Traffic Data Imputation Using Matrix Completion on Graphs

行列補完理論による車両感知器データの補完

東京大学 生産技術研究所 大口研究室（交通制御工学）
http://www.transport.iis.u-tokyo.ac.jp/
Tianyang HAN, Kentaro WADA, Takashi OGUCHI

What is Traffic Data Imputation?

- **Traffic Data.** The traffic flow data (speed and volume) with fixed temporal spatial resolution.
- **Missing.** Empty or error data record including random (e.g. malfunction) and non-random (e.g. maintenance) cases.
- **Imputation.** Recovery of missing data with partially observation.

Motivation

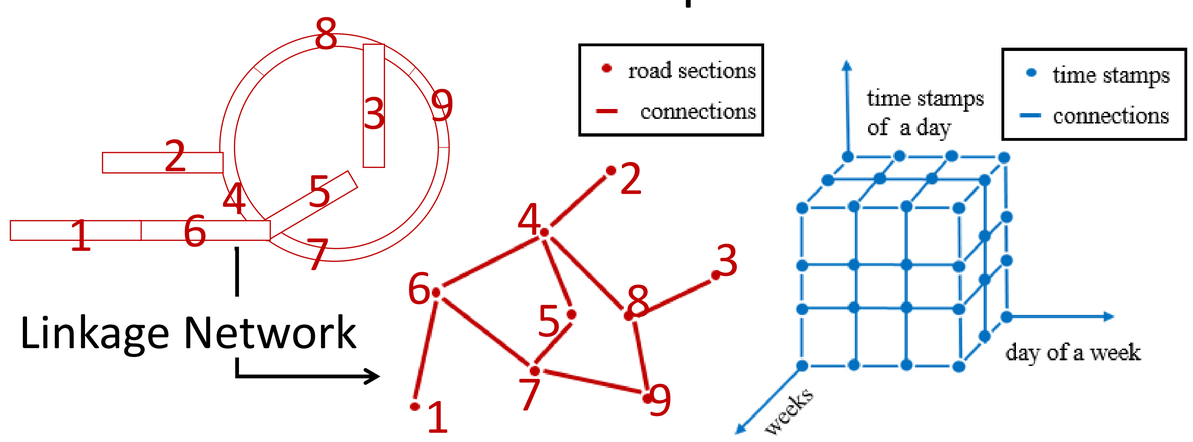
- **Unavoidable Missing Problem.** Maintenance, detector malfunction, signal distortion, low sampling rate …
- **Data Requirement.** Signal control and other transportation management system require complete traffic data.

Matrix Completion

- **Matrix Completion** (MC) solves traffic data imputation problem by low rank approximation. The missing could be recovered by same pattern observation.
  - **Rank.** Maximum independent component of a matrix.
  - **Pattern.** Approximately similar component based on partially observation.
  - **Low rank.** Minimum number of patterns based on approximation.
- **Advantage.** Derive low rank approximation automatically without prior knowledge. Simple solving with relaxation and iteration solver.
- **Limitation.** Has no explanations with pattern. The temporal spatial information from our knowledge (or understanding) is not used

How to use our knowledge?

- **Temporal Spatial Information.** Compared with the algorithm, our human being can understand the road network structure and recurrent congestion in our daily life. Such understanding could be input as prior knowledge for better imputation.
- **Matrix Completion on Graphs.** Consider the relationship among data as graphs (See below). We can add graph smoothness assumption to the low rank minimization problem.

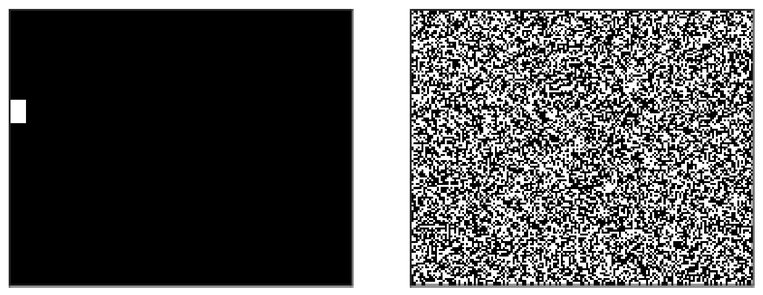


- The smoothness assumption supposes the similarity between connected pairs of data. For example the neighboring road sections have the same traffic status.

Results from Numerical Experiments

- **Data source.** Detector data (5min resolution) from Tokyo Metropolitan Expressway (首都高).
- **Performance.**
  - Applicable for Large-scale data
  - Significant in large-missing-rate case and non-random missing case. (Compared with ordinary matrix completion, see the result with relative error (%) on volume data)

Model	10%	20%	30%	40%	50%	MNR
with graph	9.43	**11.43**	**13.15**	**13.26**	**14.71**	**16.98**
without graph	**8.49**	11.69	13.24	14.77	17.64	17.46



non-random missing　　random missing $\phi = 0.5$

- **Future work.** (1) Combined model for accuracy and robustness (decomposition methods and Bayesian inference); (2) Different road topology; (3) Data characteristics