Internet Interventions xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Internet Interventions

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

Predicting therapy success for treatment as usual and blended treatment in the domain of depression

Ward van Breda[a,*], Vincent Bremer[b], Dennis Becker[b], Mark Hoogendoorn[a], Burkhardt Funk[b], Jeroen Ruwaard[c], Heleen Riper[c]

[a] Department of Computer Science, VU University, Amsterdam, The Netherlands
[b] Institute of Information Systems, Leuphana University, Luneburg, Germany
[c] Department of Clinical Psychology, VU University, Amsterdam, The Netherlands

ARTICLE INFO

Keywords:
Prediction
Therapy success
E-health
Depression
Classification

ABSTRACT

In this paper, we explore the potential of predicting therapy success for patients in mental health care. Such predictions can eventually improve the process of matching effective therapy types to individuals. In the EU project E-COMPARED, a variety of information is gathered about patients suffering from depression. We use this data, where 276 patients received treatment as usual and 227 received blended treatment, to investigate to what extent we are able to predict therapy success. We utilize different encoding strategies for preprocessing, varying feature selection techniques, and different statistical procedures for this purpose. Significant predictive power is found with average AUC values up to 0.7628 for treatment as usual and 0.7765 for blended treatment. Adding daily assessment data for blended treatment does currently not add predictive accuracy. Cost effectiveness analysis is needed to determine the added potential for real-world applications.

1. Introduction

Nowadays, individuals that suffer from mental health problems face a range of different treatment types to choose from such as treatment as usual (TAU) and blended treatment (BT). Making decisions regarding suitable treatment types on an individual level is a challenging problem. In online interventions, an increasing amount of data is collected including socio-demographic aspects, recurring symptomatic questionnaires, Ecological Momentary Assessments (EMA), and outcome related factors. Machine learning techniques can support the decision-making process of therapists and practitioners by extracting valuable information from this wealth of data and providing crucial input regarding therapy success and symptom development. In this context computerized systems can potentially improve the care and accomplishment of practitioners in this field (Garg et al., 2005).

Therefore, we investigate possibilities of predicting therapy success using data from the EU funded project *E-COMPARED*, in which the effectiveness of TAU and BT are investigated. We define therapy success based on the Patient Health Questionnaire-9 (PHQ-9, (Kroenke et al., 2001)). This is an internationally acknowledged and validated questionnaire that measures the presence and extent of depression. We utilize baseline measures of individuals for TAU and BT and further evaluate if involving EMA data can lead to increased prediction performance in BT. For this purpose, we take advantage of multiple statistical models, preprocessing steps, and feature selection methods. In the case of accurate predictions at intake, more tailored therapy types can potentially be offered, more effective treatment recommendation can be provided, enhanced decision-support tools can be developed, and even health care costs can eventually be reduced.

Related research is scarce due to the fact that the field of predictive modeling in e-mental health is still young. Good examples of relevant work in the context of this paper are Both et al. (2009), where the design and analysis of an ambient intelligent system is described that offers support during therapy of patients that recover from uni-polar depression; and in Duppong Hurley et al. (2015), where changes in therapeutic alliance are investigated to have power to predict therapy outcome in youth with a disruptive-behavior diagnosis.

In the following sections, we introduce the data we utilize, details of the experimental setup and illustrate our results. We finalize the paper by discussing the outcomes, limitations, and future research opportunities.

2. Data

The data utilized for our approach consists of several sub-datasets, each of which represents a certain area of information. For each patient, demographic information is gathered. Additionally, information about current treatment is collected such as current medication usage and

* Corresponding author.

http://dx.doi.org/10.1016/j.invent.2017.08.003
Received 4 July 2017; Received in revised form 16 August 2017; Accepted 16 August 2017

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?

psychotic symptoms. Furthermore, validated questionnaires are used to measure psychiatric disorders (M.I.N.I.; Sheehan et al., 1998), depressive symptomatology (QIDS-SR16; Trivedi et al., 2004), severity of depression (PHQ-9 Kroenke et al. (2001)), and generic health status (EQ-5D-5L; Herdman et al., 2011). In addition, information about treatment preferences is available. The TAU and BT have a duration of three months. At the start of the therapy, a baseline measurement is conducted covering all of the measurements described above and follow-up severity of depression (PHQ-9) measurements are repeated at three months, six months, and twelve months.

In BT, additional EMA information is gathered on a daily basis, which is one of the big difference with the usual treatment a patient is given in TAU. Specifically, patients can enter EMA data when reminded by the application, or whenever the patient wants. As displayed in Table 1, the EMA data consists of seven variables such as mood level, worry level, etc, which are measured on an interval scale of [1, 10] (for more information about EMA data and EMA data collection, see Shiffman et al., 2008).

There are 780 unique patients suffering from major depressive disorder (DSM-IV) in the dataset for which the intake data is complete. As the data is currently being processed within the project, the number of usable patients declines over time as the data of the follow ups has not been processed yet. Specifically, at three-months 555 patients, at six-months 215 patients, and at twelve-months 116 patients present for the purpose of predicting therapy success.

The total number of raw features collected for each patient is 119, which are mostly categorical in nature. The BT dataset consists of additional EMA data such as the day number, date, time, schedule, and rating related to the different EMA questions. Because of the format of the EMA data, data transformation is required to merge it to the other sub-datasets.

Exploring the sub-datasets for missing data reveals that data regarding current treatment, psychiatric disorders, and depressive symptomatology contain missing values (due to conditional follow-up questions that were not relevant in some situations). Furthermore, there are missing values in the data regarding severity of depression at the three months measurement (five patients), six month measurement (three patients), and twelve month measurement (82 patients). Moreover, missing values are found in the data about treatment preferences, especially about willingness to carry a smart phone.

3. Methods

3.1. Data selection

Because we have two therapy types with different patient for each therapy type, we divide the data in the TAU dataset and the BT dataset. Furthermore, due to data inconsistencies, we need to exclude information. In the TAU dataset, one participant was excluded because the age of the participant is unknown, resulting in 276 patients in this dataset. In the BT dataset, 56 patients were excluded due to merging procedures with the EMA dataset and an insufficient number of observations in EMA (less than seven days). Therefore, we are left with 227 patients in the BT dataset.

Table 2
The number of successful and unsuccessful therapy effects in the TAU and BT datasets.

Treatment	Nr patients	Unsuccessful	Successful
TAU	276	231	45
BT	227	169	58

3.2. Target feature

Because we seek to have as much data as possible to increase chances of accurate predictions, we choose to compare the three-month follow-up measurement with the baseline measurement. We use this difference in measurement to engineer the target feature that represents therapy success for each patient. The PHQ-9 consists of nine questions which can be scored from zero to three. The total of the scores is used to express the extent of current depressive symptoms. Based on the rationale found in McMillan et al. (2010), we define the therapy to be a success in the following cases:

1. $PHQ-9_{post} <= 9$ and $\Delta PHQ-9_{pre\_post} > 50\%$
2. $PHQ-9_{pre} >= 5$ and $PHQ-9_{post} <= 4$.

As displayed in Table 2, this results in 45 successful therapy effects in the TAU dataset and 58 successful therapy effects in the BT data.

3.3. Preprocessing categorical features

A lot of the questionnaire outcomes are categorical. We have a total of 107 categorical features (besides the 12 non-categorical features). These categorical features can be handled in different ways.

One strategy to transform the data is to use dummy encoding, which generates binary variables that indicate the presence or absence of specific categorical values (see e.g. Hardy, 1993). This procedure considers each possible answer and also takes into account the presence of missing values as separate variables. The downside, however, is the increasing amount of features in the datasets. Using the approach sketched above results in 408 dummy features for the TAU dataset and 407 features for the BT dataset. This preprocessing strategy is referred to as the binary encoding condition.

An alternative approach we consider to to transform the categorical values into continuous features. This approach is reasonable because many of the questionnaires include responses with a certain order such as time related questions (e.g., one to six months, etc), extent related questions (e.g. no pain, slight pain, etc.), and confirmation questions (yes, no). However, not all features are compatible. From the 107 categorical variables, we create 45 and 47 continuous features for the TAU and BT dataset respectively and use a binary encoding for the remaining features (resulting in 247 binary features for TAU and 225 for BT) This preprocessing strategy is referred to as the mixed encoding condition.

3.4. Preprocessing the EMA features

To explore if adding EMA data results in increased predictive accuracy, we choose to use the first seven days of EMA data in the analysis. If we are able to predict the success better after a week of therapy this might still be valuable. Possibly, the amount of bad mood scores or mood patterns during the day can help describe the probability of future therapy success. Because the EMA data in its raw form is not compatible with the BT dataset, we transform the answers of each of the seven questions into the sum, mean, slope, standard deviation, minimum, and maximum of the values over the seven days. We