

世界メッシュ統計に基づく都市の特徴付け

佐藤彰洋†

京都大学大学院情報学研究科†

この研究では、都市に存在する施設(ATM,公衆トイレ,駅,バス停など)に関する世界メッシュ統計に基づき都市を特徴付ける方法について議論する。まず、世界メッシュ統計がどのようなものであるかについて説明し、都市に存在する施設の数を用いてどのように都市を特徴付けるかを述べる。7種類の設備(ATM, 簡易宿舎, 公衆トイレ, 銀行, 郵便局, ホテル, 喫茶店)のポイントデータを用いた世界メッシュ統計を京都市内に対して作成した実証分析の結果について述べる。更に、京都とエレバンに対してOLSとRMAによる回帰分析を行い、更にt-検定を行った。OLSとRMAの結果を比較するとともに、4種類の都市について都市間の類似度を提案手法により計量した結果を述べる。

キーワード：世界メッシュ統計、OLS 回帰分析、RMA 回帰分析、空間特徴量

1. はじめに

情報通信技術の発展と情報化社会の進展の結果データ基盤上に多数のデータが蓄積されるようになってきた。情報通信システムは我々の日常生活で利用されるようになり、データ駆動型社会の新しい時代が始まりつつある。

しかしながら、データは単なるシンボル列であるため、計算なしには我々はデータから意味を理解することができない。それゆえ我々の容量を超えてデータを分析し定量化するためのソフトウェアを必要とする。特に、統計的方法はビッグデータを特徴づけするために有用である。

データの中でも空間上の位置とコンテキストを紐づけるポイントデータは近年大量に利用可能となってきた。しかしながら、大量のデータを単純に地図上にマッピングするだけでは、我々は空間の特徴を理解することができない。

本研究では、空間統計に着目しインターネット上で収集される位置情報付きのデータからメッシュ統計を作成する方法について議論する。更に、特定の設備の数に関するメッシュ統計を用いたアプリケーションソフトウェアについて述べる。

メッシュ統計のユースケースは以下5種類が挙げられる。(1)空間的な位置を考慮して異なるメッシュ統計を比較する(2)異なる複数のメッシュ統計から新しいメッシュ統計を合成する(3)いくつかのメッシュ統計を考慮し特定の場所を探索する(4)メッシュ統計を考慮し特定の特徴を示す特定の設備を見つける(5)異なるメッシュ統計間の相関を計算する、

または異なる設備と空間の特徴を比較する。

特に、この研究では特定の場所の特徴を定量化し、着目する設備の数に関するメッシュ統計間の相関を考慮することにより2つの場所の類似性を比較する方法について議論する。

2. 方法

世界メッシュ統計は雑誌統計[1]で提案した地域メッシュコード(JIS X0410)の上位互換性を有する全世界を覆うメッシュコード体系である[2, 3]。世界メッシュコードを用いることにより世界のあらゆる場所に対するメッシュ統計を作成することが可能である。本発表では世界メッシュ統計を用いて都市間の類似性を計量する方法[4]について説明する。

n 種類の設備を考える。ここで $m(i)$ 個 ($i=1, \dots, n$)の設備がそれぞれ存在していると考え、 n 種類の設備それぞれのメッシュ統計を $x(i)_k$ と表現する。ここで、 k は領域 D に含まれる世界メッシュコード w_k に対する設備 i の数を表す。設備 i と設備 j との間のメッシュ統計、 $x(i)_k$ と $x(j)_k$ に対して線形の回帰関係

$$x(j)_k = a_{ij} x(i)_k + b_{ij} + e_{ijk} \quad (1)$$

を仮定する。ここで、 a_{ij} は設備 j に対するメッシュ統計を被説明変数、設備 i に対するメッシュ統計を説明変数とした場合の回帰係数である。 b_{ij} はオフセット成分、 e_{ijk} は誤差を表す。ここで、OLS回帰では誤差を最小2乗誤差

$$E(a_{ij}, b_{ij}) = \sum_{k=1}^S (x(j)_k - a_{ij} x(i)_k - b_{ij})^2$$

により与える。 S は含まれるメッシュ総数である。この時、回帰係数はそれぞれ

$$a_{ij} = \frac{S \sum_{k=1}^S x(i)_k x(j)_k - \sum_{k=1}^S x(i)_k \sum_{k=1}^S x(j)_k}{S \sum_{k=1}^S x(i)_k^2 - \left(\sum_{k=1}^S x(i)_k \right)^2}$$

$$b_{ij} = \frac{\sum_{k=1}^S x(j)_k - a_{ij} \sum_{k=1}^S x(i)_k}{S}$$

で与えられる。明らかに $a_{ii}=1, b_{ii}=0$ である。しかしながら、一般に OLS 回帰では $x(i)_k$ を $x(j)_k$ で回帰する場合と $x(j)_k$ を $x(i)_k$ で回帰する場合とで同一の結果を与えない。すなわち $a_{ij} \neq 1/a_{ji}$ である。

この問題を解決するために RMA 回帰と呼ばれる対称性を有する誤差関数

$$E(a_{ij}, b_{ij}) = \frac{1}{2} \sum_{k=1}^S \frac{(a_{ij} x(i)_k + b_{ij} - x(j)_k)^2}{|a_{ij}|}$$

を用いた線形回帰分析

$$x(j)_k = a^*_{ij} x(i)_k + b^*_{ij} + e^*_{ijk} \quad (2)$$

がある。この回帰法では $a^*_{ij}=1/a^*_{ji}$ を満足する。この OLS 回帰と RMA 回帰との間で得られる係数間には関係があり、

$$a^*_{ij} = \frac{(\sqrt{a_{ij}^2 + 1} - 1)/a_{ji} + (\sqrt{1/a_{ji}^2 + 1} - 1)a_{ij}}{a_{ij}/a_{ji} - (\sqrt{a_{ij}^2 + 1} - 1)(\sqrt{1/a_{ji}^2 + 1} - 1)}$$

$$b^*_{ij} = \frac{a_{ij} b_{ji} + b_{ij}}{1 - a_{ij} a_{ji}} - a^*_{ij} \frac{a_{ji} b_{ij} + b_{ji}}{1 - a_{ij} a_{ji}}$$

の関係がある。最終的に都市 s と都市 t における回帰係数行列をそれぞれ $a^*(s)_{ij}, a^*(t)_{ij}$ とすると都市 s と都市 t の差異を

$$K(s, t) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (a^*(s)_{ij} - a^*(t)_{ij})^2 \quad (3)$$

により計量することができる。

3. 実証分析

都市に存在する 7 種類の設備 (ATM, 簡易宿泊施設, 公衆トイレ, 銀行, 郵便局, ホテル, 喫茶店) を用いて京都, エレバン, トリノ, 大阪の比較分析を行った。データは OSM の Web API を用いて 2017 年 1 月 23 日に収集した。これらの設備に対する 3 次世界メッシュを計算し、OLS 回帰分析から得られた回帰係数行列 a_{ij} から RMA 回帰分析により得られる回帰係数 a^*_{ij} を計算し、それぞれの都市間の差異を (3) 式を用いて求めた。表 1 と表 2 が得られた都市間の差異を表現する

行列である。

表 1 OLS 回帰係数から得られた都市間の差異

都市	京都	エレバン	大阪	トリノ
京都	0.000000	2.369843	0.888104	0.524196
エレバン	2.369843	0.000000	2.949257	2.547829
大阪	0.888104	2.949257	0.000000	0.820071
トリノ	0.524196	2.547829	0.820071	0.000000

表 2 RMA 回帰係数から得られた都市間の差異

都市	京都	エレバン	大阪	トリノ
京都	0.000000	3.902147	1.981784	0.997005
エレバン	3.902147	0.000000	2.949257	3.903612
大阪	1.981784	5.714781	0.000000	2.026900
トリノ	0.997005	3.903612	2.026900	0.000000

表 1 と表 2 より OLS 回帰係数から得られた都市間の差異と RMA 回帰係数から得られた都市間の差異とは値が大きく異なることがわかる。順位も若干ことなる。RMA 回帰で計算される 2 都市間の差異に着目してみる。表 2 に示す RMA 回帰係数から得られた都市間の差異においては、京都とトリノの類似性は京都と大阪や京都とエレバンの類似性に比べて高いことが分かる。エレバンは大阪、京都、トリノともに大きく異なる特徴を示している。大阪は京都により類似性が高くトリノとエレバンとは差異がある。

4. まとめと今後の課題

本研究では、都市の類似性を世界メッシュ統計に対する RMA 回帰係数の差異から測る方法を提案した。提案手法により都市間の差異を 7 種類の設備を用いて 4 つの都市 (京都, エレバン, 大阪, トリノ) に対して計量を行った。今後の課題としては、この分析方法をより多くの都市間について行い都市の分類を行うことが挙げられる。更に、都市に存在する設備の数や種類を変化させた場合の感応度分析を行う必要がある。

参考文献

- [1] 佐藤彰洋, 椿広計, ビッグデータ時代に必要な標準化, 統計, 2015 年 9 月号, pp. 32- 38.
- [2] Aki-Hiro Sato "Characterization of Cities Based on World Grid Square Statistics about Specific Properties", IEEE Big Data 2017 on, 2017, pp. 4146-4155.
- [3] 世界メッシュ研究所 <http://www.fttsus.jp/>
- [4] Aki-Hiro Sato, Shoki Nishimura, Hiroe Tsubaki, "World Grid Square Codes: Definition and an example of world grid square data", IEEE Big Data 2017 on, 2017, pp. 4156-4165.