

Usable Provenance for Personal Goal Assistants

See Fekete et al,

“Managing Information for Personal Goals (Vision)”

CloudDM workshop at ICDE’15, pp 30-33

Alan Fekete (Sydney University)

Personal Goals

- Health
 - More generally, well-being
- Education
- Finance
- Entertainment

Influencing behavior (nudges)

- Reminders
- Exhortations
- Competitions
- Rewards

- “what fitness devices do, at least for me, is make it harder to lie to myself”
<http://krugman.blogs.nytimes.com/2015/03/09/wearables-and-self-awareness-personal/>

Data-driven goal-support systems

- Data about you and others in content
- Data about you and others to choose when and how to nudge
- Data both recent and long-term

Data sources

- Devices you carry/wear
 - Eg step counter
- Pervasive devices in environment
 - Eg internet-enabled fridge
- Applications watching your online activity
 - Eg email sentiment analysis

Stores for source data

- Held by device vendor or monitoring application vendor
 - Often cloud-hosted
- Held by/on behalf of subject human
 - Often cloud-hosted
 - Cf Microsoft HealthVault, Apple HealthKit

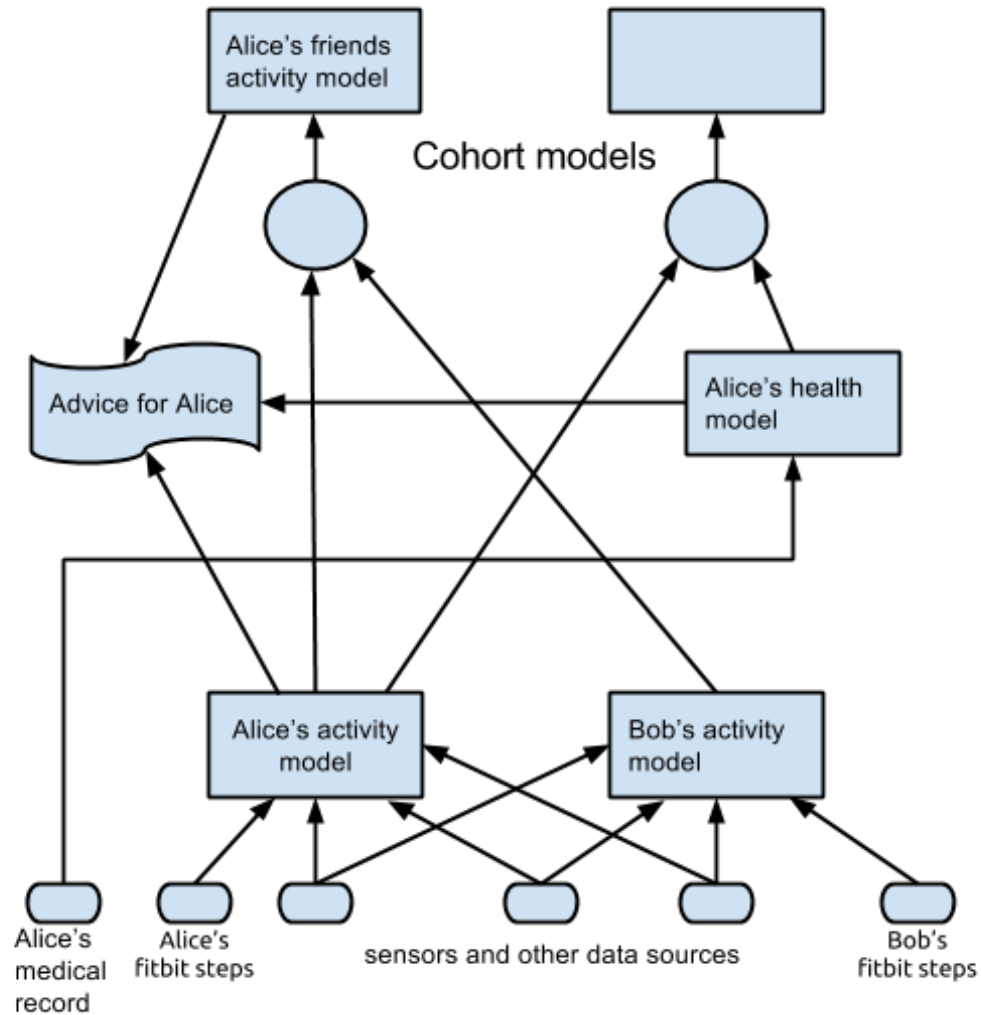
Derived data (cohort models)

- Combine data about the same aspect from many similar subjects
 - Eg Average weight of men in sedentary occupations aged 55-60
- Especially useful for competitive nudges
 - Eg leader boards

Derived data (user models)

- Combine data about many aspects of the same human subject
 - Eg Alan's weight history, food intake history, step history
- Especially useful for choosing best ways to nudge

Models from models



Characteristics of the dataspace

- Broad and deep chains of derivation
- Complex algorithms used
- Management spread among many platforms
 - APIs for data exchange
 - Integration as-you-go

Data which is about a person

- Special requirements
 - Legislative (different between jurisdictions)
 - Research ethics
 - Moral rights
- Issues of control
 - See data about you
 - Correct data about you
 - Remove data about you
- Issues of access control
 - Limit access
 - Discover access

Scrutability

- For personalized systems, it is expected that users will be able to find out why the system did something
 - Eg why was I given the recommendation to go for a walk now
- This covers both: what algorithm was used, and what data was relied on
- J. Kay and B. Kummerfeld, “Creating personalized systems that people can scrutinize and control: Drivers, principles and experience,” *ACM TiiS*, vol. 2, no. 4, p. 24, 2012.

Usable tracking and control of data flow

- Why it's hard
 - So many data items
 - So many places where the data is used
 - So much complexity in the derivation algorithms
 - So many federated, autonomous platforms

Vision

- Efficient tracking of data flow
- Separate important from unimportant flows
- Presentation that users can understand

Initial ideas

- Treat data in categories with different granularity
 - Eg recent data is observation-by-observation, older data is treated by category, oldest data is forgotten except in aggregate
- Derivation algorithms could come with information on sensitivity to source data
- Track data flow between platforms with summary metadata (like vector clocks)