Sensor Data Streams

Stephen Voida, Donald J. Patterson, and Shwetak N. Patel

Presenters: Kevin Li, Kyungho Lee, Ren-Jay Wang
What are Sensor Data Streams?

- “The increased availability of inexpensive sensors, tremendous processing capabilities (even in mobile devices), high-bandwidth wireless networks, and vast quantities of data storage have made it much more practical to continuously collect streams of low-level data about people and their environments.”

- Data is collected
  - Extensively
  - Automatically
  - Quickly
  - Over long periods of time
Examples of sensors
History of Sensor Data Streams

- Evolved from quantitative data methods employed in experimental psychology
- Decreasing cost and increasing availability allow more sensors to observe a single phenomenon
- Improvements in data analysis allow easier interpretation.
- Experience Sampling Method - not as useful
What kinds of questions can we answer?

- Who are people with?
- What tools or resources do they use throughout the day?
- When do people normally do certain tasks?
- Where do people go over the course of a day?
- How healthy are a person’s behaviors?

What is missing?
What kinds of questions can we answer? (2)

- Egocentric methods: Focused on an individual
- Group-centric methods: Focused on group and person-to-person interactions
- Space-centric methods: Focused on a certain space or area regardless of particular inhabitants
  - IMS: Infrastructure Mediated Sensing = piggybacking
Why use sensor data streams?

- I.e. How is it different from other methods of sensing?
- Very fine-grained events
- “Distinguished by source...and their level of interpretive fidelity”
- Data is timestamped and originates from environment and user interactions
Sensor data streams vs Context-aware Computing

- In context-aware computing, the user’s implicit behavior serves as an auxiliary input to the system
- Studying user behaviors vs developing interactive systems
- Post-experiment analysis vs real time
- Generate user model vs predict user behavior with existing model
Limitations of the approach

- “[S]ensor data generally does a poor job of answering questions of why things have happened in the real world.”
- Clear understanding of the phenomena is required
- Additional tools to process and interpret raw data may be required
- Difficult to make changes once begun
- Can be expensive, unreliable
- Noisy output
- Large amounts of output
- Technical complexity
Privacy and Ethical Concerns

- Compared to human observers, sensors are far less intrusive
- Could collect data on human participants and passers-by without their knowledge
- Must inform participants of the scope of the data collection
  - Provide sample data sets when requesting consent
  - Possibly allow participants to revoke consent to be recorded
Building, Selecting, and Acquiring Sensors

- Two main types of sensors
  - Sensors that are worn or carried
  - Sensors that are deployed in a particular space and record actions in that space
- Take into account the degree of intrusiveness
Determining Frequency and Fidelity of Data

- Tradeoffs between higher quality data and storage/bandwidth, processing, and power requirements
- Try to vary fidelity based on other input
  - GPS readings only when accelerometer detects motion
  - Transmit data only when at Wi-Fi access point
- Privacy is another issue, behaviors can be reconstructed even from low-fidelity data
  - Using video cameras as motion detectors can create negative reaction even if participants are assured no video is being collected
Installing the Sensors

- When covering large spaces, there are added complexity challenges with the installation and maintenance of sensors
  - High labor costs during installation
  - Difficulty managing a large network of sensors
  - Handling power sources (e.g. replacing batteries, etc)
- Wizard-of-Oz method to mitigate risk
  - Have researcher visit site in person
  - Manually record the same data that would be expected of automated sensors
  - Assess if results are promising and if sensor deployment is worth the cost
Storing the Collected Data

- Two most common options for capturing data streams:
  - Record data at a device connected to the source, periodically collect a copy of the collected data
    - More technically straightforward
    - Does not require a constant network connection
  - Continually transmit data in real time through a network connection
    - Easier to measure volume of data
    - Can dynamically add storage capacity and maintain backups
    - Have to consider battery life and transmission costs
Making Sense of Data Collected

- Standard statistical methods can be used
  - Identifying correlations between conditions and observed behaviors
- Machine learning approaches may be more valuable due to the volume and noise inherent in sensor data
- Oftentimes, want to classify the data stream into segments or identify particular events in data
User Modeling and Event Detection

- Identify the data that will serve as inputs to the classifier
- Establish the classification categories or metric to be detected
  - Classification task
    - Map to discrete number of categories
  - Regression task
    - Map to a continuous real-valued metric
- “Gold standard” data
- Different hardware can produce different results due to calibration, sensitivity, etc.
Validating User Models

- **Cross-validation**
  - Break up gold-standard data into subsets
  - Use different permutations of these subsets as train and test sets

- **Data points must be independent**
  - Sensors recording a running average of previous data points

- **Special cases require separating data in unique ways**
  - Gesture data should be separated such that different individuals are in different folds
What to Report in a Study Using Sensors

- Report in sufficient detail such that another research could reproduce the study
  - Hardware
  - Experimental setup
  - Participant knowledge
  - Experimental execution
  - Software infrastructure
  - Analysis
  - Reasoning behind these choices
Summary: Research with Data Stream

1. Forming Research Goals/Questions
2. Building Sensors and Acquiring Data with Sensors
3. Analyzing Data and Building a Model
4. Event Detection
5. Making Sense of Data Stream
Motivation and Research Questions
- WiFi signals are available almost everywhere and they are able to monitor surrounding activities
- How to build a robust human activity recognition model using commercial WiFi devices?

Sensors
- Wireless router NETGEAR JR6100 and TP-Link TL-WDR7500

Data Stream
- Channel State Information (CSI) from 5Ghz frequency band with 20MHz bandwidth channels

Figure 2: Multi-paths caused by human movements
[MobiCom2015] Understanding and Modeling of WiFi Signal Based Human Activity Recognition, Wei Wang and Alex X. Liu

- Research Questions
  - How to build a robust human activity recognition model using commercial WiFi devices?

- Evaluation (what to report in a study using sensors)
  - To demonstrate the robustness of the proposed model, the authors used a 10-fold cross-validation to measure the accuracy of Activity Detection and Activity Recognition rate
  - No user studies
  - The authors addressed the weakness of the proposed model

Real-Time Data Stream from Sensors

Channel State Information

Event Detection
Activity Recognition

Output
9 Different Activity
e.g. Running, Walking, Sitting down...

Figure 15: Accuracy in different environments

- Motivation and Research Questions
  - Recognition of complex activities of objects in videos requires effective representations for further analysis
  - How to model motion trajectories in the spatio-temporal domain using 3D flow fields generated from a GPRF?

- Sensors
  - Traffic camera installed at a great distance about 640 meters (2,100 feet)

- Data Stream
  - Hours of RGB data
GPRFs for other data sets
Each video has different patterns and frame rates

www.justin.tv/adamsblock
OceanCity data (Kitware)
UCF Crowd Sementation data

- Research Questions
  - How to model motion trajectories in the spatio-temporal domain using 3D flow fields generated from a GPRF?

- Evaluation (what to report in a study using sensors)
  - To demonstrate the effectiveness of the proposed model, the authors compared its results to a well-known trajectory matching algorithm such that DDTW
  - No user studies
  - The authors addressed the limitations about the model as future work

<table>
<thead>
<tr>
<th></th>
<th>GPRF</th>
<th>DDTW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set</td>
<td>Accuracy</td>
<td>Precision</td>
</tr>
<tr>
<td>Adam</td>
<td>0.9555</td>
<td>0.9843</td>
</tr>
<tr>
<td>Ocean</td>
<td>0.966</td>
<td>0.8782</td>
</tr>
<tr>
<td>UCF</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>CLIF</td>
<td>0.9929</td>
<td>0.9488</td>
</tr>
</tbody>
</table>

Table 2: Trajectory recognition using complete trajectories. Each value denotes the average of each class. Our proposed approach works better in most of datasets except UCF data, which has small number of very distinctive motion patterns. See Fig. 8
[CHI2010] Skinput: appropriating the body as an input surface, Chris Harrison et al.

- **Motivation and Research Questions**
  - The human body as an input device is appealing not only because we have roughly two square meters of the skin, but also it is easily accessible by our hands
  - How to build a novel interface using the skin as an input surface with acoustic transmission?

- **Sensors**
  - A Customized, wearable bio-acoustic sensing armband

- **Data Stream**
  - Spectrum of frequencies (resonance of the skin) transmitted through the arm with 10 channels at 77Khz

---

**Figure 2.** Transverse wave propagation: Finger impacts displace the skin, creating transverse waves (ripples). The sensor is activated as the wave passes underneath it.

**Figure 3.** Longitudinal wave propagation: Finger impacts create longitudinal (compressive) waves that cause internal skeletal structures to vibrate. This, in turn, creates longitudinal waves that emanate outwards from the bone (along its entire length) toward the skin.
[CHI2010] Skinput: appropriating the body as an input surface, Chris Harrison et al.

- **Research Questions**
  - How to build a novel interface using the skin as an input surface (with acoustic transmission)?

- **Evaluation (what to report in a study using sensors)**
  - To demonstrate the effectiveness of the interface, the authors tested it with possible location combinations (Fingers, Whole Arm, Forearm)
  - Conducted user studies with 13 participants
  - Used the mean and standard deviation of the accuracy to compare the results

---

**Real-Time Data Stream from Sensors**

*Acoustic Data (wave) in 10 Channels*

---

**Event Detection**

*Tap Resonance Recognition*

---

**Output**

*Location of Finger taps on the arm*

---

**Figure 9.** Higher accuracies can be achieved by collapsing the ten input locations into groups. A-E and G were created using a design-centric strategy. F was created following analysis of pre-location accuracy data.
[CHI2016] TimeAware: Leveraging Framing Effects to Enhance Personal Productivity, Young-Ho Kim et al.

- **Motivation and Research Questions**
  - The authors argue that leveraging frame effects will help people enhance their productivity
  - How to build a self-monitoring system for capturing and reflecting on personal computer use to explore the RQ?

- **Sensors**
  - Laptops and Desktops

- **Data Stream**
  - Names of active applications at a 5 minute-interval via RescueTime’s API over 8 weeks (4874 hours of data)

Figure 2. TimeAware information dashboard for the Positive Framing condition: (A) the calendar navigator, (B) the summary panel, (C) the history chart, (D) the hourly trends panel, and (E) the top 10 activities panel.
and a visualization dashboard.
[CHI2016] TimeAware: Leveraging Framing Effects to Enhance Personal Productivity, Young-Ho Kim et al.

- **Research Questions**
  - How to build a self-monitoring system for capturing and reflecting on personal computer use?

- **Evaluation (what to report in a study using sensors)**
  - To evaluate the effects of different framing (positive and negative), the authors conducted longitudinal user studies with 24 people over 8 weeks.
  - The authors found how different framing affected productivity, engagement and self-awareness in two different groups.
  - Used t-test and post-hoc comparisons using Holm-Bonferroni correction to seek significant differences.
Conclusion

- Using sensor data streams is beneficial if researchers seek for strong and solid empirical evidences.
- Forming right research questions is still crucial to collect very specific data with sensors.
- Trying commercial sensors that fit your scope first, then considering building customized sensors.
- Participants should be fully and appropriately informed about the scope and nature of the data collection!
References