Data-Driven Evaluation of In-Vehicle Information Systems: Supplementary Material

Patrick Ebel

Köln, 2023

Preface

The material in this document is supplementary to Patrick Ebel's PhD thesis "Data-Driven Evaluation on In-Vehicle Information Systems" and is taken directly or with minor modifications from the following previously published works:

[1] P. Ebel, F. Brokhausen, and A. Vogelsang, "The Role and Potentials of Field User Interaction Data in the Automotive UX Development Lifecycle: An Industry Perspective," in *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Virtual Event DC USA: ACM, Sep. 2020, pp. 141–150

[2] P. Ebel, J. Orlovska, S. Hünemeyer, C. Wickman, A. Vogelsang, and R. Söderberg, "Automotive UX design and data-driven development: Narrowing the gap to support practitioners," *Transportation Research Interdisciplinary Perspectives*, vol. 11, p. 100455, Sep. 2021

[3] P. Ebel, C. Lingenfelder, and A. Vogelsang, "Visualizing Event Sequence Data for User Behavior Evaluation of In-Vehicle Information Systems," in *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Leeds United Kingdom: ACM, Sep. 2021, pp. 219–229

[4] P. Ebel, M. Berger, C. Lingenfelder, and A. Vogelsang, "How Do Drivers Self-Regulate their Secondary Task Engagements? The Effect of Driving Automation on Touchscreen Interactions and Glance Behavior," in *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Seoul Republic of Korea: ACM, Sep. 2022, pp. 263–273

[5] P. Ebel, K. J. Gülle, C. Lingenfelder, and A. Vogelsang, "ICEBOAT: An Interactive User Behavior Analysis Tool for Automotive User Interfaces," in *The Adjunct Publication of the 35th Annual ACM Symposium on User Interface Software and Technology*, Aug. 2022

[6] P. Ebel, C. Lingenfelder, and A. Vogelsang, "On the forces of driver distraction: Explainable predictions for the visual demand of in-vehicle touchscreen interactions," *Accident Analysis & Prevention*, vol. 183, p. 106956, Apr. 2023

Under Review

[7] P. Ebel, C. Lingenfelder, and A. Vogelsang, "Multitasking while Driving: How Drivers Self-Regulate their Interaction with In-Vehicle Touchscreens in Automated Driving," *International Journal of Human-Computer Interaction*, 2023, *Extended version of [4], Accepted*

[8] P. Ebel, K. J. Gülle, C. Lingenfelder, and A. Vogelsang, "Exploring Millions of User Interactions with ICEBOAT: Big Data Analytics for Automotive User Interfaces," in *AutomotiveUI '23: 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Ingolstadt, Germany, 2023, *Submitted to the Full Paper Track*

Contents

1 Chapter 4

1.1 Interview Guideline

Introduction (10 minutes)

1. Personal introduction
2. Research goals
3. Definition of field user interaction data
4. Interview structure
5. Resolve follow-up questions

Interview

Pre-Design Stage (15 minutes)

Question 1/9 - State-of-the-art: Where in the pre-design stage do you incorporate (field) user interaction data in your decisions?
Question 2/9 - Challenges: What are the most time consuming and expensive tasks the pre-design stage?
Question 3/9 - Potentials/Opportunities: What information regarding the user, the system and their interaction would benefit the pre-design stage?

Design Stage (15 minutes)

Question 4/9 - State-of-the-art: Where in the design stage do you incorporate (field) user interaction data in your processes?
Question 5/9 - Challenges: What challenges do you face when evaluating the quality and/or usability of a design during the design stage?
Question 6/9 - Potentials/Opportunities: Where do you see potential, that the usage of field user interaction data might ease the design process?

Post Design Stage (15 minutes)

Question 7/9 - State-of-the-art: Do you gather feedback from the field users after the design was deployed in a product, and if so how?
Question 8/9 - Challenges: What are the challenges in collecting feedback from the users after deployment to the end user?
Question 9/9 - Potentials/Opportunities: What additional information regarding field user interaction would benefit the post design phase?

2 Chapter 6

2.1 Interview Questions and Results

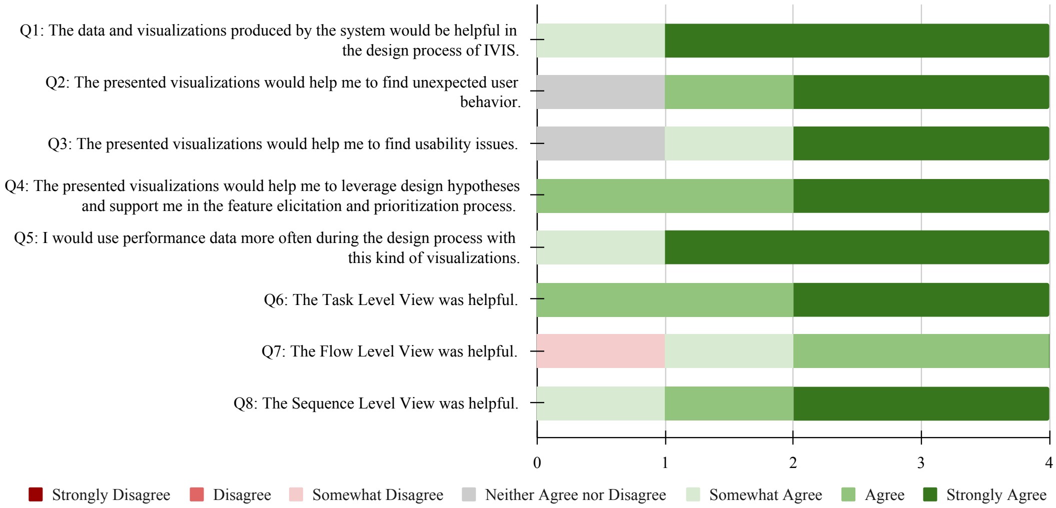


Figure 2.1: Horizontal bar chart that indicates the agreement of the study participants with the respective statements.

3 Chapter 7

3.1 Storyboard

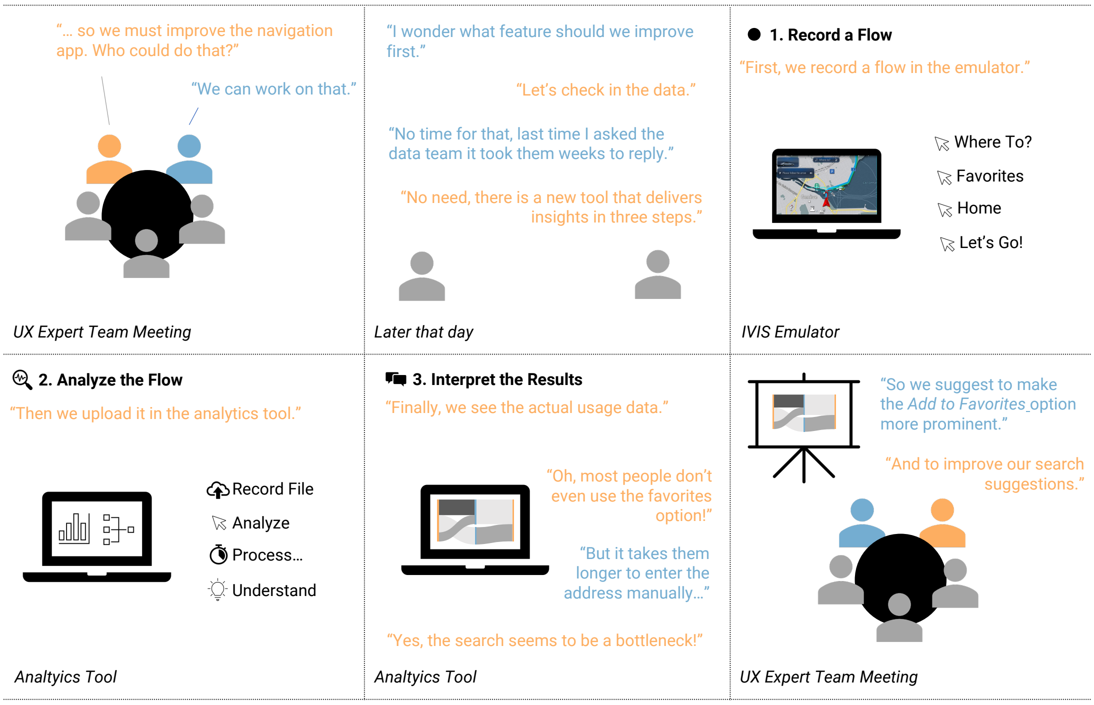


Figure 3.1: Storyboard as derived from the initial interview study.

4 Chapter 8

4.1 Models

Table 4.1: Mixed-effects models for the mean glance duration according to speed and road curvature.

	Dependent variable:	
	Mean Glance Duration	
	Model 5	Model 6
Constant	7.28^{***} (0.02)	7.27^{***} (0.02)
50-100	-0.06^{***} (0.01)	
100+	-0.07^{***} (0.01)	
curved		-0.14^{***} (0.01)
Akaike Inf. Crit.	44,429.86	44,178.02
Bayesian Inf. Crit.	44,479.98	44,219.78
Note:		*p<0.05; **p<0.01; ***p<0.001

Table 4.2: Mixed-effects models for the interaction probability with Keyboard, CoverFlow, Slider, RemoteUI, ControlBar, Other, and Unknown UI elements. Note: In contrast to the models presented in the paper we did not include the car type as random effect since it led to a singularity warning. This warning is often associated with an overfitted model as the random effect structure might be too complex to be supported by the data. This in turn might be due to the small amount of interaction with these UI elements.

	Dependent variable:					
	Keyboard	CoverFlow	Slider	RemoteUI	ControlBar	Other
			conv. error			
Intercept	−8.46*** (0.19)	−9.35*** (0.22)	−8.88*** (1)	−10.33*** (0.23)	−10.37*** (0.30)	−0.40*** (0.03)
ACC	0.07 (0.26)	0.43 (0.30)	0.02*** (1)	−0.01 (0.30)	0.31 (0.34)	0.06 (0.07)
ACC+LKA	0.21 (0.14)	0.39* (0.16)	0.18 (0.12)	−0.31* (0.15)	0.14 (0.22)	−0.17*** (0.03)
50-100	−0.19 (0.12)	−0.13 (0.14)	0.24*** (1)	−0.07 (0.15)	0.32 (0.18)	0.07* (0.03)
100+	−0.32* (0.14)	−0.31 (0.16)	0.23* (0.11)	0.05 (0.17)	0.37 (0.22)	0.09* (0.04)
curved	−0.15 (0.13)	−0.38* (0.16)	−0.33* (0.15)	−0.14 (0.15)	−0.36 (0.20)	−0.06 (0.04)
Akaike Inf. Crit.	8,886.07	6,982.70	7,282.68	6,270.20	3,965.26	41,098.95
Bayesian Inf. Crit.	8,944.55	7,041.17	7,341.16	6,328.68	4,023.74	41,157.43

Note: conv. error: Model failed to converge, *p<0.05; **p<0.01; ***p<0.001

4.2 Calibration Plot of Model 4

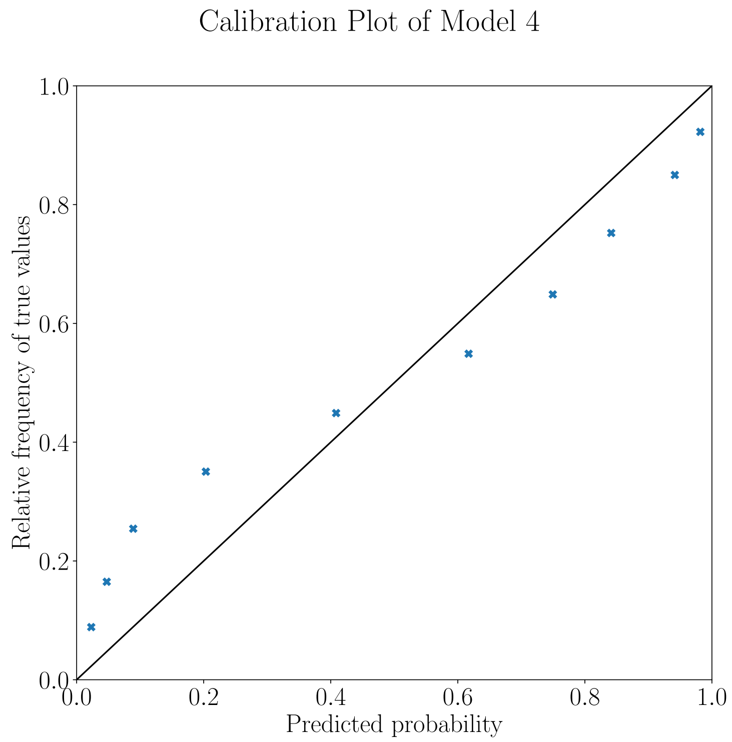


Figure 4.1: Calibration plot of Model 4 (generalized linear mixed-effects model). Marginal $R^2 = 0.099$, Conditional $R^2 = 0.347$

4.3 “True vs. Predicted” Plot of Model 2

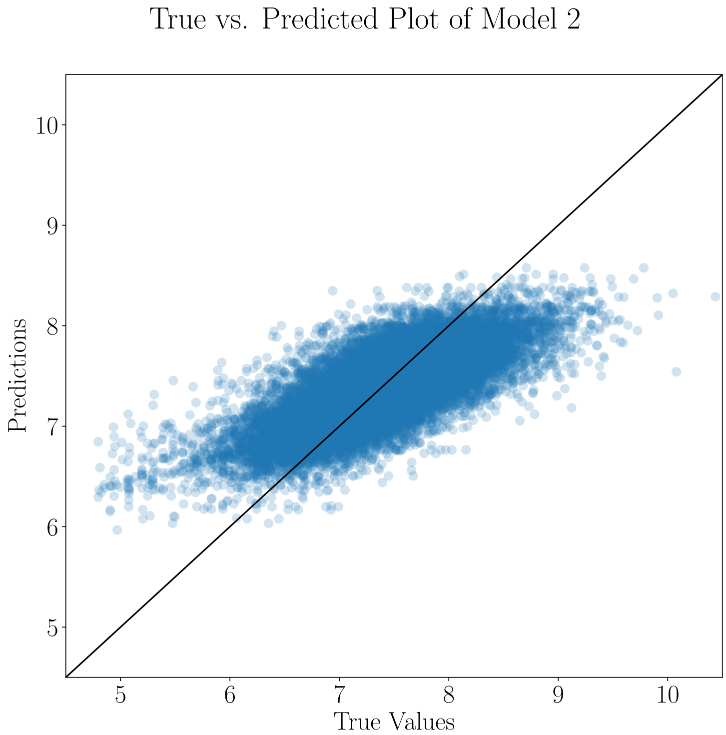


Figure 4.2: True values of the mean glance duration plotted against the predictions of Model 2 (linear mixed-effects model). All values are given on a logarithmic scale. Marginal $R^2 = 0.091$, Conditional $R^2 = 0.441$

5 Chapter 9

5.1 Hyperparameter Optimization

In the following we report the results of the hyperparameter optimization for each of the individual models.

Random Forest Models

The Implementation and the descriptions based on the scikit-learn python package.

n_estimators = [100, 200, 400, 800, 1200, 1600, 2000] – The number of trees in the forest.
max_features = ['auto', 'sqrt'] – Number of features to consider when looking for the best split
max_depth = [10, 20, 30, 40, 60, 80, 100] – Maximum depth of the tree.
min_samples_split = [2, 5, 10] – Minimum number of samples required to split an internal node.
min_samples_leaf = [1, 2, 4] – Minimum number of samples required to be at a leaf node.
bootstrap = [True, False] – Whether bootstrap samples are used when building trees.

Table 5.1: Sets of best performing parameters for the Random Forest models.

Feature	Long Glance Prediction	TGD Prediction
n_estimators	200	1600
max_features	auto	auto
max_depth	10	60
min_samples_split	5	2
min_samples_leaf	2	4
bootstrap	True	True

XGBoost Models

The Implementation and the descriptions are based on the XGBoost python package.

n_estimators = [20, 100, 500, 1000, 5000, 10000, 20000] – Number of boosting rounds.
subsample = [0.2,0.4,0.6,0.8,1] – Subsample ratio of the training instance.
max_depth = [5,10,50,100] – Maximum tree depth for base learners.
learning_rate = [05, 0.001, 0.01, 0.1, 1] – Boosting learning rate (xgb's "eta")
colsample_bytree = [0.2,0.4,0.6,0.8,1] – Subsample ratio of columns when constructing each tree.
colsample_bylevel = [0.2,0.4,0.6,0.8,1] – Subsample ratio of columns for each level.

Table 5.2: Sets of best performing parameters for the XGBoost models.

Feature	Long Glance Prediction	Total Glance Duration Prediction
n_estimators	5000	5000
subsample	0.6	0.8
max_depth	10	10
min_child_weight	4	10
learning_rate	0.01	05
colsample_bytree	0.2	0.6
colsample_bylevel	0.2	1

Feedforward Neural Networks

The Implementation and the hyperparameter optimization was performed using the Keras API. It needs to be noted that the different hyperparameter combinations did not show large differences in their predictive performance.

n_hidden_layers = [1,2,3,4,5] – Number of layers between the input and output layer of the neural network

n_neurons = [32, 64, 128, 256, 512]) – Number of neurons per layer.

activation = ["relu", "sigmoid"] – The activation function of the neurons in the respective layer.

drop_out = [0,0.1,0.2,0.3] – The probability at which random units are set to zero during training.

learning_rate = [0.01,0.001,01] – Initial learning rate of the ADAM optimizer.

Table 5.3: Sets of best performing parameters for the FNN models.

Feature	Long Glance Prediction	Total Glance Duration Prediction
n_hidden_layers	3	1
learning_rate	01	0.001
n_neurons layer 1	512	512
activation layer 1	sigmoid	relu
drop_out layer 1	0.3	0.1
n_neurons layer 2	64	-
activation layer 2	relu	-
drop_out layer 2	0.1	-
n_neurons layer 3	256	-
activation layer 3	sigmoid	-
drop_out layer 3	0.1	-

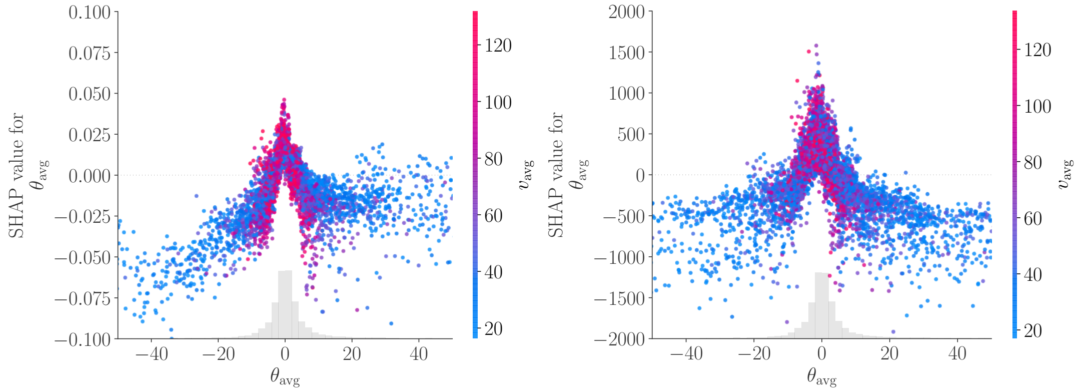
5.2 Dataset Summary Statistics

Table 5.4 provides an overview of the statistics of all features used in the models.

Table 5.4: Dataset Summary Statistics

Statistic	Mean	St. Dev.	Min	$Q_1(25)$	Median	$Q_3(75)$	Max
Interactions	4.431	4.993	1	1	3	5	41
Tap Gestures	3.814	4.495	0	1	2	5	40
Drag Gestures	0.363	1.324	0	0	0	0	31
Multitouch Gestures	0.240	1.108	0	0	0	0	26
MGD in ms	1,441.491	929.736	120	960	1,241	1,659	26,801
Number of glances	4.367	4.998	1	1	3	6	50
Long Glances	0.569	1.102	0	0	0	1	13
TGD in ms	5,742.751	7,049.487	120	1,590.500	3,499	7,354	262,416
Average speed in km/h	70.516	36.935	0.633	40.881	66.230	96.567	209.883
ACC active	0.206	0.404	0	0	0	0	1
SA active	0.099	0.299	0	0	0	0	1
AppIcon interactions	0.196	0.586	0	0	0	0	13
CoverFlow interactions	0.038	0.434	0	0	0	0	16
Unknown interactions	0.049	0.450	0	0	0	0	23
Other interactions	0.731	1.568	0	0	0	1	37
List interactions	0.518	1.652	0	0	0	0	31
Tab interactions	0.385	1.335	0	0	0	0	35
ControlBar interactions	0.012	0.135	0	0	0	0	4
Button interactions	0.640	1.515	0	0	0	1	36
Homebar interactions	0.892	2.032	0	0	0	1	36
Slider interactions	0.015	0.228	0	0	0	0	9
ClickGuard interactions	0.058	0.313	0	0	0	0	8
PopUp interactions	0.030	0.232	0	0	0	0	9
Keyboard interactions	0.184	1.435	0	0	0	0	31
Map interactions	0.508	2.181	0	0	0	0	39
RemoteUI interactions	0.173	1.179	0	0	0	0	31
Browser interactions	0.002	0.093	0	0	0	0	8

5.3 Steering Wheel Feature Dependence Plot



(a) Long Glance Prediction Sample.

(b) Total Glance Duration Prediction Sample.

Figure 5.1: Feature dependence plots of the steering wheel angle and its interaction with the vehicle speed for the long glance classification and TGD model.