Interaction Challenges for N-Of-One Experiments based on Mobile Sensing Data

FLORIAN BEMMANN, LMU Munich, Germany

DANIEL BUSCHEK, Department of Computer Science, University of Bayreuth, Germany

Mobile Sensing data collected by smartphones contains a rich body of information. Researchers and companies already leverage that for research and businesses, respectively. However, the end-user from whom the data originates is not enabled to do so yet. We show how smartphone applications could enable users to leverage their data themselves. In N-of-1 experiments users conduct research questions on themselves. By passively tracking behavioral and contextual data, enriching it with a self-reported variable of interest (e.g. mood, stress, cognitive load), and using appropriately presented statistical and/or Machine Learning methods they could reflect on and learn about the factors that might influence personal variables of interest. To make progress towards this vision of empowered end-users, we outline a Mobile Sensing app for N-of-1 experiments, point out relevant challenges, and propose ideas and future research directions to solve them. More broadly, we envision mobile sensing-based machine learning applications to spread in the area of personal informatics.

Additional Key Words and Phrases: personal informatics, interactive machine learning, explainable ai

ACM Reference Format:

Florian Bemmann and Daniel Buschek. 2022. Interaction Challenges for N-Of-One Experiments based on Mobile Sensing Data. In *CHI '22 Workshop: Grand Challenges in Personal Informatics and AI, May 11, 2022, Online*. ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 MOBILE SENSING AND N-OF-ONE TRIALS IN PERSONAL INFORMATICS

Ubiquitous and mobile technology tracks various kinds of user behavior data, e.g., location [10], physiological data [8], and mobile behavior [9]. These tracking features make it possible to build adaptive and intelligent user interfaces providing the user with information right when they are needed, e.g. Kunzler [4] who studied context-aware notifications. Researchers use such data to conduct studies in various disciplines e.g. in psychology [3, 11] where such data replaces self reports.

Unfortunately the user is only rarely included in the process, and does not draw a direct benefit of such data. The community around data science [2, 7] demands to include the human in every step that concerns their data [2]. Beyond just informing people about *who* collects *what* data, developers should go one step further and also answer *how* data is used [2]. Personal informatics applications are one way to allow the user to make use of their data. Being fueled with such mobile sensing data, users can analyze their own data and derive insights that are ideally useful to them directly. In self-experiments [1] (also called N-of-One trials [5]) people investigate on relationships in their behavior themselves. Depending on the experiment question, one introduces an independent variable (e.g. a treatment that is assigned randomly throughout the time-series), or labels an event of interest as dependent variable. This concept initially comes from the medical context and recently also finds application in the domain of personal informatics using self-tracking data.

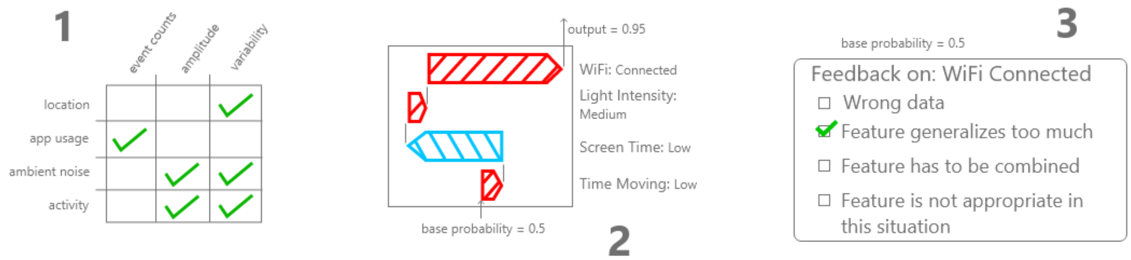

Fig. 1. Sketched approaches to three interaction challenges: (1) Feature Selection via a matrix that connects datatypes with kinds of feature representation, (2) explanation of feature importance, and (3) a modality to give feedback on the reasons for a wrong prediction.

In this workshop paper we present our vision of a mobile sensing smartphone app that brings interactive machine learning to end users. We thereby target interested users with a basic understanding of data analysis and information mining, e.g. from the Quantified Self and Personal Informatics community. We argue that there would be manifold benefits if users would be enabled to dig into their data themselves, especially regarding privacy, empowerment, and education on technology. In the workshop at CHI we would like to discuss potential solutions to two categories of challenges that we consider especially critical: We (1) argue that appropriate *interaction concepts* need to be developed to enable interacting with machine learning on mobile sensing capable devices, and (2) identify the *trade-off between technical feasibility and practical relevance* of single-user machine learning models as major challenge. We structure the remainder of this paper into these two challenges.

2 INTERACTION CHALLENGES

A main challenge towards an interactive machine learning application for personal informatics lies in the interaction. High dimensional data and complex machine learning processes have to be brought onto the limited screen space of a smartphone, in a way that is understandable and usable for non-experts. We subdivide this challenge into three tasks for each of which an interaction concept is needed. Our ideas are visualized in Figure 1.

Feature Selection. The aim of this stage is that the user should be able to select which features to record and thereby include into the model training. Mobile sensing data usually is very high dimensional, having more features than instances [11]. This raises the challenge that simply listing all features in a multiple choice like list becomes impossible. To reduce that complexity in a structured way, we suggest to group the features by two dimensions: Their originating data type (e.g. location, app usage, heartrate sensor, ...) and their kind of representation (e.g. count of events, amplitude, variability, ...). Applying this structure to the mobile sensing data, the available features could be visualized as matrix, having the data type on one axis and the kind of representation on the other. The functionality that is intended to be supported can then be implemented in the matrix cells. Turning features on and off could be realized by a tap, further configuration options could be wrapped into a popup that opens when taping a cell.

Model Explanation. After training and evaluating a model, the results have to be presented to the user. In the ideal case where a model could be built that predicts the desired target variable better than random, the app should explain which features contribute to that prediction to which extent. This information can help the user to draw insights, i.e.

which recorded behavior might relate to the (predicted) variable of interest. Understanding and explaining (1) *which* and (2) *how* each feature contributes to a prediction is major challenge in machine learning research. We suggest to use Shapley Values [6]. They indicate feature attributions, by regarding each feature as a force. Starting from a baseline score, each feature's force either increases or decreases the output. Magnitude and direction of each force can be interpreted by the user. In a plot showing each force's bar in a separate line one could visualize the top ten features of a prediction, which should be sufficient for most use cases. The visualization should stick to the feature grouping used in the feature selection stage.

Feedback Loop. Machine Learning models practically never work perfectly. A more or less high amount of erroneous predictions remains. In the context N-of-One experiments on ones own data we have the opportunity that the one who understands the context and origin of that data best, the data-producing user, can directly interact with the system. Furthermore, in the personal informatics context the user is usually highly motivated to contribute with their domain knowledge. Thus the challenge is to provide an interface that allows the user to report erroneous conclusions (e.g. features that received a high weight by the algorithm, but are not appropriate in a specific situation for some reason that the model cannot understand). The user should be able to tell the system what it should do different in this specific situation, and how it can identify similar cases. As a first step, we suggest a rule based approach. A rule that combines one or more feature(s) with simple operators (i.e. and, or, equals, is larger than, ...) specify the situation, an expression (i.e. halve feature X) define an action.

3 FEASIBILITY OF SERIOUS N-OF-ONE EXPERIMENTS IN PRACTICES

An open question is the feasibility of relevant N-of-One experiments in practice. Easy machine learning tasks are technically possible with satisfying accuracy but might not be interesting for the user. For example, in a recent prototype we showed that a personal model that predicts whether the user is at home our outdoors can be trained on-device with a few days of data easily. However this might not be an interesting prediction to personal informatics affine users. On the other hand, tasks about real questions that the user wants to investigate might not be solvable with the limited amount of samples. A interesting objective could be to investigate the causes of stress in everyday life. However it is doubtful whether the amount of self-labeled ground truth and according mobile sensing data is sufficient to yield satisfying insights.

Fortunately this kind of data collection also has advantages that might affect the performance of machine learning models beneficially: When modeling behavior of a single user, the data and model do not need to capture and account for variability between people. That should support models to fit with less data points respectively. Furthermore, interactivity could be leveraged to collect feedback from the user, to improve the trained model. Field studies have to show whether it is possible to find a trade-off between relevance for users and feasibility, and how much the benefits of one-user data compensate for the lack of large sample sizes.

4 CONCLUSION

We envision mobile sensing smartphone apps that incorporate interactive machine learning in the domain of personal informatics. Users could thereby conduct N-of-One experiments, generating insights on a desired target variable in relation to passively sensed behavior and context data from their own life. However the complexity of machine learning applications makes the development of suitable interaction concepts for end users challenging. We propose starting points for interaction design solutions for three steps in a mobile sensing personal informatics application.

ACKNOWLEDGMENTS

This project is partly funded by the Bavarian State Ministry of Science and the Arts and coordinated by the Bavarian Research Institute for Digital Transformation (bidt).

REFERENCES

[1] Nediyana Daskalova, Eindra Kyi, Kevin Ouyang, Arthur Borem, Sally Chen, Sung Hyun Park, Nicole Nugent, and Jeff Huang. 2021. Self-E: Smartphone-Supported Guidance for Customizable Self-Experimentation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[2] Gabriella M Harari. 2020. A process-oriented approach to respecting privacy in the context of mobile phone tracking. *Current Opinion in Psychology* 31 (Feb. 2020), 141–147. https://doi.org/10.1016/j.copsyc.2019.09.007

[3] Gabriella M Harari, Sandrine R Müller, Clemens Stachl, Rui Wang, Weichen Wang, Markus Bühner, Peter J Rentfrow, Andrew T Campbell, and Samuel D Gosling. 2020. Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of personality and social psychology* 119, 1 (2020), 204.

[4] Florian Künzler. 2019. Context-aware notification management systems for just-in-time adaptive interventions. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 435–436. https://doi.org/10.1109/PERCOMW.2019.8730874

[5] Elizabeth O Lillie, Bradley Patay, Joel Diamant, Brian Issell, Eric J Topol, and Nicholas J Schork. 2011. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personalized medicine* 8, 2 (2011), 161–173.

[6] Scott M. Lundberg, Gabriel G. Erion, and Su-In Lee. 2018. Consistent Individualized Feature Attribution for Tree Ensembles. *CoRR* abs/1802.03888 (2018). arXiv:1802.03888 http://arxiv.org/abs/1802.03888

[7] S. C. Matz, M. Kosinski, G. Nave, and D. J. Stillwell. 2017. Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences* 114, 48 (2017), 12714–12719. https://doi.org/10.1073/pnas.1710966114 Publisher: National Academy of Sciences _eprint: https://www.pnas.org/content/114/48/12714.full.pdf.

[8] Sarah Prange, Sven Mayer, Maria-Lena Bittl, Mariam Hassib, and Florian Alt. 2021. Investigating User Perceptions Towards Wearable Mobile Electromyography. In *Human-Computer Interaction – INTERACT 2021*. Springer International Publishing, Cham, 339–360. https://doi.org/10.1007/978-3-030-85610-6_20

[9] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-Scale Assessment of Mobile Notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3055–3064. https://doi.org/10.1145/2556288.2557189

[10] Albrecht Schmidt, Michael Beigl, and Hans Gellersen. 1999. There is more to context than location. *Computers & Graphics* 23, 6 (1999), 893–901. https://doi.org/10.1016/S0097-8493(99)00120-X

[11] Clemens Stachl, Quay Au, Ramona Schoedel, Samuel D Gosling, Gabriella M Harari, Daniel Buschek, Sarah Theres Völkel, Tobias Schuwerk, Michelle Oldemeier, Theresa Ullmann, et al. 2020. Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences* 117, 30 (2020), 17680–17687.