Understanding and Controlling Instrumented Physical Systems: Modeling is Complex, but Optimization is Easy
Bradley Bennett, Alberto Cerpa, Jessica Feng, Farinaz Koushanfar, Richard Park, Thomas Schoelhammer, Jennifer L. Wong, and Miodrag Potkonjak
Group for Stochastic Algorithmics  http://www.cs.ucla.edu/~miodrag

Introduction: Data Integrity and Computational Sensing

We address two of the canonical problems in sensor networks:
• Data integrity
• Computational sensing.

Due to the large scale and distributed nature of sensor networks, their heterogeneous node structure, cost and power constraints, operation in unpredictable and unconditioned (and often harsh) environmental surroundings and inherent unreliability of sensors with embedded sensors often collect data with errors, faults and missing samples.

Problem Description: Inter-sensor Modeling and Prediction

• Scientists and application engineers
  – Analysis and synthesis, application
  – Statistical data-driven models
• Sensor network engineers
  – Modeling and management
  – Energy and cost efficient and secure sensor networks
• Combinatorial Domain-based Statistical Modeling
  – Flexibility of the combinatorial modeling
    – Nodes: Different error norms, outlier elimination, robust regression, etc.
    – Edges: Maximum/minimum slope, adding and removing ordering constraints, etc.
    – Paths: Unimodular, convex, locally monotonic, number of break points, etc.
    – Optimization-friendly
• Density estimation, consistency-based techniques

Experimental Results: Modeling, Prediction, and Optimization

Inter-sensor Modeling

• Data from sensors X and Y, find the model \( y = f(x) \)
• Hidden covariate problem captured by isotonicity constraint:
  – For model \( f: x_1 < x_2 < \ldots < x_K \)
• Univariate CIR (Combinatorial Isotonic Regression):
  – Given data \((x_i, y_i, \omega_i), i=1,\ldots,K\)
  – Given an error measure \( \epsilon_i \) and \( x_1 < x_2 < \ldots < x_K \)
  – \( \epsilon_i \) isotonic regression is set \((x_i, y_i), i=1,\ldots,K\) s.t.
  – Objective function: \( \min \epsilon_i(x_i, y_i, \omega_i) \)
  – Constraints: \( y_1 \leq y_2 \leq \ldots \leq y_K \)

Modeling

Linear fit vs. CIR fit and its bounds

Prediction error over all nodes

Limiting number of parameters - AIC criteria

Combinatorial Domain-based Statistical Modeling

<table>
<thead>
<tr>
<th>Mean Error Bound</th>
<th>Linear</th>
<th>Nonpar</th>
<th>CIR</th>
<th>Lim CIR</th>
<th>Average CPLEX Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>2%</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>0.31</td>
</tr>
<tr>
<td>3%</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>1.16</td>
</tr>
<tr>
<td>4%</td>
<td>4</td>
<td>4</td>
<td>10</td>
<td>9</td>
<td>1.16</td>
</tr>
<tr>
<td>5%</td>
<td>6</td>
<td>8</td>
<td>13</td>
<td>12</td>
<td>0.98</td>
</tr>
<tr>
<td>6%</td>
<td>7</td>
<td>12</td>
<td>24</td>
<td>17</td>
<td>3.22</td>
</tr>
</tbody>
</table>

Three phase approach

Phase-1: Modeling
  – Statistical data-driven models
  – Unique features of computational sensor modeling

Phase-2: Prediction
  – Constraint manipulation

Phase-3: Fusion and analysis
  – Formulate data integrity as an optimization problem
  – Objective function: Minimize the discrepancies between sensor readings and the models
  – Constraints: Model constraints and user’s specified constraints

Limiting number of parameters - AIC criteria