KRIHS seminer

Probabilistic Tensor Decomposition for Human Mobility[*1]

Shonosuke Sugasawa

Center for Spatial Information Science, The University of Tokyo

[*1] Joint work with Shinichiro Shirota (Department of Biostatistics, UCLA)

Introduction

Human mobility data contains complex stochastic structure.

- Tensor Data

  Divide the whole space into a finite number of grids and count the number of people with certain categories in each grid at specific time.

  $\Rightarrow$ Multi-dimensional array data (= Tensor data)

- Goal: Find underlying dependence structures

  $\Rightarrow$ Approximate the observed tensor by a parsimonious (low rank) tensor representation (= Tensor Decomposition)

  We use non-negative Tucker decomposition.

  Related literatures: Fan et al., (2014), Sun and Axhausen (2016) etc...

Data

We divide Tokyo metropolitan area into 10 × 10=100 grids and count the number of people with different types of jobs at two time (00:00 and 12:00).

→ 3-dimensional tensor

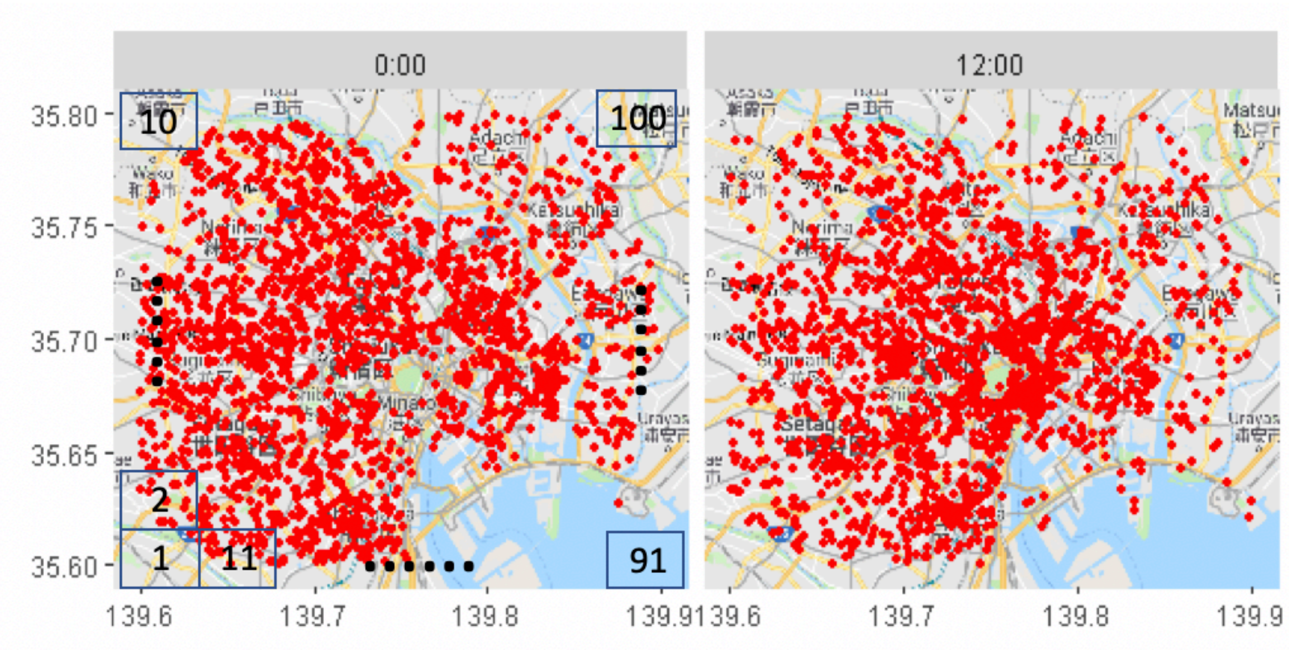

Figure1 Mobility pattern of *transportation job* at 00:00 (left) and 12:00 (right).

Non-negative Tucker decomposition

- G: core tensor A, B, C: base matrices
- Core tensor = weight (sum of all the elements is 1)

  ⇒ By investigating the elements in A, B, C whose weight is large, we can see the dependence structure.

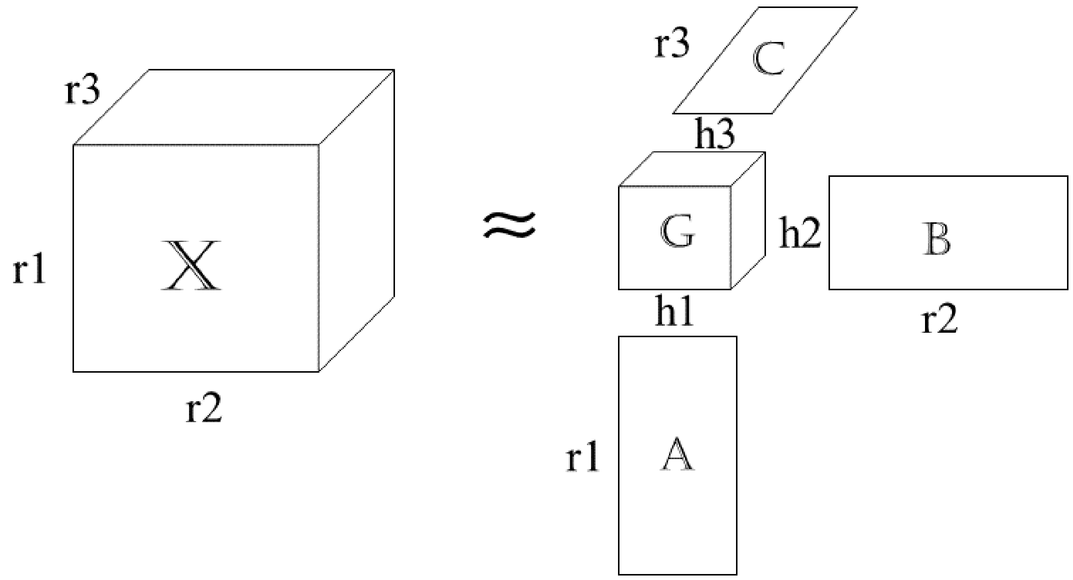

Figure2 Image of Tucker decomposition

Non-negative Tucker decomposition

$$x_{c_1 \ldots c_m} = \sum_{k_1=1}^{h_1} \cdots \sum_{k_m=1}^{h_m} g_{k_1 \ldots k_m} u^{(1)}_{c_1 k_1} \times \cdots \times u^{(m)}_{c_m k_m}$$

- $x_{c_1 \ldots c_m}$: element of observed tensor
- $g_{k_1 \ldots k_m}$: element of core tensor

- In our application, $m = 3$ (3-dimensional tensor), $r_1 = 100$, $r_2 = 100$, $r_3 = 2$ (*transportation* and *sales*)

  $\Rightarrow \quad h_1 = 16, h_2 = 16, h_3 = 2$

Results

Results for *transportation workers*

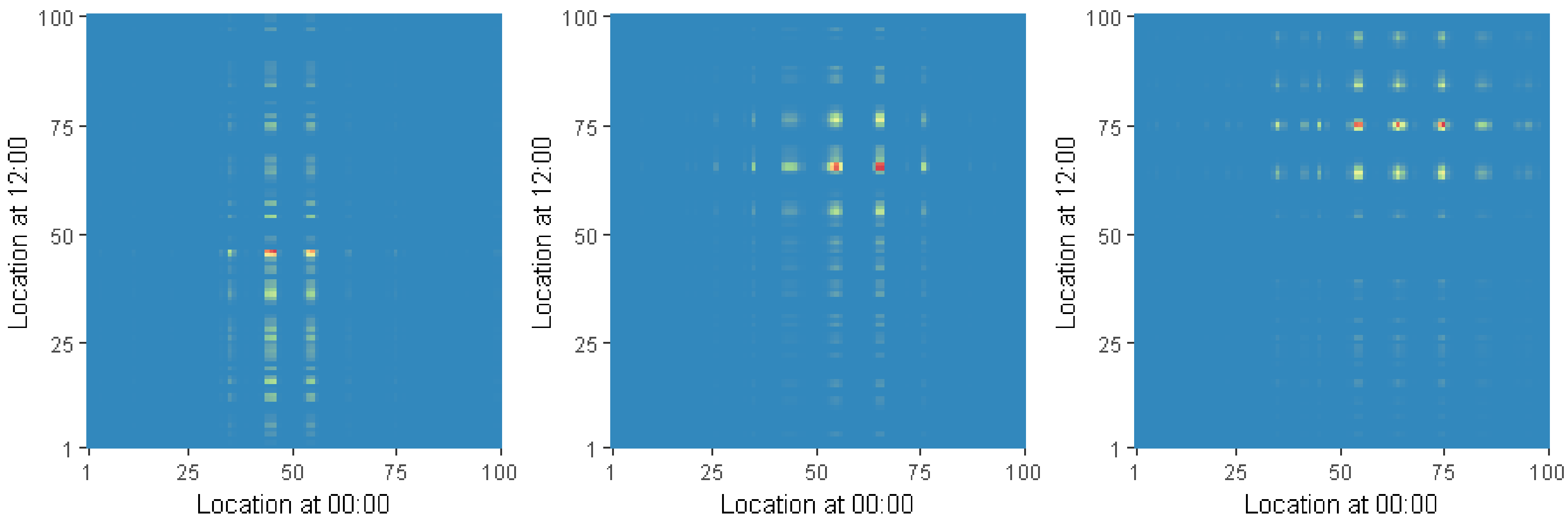

Figure3 Top 3 basis

- Base with largest weight : People are concentrated at the center of Tokyo at 00:00, but they spread at 12:00.
- Based with second and third largest weights: People remain in the east of Tokyo.

Results

Results for *sales workers*

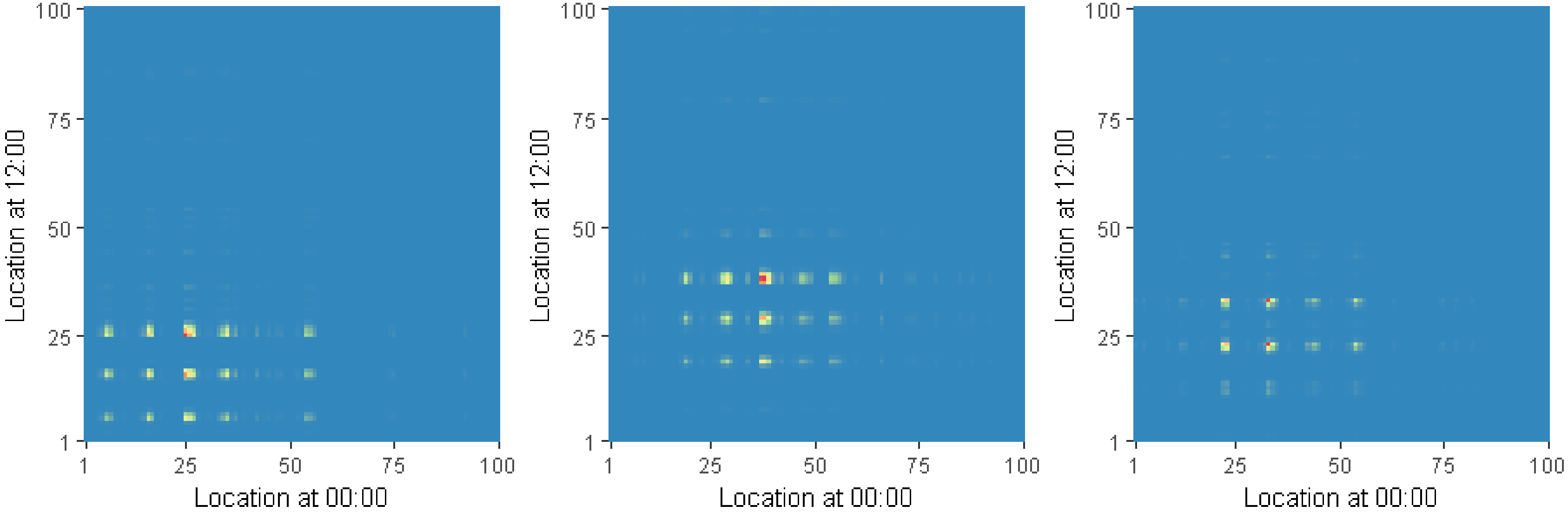

Figure4 Top 3 basis

- People are more concentrated at 12:00 than at 00:00. (Commuting?)

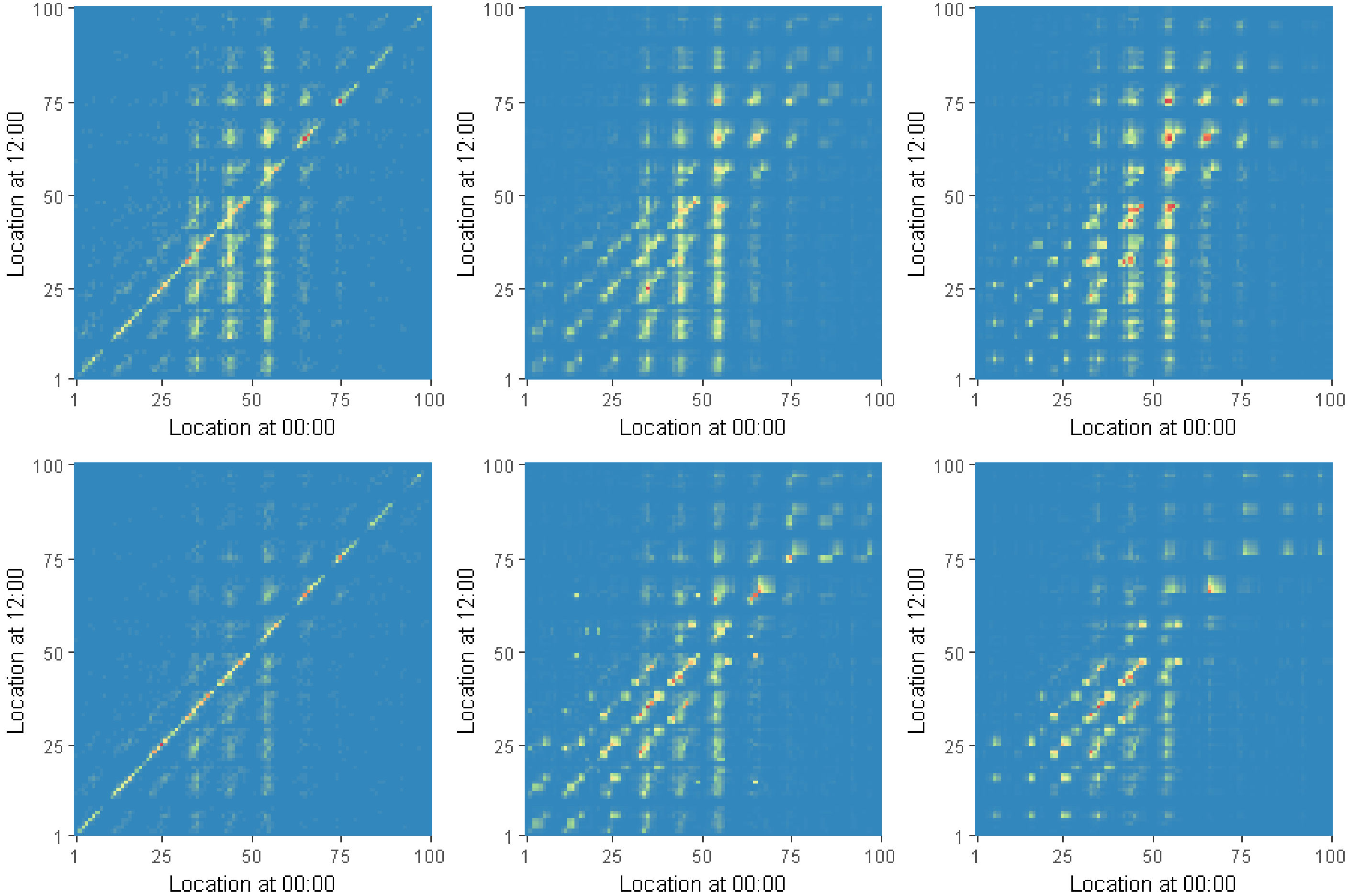

Figure5 (left) observed tensor, (center) approximated tensor, (right) approximated tensor based on bases with top 10 weights; (upper) transportation and (lower) sales workers.

Conclusion and Future Work

- Using tensor decomposition, we can reveal the difference of human mobility among categories.

  "Statistical" comparison of human mobility will be a future work.

- Scalability

  We can use other information (Age and Purpose etc...), but the size of observed tensor will be large.

- Clustering

  Using Dirichlet process, we may introduce clustering structure into a core tensor. (Automatic determination of the size of core tensor.)