A METHODOLOGY FOR EXTRACTING TEMPORAL PROPERTIES FROM SENSOR NETWORK DATA STREAMS

Dimitrios Lymberopoulos, Microsoft Research
Athanasios Bamis and Andreas Savvides, Yale University
Spatiotemporal Sensor Streams

- Reliable sensor data collection already shown
  - Wireless sensor networks (vibration, humidity, etc.)
  - Mobile phone sensors (GPS, accelerometer, etc.)
- Spatial context is usually easy to extract
  - Closely tied to sensor’s location and/or type
- Temporal context is hard to extract
  - When and for how long does an event take place?
  - Semantic interpretation of events depends on this
Assisted Living
Spatiotemporal Mobile Dataset

- Sequence of timestamped locations

When and for how long is a location visited? When and for how long is an activity taking place?
Beyond Assisted Living

- **When and for how long...**

  ... is this road congested?

  ... is this data center’s workload high?

  ... do I visit a specific location?
Challenges

- Time is continuous
  - Appropriate discretization is needed
- Multiple temporal scales may co-exist
  - There might not be a unique answer.
- Temporal context depends on the source of the sensing events
  - e.g. Different people sleep at different times
- Sensor-specific
  - Temporal properties are different across sensors
Approach Overview

Discrete spatial events with continuous time information
Approach Overview

Discrete spatial events with continuous time information

Event Type Classification
Approach Overview

Discrete spatial events with continuous time information

Event Type Classification

Temporal Organization in Episodes of duration $t$
Approach Overview

Discrete spatial events with continuous time information

Event Type Classification

Temporal Organization in Episodes of duration $t$

Event Clustering Over Episodes
Approach Overview

Discrete spatial events with continuous time information

Event Type Classification

Temporal Organization in Episodes of duration $t$

Event Clustering Over Episodes

Temporal Conditioning

Discrete spatiotemporal events
Sensor Stream Model

\{ E_i(t_i, d_i) \}

- Sensor event $E_i$
  - Location, activity, road traffic, server workload, etc.
- $E_i$ is represented as a one-dimensional line segment
Cluster frequently similar line segments across episodes

1. Quantify similarity between two line segments
2. Extend to clusters of line segments
3. Cluster construction
Pairwise Density

\[ D_{ij} = 1 + \frac{T - d_{ij}}{T}, \quad d_{ij} = |t_i - t_j| + |(t_i + d_i) - (t_j + d_j)| \]

\[ D_{ti} = 1 + \frac{d}{d_i + d_j - d} \]

\[ D_{tj} = 1 + \frac{d_i}{d_j} \]

\[ D_{tj} = 1 - \frac{d}{d_i + d_j + d} \]
Cluster Quality Metrics

- The quality of a random cluster $C_k$ with $n_{Ck}$ instances (out of total $n$) is:

$$Q_{C_k} = \begin{cases} \frac{\sum_{i<j} D_{ij}}{\binom{n}{2}} \times \frac{n_{C_k}}{n} & \text{if } n_{C_k} \geq 2 \\ \frac{1}{n} & \text{if } n_{C_k} = 1 \end{cases}$$

- Construct the clusters by maximizing the average cluster quality

$$\frac{1}{N_C + n} \times \sum_{i=1}^{N_C} Q_{C_i}$$
Cluster Construction

- Greedy approach
  - Identical to Agglomerative Clustering

1. Every line segment is a cluster
2. Examine all possible mergings of pairs of clusters
3. Merge the two clusters that maximize average cluster quality
Experimental Results

- Simulation results
  - Verify the properties of the algorithm on artificially generated data
  - Controlled input -> verifiable output

- Assisted living deployment data
  - 30-day dataset from one of our assisted living deployments
  - Find when and for how long each detected activity takes place on a daily basis.
Simulation Results

- Mixture of gaussian, binomial and random distributions
Assisted Living Data

clusters of size > 1 for "Sleep" activity
Assisted Living Data

Clusters of size > 1 for "Hangout" activity
## Discovered Temporal Context

<table>
<thead>
<tr>
<th>Activity Cluster</th>
<th>Average Start Time</th>
<th>Average Duration (minutes)</th>
<th>Norm. Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Sleep</td>
<td>11:00:00PM</td>
<td>301</td>
<td>0.17</td>
</tr>
<tr>
<td>Regular Sleep</td>
<td>11:54:37PM</td>
<td>458</td>
<td>0.39</td>
</tr>
<tr>
<td>Early Morning Sleep</td>
<td>06:34:38AM</td>
<td>172</td>
<td>0.19</td>
</tr>
<tr>
<td>Morning Sleep</td>
<td>07:55:09AM</td>
<td>104</td>
<td>0.12</td>
</tr>
<tr>
<td>Night Sleep</td>
<td>04:21:30AM</td>
<td>322</td>
<td>0.1</td>
</tr>
<tr>
<td>Regular Out</td>
<td>03:16:57PM</td>
<td>174</td>
<td>0.64</td>
</tr>
<tr>
<td>Late Out</td>
<td>04:36:26PM</td>
<td>191</td>
<td>0.21</td>
</tr>
<tr>
<td>Morning Hangout</td>
<td>10:26:03AM</td>
<td>284</td>
<td>0.23</td>
</tr>
<tr>
<td>Afternoon Hangout1</td>
<td>05:50:52PM</td>
<td>85</td>
<td>0.18</td>
</tr>
<tr>
<td>Afternoon Hangout2</td>
<td>06:50:53PM</td>
<td>36</td>
<td>0.11</td>
</tr>
<tr>
<td>Evening Hangout2</td>
<td>09:36:00PM</td>
<td>103</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Enabling Accurate Models
Conclusions

- Data driven approach to extracting the temporal properties of sensor streams
  - Concurrently mine time, duration and frequency across episodes

- Key aspects
  - Represent events as line segments over time
  - Introduce accurate density metrics

http://bscope.eng.yale.edu
Questions?

dlymper@microsoft.com
Backup Slides
Deployment Data

Hangout Activity

average start times and durations

average start times and durations

frequency # of lines in cluster / total # of lines

normalized cluster quality

hour of the day
Comparison Results

![Comparison Results Graph]

- **Proposed Algorithm**
- **GA Agglomerative**
- **k-means**

Categories:
- Sleep
- Hangout
- Out

Graph shows the global quality for each category.
Cluster Quality Convergence

![Graph showing the global quality convergence over the number of cluster mergings.](#)