Spatio-temporal Range Searching
Over Compressed Kinetic Sensor Data

Sorelle A. Friedler
Joint work with David Mount
University of Maryland, College Park
Motivation

- **Kinetic data**: data generated by moving objects
- Sensors collect data
- Large amounts of data
- Collect and perform lossless compression
- **Goal**: Retrieve without decompressing
- **Long Term**: Analyze
Motivation

- **Computer Science**
  - Graphics: Image and video segmentation, animation
  - Databases: Maintenance over time
  - Sensor Networks: Data analysis
  - Cell phone users: Motion data analysis
    - 4.6 billion subscribers worldwide (in 2009)
    - 4.1 billion text messages per day in the US (in 2009)

- **Biology**
  - Mathematical ecology: Migratory paths, invasive species
  - Genomic data analysis: HIV strain analysis

- **Engineering**
  - Traffic patterns and identification
Our Framework

- Detection region around each sensor (stationary sensors)
- Point motion unrestricted
- No advance knowledge about motion
- Each sensor reports the count of points within its region at each synchronized time step
- \textit{k-local}: Sensor outputs statistically dependent only on \textit{k} nearest neighbors
Data Collection

Data based on underlying geometric motion

Sensor data streams

<table>
<thead>
<tr>
<th>X_1</th>
<th>X_2</th>
<th>X_3</th>
<th>X_4</th>
<th>X_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

time
Range Searching: Our Problem

Compress and preprocess the data so as to perform...

- **Temporal range query**: Given a time interval, return an aggregation of the counts over that time interval.
  
  - aggragation type: sum
  
  \[ t: 1 2 3 4 5 6 7 8 9 10 11 \]
  
  \[ X: 0,0,4,4,5,4,3,3,1 1, 0 \]
  
  \[ 00110123... \]
  
  \[ 00223101... \]
  
  \[ 11122021... \]
  
  \[ 4 + 6 = 10 \]

- **Spatio-temporal range query**: Given a time interval and spherical spatial region, return an aggregation of the counts over that time interval and within that region.
Lempel-Ziv Dictionary Compression [LZ78]

Create a trie while scanning through a string. The compressed string contains pointers to this dictionary.

(LZ78 is an optimal entropy encoding algorithm.)
Temporal Range Searching

- Create trie with accompanying pointers
- Annotate trie with aggregate values and word start times
- Given a temporal range \([t_0, t_1]\) find the anchor points \(s^0\) and \(s^1\) such that \(s^0 \leq t_0\) and \(s^1 \geq t_1\) (binary search)
- Use stored prefixes, words, and subtraction of prefixes to find aggregates

<table>
<thead>
<tr>
<th>Query Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>overlapping query: ([4,7])</td>
</tr>
<tr>
<td>(2 + 3 + 2 = 7)</td>
</tr>
<tr>
<td>internal query: ([10,10])</td>
</tr>
<tr>
<td>(3 - 1 = 2)</td>
</tr>
</tbody>
</table>
Spatio-temporal Range Searching

Overview

- Cluster sensors into regions based on locality
- Compress clusters separately
- Associate each cluster with a temporal range structure
- Search clusters hierarchically with a quadtree variant
Results

<table>
<thead>
<tr>
<th></th>
<th>Temporal</th>
<th>Spatio-temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprocessing time</td>
<td>$O(\text{Enc}(X))$</td>
<td>$O(\text{Enc}(X))$</td>
</tr>
<tr>
<td>Query time</td>
<td>$O(\log T)$</td>
<td>$O((1/\varepsilon^{d-1}) + \log S \log T)$</td>
</tr>
<tr>
<td>Space</td>
<td>$O(\text{Enc}(X))$</td>
<td>$O(\text{Enc}(X) \log S)$</td>
</tr>
</tbody>
</table>

- $X$: The set of sensor system observations
- $\text{Enc}(X)$: The encoded size (in bits) of the compressed data
- $T$: The total time over which data was collected
- $S$: The total number of sensors
- $d$: The dimension of the sensor space
- $\varepsilon$: An error parameter (for approximate range searching)

First range searching bounds over compressed data
Experimental Results: Locality

C. R. Wren, Y. A. Ivanov, D. Leigh, and J. Westbues.
The MERL motion detector dataset: 2007 workshop on massive datasets.
Technical Report TR 2007-069,
C. R. Wren, Y. A. Ivanov, D. Leigh, and J. Westbues.
The MERL motion detector dataset: 2007 workshop on massive datasets.
Technical Report TR 2007-069,
Experimental Results: Time

C. R. Wren, Y. A. Ivanov, D. Leigh, and J. Westbues.
The MERL motion detector dataset: 2007 workshop on massive datasets.
Technical Report TR 2007-069,
Thank you!
Questions?
(sorelle@cs.umd.edu)