Digital Footprints of Sensation Seeking: A Traditional concept in the Big Data Era

Big Data in Psychology 2018, Trier

Ramona Schoedel 1, Quay Au 2, Sarah Theres Völkel 3, Florian Lehmann 3, Daniela Becker 3, Markus Bühner 1, Bernd Bischl 2, Heinrich Hussmann 3, & Clemens Stachl 1

1 Department of Psychology, Psychological Methods and Assessment, Ludwig-Maximilians-Universität München
2 Department for Statistics, Computational Statistics, Ludwig-Maximilians-Universität München
3 Institute of Informatics, Ludwig-Maximilians-Universität München
Automated Trait Recognition

Prediction of traits from everyday digital technology usage

- **Social network data** (e.g. Kosinski, Stillwell, & Graepel, 2013)
- **Smartphone data** (Chittaranjan, Blom, & Gatica-Perez, 2013; Montjoye et al., 2013)
Sensation Seeking

• seeking varied, novel, complex, and intense sensations and experiences
• willingness to take physical, social, legal, and financial risks (Zuckerman, 1994)

• Focus of previous research:
  • unsocialized expression of sensation seeking (Roberti, 2004)
  • high risk activities (Zabel, Christopher, Marek, Wieth, & Carlson, 2009; Jack & Ronan, 1998)
  • self-reported behavior (Dahlen, Martin, Ragan, & Kuhlmann, 2005; Leung, 2008)
Smartphone Sensing

- Socialized expression
  - data about mobility, everyday activities and habits

- Everyday manifestation
  - digital behavior partly replaces “analog” behavior (Mayer-Schönberger & Cukier, 2013)

- Objective behavioral data
  - collection of extensive records of individual behavior (Harari et al., 2016)
    - efficient
    - unobtrusive
Can individual Sensation Seeking scores be reliably predicted from data collected via Smartphone Sensing?
PhoneStudy Research App

Data collection
October 2017 – January 2018
30 days of data logging per individual

Sample
N = 260
68% women
average age of 24 (SD = 8.82, RANGE = 18 – 72)

Data logging (GPS, app usage, phone calls)

Self-report Questionnaires
Features

Identification of behavioral correlates of Sensation Seeking

- Gaming
- Risky driving app usage
- Traveling app usage
- Aversion of low-risk/monotonous sports
- Taking financial risks/Trading
- Risky recreational activities
- Lack of planning
- Entertainment
- Social stimulation
- Dating
- Contacts
- General activity
- Circadian rhythm
- Phone usage
- Mobility

Quantification of behavioral categories

- mean frequency
- mean duration
- variation of frequency
- variation of duration
- ratio of certain behavioral category and overall smartphone usage
- maximum distance covered
- mean duration
- total distance covered
- radius of gyration
- entropy
- response rate

222 features
Criterion

• Assessed by the Impulsive Sensation Seeking Scale (ZKPQ-III-R; Zuckerman, 2002)

• True or False?
  • “I am an impulsive person”
  • “I usually think about what I am going to do before I do it”

• 19 items

• Cronbach’s $\alpha = 0.83$
Benchmark Experiment

• Comparison of:
  - featureless learner
  - random forest
  - extreme gradient boosting
  - support vector machine with RBF Kernel
  - elastic net

• Resampling:
  - Outer: 10 x 10-fold CV
  - Inner: Holdout

• Statistical Software R (mlr package, Bischl et al., 2016)
Descriptive Statistics

Sensation Seeking
RANGE = 0 - 19
M = 7.91, SD = 4.22

1263 daily events per person per day

2205 different apps

Top 10 used apps:
- Whatsapp
- Facebook
- Google Chrome
- Instagram
- Snapchat
- Spotify
- Jodel
- YouTube
- Samsung Internet Browser
- Google Maps
Benchmark experiment

<table>
<thead>
<tr>
<th>Rank</th>
<th>Method</th>
<th>MSE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>featureless learning</td>
<td>17.83</td>
<td>-0.04</td>
</tr>
<tr>
<td>2</td>
<td>random forest</td>
<td>16.03</td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td>extreme gradient boosting</td>
<td>16.71</td>
<td>0.02</td>
</tr>
<tr>
<td>4</td>
<td>support vector machine</td>
<td>17.35</td>
<td>-0.02</td>
</tr>
<tr>
<td>5</td>
<td>elastic net</td>
<td>17.43</td>
<td>-0.01</td>
</tr>
</tbody>
</table>
### Top 10 Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Permutation-based Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean frequency of missed calls per day</td>
<td>0.62</td>
</tr>
<tr>
<td>Entropy of contacts for outgoing calls</td>
<td>0.51</td>
</tr>
<tr>
<td>Entropy of contacts for missed calls</td>
<td>0.41</td>
</tr>
<tr>
<td>Variation of frequency of outgoing calls per day</td>
<td>0.32</td>
</tr>
<tr>
<td>Mean time of the last event on Friday/Saturday</td>
<td>0.21</td>
</tr>
<tr>
<td>Variation of the time of the first event from Monday to Friday</td>
<td>0.17</td>
</tr>
<tr>
<td>Mean number of intended events during night on Friday/Saturday</td>
<td>0.14</td>
</tr>
<tr>
<td>Mean radius of gyration during night on Friday/Saturday</td>
<td>0.14</td>
</tr>
<tr>
<td>Mean time of the last event on Sunday</td>
<td>0.14</td>
</tr>
<tr>
<td>Mean frequency of outgoing calls per day</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Partial dependence plots

Mean frequency of missed calls per day

Mean number of intended events on Friday/Saturday night
Conclusion & Contribution

• Random forest model as winner
• but low overall prediction performance

Limitations & Outlook

• Ambiguous meta-data versus individual privacy rights?
• Sample: composition and size

• Self-reported trait scores as ground truth?
Thank you!

Questions or comments?

Please contact me at Ramona.Schoedel@psy.lmu.de
References


Resampling

Inner: Holdout CV

Outer: 10-fold CV

10 times
App categories & Sensation Seeking