Idiographic prediction of physical pain and loneliness

true

16. May 2022

Contents

Study Information

Title

Idiographic prediction of physical pain and loneliness

Description

A longstanding goal of psychology is to describe (e.g., Titchener, 1898), predict (e.g., Meehl, 1954), and explain (e.g., Fodor, 1968) the things people do and experience. Despite this persisting emphasis, accurately predicting future socioemotional behaviors and experiences remains elusive. Indeed, most of the existent research on prediction examines broad life outcomes (e.g., Beck & Jackson, 2022; Joel et al., 2020). While such broad life outcomes result from accumulating behaviors and experiences (e.g., Hampson, Goldberg, Vogt, & Dubanoski, 2007), how predictable those behaviors are is largely unknown.

However, one recent investigation (Beck & Jackson, in press) found that personalized, machine-learning based prediction models of loneliness, procrastination, studying, interacting with others, arguing with others, feeling sick, and feeling tired performed well on hold-out test data. In other words, despite more than a century of difficulty in predicting momentary behaviors and experiences from trait personality measures or situation-based labratory experiments, more recent research using ambulatory assessment and mobile sensing provide promise for overcoming such challenges.

The present study aims to extend the findings of Beck & Jackson (in press) in a sample of older adults in the domains of physical pain and loneliness. Physical pain and loneliness have been demonstrated to have a substantial monetary burden, with, for example, loneliness an social isolation accounting for $6.7 billion in additional Medicare spending annually and that the annual cost of chronic pain in the United States is as high as $635 billion a year. Thus, idenitifying the unique antecedents to these experiences is a critical first step in determining pathways for reducing loneliness and physical pain among older adults.

Hypotheses

Rather than hypotheses, the present study addresses three key research questions.

1. To what extent can idiographic predictive models be built? In other words, what is the out-of-sample prediction error for each person?

2. Are there individual differences in out-of-sample prediction error across people?
   2a. Are some people not able to be predicted while others can be predicted almost perfectly?
   2b. Do these differences "cluster" according to the types of features used? In other words, for some people, are situations better predictors of physical pain and loneliness while for others psychological characteristics like personality and affect? Are still others best predicted by both?

3. How much heterogeneity is there in the top predictors of physical pain and loneliness across people?

Design Plan

Study type

Observational Study. Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, natural experiments, and regression discontinuity designs.

Blinding

Although the data from these study come from the first wave of a longitudinal study.

Study design

The present study is a sub-project of the broader Einstein Aging Study, which is an ongoing longitudinal study of the aging brain. The EAS began in 1980 and has enrolled more than 2,600 participants since then. Data are available through application at http://www.einstein.yu.edu/departments/neurology/clinical-research-program/eas/data-sharing.aspx.

The sub-project uses a subset of EAS participants as well as newly recruited participants. In addition to regular EAS follow-ups, these participants completed additional trait and ESM measures of personality and cognition using a measurement burst design. Essentially, this means that although participants are continuously recruited, follow-ups occur approximately one year after each completion. At each burst, participants complete 14 days of ESM surveys, with 6 beeps a day (i.e. max assessments = 84).

Randomization

No randomization was involved in this study.

Sampling Plan

Existing data

Registration prior to accessing the data. As of the date of submission, the data exist, but have not been accessed by you or your collaborators. Commonly, this includes data that has been collected by another researcher or institution.

Explanation of existing data

These data were collected as a subproject of the Einstein Study of Aging by individuals not directly involved in the development of and analysis of the proposed project who have allowed us access to their data. They have previously published on subsets of these data and are currently being used to examine general diurnal trends, but the authors of this proposal and study maintainers remind blind to those results and the questions proposed in this registration have been evaluated for overlap.

Data collection procedures

A total of 14,137 potential participants were contacted via phone. From these calls, 597 participants completed successful phone screens, 517 of which were deemed eligible for participation and had in-home visits scheduled. Of these, 52 participants dropped out of the study before meeting, leaving 465. Of these 465, 150 were EAS participants and 315 were newly recruited participants. Among these, 66 EAS participants and 258 new participants were included in the final sample of the study for a total of 324 participants.

Sample size

Approximately 300 participants have completed Burst 1 and will be our target population.

Sample size rationale

Between-person sample sizes are based on the number of individuals who have completed Burst 1 to date (beg. 2017) and limited by aims of the parent grant and project.

Within-person sample sizes are constrained by the 15 day x 5 beeps/day study design of the ESM study.

Stopping rule

The stopping rule is determined by the last date of data cleaning by the research team who manages the data (last collection, January 2020).

Variables

A list of all variables, both measured and created, are included in a codebook attached with this preregistration.

Measured variables

Measured variables include:
- 15 psychological indicators
- 19 binary situation indicators
- 12 behavior indicators - 2 outcome indicators

In addition, we have time stamps for all observations, which will be transformed into time indicators as detailed in the Transformations section.

Indices

In addition, we will include a set of temporal variables, constructed from time stamps:

- linear trends (1 variable)
- quadratic trends (1 variable)
- cubic trends (1 variable)
- sinusoidal and cosinusoidal utradian and circadian cycles (4 total variables)
- time of day dummy codes (4: morning, midday, evening, night) (4 total variables)
- day of the week dummy codes (7 variables: Sunday - Saturday, yes/no)

Together, this results in a total of 70 indicators.

Analysis Plan

The present study will test three methods of machine learning classification models, some of which have been used for idiographic prediction in other studies (Fisher & Soyster, 2019; Kaiser & Butter, 2020):
1. Elastic Net Regression (Friedman, Hastie, & Tibshirani, 2010)
2. The Best Items Scale that is Cross-validated, Correlation-weighted, Informative and Transparent [BISCWIT; Elleman, McDougald, Condon, & Revelle (2020)]

In each of these methods, we will use 10-fold cross-validation on the training set, which will be comprised of the the first 75% of the time series, and hold out the remaining 25% of the data set for the test set.

Because we have a large number of indicators to test, each of the methods used have variable selection features and, in some instances, other methods for reducing overfitting, as detailed below. To both reduce the number of indicators used in each test and to test which group of indicators are the most predictive of physical pain and physical pain, we will also test these in several sets:

- Behavioral indicators (activities) (12)
- Psychological indicators (personality + affect) (15)
- Situation indicators (binary, location and other people around) (19)
- Full set (psychological + behavioral + situations) (46)

We will additionally test each of these with and without the 18 timing indicators (detailed more below in the Transformations section).

Each method will use rolling origin validation (initial set to n/3, assess = 5, skip = 1, cumulative = T).

1. Elastic Net Regression

Elastic Net Regression proceeds from the observation that typical OLS-based regression minimizes bias but may have great variance. Using L1 (Ridge) and L2 (LASSO) approaches, which apply penalties to model estimates, elastic net attempts to balance the trade-off between bias and variance by choosing the best penalties that minimize an information criterion or prediction error. Together, these both shrink coefficients and help with feature selection by forcing some of the coefficients to be zero. Because there are a large number of values the regularization parameter λ can take on, the typical solution is to use a method like k-fold cross-validation to test a number of λ values and choose the one with the one that matches a criterion like minimizing prediction error.

In the present study, we will use the `glmnet()` function from the `glmnet` package through the `tidymodels` package with rolling origin validation folds, and use mean absolute error (mae) to tune the hyperparameters.

2. BISCWIT

The Best Items Scale that is Cross-validated, Correlation-weighted, Informative and Transparent is a correlation-based machine learning technique. The technique proceeds as follows:

- pairwise correlations between predictors and outcome(s) using cross-validation with k-folds
- items with the highest cross-validated correlations are retained
- retained items are correlation-weighted in a sum score

In the present study, we will use the `bestScales()` function from the `psych` package to create the models and will use the `scoreWtd()` function to extract the correlation weighted scores.

3. Random Forest

Random forest models are a variant of decision tree classification algorithms that additionally draw on bagging (emsemble) methods. The name itself hints at how it different from classic decision trees; we have a forest instead of a tree. In so doing, random forest models draw upon a large number of decision trees of varying depth that are then aggregated. The random part comes from two sources. First, each tree in the forest in trained on a data set drawn with replacement from the training dataset (hence also: random) as part of the bagging procedure. Observations left out of each model are termed out-of-bag observations (and collectively, the out-of-bag dataset). These are used to evaulate the performance of the tree. Second, the features used in each tree are also randomly drawn from the full of features. Each of these trees, based on their data generates a prediction given new data. The final prediction is based off the ensemble of trees – that is, the decision made by a majority of the trees. Importantly, because of the random feature selection part of the procedure, random forest can also provide estimates of variable importance, indicating which features are critical to making a less error-prone classification.

Because random forest using bagging (i.e. bootstrapping + aggregation), we will have to perform a series of steps that make bootstrapping appropriate with time series data: differencing, Box-Cox transformations, and time-delay embedding. Essentially, differencing and Box-Cox transformations stabilize the mean and variance, respectively, to make the time series stationary (Priestley, 1988), and time-delay embedding quite literally embeds the sequence of the time series into the predictor variables, which in effect preserves the order of the times series (Von Oertzen & Boker, 2010). We can easily backtransform forecasts to their original scale.

In the present study, we will use `train()` the `caret` package in `R` to build the random forest models (`method = "rf"`) using 10-fold cross validation. We will use the `varImp()` function to extract variable importance.

4. Summarizing Results

The results from each of these methods will be used to answer Questions 1-3. For each, we will test each combination of the features listed above, both with and without timing effects. Where possible, we will also attempt to use the `tidymodels` package in `R` as a wrapper for functions in the `glmnet` and `caret` packages in the hopes of making syntax more accessible.

Questions 1 and 2 are highly related. Question 1 is more concerned with the aggregated prediction error across people – in other words, can we predict generally predict when participants are lonely or procrastinating above chance levels. Assuming there is some degree prediction achieved across people, Question 2 then asks about the range of individual differences in this. For example, are there some people we can't predict but others we can?

Both questions 1 and 2 will draw on two forms of predictive tests: mean absolute error and area under the curve (AUC). See the Inference Criteria for more details.

Finally, Question 3 will examine which variables contribute the most to the prediction models for each person separately. For each method, we will extract the top 10 variables for each person, and run a number of descriptive tests (plotting, percentages, etc.) to examine the range of individual differences in what indicators were contributing to prediction across people. To the extent that these indicators differ widely across people, this highlights the experiential heterogeneity in how loneliness and physical pain arise in these participants lives, which has implication for using nomothetic prediction of such experiences.

Statistical models

The statistical models used in the present study represent three machine learning classification methods – elastic net GLM, BISWIT, and Random Forest – each of which were described in detail above.

Transformations

The main data transformations are of the date/time variable to a number of indicators.

Specifically, we will perform the following steps:
1. Create beep time of day dummy codes (morning, midday, evening, night) from time stamp data
2. Create day of the week dummy codes (7) from the date/time object using `lubridate::wday()`
3. Create a cumulative time variable (in hours) from first beep (not used in analyses)
4. Standardize the cumulative time variable to capture the linear trend
5. Use the linear trend variable to create quadratic and cubic trends
6. Use the cumulative time variable to create 1 and 2 period sine and cosine functions across each 24 period (e.g., 2 period sine = $sin(\frac{2\pi}{12} * cumtime)$ and 1 period sine = $sin(\frac{2\pi}{24} * cumtime)$)

We will also transform our outcome variables – loneliness and physical pain (0/1) – to lagged outcomes, such that the features will then be predicting next time point loneliness or physical pain (t + 1) from current time point features (t).

Inference criteria

Inference criterion will be based on classification error and area under the receiver operating characteristic curve. Classification error is a simple estimate of the percentage of the test sample that was correctly classified by the model. In addition, the ROC or AUC will capture the trade-off between sensitivity and specificity across a threshold. In the the present study, we will use an AUC threshold of .5, which indicates binary classification at chance levels. ROC plots 1 - specificity (i.e. false positive rate: false positives / (false positives + true negatives)) against sensitivity (i.e. true positive rate: true positives / (true positives + false positives)).

Data exclusion

Participants with too little data at an ESM wave (N < 40) will be excluded.

Missing data

Missing data will be treated at the obsevation level, using all available information for each participant.

Exploratory analyses (optional)

Arguably, all analyses in the present study are exploratory as we make no formal directional hypotheses.

References

Beck, E. D., & Jackson, J. J. (2022). A mega-analysis of personality prediction: Robustness and boundary conditions. *Journal of Personality and Social Psychology*, *122*(3), 523–553. Retrieved from https://doi.org/10.1037/pspp0000386

Beck, E. D., & Jackson, J. J. (in press). Personalized behavior prediction: An idiographic person-situation test. *Psychological Science*.

Elleman, L. G., McDougald, S. K., Condon, D. M., & Revelle, W. (2020). That takes the BISCUIT: Predictive accuracy and parsimony of four statistical learning techniques in personality data, with data missingness conditions. *European Journal of Psychological Assessment*, *36*(6), 948.

Fisher, A. J., & Soyster, P. (2019). *Generating accurate personalized predictions of future behavior: A smoking exemplar.*

Fodor, J. A. (1968). *Psychological explanation: An introduction to the philosophy of psychology.*

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22.

Hampson, S. E., Goldberg, L. R., Vogt, T. M., & Dubanoski, J. P. (2007). Mechanisms by which childhood personality traits influence adult health status: Educational attainment and healthy behaviors. *Health Psychology*, *26*(1), 121.

Joel, S., Eastwick, P. W., Allison, C. J., Arriaga, X. B., Baker, Z. G., Bar-Kalifa, E., … others. (2020). Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies. *Proceedings of the National Academy of Sciences*, *117*(32), 19061–19071.

Kaiser, T., & Butter, B. (2020). *Time-lagged prediction of food craving with elastic net regression and BISCWIT: Prediction performance of qualitative distinct predictor types.*

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.*

Priestley, M. B. (1988). Non-linear and non-stationary time series analysis. *London: Academic Press.*

Titchener, E. B. (1898). The postulates of a structural psychology. *The Philosophical Review*, *7*(5), 449–465.

Von Oertzen, T., & Boker, S. M. (2010). Time delay embedding increases estimation precision of models of intraindividual variability. *Psychometrika*, *75*(1), 158–175.