



Preface to the special issue on harnessing personal tracking data for personalization and sense-making

Frank Hopfgartner¹ · Judy Kay² · Amon Rapp³

Published online: 1 July 2020
© Springer Nature B.V. 2020

Abstract

Increasingly, people are making use of diverse digital services that create many types of personal data. The most recent addition to such services are self-tracking devices that are capable of creating very detailed personal activity records. The focus of this special issue is to explore how such activity records can be exploited to provide user-centric personalization services.

Keywords Self-tracking · Personalization · Special issue

1 Introduction

In recent years, growing numbers of people have taken up digital tracking to collect a variety of personal data via diverse tools, using devices ranging from desktops to smartphones, from ubiquitous system devices to wearable devices (Rapp et al. 2016). Examples include keeping records of social interactions, emails, and social media status updates, physiological and emotional status, or activities such as viewing television, use of time in general, driving habits, work productivity, monitoring

✉ Frank Hopfgartner
f.hopfgartner@sheffield.ac.uk

Judy Kay
judy.kay@sydney.edu.au

Amon Rapp
amon.rapp@unito.it

¹ University of Sheffield, Sheffield, United Kingdom

² University of Sydney, Sydney, Australia

³ University of Torino, Turin, Italy

environmental conditions, and so on. This phenomenon of self-tracking has had a vanguard group of early adopters, the so-called “Quantified Self” (QS) movement, an Internet community focusing on self-quantification through technological aids.¹ However, with the growing availability of personal data trackers, this phenomenon is now spreading to a far wider audience than the QS community (Rapp and Cena 2016). The number of tracking devices reached 225 million units in 2019 and various reports suggest that these numbers will double within the next 3 years (Gartner 2018). This makes it timely to tackle the core challenges that people face in making effective use of their personal tracking data.

This special issue brings together research that aims to transform tracking data into user models that can support personalization of software and open user model interfaces that enable individuals to self-reflect, self-monitor and plan how to achieve their long term goals. Such use of personal data to create user models creates three key challenges. First, it calls for exploration of effective ways to collect relevant personal information, by harnessing the emerging devices that make this increasingly easy to do across many spheres of life. Secondly, it deals with the multiplicity of challenges for transforming that data into user models that people can control and that address challenges they face. The third challenge concerns sharing partial data collections of certain aspects of one’s life, potentially with a worldwide audience, which can be used to create aggregate user models.

In this preface, we first summarize in Sect. 2 the manuscripts that have been accepted as part of this special issue. Section 3 then discusses the next steps ahead.

2 Accepted articles

We accepted a total of eight out of sixteen submitted manuscripts as part of this special issue. In the remainder of this section, we summarize the main research challenges addressed in these articles.

Karatzolou et al. (2020) explore how users’ personality and emotional state of mind can be exploited to predict locations. The premise behind their research is that users’ prior experience at specific locations combined with other contextual information are strong cues that can be exploited by personalization services. This work explores the ways that advances in wearable computing make it possible to predict users’ state of mind based on personal tracking data.

In this study, users’ signals are modeled using an ontology. In order to evaluate this ontological user profile, the authors first had to create a suitable dataset containing information of users visiting various locations, their reason for visiting these locations, as well as their emotional state at the time of their visit. For this, they created an app that allowed users anonymously to provide this information. In addition, users were asked to describe their personality to complete the data gathering task. The data of 13 users who contributed their data was then used to evaluate different location prediction methods.

¹ The QS movement’s main hub of communication is <http://quantifiedself.org>.

Focusing on exploring the challenge of training predictive models based on self-tracking and sensor data, Garcia-Ceja et al. (2019) argue that due to the diverse nature of data collected and user behaviors captured, such user models cannot rely on generic machine learning techniques. Therefore, they argue for user-dependent models that consider different modalities and expected user behaviors. Such user-dependent models, however, need to be built from the limited amount of data might be available.

In order to achieve this, the authors propose an approach for modeling the user's activity and for emotion detection based on deep transfer learning and data augmentation. Two publicly available datasets are used in their work: an activity recognition dataset consisting of accelerometer data recorded from six subjects while performing everyday activities and an emotion recognition dataset consisting of emotional utterances spoken by ten subjects.

Boratto et al. (2019) observe that users of self-tracking devices are motivated to share their fitness achievements on social media platforms. In their work, they study whether users' running records and contextual information can be exploited to predict whether users will share their runs on Facebook. As they argue, being able to predict users' motivation will be helpful for eCoaching apps that aim to keep users engaged and active.

They approach this by first creating a dataset consisting of approximately six months of workout data and local weather data that have been extracted from third party services. The workout data was recorded by users of an eCoaching app that generates tailored workout plans based on users' fitness levels and objectives. Only data of users who shared at least one of their runs on Facebook were considered. Each recorded run consisted of distance covered by the runner, workout duration, average speed, calories burnt, as well as information indicating whether it was shared on the users' Facebook page. This was augmented with meteorological data, including local temperature, dew point, humidity, and air pressure. In order to predict users' online sharing behavior, the authors then study various classification algorithms and evaluate them via 10-fold cross-validation.

Kraaij et al. (2019) study opportunities emerging from self-tracking devices to counter negative health conditions caused by sedentary lifestyles. In this article, they summarize several research studies that were carried out as part of a nationally funded collaborative research project aiming for the design of "user-centered ICT applications for self-management of vitality in the domain of knowledge workers". In particular, the project focused on the development of m-health applications that exploit self-tracking data to improve users' well-being.

The authors provide a comprehensive overview of the steps involved in the development of such m-health solutions. They started by conducting focus group discussions to understand users' expectations on data privacy. Then, to gain insights about knowledge workers' every-day activities and stress levels, they asked 25 users to perform common work activities such as writing a report, preparing a presentation, reading e-mails, etc. Aiming for a more authentic work context, the subjects' were confronted with several "stressors" such as working under time pressure of e-mail interruptions. All interactions with the computer, subjects' facial expressions, body gestures, and biometrics were recorded,

resulting in a multi-modal dataset. This dataset was then used for further research on visual analytics of work behavior data, work behavior classification, and classification of physical activities. Next, the authors introduce three different tools that advise users on their daily behavior. The paper concludes with a detailed discussion on lessons learned and a reflection on key limitations they encountered.

Gasparetti et al. (2019) study one of the main motivations of many self-trackers, namely the promise to lose weight by observing their behavior. As the authors point out, most activity tracking devices do not go beyond basic features such as monitoring steps taken or calories burned. Addressing this limitation, the authors develop personalized weight loss strategies based on the analysis of a large dataset of self-tracking data.

This dataset has activity data for over 10,000 users of health-monitoring devices (wristband activity trackers and digital scales) captured over a period of 1 year, as well as demographic information. The authors analyze this data to identify collective behavior clusters and to identify upward or downward activity trends. This is then used as the basis for a health recommender system that employs reinforcement learning to provide personalized recommendations.

Musto et al. (2020) point out that due to the heterogeneous nature of self-tracking devices and the many digital systems that people use, users' tracking data comes in many different forms and types. Their paper presents a platform called the Holistic User Model (HUM) that was designed to enable users to bring together their personal data from heterogeneous data sources to create a comprehensive user model, e.g., data extracted from social media, wearable devices and mobile phones each capture data about diverse aspects of users' lives. They present a pipeline that extracts personal data from diverse sources, processes this data, and stores it, then makes it available to the user at an interface that can support personal tracking and reflection that integrates multiple data sources.

They present the Myrror implementation of the HUM platform. They show how this can model diverse aspects, including demographic information, personal interests, their moods and emotions, psychological aspects, physical activities, social connections, and physical states. They explain its design and rationale to integrate data from social media (Twitter, Facebook, LinkedIn and Instagram), mobile phones (Android), and FitBit and this is processed with a range of NLP methods. The user profiling step of the pipeline then maps this data to the holistic user model. All this is made available to the user in an interface for tracking trends and seeing snapshots of interest profiles. There is careful design of the privacy control and management. The authors report a study of 40 users who used the Myrror over four weeks. This demonstrates the usability of the system that that participants found value in the services it provided.

Sanchez et al. (2019) tackle the important challenge people face in managing data access and sharing settings for wearable fitness trackers. To do this, the authors established an ontology to model relevant aspects of the data collected by a tracker aligned with the European Union's General Data Protection Regulations (GDPR). They then present a data-driven approach to recommend privacy settings for fitness trackers.

Paraschiakos et al. (2020) focus their work on activity recognition from accelerometer data. They present the Growing Old TOgether Validation (GOTOV) study for monitoring physical activity, which aims to serve multiple mobility and healthy aging studies in older adults. The authors generated a new activity recognition dataset, using multiple wearable sensors in a population of individuals over 60 years old. Then, they developed LARA, a method to learn robust and accurate activity recognition models, and deliver a sensor set-up analysis focusing on which body locations are the most efficient to monitor and predict physical activity. Finally, the authors present an activity recognition model that can be used in free-living data in order to recognize general physical activity patterns that can be associated with physiological health parameters.

3 Steps ahead

In this special issue, several of the articles deal with the challenge of creating user profiles and personalization services based on self-tracking and other heterogeneous user data. The broad promise of self-tracking devices is promoted as a way for people to do both short term tracking, for example within each day, as well as long term tracking, with reflection on the data. It is argued that such reflection and self-monitoring can help people achieve their overall fitness goals, be these to improve or maintain their levels of healthy activity. This health and fitness is strong focus in the research articles of this special issue: Kraaij et al. (2019) provide an overview of several research studies conducted to counter sedentary lifestyles. Similarly, Gasparetti et al. (2019) focus on developing personalized weight loss strategies for self-trackers. Looking at the specific case of runners, Boratto et al. (2019) explore their motivation to share summaries of completed runs on social media. Sanchez et al. (2019) deal with the needs to support people in managing privacy of their data. Paraschiakos et al. (2020) focus on self-tracking data to develop models aimed at recognizing older adults' activities, which may facilitate healthy aging.

The second main focus of this special issue is on the importance of users' emotions and current mental state. Garcia-Ceja et al. (2019) build user-dependent prediction models and test them using datasets consisting of accelerometer data and emotional utterances. Karatzolou et al. (2020) study how users' emotional state of mind can be incorporated when predicting locations. Musto et al. (2020) support modeling emotions based on diverse forms of long term data.

A third observation of this special issue is the need for user models that are capable of dealing with heterogeneous data sources. Here, Musto et al. (2020) provide an example case study in which they combine signals from different sources in one "holistic user model".

Thinking of further challenges, we argue that user models can now be expanded to make use of a variety of information that could be used to model the user's attitudes, emotions, tastes, physiology, movements, everyday behaviors, habits, working and learning performances, media uses, and preferences (Cena et al. 2019). Such information may create rich "QS user models", i.e., users models based on personal tracking data. In principle, these could model diverse aspects of the users' real and

digital life and be turned into life-long and holistic digital mirrors. However, some important research questions arise:

- How can we merge these heterogeneous data to obtain a comprehensive, semantic, and dynamic representation of the diverse aspects of users?
- How can we create reasoning tools for such data to create meaningful QS user models that can drive personalization?
- And can we enable predictions about the users' behavior, health, and objectives?
- How can we mine such data to detect and model trends in time and unexpected correlations among different aspects of their life?
- How can all of these be done in ways that match individuals' privacy preferences.

On the other hand, modeling all this information in a comprehensive representation of the user may create opportunities for new forms of personalization embedded in daily life, such as real-world, context-aware, just-in-time recommendations and services, tailored on the users' changing state, goals, and environments. Here further research questions appear to be fundamental:

- How can we deal with different, including conflicting, data sources to create user models to drive recommendations?
- How can we exploit information coming from self-tracking in personalized systems?
- How can we adapt applications by taking account of the continuous flux of data about diverse aspects of the user?
- How can we create user models that address people's privacy concerns and ensure people can understand the ways their user model was created and may be used?

Finally, when these enriched QS user models have suitable interfaces, they could become a form of lifelong and life-wide user model (Kay and Kummerfeld 2013) facilitating meta-cognitive processes of self-reflection, self-monitoring and planning, based on long-term user models.

References

- Boratto, L., Carta, S., Ibba, F., Mulas, F., Pilloni, P.: Modeling real-time data and contextual information from workouts in eCoaching platforms to predict users' sharing behavior on Facebook. *User Model. User Adapt. Interact.* (2019). <https://doi.org/10.1007/s11257-019-09229-4>
- Cena, F., Likavec, S., Rapp, A.: Real world user model: evolution of user modeling triggered by advances in wearable and ubiquitous computing: state of the art and future directions. *Inf. Syst. Front.* **21**(5), 1085–1110 (2019). <https://doi.org/10.1007/s10796-017-9818-3>
- Garcia-Ceja, E., Riegler, M., Kvernberg, A.K.: User adaptive models for activity and emotion recognition using deep transfer learning and data augmentation. *User Model. User Adapt. Interact.* (2019). <https://doi.org/10.1007/s11257-019-09248-1>

- Gartner: Gartner says worldwide wearable device sales to grow 26 percent in 2019 (2018). <https://www.gartner.com/en/newsroom/press-releases/2018-11-29-gartner-says-worldwide-wearable-device-sales-to-grow>
- Gasparetti, F., Aiello, L.M., Quercia, D.: Personalized weight loss strategies by mining activity tracker data. *User Model. User Adapt. Interact.* (2019). <https://doi.org/10.1007/s11257-019-09242-7>
- Karatzolou, A., Ebbing, J., Ostheimer, P., Hua, W., Beigl, M.: Sentient destination prediction. *User Model. User Adapt. Interact.* (2020). <https://doi.org/10.1007/s11257-020-09257-5>
- Kay, J., Kummerfeld, B.: Creating personalized systems that people can scrutinize and control: drivers, principles and experience. *ACM Trans. Interact. Intell. Syst. (TiiS)* **2**(4), 1–42 (2013)
- Kraaij, W., Verberne, S., Koldijk, S., de Korte, E., van Dantzig, S., Sappelli, M., Shoaib, M., Bosems, S., Achterkamp, R., Bonomi, A., Schavemaker, J., Hulsebosch, B., Wabeke, T., Vollenbroek, M., Neerinx, M., van Sinderen, M.: Personalized support for well-being at work: an overview of the SWELL project. *User Model. User Adapt. Interact.* (2019). <https://doi.org/10.1007/s11257-019-09238-3>
- Musto, C., Polignano, M., Semeraro, G., de Gemmis, M., Lops, P.: Myrror: a platform for holistic user modeling merging data from social networks, smartphones and wearable devices. *User Model. User Adapt. Interact.* (2020). <https://doi.org/10.1007/s11257-020-09272-6>
- Paraschiakos, S., Cachucho, R., Moed, M., Heemst, D.V., Mooijaart, S., Slagboom, E., Knobbe, A., Beekman, M.: Activity recognition using wearable sensors for tracking the elderly. *User Model. User Adapt. Interact.* (2020). <https://doi.org/10.1007/s11257-020-09268-2>
- Rapp, A., Cena, F.: Personal informatics for everyday life: how users without prior self-tracking experience engage with personal data. *Int. J. Hum. Comput. Stud.* **94**(1), 1–17 (2016). <https://doi.org/10.1016/j.ijhcs.2016.05.006>
- Rapp, A., Cena, F., Kay, J., Kummerfeld, B., Hopfgartner, F., Plumbaum, T., Larsen, J.E., Epstein, D.A., Gouveia, R.: New frontiers of quantified self 2: going beyond numbers. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, UbiComp '16*, pp. 506–509. Association for Computing Machinery, New York, NY, USA (2016). <https://doi.org/10.1145/2968219.2968331>
- Sanchez, O., Torre, I., He, Y., Knijnenburg, B.: A recommendation approach for user privacy preferences in the fitness domain. *User Model. User Adapt. Interact.* (2019). <https://doi.org/10.1007/s11257-019-09246-3>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Frank Hopfgartner is Senior Lecturer in Data Science and Head of the Information Retrieval Research Group at the University of Sheffield, United Kingdom. His main research interest is in the field of interactive information access and self-tracking.

Judy Kay is Professor of Computer Science and leads the Human-Centered Technology Research Cluster at the University of Sydney, Australia. Her research aims to create new technologies for human–computer interaction, particularly novel interfaces and scrutable user modeling.

Amon Rapp is Assistant Professor at the Computer Science Department, University of Torino, Italy, where he leads the Smart Personal Technology Lab at ICxT. His scientific research is situated within the area of human–computer interaction. It focuses on the investigation of the effects of self-tracking, behavior change and gaming technologies on people's everyday lives.