

## テンソル分解による人の動きの統計解析

城田 慎一郎<sup>1</sup>, 菅澤 翔之助<sup>2</sup>

<sup>1</sup>Department of Biostatistics, UCLA, <sup>2</sup>東京大学 空間情報科学研究センター

連絡先: <shinichiro.shirota@gmail.com> Web: <https://sites.google.com/site/shinichiroshirota/>

(1) **動機:** 近年, GPS などの空間データを用いた人の動き・都市に関する様々な研究が行われている。一方で, これら人の動きの都市間の構造の比較や, 各種カテゴリーカルな個人の属性との依存構造の比較などに関して, 統計学的手法はまだまだ発展途上である。特に GPS データに紐づいた各種カテゴリーカルな属性情報との依存構造の推定に関しては, 社会科学や都市計画において, 重要な研究課題である。本研究の目標は, 機械学習を用いた人の動きの特徴量の自動抽出と, その都市間・異なる年代間における比較を可能にする統計的検定手法の開発である。具体的には, 各種カテゴリーカルな属性と人の動きの特徴量との依存関係に柔軟な確率的依存構造を導入し, データからその構造を推定することで属性ごとの人の動きの定量的な特徴付けを行うことが可能な統計的手法を開発する。

(2) **方法:** 「人の流れプロジェクト」で提供される, パーソントリップ調査を基にした, 各都市における人の軌跡データを用いる。この軌跡データは時間・空間情報と共に様々なカテゴリーカルな属性情報(職業や年齢など)も付随しているため, テンソル(高次元配列)として扱うことを考える。近年, ネットワーク理論の観点から, 人の動きの特徴量として, 人の動きの最小単位を構成する *motifs* (モチーフ)と呼ばれる要素を構築する方法論が展開されている。従来, このモチーフは研究者が自ら構成・提案する必要があったが, 近年, 機械学習的なアプローチを導入してデータからモチーフを構成する方法が提案されている。本研究では, テンソル解析の枠組みの下で, これらのモチーフをデータから自動抽出する統計・機械学習的アルゴリズムを開発する。具体的には, 非負タッカー分解と呼ばれるテンソル解析の手法を用いて, 解釈可能なモデルとその高速な推定手法を構築する。また, その手法を人の軌跡データに適用し, その有用性を示す。

(3) **結果:** 東京都市圏を  $10 \times 10 = 100$  のグリッドに分割し, 職種カテゴリー・時刻・グリッドごとに該当する人数の相対頻度を計算し, テンソルデータを作成する。このデータに対して提案手法を適用し, 人の流れを特徴付ける構造を推定した。図1は, 職業カテゴリーが運輸業従事者である人の軌跡データの時刻0:00および時刻12:00の様子を図示したものである。また, データから計算された各グリッドの相対頻度

およびモデルから推定された相対確率を2つの時刻で比較したものが図2である。対角部分の値が大きくなっているのは, 2つの時刻で人数の相対頻度があまり変わらないことを表す。また, X軸の50前後から縦に値が大きくなっているが, 50番前後のグリッドは東京の中心部に位置しているため, 深夜の時刻0:00よりも昼の時刻12:00の方が東京中心部の相対頻度が大きくなっていることがわかる。様々な属性による統計的な比較やより細かいグリッドを用いて推定した結果などは当日報告する。

### (4) 使用したデータ:

- ・「2008年東京都市圏の人の流れデータセット」, 東京大学空間情報科学研究センター
- ・「2011年中京都市圏の人の流れデータセット」, 東京大学空間情報科学研究センター

(5) **謝辞:** 本研究は東大CSIS共同研究No.816の成果の一部として実施した。ここに記して謝意を表したい。

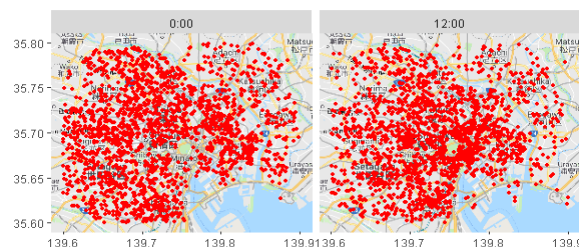


図1: (左)時刻0:00と(右)時刻12:00における運輸業従事者の位置。

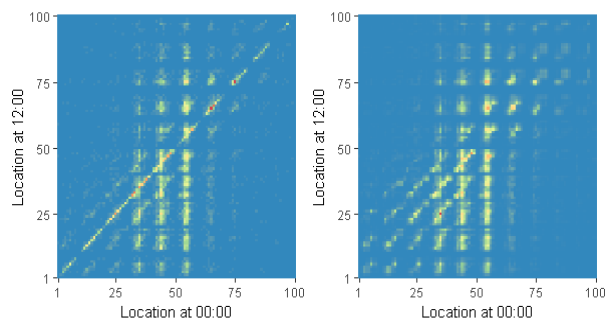


図2: 時刻0:00と時刻12:00の運輸業従事者の100グリッド上の相対割合の比較。(左)データから算出される相対的頻度, (右)テンソル解析による推定割合。