ST-MVL: Filling Missing Values in Geo-sensory Time-series Data

Xiwen Yi\textsuperscript{1,2}, Yu Zheng\textsuperscript{2,1}
Southwest Jiaotong University, Chengdu, China
Microsoft Research, Beijing, China

Motivation
Data missing is a very common phenomenon in LOT data
- Due to communication or device errors
- Affect real-time monitoring and further data analytics

Goal
Filling the missing values in a collection of geo-sensory time series data using collective information:
- Data of a sensor
- Its neighborhoods

A fundamental problem
Challenges
- Random missing and block missing
- Lose readings of multiple sensors simultaneously
- Lose readings of a sensor at consecutive timestamps
- Hard to find stable inputs for a model
- Readings changing over time and location non-linearly
- Not handled by simple interpolations

Overview
Integrating four perspectives
- Spatial and Temporal perspectives
  Spatial neighbors
  Temporally adjacent time intervals
- Global and local perspectives
  Global: Long-term patterns
  Local: Recent context

Methodology
Result
Multi-view Learning

Global spatial view: Inverse Distance Weighting (IDW)
Global temporal view: Simple Exponential Smoothing (SES)
Local spatial view: User-based Collaborative filtering (UCF)
Local temporal view: Item-based Collaborative filtering (ICF)

Multi-view learning
- Four views combination by linear least square
  \( \hat{\nu}_{\text{mvl}} = w_1 \cdot \hat{\nu}_{gs} + w_2 \cdot \hat{\nu}_{gt} + w_3 \cdot \hat{\nu}_{ls} + w_4 \cdot \hat{\nu}_{lt} + b \)

Evaluation
Comparison among different methods (based on PM2.5)