

Enabling an Ecosystem of Personal Behavioral Data

Jason Wiese, Jason Hong, John Zimmerman
Carnegie Mellon University
{jwwiese, jasonh, johnz}@cs.cmu.edu

Every computational system a person interacts with has the potential to keep a detailed log of that person's behavior. The potential of long-term personal logs across many data sources promises a breadth of new service opportunities for improving people's lives through deep personalization, tools to manage aspects of their personal wellbeing, and services that support identity construction. For example, applications might use data from communication logs to automate or partially automate the process of specifying sharing preferences by inferring the emotional closeness and life facet of the user's social relationships with other people [3, 5]. Opportunities exist for using this data through tools to manage aspects of the user's personal wellbeing like physical activity (i.e. observing movement with accelerometers and location) [2] or mental health (i.e. observing a user's calls and text messages) [1].

However, today a person's data is fragmented across a variety of service providers and resides in many different formats that were not designed to work together, which makes any attempts at a holistic, egocentric view of a user's data very difficult [4]. The siloed manner in which this data is currently being collected creates significant challenges for realizing the potential of this personal behavioral data. This siloed nature significantly raises the level of effort required to build such applications. There is no standard or reusable process for dealing with personal behavioral data. Application developers and researchers who wish to build on top of personal behavioral data typically follow ad-hoc processes that require lots of redundant effort, and these efforts are usually not reusable. This process could be drastically simplified.

This abstract argues that there are a broad variety of opportunities that we are missing because of the haphazard manner in which personal behavioral data is collected and managed, and that a variety of new applications would become viable if we could make it dramatically easier to capture and make use of personal behavioral data. To this end, we describe the concept of a *Personal Behavioral Data Service*.

The central concept behind this personal behavioral data service is simple: personal behavioral data should be unified on the level of the individual user (instead of siloed within individual services). The fundamental goal is to factor out as much redundant effort as possible for dealing with personal behavioral data. Figure 2 shows a rough design sketch for this kind of behavioral data service.

Such a service would need to support aggregating data, collecting ground truth, and building inferences. Furthermore, the service should also provide facilities by which users can give permission to client applications to access their data (raw or summarized) and inferences made with their data, which developers could access through some kind of API.

By employing this framework, application developers who do not have the resources to construct their own inferences can still make use of existing inferences and even help to improve them through the user's feedback. This process also creates the potential for consistency across independent applications and services: whether an inference is right or wrong, it is right or wrong across all of the user's various applications and services. If an incorrect inference is corrected once, it is corrected everywhere, globally improving the user experience.

There are a number of major challenges that must be solved to realize this vision of a personal behavioral data service:

Aggregating Raw Data – It should be easy to add additional sources of data to the personal behavioral data service. Security and data freshness are also important concerns to address.

Unified Representation of Raw Data – Raw data should be easily interconnected across different sources, but should not be so rigid that it sacrifices the granularity or completeness of the data.

Incentives for Data Providers – We want to encourage many behavioral data sources to aggregate data within the service.

Getting Ground Truth for Inference Models – Ground truth labels are essential to constructing inference models based on personal behavioral data. This challenge of collecting ground truth needs to account for the user's attention and needs.

Building Inference Models – What is the process for adding new models? What is the process for creating a new model? Where does the model run? Can we guarantee that the models are secure and do not leak data? How do we handle a model's uncertainty about an inference that it makes?

Useful Developer Tools – How should the API be designed, and what are some ways that we can make developer's lives easier?

Managing Privacy and Security Issues – Even if we assume that the service itself is completely secure from attackers (which is certainly not guaranteed), many challenges remain.

REFERENCES

- [1] Bardram, J.E. et al. 2012. The MONARCA self-assessment system. *In Proc. IHI '12*
- [2] Consolvo, S. et al. Activity sensing in the wild: a field trial of ubifit garden. *In Proc. CHI '08*.
- [3] Min, J.-K. et al. Mining smartphone data to classify life-facets of social relationships. *In Proc. CSCW '13*.
- [4] Odom, W. et al. Fragmentation and transition: understanding perceptions of virtual possessions among young adults in Spain, South Korea and the United States. *In Proc. CHI '13*.
- [5] Wiese, J. et al. Are you close with me? are you nearby? *In Proc. UbiComp '11*.

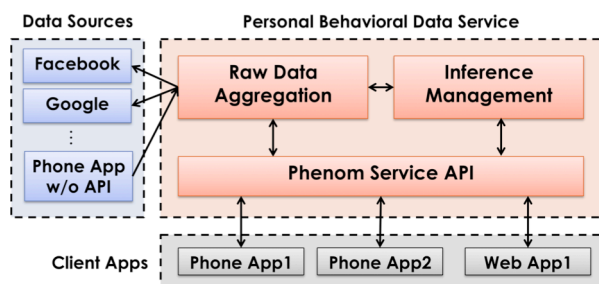


Figure 1: A basic system architecture for a personal behavioral data service, with a uniform interface for storing and accessing personal behavioral data.