Sparse GPS Trajectory Data Compression and Recovery based on Compressed Sensing

Hong Yang, Zhenyu Wang
Department of Modeling, Simulation and Visualization Engineering
Old Dominion University
November 20, 2015
Motivations

- Massive GPS/Smartphone trajectory data
- Privacy concern
- Data needs/applications
- Storage/processing issues
Challenges for Over-Compressed Data

Only Origin and Destination are available. (i.e. current NYC Taxi trip data)

Origin, Destination and if given some points in between
What if we provide more data points?

A better representation of the route?
What if we provide more data points?

- Different compressibility of the GPS trajectory may have different effects on different networks.
Existing Research / Practices

• Several classical methods using spatial and temporal dimensions to compress data
  – Uniform sampling
  – Douglas-Peucker algorithm
  – Bellman’s algorithm
  – STTrace algorithm

• Other new methods using dimensions like sparsity and category
  – Greedy matching pursuit algorithm (GMP)
  – Compressed sensing (CS)
  – Coupled Hidden Markov Models
Methodology: Douglas-Peucker (DP) Algorithm

- DP - Using the spatial information

- Step 1: Link nodes 1 and 8
- Step 2: Identify node with maximum distance (Node 4)
- Step 3: Iteration
Issues of DP Algorithm

- Threshold of DP and the compression rate

**Compression for highway data**

**Compression for local street data**

Some Issues:
- Need to sample all the data at the beginning
- Hard to deal with loop in the trajectory
Methodology: Compressed Sensing

- CS – using the sparsity information

\[\begin{align*}
y &= \begin{bmatrix} A \\ x \end{bmatrix}
\end{align*}\]

The sparsity of the trajectory in the sparse basis is the key point $\text{sparse}(A) > 2k$.

The recovery of the 2-norm of $x$ is a convex problem.

\[A = \begin{bmatrix}
a_{11} & a_{12} & 1 & 1 & a_{15} \\
a_{21} & a_{22} & 1 & 1 & a_{25} \\
a_{31} & a_{32} & 1 & 1 & a_{35} \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0
\end{bmatrix},
\]

\[x_1 = 1, x_2 = 0, y = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.
\]
Methodology: Compressed Sensing

- $y = Ax = sample\_matrix \times basis\_matrix \times x$
  
  $= sample\_matrix \times Tr$

  $y$: sampled data ($m \times 1$)  
  $sample\_matrix$: $m \times n$

  $x$: $n \times 1$  
  $basis\_matrix$: $n \times n$  
  $Tr$: raw data ($n \times 1$)

- Choose suitable $sample\_matrix$ and $basis\_matrix$ to get a sparse representation $x$ of $y$

- The mode (BCS)
Methodology: Compressed Sensing (CS)

- Framework of CS in GPS data compression

\[ X \rightarrow X_c \rightarrow \text{Compressed recovery} \rightarrow X^\sim \rightarrow \text{GPS data} \]

\[ Y \rightarrow Y_c \rightarrow \text{Compressed recovery} \rightarrow Y^\sim \]

\[ \text{Down sampling uniformly} \]

\[ \text{MSE} \]

HY@ODU
Simulation Results

• Test Scenarios

• Data and Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_1$</td>
<td>The interval of trajectory observation (highway)</td>
<td>3s</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>The interval of trajectory observation (local street)</td>
<td>10s</td>
</tr>
<tr>
<td>$N$</td>
<td>The length of the processed sequence</td>
<td>60</td>
</tr>
<tr>
<td>$n$</td>
<td>The number of Monte Carlo simulations</td>
<td>100</td>
</tr>
<tr>
<td>$\theta$</td>
<td>The dB value of Gaussian noise</td>
<td>15</td>
</tr>
</tbody>
</table>

• Performance measure

$$\text{Error}(x, y, x^\sim, y^\sim) = \frac{\| (x^\sim, y^\sim) - (x, y) \|_2}{\| (x, y) \|_2}$$
Simulation Results

- Performance of DP vs. CS with compression rate = 0.5

<table>
<thead>
<tr>
<th>Data</th>
<th>Algorithm</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle 1 (highway)</td>
<td>DP</td>
<td>2.7*10^{-6}</td>
</tr>
<tr>
<td>Vehicle 1 (highway)</td>
<td>BCS</td>
<td>3.4*10^{-11}</td>
</tr>
<tr>
<td>Vehicle 2 (local street)</td>
<td>DP</td>
<td>0.1291</td>
</tr>
<tr>
<td>Vehicle 2 (local street)</td>
<td>BCS</td>
<td>9.1*10^{-10}</td>
</tr>
<tr>
<td>Vehicle 3 (highway)</td>
<td>DP</td>
<td>3.9*10^{-6}</td>
</tr>
<tr>
<td>Vehicle 3 (highway)</td>
<td>BCS</td>
<td>1.1*10^{-11}</td>
</tr>
<tr>
<td>Vehicle 4 (local street)</td>
<td>DP</td>
<td>0.1825</td>
</tr>
<tr>
<td>Vehicle 4 (local street)</td>
<td>BCS</td>
<td>7.7*10^{-10}</td>
</tr>
<tr>
<td>Vehicle 5 (highway)</td>
<td>DP</td>
<td>3.3*10^{-6}</td>
</tr>
<tr>
<td>Vehicle 5 (highway)</td>
<td>BCS</td>
<td>1.06*10^{-11}</td>
</tr>
<tr>
<td>Vehicle 6 (local street)</td>
<td>DP</td>
<td>0.2132</td>
</tr>
<tr>
<td>Vehicle 6 (local street)</td>
<td>BCS</td>
<td>4.3*10^{-10}</td>
</tr>
</tbody>
</table>

- Performance with Gaussian noise of 15dB in highway

<table>
<thead>
<tr>
<th>Compression rate</th>
<th>Scenario</th>
<th>DP Error</th>
<th>BCS Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>highway</td>
<td>0.1040</td>
<td>0.0697</td>
</tr>
<tr>
<td>0.2</td>
<td>highway</td>
<td>0.1192</td>
<td>0.1041</td>
</tr>
</tbody>
</table>
Simulation Results

- Trace and recovered trace by DP & BCS (compression = 0.5)
Discussion

- Loss/Distortion of information (acceleration, speed, travel time, etc.)

10 real points on local street

CS Approach

10 real points on highway

CS Approach
Concluding Remarks

• Raw GPS data can be represented relatively well by using appropriate compression techniques

• The proposed BCS approach can achieve relatively higher compression rate but maintain a better performance

• Despite the complexity, BCS approach does not require to store all raw data before sampling

• Compression means information loss/distortion (Consider trade off between compression rate and information change)
References


• Bellman, R. On the approximation of curves by line segments using dynamic programming, Commun. ACM, Vol. 4, No. 6, 1961, pp. 284,


• Zhang, B., X. Cheng, N. Zhang, Y. Cui, Y. Li, and Q. Liang, Sparse target counting and localization in sensor networks based on compressive sensing, in Proc. IEEE INFOCOM, 2011


Thank You Very Much!

Hong Yang, Ph.D.
Assistant Professor
Dept. of Modeling, Simulation & Visualization Engineering
Old Dominion University
Norfolk, VA
Email: hyang@odu.edu