2}$ are input and observation matrices, respectively. $\mu_\mathbf{y}$ represents a baseline value. We assume the observation noise and the transition noise to be Gaussian, i.e. $\epsilon_\mathbf{x}(t) \sim \mathcal{N}(0, S_\mathbf{x})$ and $\epsilon_\mathbf{y}(t) \sim \mathcal{N}(0, S_\mathbf{y})$. Here, $\mathbf{x}(t)$ represents the self-reported well-being score and $\mathbf{y}(t)$ are the self-reported data based on questionnaires. The input $\mathbf{u}(t)$ is the target intervention we want to study, e.g., whether the patient did any treatment, and/or certain environment factor of interest, e.g., daylight time.

In conventional detection or prediction methods for self-reported data, usually certain self-reported information is used as ground truth (i.e. labels) and other measurements as features, to which standard classifiers are applied. The model in (1) could also be used for self-reported well-being score prediction since, given previous states and observations $\mathbf{x}(t-1)$ and $\mathbf{y}(t-1)$, $\mathbf{x}(t)$ and $\mathbf{y}(t)$ could be predicted. Compared to a classification model, the advantage of the proposed state model lies in the fact that self-reported data are used as observations rather than ground truth, allowing more flexibility in processing inconsistent self-reports. The estimation of the model will be introduced in the following section.

B. Model Estimation

1) Expectation Maximization Method: Given the observation $\mathbf{y}(t)$ and input $\mathbf{u}(t)$, the aim is to estimate the probability distribution of $\mathbf{x}(t)$ and the model parameters, which can be accomplished by the Expectation Maximization (EM) method [9].

The EM method includes two steps, the expectation step (E-step) and the maximization step (M-step). In the E-step, a Kalman filter is applied to the current model $\Theta = \{A, B, C, S_\mathbf{x}, S_\mathbf{y}, \mu_\mathbf{y}\}$ to find the maximum-a-posteriori estimate of the latent state $\mathbf{x}(t)$ and its covariance structure as

$$\mathbf{x}_T = \arg\max_{\mathbf{x}_T} \mathcal{L}(\Theta; \mathbf{x}_T, \mathbf{y}_T, \mathbf{u}_T), \quad (4)$$

where $\mathbf{x}_T$ is the sequence consisting of $\{\mathbf{x}(t),\ t = 1...T\}$, $\mathbf{y}_T$ and $\mathbf{u}_T$ are similarly defined, and $\mathcal{L}(\Theta; \mathbf{x}_T, \mathbf{y}_T, \mathbf{u}_T)$ is logarithm of the joint probability

$$\mathcal{L}(\Theta; \mathbf{x}_T, \mathbf{y}_T, \mathbf{u}_T) = \log \Pr(\mathbf{x}_T, \mathbf{y}_T | \mathbf{u}_T, \Theta). \quad (5)$$

With the estimates yielded by the Kalman filter, the new Θ can be obtained by taking derivatives with respect to Θ of the expected value of $\mathcal{L}(\Theta; \mathbf{x}_T, \mathbf{y}_T, \mathbf{u}_T)$ in the M-step.

2) Sequential EM and Non-uniformly Sampled Observations: In practical studies, measurements may not be collected uniformly and continuously, which makes conventional classifiers less appropriate. In the proposed method, the observations could be used in a sequential manner, which does not require complete measurements.

Assuming that the observations are conditionally independent of each other and $S_\mathbf{y}$ is a diagonal matrix, Kalman filtering could be implemented in a sequential manner in the E-step [10]. In the Kalman filtering, the current state is predicted from the previous state, denoted as $\mathbf{x}^{t-1}(t)$, as in the following

$$\mathbf{x}^{t-1}(t) = A\mathbf{x}^{t-1}(t-1) + B\mathbf{u}(t) \quad (6)$$
$$V^{t-1}(t) = AV^{t-1}(t-1)A^\top + S_\mathbf{x}. \quad (7)$$

From the state prediction $\mathbf{x}^{t-1}(t)$, the sequential Kalman filter uses $\mathbf{y}(t)$ to update the state one dimension at the time, as follows:

$$\mathbf{k}_i(t) = V_{i-1}^{t-1}(t)\mathbf{c}_i^\top(\mathbf{c}_i V_{i-1}^{t-1}(t)\mathbf{c}_i^\top + s_{\mathbf{y}_i})^{-1} \quad (8)$$
$$\mathbf{x}_i^t(t) = \mathbf{x}_{i-1}^t(t) + \mathbf{k}_i(t)(\mathbf{y}_i(t) - \mathbf{c}_i\mathbf{x}_{i-1}^t(t) - \mu_{\mathbf{y}_i}) \quad (9)$$
$$V_i^t(t) = V_{i-1}^t(t) - \mathbf{k}_i(t)\mathbf{c}_i V_{i-1}^t(t) \quad (10)$$

where $i = 1, ..., n_\mathbf{y}$ is the dimension index of $\mathbf{y}(t)$, and $\mu_{\mathbf{y}_i}$ and $s_{\mathbf{y}_i}$ are the i-th element and i-th diagonal element of $\mu_\mathbf{y}$ and $S_\mathbf{y}$, respectively. $\mathbf{k}_i$ and $\mathbf{c}_i$ correspond to the i-th column and row of the Kalman gain matrix and of the observation matrix C, respectively. At each time t, (8) to (10) is only applied to valid observation(s) $\mathbf{y}_i(t)$.

Similarly, in the M-step, the expectation value of $\mathcal{L}(\Theta; \mathbf{x}_T, \mathbf{y}_T, \mathbf{u}_T)$ is calculated only with the valid observations. With the assumptions of Gaussian initialisation and Markov property, (5) in the case of non-uniformly sampled observations could be written as:

$$\begin{aligned}
&\mathcal{L}(\Theta; \mathbf{x}_T, \mathbf{y}_T, \mathbf{u}_T) = \\
&-\frac{1}{2}\sum_{t=1}^{T}\sum_{i}^{n_y}\Big\{\delta_{i,t}\big[\mathbf{y}_i(t) - \mathbf{c}_i\mathbf{x}(t) - \mu_{\mathbf{y}_i}\big]^2 s_{\mathbf{y}_i}^{-1} + \log s_{\mathbf{y}_i}\Big\} \\
&-\frac{1}{2}\sum_{t=1}^{T}\Big\{\big[\mathbf{x}(t) - A\mathbf{x}(t-1) - B\mathbf{u}(t)\big]^\top \\
&\quad S_\mathbf{x}^{-1}\big[\mathbf{x}(t) - A\mathbf{x}(t-1) - B\mathbf{u}(t)\big]\Big\} \\
&-\frac{1}{2}\big[\mathbf{x}(0) - \mu_0\big]^\top S_\mathbf{x}^{-1}\big[\mathbf{x}(0) - \mu_0\big] \\
&-\frac{T}{2}\log|S_\mathbf{x}| - \frac{1}{2}\log|S_0| - \text{const.},
\end{aligned} \quad (11)$$

where $\delta_{i,t}$ is a binary indicator of whether $\mathbf{y}_i$ is available at time t. The process of model estimation using EM is shown in Algorithm 1. Details on the Kalman smoother can be found in [9].

III. Experimental Study

A. Simulation

To evaluate the model performance with non-uniformly sampled observations and different initializations during

Algorithm 1: EM algorithm for LDS estimation with non-uniformly sampled observations.

```
Input: System observation y_T and input u_T;
Output: System parameters Θ and state inference x_T.
begin
    Initialise system parameters as Θ^k|_{k=0};
    repeat
        for t = 1 : T do
            Calculate x^{t-1}(t), V^{t-1}(t) with (6) and (7);
            Set x_0^t(t) = x^{t-1}(t) and V_0^t(t) = V^{t-1}(t);
            for i = 1 : n_y do
                if y_i(t) is valid then
                    Compute x_i^t(t) and V_i^t(t) using (8)–(10);
                else
                    x_i^t(t) = x_{i-1}^t(t), V_i^t(t) = V_{i-1}(t);
                end
            end
            Obtain x^t(t) = x_{n_y}^t(t) and V^t(t) = V_{n_y}(t);
        end
        for t = T : 1 do
            Obtain x^T(t) and V^T(t) with Kalman smoothing;
        end
        Calculate Θ^k taking derivatives with respect to Θ of
        the expected value of L(Θ; x_T, y_T, u_T) in (11);
        k = k+1.
    until convergence;
end
```

model estimation, a simulation study was performed with the following choice of parameters:

$$A = \begin{bmatrix} 0 & 1 \\ -0.8 & 1.25 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ -0.5 \end{bmatrix}, \quad C = \begin{bmatrix} 0.2 & 0.8 \\ 0 & 0.5 \\ 0.05 & 0.5 \end{bmatrix}.$$

The input $\mathbf{u}_T$ ($T = 100$) was a randomly generated binary sequence. The state $\mathbf{x}_T$ and observation $\mathbf{y}_T$ were generated from (6) and (7). To simulate the non-uniformly sampled measurements, for each observation dimension i, a random number N_i between 40 and 60 was generated, and N_i random samples were removed from $\mathbf{y}_T$.

The results obtained by fitting models with 10 random initializations of Θ as well as the ground truth are shown in Fig. 1. The ground truth state $\mathbf{x}(t)$ and observations $\mathbf{y}(t)$ are represented with red circles, the maximum-a-posteriori sequences inferred by the model are drawn with dashed lines. It is apparent that the results obtained with different initializations are quite consistent. That is, our LDS model estimation is robust against non-uniformly sampled observations.

B. Self-reported Well-being Score Modelling

To further validate the proposed method, we applied the LDS model to actual self-reported well-being score modelling. This experimental work is part of the pilot study of the project NEVERMIND. In the experiment, the 3 self-reported well-being scales correspond to the questions "*How are you feeling today?*", "*How was your sleep?*" and "*How was your day?*". Note that the questions used to get the perceived well-being scores are used in NEVERMIND only as an early screening tool, which may then trigger the use of validated clinical questionnaires and interviews. The self-reporting system is embedded in a mobile phone app, enabling users to answer the question by using a sliding scale going from 1 to 6. For all 3 scales, the smaller the value, the better the reporter's condition. Requests seeking user's input are shown once a day. Two female participants, 43-year-old and 58-year-old, both diagnosed with cancer, took part in the experiment for approximately 60 days.

In this study, the treatment log is used as $\mathbf{u}(t)$ and the scales as $\mathbf{y}(t)$. Due to the limited available data-size, the LDS model is estimated using all available data. With the estimated model and inferred states, the future scales are predicted using data from the previous day. The modelling and prediction results for two participants are illustrated in Figure 2, where the abscissas are the days since the beginning of the trial and the ordinates are the scale values. The actual observations are represented with red circle markers, while the predicted observations are represented with blue lines and green bars indicating its standard deviation.

As shown in the figure, the recorded sequences of scales are quite sparse. When there is a sequence of missing observations, the prediction variance grows bigger and bigger, e.g., from day 21 to 32 for participant 2. In other words, the longer the time interval without available observations, the lower the confidence of the prediction made by the model. The time-varying confidence is a clear advantage of the proposed method over conventional classifiers as it helps the final user gauge the reliability of the predictions over time. Yet, the model can reliably capture the mood dynamics if the observations are frequent enough. Generally the predictions are consistent with the true observations. Compared to conventional classification models, where the number of features is usually fixed, the proposed method is more flexible in dealing with irregularly sampled observations/features.

Note that in the current study the prediction is causal but the model identification is conducted in non-causal manner using all the data available. In a practical implementation the model would need to be trained only using previous observations. In our future work, the causal model will be evaluated with comprehensive studies. Moreover, in this study, only self-reported scales are used as observations, whereas additional kinds of measurements, such as heart rate variability and voice prosody, will be included in the future to model the patients mood dynamic.

IV. CONCLUSION

In this work, we proposed an approach for self-reported well-being score modelling and prediction based on the theory of linear dynamic systems. Compared to conventional classification methods, the proposed method is more appropriate for processing data sampled in a non-uniform and irregular manner. As a proof of the methodological concept, we evaluated it with synthetic data as well as through an experimental study including two participants who have been diagnosed with cancer. Importantly, in such an experimental study, self-reported well-being scores were highly irregular,

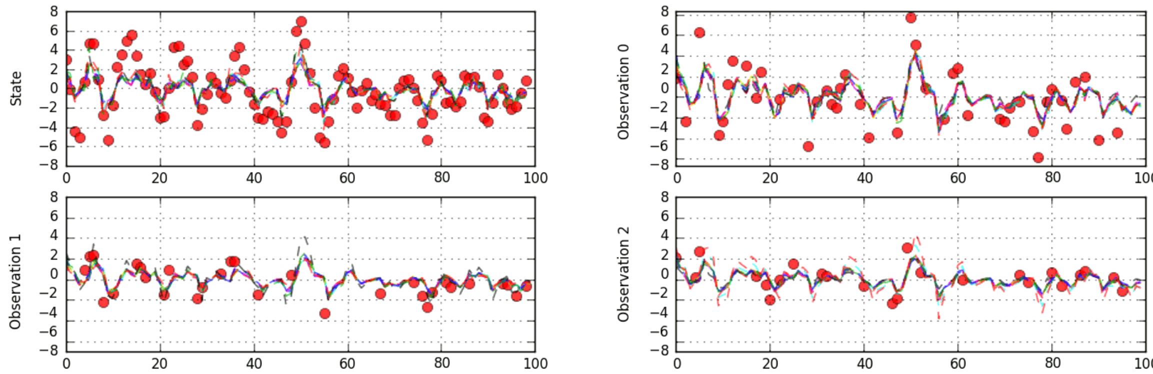
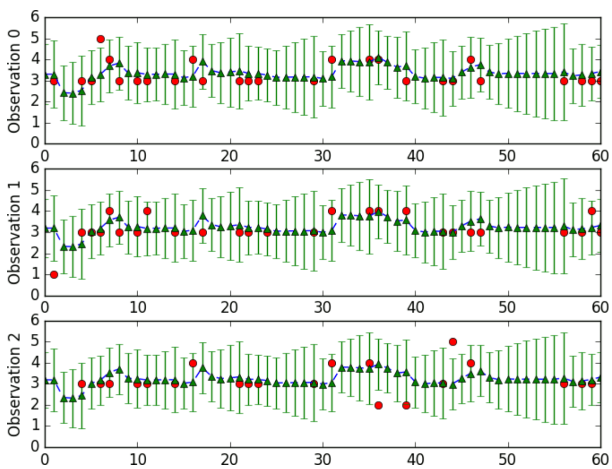
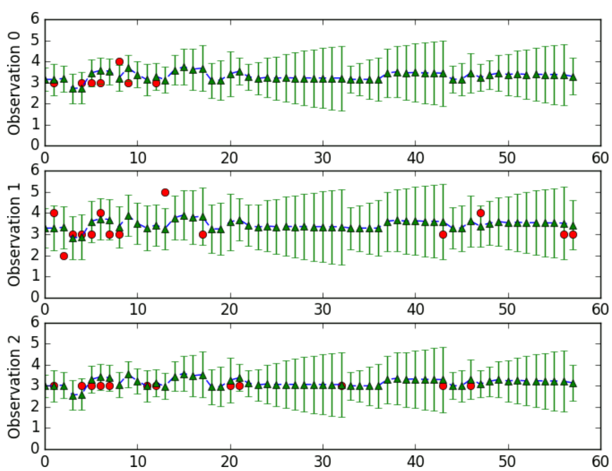


Fig. 1. Simulation study of LDS with 10 random initializations of the EM algorithm for model fitting and state estimation from non-uniformly sampled observations.



(a) Participant 1



(b) Participant 2

Fig. 2. Self-reported well-being score modelling and prediction over the 60-day trial (see text for details).

and observations were unavailable for more than half of the sampling points. Nonetheless, the proposed model could track the participants' score dynamics to a significant extent without requiring the use of a prefixed and constant number of features.

As the current model is implemented in non-causal manner, in our future endeavours a causal version of the model will be evaluated with comprehensive studies, and implemented for real-time prediction.

ACKNOWLEDGEMENT

The research leading to these results has received partial funding from the European Commission Horizon 2020 Programme under grant agreement no. 689691 "NEVERMIND".

REFERENCES

[1] W. Carswell, P. J. McCullagh, *et al.*, "A review of the role of assistive technology for people with dementia in the hours of darkness," *Technology and Health Care*, vol. 17, no. 4, pp. 281–304, 2009.

[2] N. Jaques, S. Taylor, *et al.*, "Predicting students' happiness from physiology, phone, mobility, and behavioral data," in *Affective computing and intelligent interaction (ACII), 2015 international conference on*. IEEE, 2015, pp. 222–228.

[3] G. Valenza, M. Nardelli, *et al.*, "Predicting mood changes in bipolar disorder through heartbeat nonlinear dynamics," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1034–1043, 2016.

[4] C. A. Depp, B. Mausbach, *et al.*, "Mobile interventions for severe mental illness: design and preliminary data from three approaches," *The Journal of nervous and mental disease*, vol. 198, no. 10, p. 715, 2010.

[5] D. A. Axelson, M. A. Bertocci, *et al.*, "Measuring mood and complex behavior in natural environments: use of ecological momentary assessment in pediatric affective disorders," *Journal of child and adolescent psychopharmacology*, vol. 13, no. 3, pp. 253–266, 2003.

[6] B. N. Uchino, J. T. Cacioppo, and J. KKiecolt-Glaser, "The relationship between social support and physiological processes: a review with emphasis on underlying mechanisms and implications for health." *Psychological bulletin*, vol. 119, no. 3, p. 488, 1996.

[7] I. A. van de Leemput, M. Wichers, *et al.*, "Critical slowing down as early warning for the onset and termination of depression," *Proceedings of the National Academy of Sciences*, vol. 111, no. 1, pp. 87–92, 2014.

[8] A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[9] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Technical Report CRG-TR-96-2, University of Totronto, Dept. of Computer Science, Tech. Rep., 1996.

[10] D. Willner, C. B. Chang, and K. P. Dunn, "Kalman filter algorithms for a multi-sensor system," in *1976 IEEE Conference on Decision and Control including the 15th Symposium on Adaptive Processes*, Dec 1976, pp. 570–574.