Accuracy-aware Data Modeling in Sensor Networks

Ryo Sugihara **Andrew A. Chien**

Computer Science and Engineering Dept, Center for Networked Systems, University of California San Diego

{ryo, achien}@ucsd.edu

Motivation

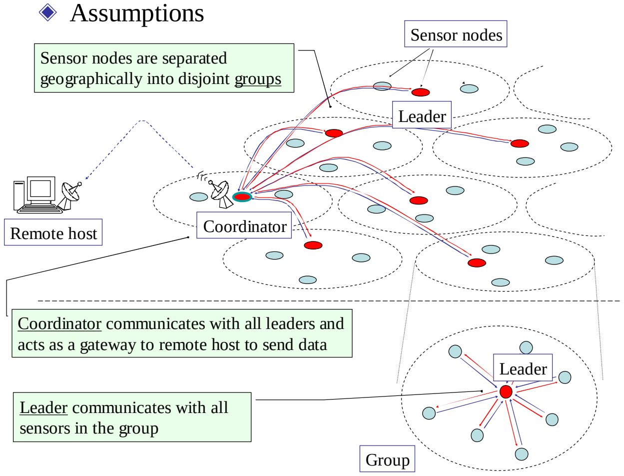
- Accuracy of data in sensor networks
  - Should be under control of applications
- Existing approaches: best-effort
  - No guarantee on accuracy

Problem: Pick an <u>efficient data representation</u> that meets the <u>specified accuracy requirement</u>

→ **Accuracy-aware data modeling**

The Problem

- Assumptions

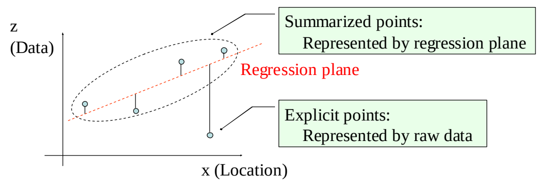


- Application is interested in an environmental attribute (e.g. temperature) over the entire space
- Application has an accuracy requirement
  - Expressed as an "Error bound"

- Optimization
  - Objective: Compact representation of data (@ Coordinator)
  - Constraint: Must satisfy accuracy requirement (Represented within the error bound ε)

Approach

- Hybrid: Regression Plane + Explicit Points



<u>Key Idea</u>: Choose cutoff to meet accuracy requirement

→ Produces efficient representation

Algorithm

Data summarization and filtering

For each group: do iteratively (@ Leader)

1. Calculate regression plane for data

$$f_K(x,y) = ax + by + c$$

where $a = \frac{\sigma_{xz}^2\sigma_{yy}^2 - \sigma_{xy}^2\sigma_{yz}^2}{\sigma_{xx}^2\sigma_{yy}^2 - \sigma_{xy}^4}$ $b = \frac{\sigma_{xx}^2\sigma_{yz}^2 - \sigma_{xy}^2\sigma_{xz}^2}{\sigma_{xx}^2\sigma_{yy}^2 - \sigma_{xy}^4}$ $c = \bar{z} - a\bar{x} - b\bar{y}$

$$\bar{A} = \frac{1}{N}\sum_{i=1}^{N} A_i \quad \sigma_{AB}^2 = \frac{1}{N}\sum_{i=1}^{N}(A_i - \bar{A})(B_i - \bar{B}) \quad \left(= \overline{AB} - \bar{A}\bar{B}\right)$$

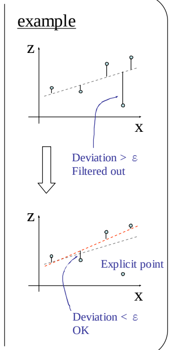
2. Filter out a data

Data with the maximum deviation

$$m = \arg\max_i \underbrace{\left| f_K(x_i, y_i) - z_i \right|}_{(\equiv d_i)}$$

If $d_i \le \varepsilon$, break the loop

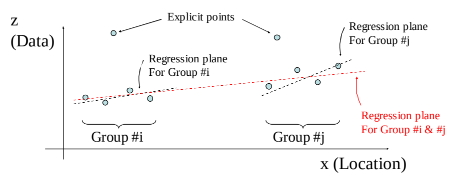
Otherwise eliminate sensor #m and repeat



Combining regression planes

Iteratively combine regression planes (@ Coordinator)

Using approximation similar to above



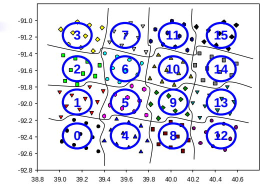
Collecting explicit data

Coordinator requests explicit data for each group leader

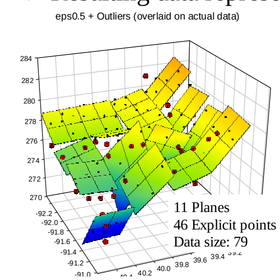
Leader sends raw data back to coordinator

Simulation Experiment

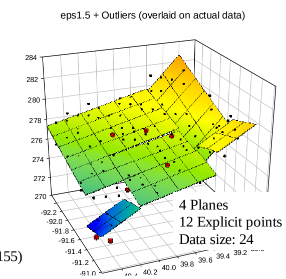
- Simulation data:
  - Environmental data (Brightness temperature)
  - 155 sensors, 16 groups



- Resulting data representation:



(Sensor #: 155)

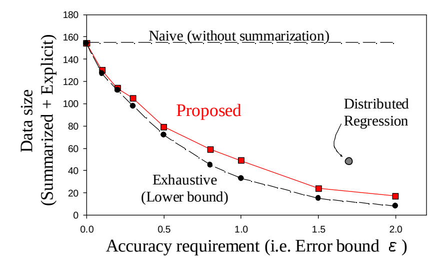


<u>Tight requirement</u>
Accuracy requirement $\varepsilon = \pm 0.5$

<u>Loose requirement</u>
Accuracy requirement $\varepsilon = \pm 1.5$

- Tighter accuracy requirement (left) produces more complex representation

- Results



- Flexible tuning of accuracy
- Fairly good performance

Conclusion & Future Work

- Accuracy-efficiency trade-off
  - Assuring accuracy is first-priority
    - Previous approaches focus on efficiency only
  - Our approach provides a new spectrum of flexibility
- Future work
  - Scalability: Fully-distributed algorithm (Ongoing)

Supported in part by the National Science Foundation under awards NSF Cooperative Agreement ANI-0225642 (OptIPuter), NSF CCR-0331645 (VGrADS), NSF ACI-0305390, and NSF Research Infrastructure Grant EIA-0303622. Support from the UCSD Center for Networked Systems, BigBangwidth, and Fujitsu is also gratefully acknowledged. Support from IBM Japan is also gratefully acknowledged.