**Health studies**

Collecting phenotyp and behavioral data of the participants to find correlations between those and health issues.
- Until recently mostly relying on diaries, written by the participants
- But often incomplete or difficult to get highly motivated people

The solution:
- Use people’s smartphones to automatically gather the contextual information
- Including their location, activity and environment
- Data is aggregated anonymously and fused with other sources, like pollution maps in our example.

**Challenges**

Leveraging user’s own devices as part of the “infrastructure”
- Rich sensors environment (internal + external + virtual)

Competing for system resources like CPU, memory, battery, ...
- ... but also user’s attention
- The goal is to minimize the load on the device and the user.

Two approaches:
- **Adaptively enable sensors**
  - Driven by application need and sensor availability
  - Collaborate with other devices
- **Continuously sense everything**
  - Richer dataset, training data
  - Performing data-mining offline

**TinyGSN**

TinyGSN is an Android background application with a front-end to change the parameters and select the sensors to use. It is based on the same principles as GSN (Global Sensor Networks): wrapper, virtual sensors, streamElements. A scheduler, optimized for gathering continuous location without depleting the battery, is managing the Android Services and Alarms to allow the device to sleep between the measurements.

**Sensor scheduling**

Based on a state machine, reflecting the user’s state.
- Turns off the GPS when not needed:
  - unavailable for a certain period (indoor)
  - user is not moving
  - WiFi access point doesn’t change
- Schedules the other sensors according to the state of the user:
  - moving or stationary, indoor or outdoor

**Data Processing**

Use the smartness of the device to reduce communication, processing and storage load.

**The idea**: aggregating the data into meaningful symbols and perform the usual processing and machine learning tasks on them. [2]

Several levels of symbolic representation (abstraction level).
- Constraints: unsupervised, online, limited memory and computation.

Symbols mapped to semantic meaning (not always)

**Level-0**: from global data distribution
- Using clustering, quantiles, expert knowledge,…
- Suitable for data-mining

**Level-I**: patterns composed of level-(i-1) symbols [3]
- Using online state recognition
- Maintain the recent history in a transition matrix
- Predictability to infer state changes

---

**References**