# DETECTING DIFFICULTY IN COMPUTER-ASSISTED SURVEYS THROUGH MOUSE MOVEMENT TRAJECTORIES: A NEW MODEL FOR FUNCTIONAL DATA CLASSIFICATION

**Amanda Fernández-Fontelo**, Felix Henninger, Pascal J. Kieslich, Frauke Kreuter and Sonja Greven

# Outline



- Motivation
- Data preparation
- Statistical modelling
- Results
- Conclusions and further research

# Motivation

# Mouse-tracking



- **Psychological research** (cf. reviews by Freeman, 2018; Stillman et al., 2018)
  - Cursor movements recorded in specific setup while participants decide between options
  - Various analyses approaches for detecting tentative commitments and conflict
- **Survey research** (e.g., Stieger & Reips, 2010; Horwitz et al., 2016)
  - Cursor movements recorded as additional paradata in web surveys
  - Specific movement patterns as indicators for data quality and difficulty
- **Project goals**
  - To improve data quality in web surveys
    - Be able to detect difficulty in web surveys to correct the potential measurement errors derived from this difficulty afterwards
  - Bring together and extend analyses approaches and make use of all the available information in the movements

# Survey



- **Survey „Herausforderungen am deutschen Arbeitsmarkt"**
  - Conducted in 2016 in collaboration with the Institute for Employment Research (IAB) in Nürnberg
  - Respondents from a previous survey (who had agreed to this) were contacted again via mail with a 5 € incentive (response rate: 76.8 %)
  - Respondents included: employed, unemployed, job seekers, recipients of unemployment benefit, active labor market program participants
- **Sample size**
  - 1213 participants completed the survey
  - 886 of those reported using the mouse
  - 853 of those continuously had mouse-tracking data available
  - Sample size may vary for each question depending on additional criteria

# Survey overview



Consent to link with administrative data at IAB

Employment and opinion questions

- Some questions with manipulation with respect to the difficulty
  → Do manipulations systematically influence cursor movements?

Psychological questionnaires

Demographic information

# Manipulation of **employment detail** question: Simple vs. complex language (n=551)



UNIVERSITÄT MANNHEIM

Institut für Arbeitsmarkt- und Berufsforschung
Die Forschungseinrichtung der Bundesagentur für Arbeit
IAB

**Sind Sie derzeit...**

Falls Sie momentan in mehreren Beschäftigungsverhältnissen stehen, denken Sie bitte bei dieser und den folgenden Fragen an Ihre hauptsächliche Beschäftigung. Das ist diejenige Erwerbstätigkeit, in der Sie die meisten Stunden arbeiten.

- angestellt bei einem privaten, profitorientierten Unternehmen
- angestellt bei einem privaten, nicht profitorientierten Unternehmen
- angestellt im öffentlichen Dienst des Bundes
- angestellt im öffentlichen Dienst eines Bundeslandes
- angestellt im öffentlichen Dienst einer Kommune
- selbständig in einem freien Beruf
- selbständig in Handel, Gewerbe, Industrie, Dienstleistung, Landwirtschaft
- Freier Mitarbeiter/Freie Mitarbeiterin
- Mithelfende/r Familienangehörige/r

Weiter

UNIVERSITÄT MANNHEIM

Institut für Arbeitsmarkt- und Berufsforschung
Die Forschungseinrichtung der Bundesagentur für Arbeit
IAB

**Sind Sie derzeit...**

Falls Sie momentan in mehreren Beschäftigungsverhältnissen stehen, denken Sie bitte bei dieser und den folgenden Fragen an Ihre hauptsächliche Beschäftigung. Das ist diejenige Erwerbstätigkeit, in der Sie die meisten Stunden arbeiten.

- angestellt bei einem privaten, profitorientierten Unternehmen oder einer Privatperson, das/die regelmäßig Löhne oder Gehälter auszahlt
- angestellt bei einem privaten, nicht profitorientierten Unternehmen, welches gemeinnützig ist oder von Steuervergünstigungen profitiert
- angestellt im öffentlichen Dienst des Bundes (Arbeitgeber ist die Bundesrepublik)
- angestellt im öffentlichen Dienst eines Bundeslandes (Arbeitgeber ist ein Bundesland)
- angestellt im öffentlichen Dienst einer Kommune (Arbeitgeber ist zum Beispiel eine Gemeinde, eine Stadt, ein Landkreis, ...)
- selbständig in einem freien Beruf (z.B. Arzt/Ärztin, Rechtsanwalt/Rechtsanwältin oder Architekt/in)
- selbständig, jedoch nicht in einem freien Beruf im Bereich Handel, Gewerbe, Industrie, Dienstleistung, Landwirtschaft
- Freier Mitarbeiter/Freie Mitarbeiterin, der aufgrund eines Dienst- oder Werkvertrags für ein Unternehmen Aufträge persönlich ausführt
- Mithelfende/r Familienangehörige/r ohne Bezahlung

Weiter

# Manipulation of **employee level** question: Sorted vs. unsorted options (n=501)



UNIVERSITÄT MANNHEIM

Institut für Arbeitsmarkt- und Berufsforschung
Die Forschungseinrichtung der Bundesagentur für Arbeit

**Sind Sie Angestellte/r mit . . .**

- ausführender Tätigkeit nach allgemeiner Anweisung (z.B. Verkäufer/in, Datentypist/-in, Sekretariatsassistent/-in, Pflegehelfer/-in)
- qualifizierter Tätigkeit, die Sie nach Anweisung erledigen (z.B. Sachbearbeiter/-in, Buchhalter/in, technische/r Zeichner/in)
- eigenständiger Leistung in verantwortlicher Tätigkeit oder mit Fachverantwortung für Personal (z.B. wissenschaftliche/r Mitarbeiter/in, Prokurist/in, Abteilungsleiter/in oder Meister/in im Angestelltenverhältnis)
- umfassenden Führungsaufgaben und Entscheidungsbefugnissen (z.B. Direktor/in oder Geschäftsführer/in, Mitglied des Vorstands)

Weiter

UNIVERSITÄT MANNHEIM

Institut für Arbeitsmarkt- und Berufsforschung
Die Forschungseinrichtung der Bundesagentur für Arbeit

**Sind Sie Angestellte/r mit . . .**

- eigenständiger Leistung in verantwortlicher Tätigkeit oder mit Fachverantwortung für Personal (z.B. wissenschaftliche/r Mitarbeiter/in, Prokurist/in, Abteilungsleiter/in oder Meister/in im Angestelltenverhältnis)
- ausführender Tätigkeit nach allgemeiner Anweisung (z.B. Verkäufer/in, Datentypist/-in, Sekretariatsassistent/-in, Pflegehelfer/-in)
- umfassenden Führungsaufgaben und Entscheidungsbefugnissen (z.B. Direktor/in oder Geschäftsführer/in, Mitglied des Vorstands)
- qualifizierter Tätigkeit, die Sie nach Anweisung erledigen (z.B. Sachbearbeiter/-in, Buchhalter/in, technische/r Zeichner/in)

Weiter

# Manipulation of **education level** question: Sorted vs. unsorted options (n=548)



UNIVERSITÄT MANNHEIM

Institut für Arbeitsmarkt- und Berufsforschung
Die Forschungseinrichtung der Bundesagentur für Arbeit
IAB

**Welchen höchsten allgemeinbildenden Schulabschluss haben Sie?**

- Sonderschulabschluss, Abschluss der Förderschule
- Hauptschulabschluss, Volksschulabschluss
- Polytechnische Oberschule (POS) Abschluss 8. Klasse (DDR Abschluss)
- Polytechnische Oberschule (POS) Abschluss 10. Klasse (DDR Abschluss)
- Realschulabschluss, Mittlere Reife, Fachoberschulreife
- Fachhochschulreife (z.B. Abschluss einer Fachoberschule)
- Abitur / allgemeine oder fachgebundene Hochschulreife
- Abschluss einer Erweiterten Oberschule (EOS) oder Berufsausbildung mit Abitur (DDR-Abschlüsse)
- Anderer deutscher Schulabschluss, und zwar
- Ausländischer Schulabschluss und zwar
- kein Abschluss

Weiter

UNIVERSITÄT MANNHEIM

Institut für Arbeitsmarkt- und Berufsforschung
Die Forschungseinrichtung der Bundesagentur für Arbeit
IAB

**Welchen höchsten allgemeinbildenden Schulabschluss haben Sie?**

- Abschluss einer Erweiterten Oberschule (EOS) oder Berufsausbildung mit Abitur (DDR-Abschlüsse)
- Sonderschulabschluss, Abschluss der Förderschule
- Fachhochschulreife (z.B. Abschluss einer Fachoberschule)
- Polytechnische Oberschule (POS) Abschluss 8. Klasse (DDR Abschluss)
- Abitur / allgemeine oder fachgebundene Hochschulreife
- Hauptschulabschluss, Volksschulabschluss
- Realschulabschluss, Mittlere Reife, Fachoberschulreife
- Polytechnische Oberschule (POS) Abschluss 10. Klasse (DDR Abschluss)
- Anderer deutscher Schulabschluss, und zwar
- Ausländischer Schulabschluss und zwar
- kein Abschluss

Weiter

# Data preparation

# Mousetrap R package



- Extraction of mouse movement trajectories from raw data
- Preprocessing of mouse-tracking with mousetrap R package (Kieslich, Wulff, Henninger, Haslbeck & Schulte-Mecklenbeck, in press)
  - Freely available from CRAN
  - Documentation: http://pascalkieslich.github.io/mousetrap/
- Preparation steps for each participant and question
  - Resampling of recorded cursor coordinates (every 10 ms) and calculation of mouse-tracking indices
  - Time-normalize trajectories

# Mouse-tracking indices



| Type | Measure | Definition |
|---|---|---|
| **Time** | RT | Time until response is given |
| | initiation time | Time until first movement is initiated |
| | idle time | Total time without movement across trial |
| **Complexity** | xpos flips | Number of directional changes along x-axis |
| | ypos flips | Number of directional changes along y-axis |
| **Derivatives** | total dist. | Euclidean distance traveled by trajectory |
| | vel. max. | Maximum movement velocity |
| | acc. max. | Maximum movement acceleration |
| **Hovers** | hovers ini. | Number of periods (> certain threshold) without movement |
| | hovers ini. time | Total time of periods without movement (w/o initiation time) |

# Statistical modelling

# Planning



- **Goal?** Improve web surveys data quality by detecting respondents' difficulty and correct potential measurement errors derived from this difficulty
- **How to detect respondents' difficulty?**
  - Gold standard learning methods: use existing learning models to classify among difficulty (e.g., complex language, unsorted choices, etc.) and non-difficulty (e.g., simple language, sorted choices, etc.) scenarios using both the indices and demographic covariates as predictors
  - Mouse movements trajectories as predictors of difficulty: generalize the ideas behind the works by Fuchs et al. (2015,2017) and Ferraty and Vieu (2003,2006)
    - Extend standard semi-metrics to cope with d-dimensional trajectories
    - Propose new semi-metrics both related to the application (e.g., distances between flips, hovers, etc.) and based on relevant curves' similarity measures (e.g., Hausdorff or Fréchet)
    - Propose an ensemble of semi-metric-based classifiers for d-dimensional trajectories classification

# Gold standard learning methods



- Questions and categories: employment detail (simple vs. complex language), employee level and education level (sorted vs. unsorted choices)
- Predictors (indices): measures of time, complexity, derivatives and hovers. Demographic covariates such as gender and age
- Indices are personalized to account for individual variabilities
- Learning classifiers:
  - Logit models
  - Decision trees
  - Tree-based random forest
  - Tree-based bagging
  - Tree-based gradient boosting
  - Support vector machines
  - Neural networks

# Functional data classifier – Ensemble



## Goal

- Use both the mouse movement trajectories and the indices to improve the predictive capabilities of the gold standard methods
- If we only use the indices (aggregated data), we could miss relevant information that is still in the mouse movement trajectories

## Idea

- Use semi-metrics and the functional k-nearest neighbors' method to create base classifiers for 2-dimensional trajectories (extend the works by Fuchs et al. (2015) and Ferraty and Vieu (2003))
- Ensemble these base classifiers to create a super classifier (stacked ensemble) that strategically combines the information of the base classifiers to improve base classifiers' predictive capability

# Semi-metrics for d-dimensional trajectories



- Semi-metrics are used to measure distances and capture specific characteristics of functions (trajectories).
- Several authors as Fuchs et al. (2015, 2017) and Ferraty and Vieu (2003, 2006) have used semi-metrics for 1-dimensional trajectories classification
  - We have both extended some of the existing ones to d-dimensional trajectories and proposed new ones
- Semi-metrics can be computed on the mouse movement trajectories and their derivatives

# Semi-metrics – examples



- **Standard semi-metrics**: semi-metrics such as $\mathbf{L^p}$, global mean, different definitions for maximum, minimum, supremum, infimum, and dynamic time warping, among others.
- **New proposals:** Fréchet, Hausdorff, Aitchison or Needleman-Wunsch, etc.
- **Application-specific distances:**
  - Euclidean distance between measures (e.g., flips, hovers, maximum velocity and acceleration, etc.) of two different trajectories $\mathbf{x_i}$ and $\mathbf{x_*}$
  - The distance between the trajectories' length is also considered as the trajectories have been time-normalized and we lose otherwise this information

# Functional k-nearest neighbours method (I)



- Suppose a learning sample $(\mathbf{x}_i, y_i)$ for $i = 1, \dots, n$ where $\mathbf{x}_i$ is a $d$-dimensional regular spaced functional data observation (e.g., 2-dimensional time-normalized mouse movement trajectory), and $y_i$ which is the class of $\mathbf{x}_i$ over $\mathcal{L} = (1, .., l)$ (e.g., in our case, we have two classes as we have simple and complex scenarios)
- Suppose we want to classify a new observation $\mathbf{x}_*$ (from the same space of functions than $\mathbf{x}_i$, $i = 1, \dots, n$)
  - We want to estimate $y^*$ (categorical variable)

# Functional k-nearest neighbours method (II)



- Functional k-nearest neighbours is an extension of the multivariate k-nearest neighbours method (Ferraty and Vieu, 2006). In our case:
  - Order the d-dimensional training observations according to the predetermined semi-metric
    $$\mathrm{D}\left(\mathbf{x}_*, \mathbf{x}_{(1)}\right) \leq \cdots \leq \cdots \mathrm{D}\left(\mathbf{x}_*, \mathbf{x}_{(\mathrm{k})}\right) \leq \cdots \leq \mathrm{D}\left(\mathbf{x}_*, \mathbf{x}_{(\mathrm{n})}\right)$$
  - Determine the neighbourhood $\mathcal{N}^{\mathrm{k}}(\mathbf{x}_*)$ with the $\mathrm{k}$ closest training observations to the new observation
    $$\mathcal{N}^{\mathrm{k}}(\mathbf{x}_*) = \left\{\mathbf{x}_{(\mathrm{j})} : \mathrm{D}\left(\mathbf{x}_*, \mathbf{x}_{(\mathrm{j})}\right) < \mathrm{D}\left(\mathbf{x}_*, \mathbf{x}_{(\mathrm{k})}\right)\right\}$$
  - The prediction of the unknown class $\mathbf{y}_*$ is given by
    $$\hat{\mathrm{y}}_* = \underset{\mathrm{l} \in \mathcal{L}}{\operatorname{argmax}}\, \omega_{\mathrm{l}}\,,$$

  where $\omega_{\mathrm{l}} = \frac{1}{\mathrm{k}} \sum_{\mathrm{x_i} \in \mathcal{N}^{\mathrm{k}}(\mathbf{x}_*)} 1_{\mathrm{y_i} = \mathrm{l}}$ is the estimated posterior probability that the new observation $\mathbf{x}_*$ belongs to the class $\mathrm{l} \in \mathcal{L}$.

# Stacked Ensemble – idea



- Base learners are naïve classifiers that are combined strategically to obtain better predictive models called ensembles
  - Tree models are the base learners in tree-based gradient boosting models (ensemble)
  - Classifiers derived from the functional k-nearest neighbours method are the base learners in our ensemble
- Different methods can be used to ensemble base learners (e.g., linear combinations or model stacking)
- Model stacking is a two steps procedure:
  - Tune base learners and get cross-validated predictions (step 1)
  - Use the cross-validated predictions from step 1 as covariates of a new classifier called *super learner* (step 2). This *super learner* is finally used to predict the category of the new observations

# Stacked Ensemble for mouse movement trajectories



- Base learners can be based on standard semi-metrics or distances between indices. Two scenarios have been considered:
  - Ensembles where indices are incorporated through semi-metrics (distances between indices)
  - Ensembles where indices are additional covariates included jointly with the cross-validated predictions in the *super learner*
- Tree-based random forest and boosting models are considered as *super learners*
- Demographic covariates such as gender and age are also included as covariates in the *super learner*

# Indices' correction, tunning parameters and learning classifiers' performance



- The indices need to be first personalized to account for individuals' variabilities. A linear model has been considered with the following variables:
  - Indices from eight questions that were not manipulated (baseline)
  - Physical positions of individuals' answer choice in the target question (baseline + position)
- Nested cross-validation is needed to:
  - Tune parameters to select the best set of parameters for each learning classifier
  - Honestly, evaluate the learning classifiers' predictive capability (performance), and select the best model among the candidates according to this performance

# Results

# Best gold standard learning models and nested-cross validated accuracies



| | n | indices' correction method (personalization) and hovers threshold | learning model | top three most important indices | nested cross-validated accuracy (smallest accuracy*) |
|---|---|---|---|---|---|
| employment detail (simple vs. complex language) | 551 | baseline<br>2000 ms | tree-based gradient boosting | response time<br>vertical flips<br>hovers | 0.6587<br>(0.5628) |
| employee level (sorted vs. unsorted answering choices) | 501 | baseline and position<br>3000 ms | tree-based gradient boosting | initiation time<br>vertical flips<br>total distance | 0.5909<br>(0.4751) |
| education level (sorted vs. unsorted answering choices) | 548 | baseline and position<br>250 ms | tree-based random forest | maximum acceleration<br>initiation time<br>hovers | 0.5895<br>(0.4634) |

*Smallest accuracy among all considered learners with the same hovers' threshold and all considered indices
**Results have been obtained with `R 3.6.1` through the library `mlr`
***The seed 2428 has been set for each question to make the analyses completely reproducible
****To speed up the models' tuning and selection, parallelization was used with 32 CPUs for all models

# Best stacked ensembles and nested-cross validated accuracies



| | n | base learners (derivative) | nested cross-validated accuracies (base learners) | Super learner | nested cross-validated accuracy (ensemble) |
|---|---|---|---|---|---|
| employment detail (simple vs. complex language) | 551 | *globMax3-x (0)*<br>*globMax2 (0)*<br>*mean (1)*<br>*Hausdorff (1)* | *0.5988*<br>*0.5807*<br>*0.5791*<br>*0.5788* | *Tree-based random forest* | 0.6552 |
| employee level (sorted vs. unsorted answering choices) | 501 | globMin3-y (0)<br>dynamic time warping (0)<br>globMin2 (0)<br>infimum 1 (0)<br>initiation time (0)<br>Manhattan (0)<br>Infimum2 (0) | 0.6985<br>0.6926<br>0.6247<br>0.5909<br>0.5728<br>0.5648<br>0.5611 | Tree-based gradient boosting | 0.7266 |
| education level (sorted vs. unsorted answering choices) | 548 | Mean (0)<br>globMin3x (0)<br>globMax3y (0)<br>Manhattan (0)<br>dynamic time warping (0) | 0.6039<br>0.5783<br>0.5604<br>0.5564<br>0.5401 | Tree-based gradient boosting | 0.6348 |

*Text in italics indicates that other ensemble candidates have been still studying to improve accuracy
**Results have been obtained with `R 3.6.1` through the library `classiMultiFunc` (under construction)
***The seed 2428 has been set for each question to make the analyses completely reproducible

# Conclusions and further work

- Personalization in gold standard models usually improves models' accuracy
- Ensemble models with both mouse movement trajectories and indices as predictors substantially improve gold standard models' accuracies
  - Employment detail: 0.6587 (gold standard) – 0.6552 (ensemble)
  - Employee level: 0.5909 (gold standard) – 0.7266 (ensemble)
  - Education level: 0.5895 (gold standard) – 0.6348 (ensemble)
- As further work, we want to use another method based on kernel functions to construct base classifiers, and consider different ways to ensemble base learners as linear combinations

# Thank you for your attention!