Internet Interventions xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.elsevier.com/locate/invent

# Predicting short term mood developments among depressed patients using adherence and ecological momentary assessment data

Adam Mikus[a], Mark Hoogendoorn[a,*], Artur Rocha[b], Joao Gama[c], Jeroen Ruwaard[d], Heleen Riper[d]

[a] Vrije Universiteit Amsterdam, Department of Computer Science, De Boelelaan 1081, Amsterdam 1081 HV, The Netherlands
[b] Centre for Information Systems and Computer Graphics, INESC TEC, Porto, Portugal
[c] University of Porto, Laboratory of Artificial Intelligence and Decision Support, Porto, Portugal
[d] Vrije Universiteit Amsterdam, Department of Clinical Psychology, De Boelelaan 1081, Amsterdam 1081 HV, The Netherlands

ARTICLE INFO

*Keywords:*
Depression
Machine learning
Short term mood
Prediction

ABSTRACT

Technology driven interventions provide us with an increasing amount of fine-grained data about the patient. This data includes regular ecological momentary assessments (EMA) but also response times to EMA questions by a user. When observing this data, we see a huge variation between the patterns exhibited by different patients. Some are more stable while others vary a lot over time. This poses a challenging problem for the domain of artificial intelligence and makes on wondering whether it is possible to predict the future mental state of a patient using the data that is available. In the end, these predictions could potentially contribute to interventions that tailor the feedback to the user on a daily basis, for example by warning a user that a fall-back might be expected during the next days, or by applying a strategy to prevent the fall-back from occurring in the first place.

In this work, we focus on short term mood prediction by considering the adherence and usage data as an additional predictor. We apply recurrent neural networks to handle the temporal aspects best and try to explore whether individual, group level, or one single predictive model provides the highest predictive performance (measured using the root mean squared error (RMSE)). We use data collected from patients from five countries who used the ICT4Depression/MoodBuster platform in the context of the EU E-COMPARED project. In total, we used the data from 143 patients (with between 9 and 425 days of EMA data) who were diagnosed with a major depressive disorder according to DSM-IV.

Results show that we can make predictions of short term mood change quite accurate (ranging between 0.065 and 0.11). The past EMA mood ratings proved to be the most influential while adherence and usage data did not improve prediction accuracy. In general, group level predictions proved to be the most promising, however differences were not significant.

Short term mood prediction remains a difficult task, but from this research we can conclude that sophisticated machine learning algorithms/setups can result in accurate performance. For future work, we want to use more data from the mobile phone to improve predictive performance of short term mood.

## 1. Introduction

Depression is a highly prevalent disorder associated with a huge loss of quality of life, increased mortality rates high levels of service cost. Earlier research has estimated the cost of depression at 177 million euros per year per 1 million inhabitants for major depression on top of 147 million euros per year for minor depression (Cuijpers et al., 2007). Depression is currently the fourth disorder worldwide in terms of disease burden (Üstün et al., 2004). A lot of developments in treatments for depression can be seen in the last decade, where a shift is taking place from the more traditional face-to-face counseling to self-help therapies or blended care settings (see e.g. Kooistra et al., 2016; Riper et al., 2010). These changes have been driven by advancements in technologies in society: Internet and mobile phones are widely available and enable more technologically supported forms of interventions. Also better and more fine-grained ways of measuring the state of patients have resulted, such as EMA (for Ecological Momentary Assessment, see e.g. Shiffman et al., 2008; Asselbergs et al., 2016): measurements that assess the mental state of a patient in context and over time, often via questions posed to the patient on the mobile phone.

EMA questions can be useful for a therapist or researcher to understand how a patient is progressing. The purpose can range from

* Corresponding author.
*E-mail addresses:* a.mikus@student.vu.nl (A. Mikus), m.hoogendoorn@vu.nl (M. Hoogendoorn).

http://dx.doi.org/10.1016/j.invent.2017.10.001
Received 6 July 2017; Received in revised form 29 September 2017; Accepted 1 October 2017

gaining a fine grained insight in the manifestation of depression before or during the start of the treatment towards fluctuations in mood and behaviors during treatment. In addition however, it also makes one wonder whether it is possible to extract patterns that allow one to create forecasts of the mental state for individual patients. Previous research has shown that prediction of mood is difficult due to large individual differences and a limited predictive value of previous measurements for the future developments (see e.g. van Breda et al., 2016b; Becker et al., 2016; van Breda et al., 2016a). There are however more advanced techniques available from the domain of artificial intelligence that might result in models that are better able to predict future developments of a patient. In addition, usage data (from both the EMA/therapeutic app and the web modules) has hardly been exploited to improve predictions.

In this paper, we will focus on predicting short term mood for the EMA data that has been collected within the E-COMPARED project. E-COMPARED stands for European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual. The protocol of the trial conducted in the project can be found in Kleiboer et al. (2016). We aim to predict the value reported for the mood of the patient on the next day and will extend previous work van Breda et al. (2016b) by (1) using the response times to EMA requests and usage data of the patient as an additional predictor, and (2) applying more sophisticated machine learning techniques in the form of recurrent neural networks. The EMA dataset we use contains data from 143 patients with at least 9 up to 425 days of EMA data. We evaluate the approaches by comparing the predictions of the model with the reported EMA values.

This paper is organized as follows. First, we will describe how the data has been collected and the setting in which we apply our machine learning algorithms. We will then describe the application of the machine learning algorithms themselves followed by the results. Finally, we will draw conclusions.

## 2. Dataset description and initial exploration

We use the EMA data collected from patients from five countries who used the MoodBuster platform in the context of the EU E-COMPARED project. In total we used the data from 143 patients (with between 9 and 425 days of EMA data) who were diagnosed with a major depressive disorder according to DSM-IV. Table 1 provides an overview of the EMA questions that were posed.

Next to the EMA data we also exploit the log table of the MoodBuster system which indicates the response times of patients, i.e. the time between the system requesting a mood rating from the user and the moment when the actual input is received. In addition, it stores a lot more information about the behavior of the user, including module completion and the amount of time spent on them, messages being exchanged (when, their frequency and the number of characters), the number of web sessions, and the number of pages passed in the module. Hence, a wealth of data.

Before diving into the details of the method to forecast the EMA rating of the next day, we performed an initial correlation analysis (using the Pearson metric) between the EMA rating of the previous day and the next day. We performed this on a per patient basis as these correlations differ greatly per patient. The results show that from 27% to 45,45% of the patients show relevant (i.e. correlation coefficient above 0.4 or below −0.4) correlations with the specific EMA values of the previous day and today's mood. For most of the patients, the highest correlation values occurred between the mood of previous day and the mood of today. These correlations are mainly positive, except in the case of worrying - where the category is rated in the opposite direction (the higher the worse).

**Table 1**
The EMA measures that are present in the dataset.

| Abbreviation | EMA question |
|---|---|
| Mood | How is your mood right now? |
| Worry | How much do you worry about things at the moment? |
| Self-esteem | How good do you feel about yourself right now? |
| Sleep | How did you sleep tonight? |
| Activities done | To what extent have you carried out enjoyable activities today? |
| Enjoyed activities | How much have you enjoyed the days activities? |
| Social contact | How much have you been involved in social interactions today? |

We have also explored the relationship between the response times and EMA ratings. To be more precise, we computed the correlations between the following values:

1. Today's response time and today's mood
2. Yesterday's response time and today's mood
3. Today's mood and tomorrow's response time

The analysis resulted in 11 patients who showed relevant correlation values. To validate the results of the analysis, simple linear regressions have been performed between the time series of mood and its corresponding response times. The regressive modeling resulted in 6 patients where the p-values were less than 0.05. Now that we gained a bit more insight into the raw data, let us move to our machine learning approach.

## 3. Machine learning approach

This section addresses the machine learning approach. First, we describe the features used to feed the recurrent neural network and how they are computed. These are based on the raw data we have just explained. Then we explain the machine learning approach and the different settings that we tried.

### 3.1. Features

As said, we aim at predicting the reported mood value of the next day based on the measurements and response times of the previous day (s). To get most out of our data, we have developed dedicated features that summarize the patients rating and behavior during previous days. These are shown in Table 2. We distinguish between base features and an extended set of features which exploits data about the response times, the web sessions and messages exchanged.

We obtain these values for all patients from the MoodBuster database and take values on a per day basis. The intensity of the EMA questions changes according to the trial protocol (Kleiboer et al., 2016) (more ratings in the start and end weeks of the treatment). Ratings are triggered randomly in the following time intervals: 1) 9 to 10 am; 2) 8 to 9 pm.

We normalized values using a scale between 0 and 1. For category type features (e.g. weekdays) we use a binary encoding, meaning that we create a feature per value and express a 1 when the values holds and a 0 otherwise. We obtain a lot of missing values as some ratings are measured less frequent and in addition, patients do not always provide ratings.

In the case of EMA mood ratings, we interpolate values if we have a gap of at most three days by considering the previous and next value and interpolating in a linear fashion. Otherwise we use the mean value for that feature of the patient over the days we do have values for.

This way, the regular input set does not contain any missing values. In the case of extended input set, due to the difficult nature of missing value imputation, binary indicators have been used as dummy variables to indicate whether the value was missing or not. Missing values have been filled with 0 (by the nature of the collected data 0 is not an option to be given by the patient).