Spatial Data Quality in the IoT Era

Management and Exploitation

Huan Li, Bo Tang, Hua Lu, Muhammad Aamir Cheema, Christian S. Jensen

2.
SID QUALITY Framework

By Hua Lu

Outline

- Quality dimensions
- Quality issues
- Means to resolve DQ issues

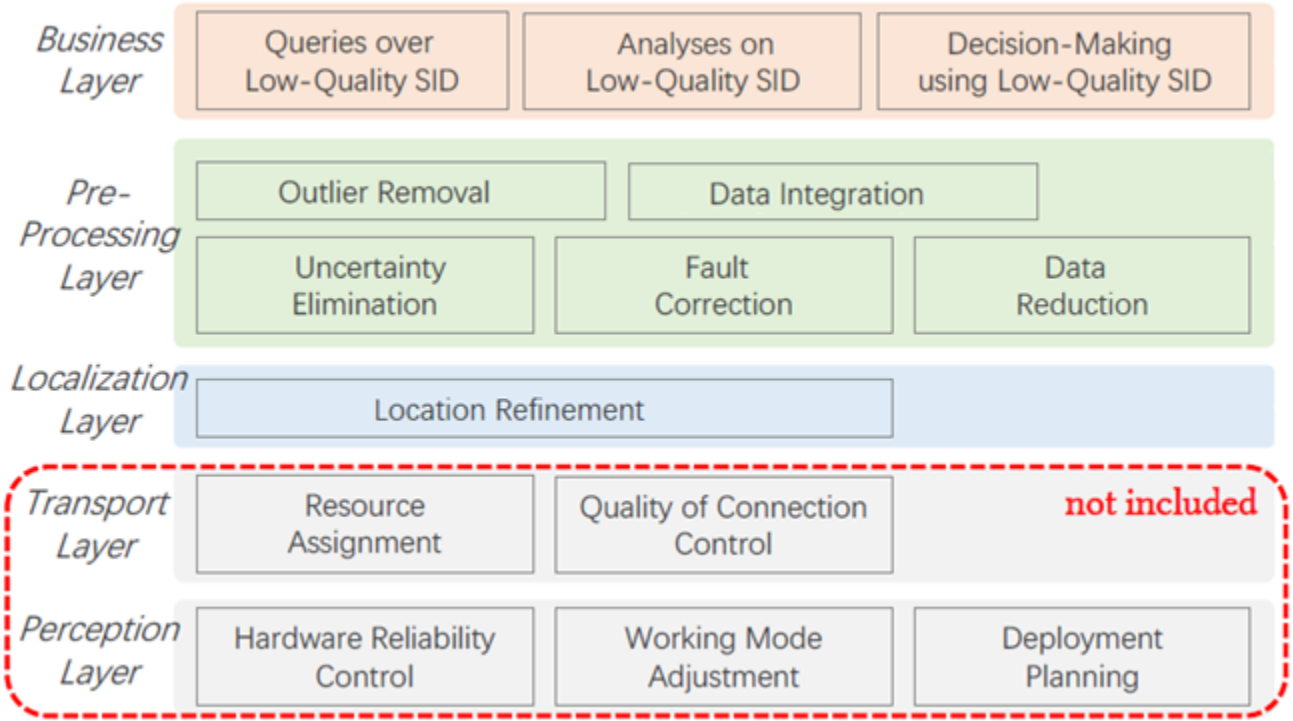
Quality Dimensions

- Data quality: How well data satisfies the consumption purpose?
- DQ dimensions: Criteria from consumers' perspective
- SID: *observations* of *realities* via IoT devices.
- SID quality dimensions:
  - Reliability aspect
    - Precision, Accuracy, Consistency
  - Comprehensiveness aspect
    - Sparsity, Space Coverage, Completeness, Redundancy
  - Usability aspect
    - Latency, Staleness, Data Volume, Truth Volume, Resolution, Interpretability

Quality Issues

SID Characteristic	Quality Issues (↓: low; ↑: high)
[omnipresent in IoT setting]	
Noisy and erroneous	↓ precision, ↓ accuracy, ↓ consistency
Temporally discrete	↑ time sparsity, ↓ completeness, ↑ staleness
Decentralized and heterogeneous	↓ consistency, ↑ latency, ↓ interpretability
Dynamic	↓ precision
Voluminous and duplicated	↑ redundancy, ↑ latency, ↑ data volume
Isolated and conflicting	↓ consistency, ↓ interpretability
[specific in spatial data domain]	
Unverifiable	↓ truth volume
Hierarchical and multi-scaled	↓ consistency, ↓ resolution, ↓ interpretability
Spatially discrete	↓ space coverage

Means to Resolve DQ Issues: Task Perspective



Localization Layer

- Key task
  - Location Refinement
- Main goals
  - ↑ precision
  - ↑ accuracy
  - ↑ resolution

Pre-processing Layer

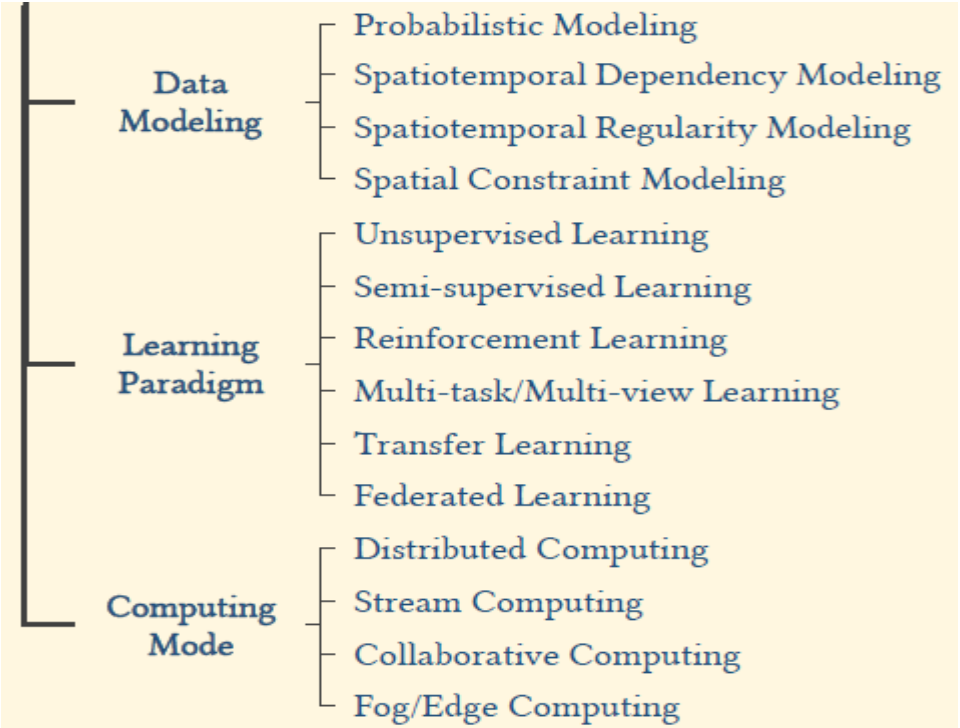
- Key tasks
  - Uncertainty Elimination
    - ↑ precision, ↑ completeness, ↑ resolution, ↓ time sparsity
  - Outlier Removal
    - ↑ precision, ↑ accuracy, ↑ consistency
  - Fault Correction
    - ↑ accuracy, ↑ consistency, ↑ completeness
  - Data Integration
    - ↑ accuracy, ↑ completeness, ↑ data volume, ↑ resolution, ↑ interpretability
  - Data Reduction
    - ↓ data volume, ↓ latency, ↓ redundancy

Business Layer

- Key tasks
  - Querying
  - Analysis
  - Decision-making
- *NB*: All over low quality SID

Means to Resolve DQ Issues: Technique Perspective

- They apply to *all* layers
  - Business
  - Pre-processing
  - Localization
- They may be combined



References

- [Karkouch et al., 2016] Data quality in Internet of Things: A state-of-the-art survey. Journal of Network and Computer Applications.
- [Li et al., 2018] Spatiotemporal distribution of indoor particulate matter concentration with a low-cost sensor network. Building and Environment.