

川口市犯罪データの空間分析

大下祐樹* 垂水共之**

Spatial analysis of Kawaguchi City crime data

Yuki Ohshimo* and Tomoyuki Tarumi**

In this paper, we built some models of the spatial data, and evaluate those, using the crime data of Kawaguchi city, Saitama prefecture in Japan. Compute the Moran'I statistics of five crime data related to thefts in the city, the value of sneakthief take the aggregation. So, we pay attention to the sneak, and make a model to estimate the incidence of the events. As a regression, we select the old-age index among regional indexes. Applying a simple regression(SR), a spatial autoregressive model(SAR), a geographically weighted regression(GWR), we evaluate these models. In GWR model, it is interesting that there are some sections in which the sign of the coefficient takes the opposite versus SR and SAR models. So, stratifying the data by the sign of it, we investigate to data precisely.

Key words: crime data, Kawaguchi city, spatial analysis, spatial autoregressive model, geographically weighted regression

1 はじめに

空き巣のような犯罪には犯罪が起きやすい地区、起きにくい地区等があり、起きやすい地区の周りの地区は他の地区より起きやすい傾向が、あまり犯罪が起きない地区の周りの地区は他の地区よりも起きにくいという集積性を示すことが多い。本研究では川口市の町丁別窃盗データ(埼玉県川口市)を用いて空き巣の空間的な集積性を分析した上で、空間自己回帰モデル、地理的加重回帰モデルを推定しそのモデルの比較を行った。

2 Moran'I 統計量

分析対象が n 個の地区(町丁別など)に分割されており、各地区でのデータを $y_i (i = 1, 2, \dots, n)$ とする。このとき、Moran'I 統計量は規則的な空間的変動の存在を示す指標であり次式で定義される。

$$I = \frac{1}{W} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sigma_y^2} \quad (1)$$

ここに、

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

w_{ij} は一次の連結性指標であり、 $w_{ij} = c_{ij} / \sum_{i=1}^n c_{ij}$ 、 c_{ij} は隣接関係を表し、 $c_{ij}=1$ (地区 i と地区 j が接している)、 0 (その他)である。また、

$$W = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$$

Moran'I 統計量が有意に正の値をとれば、集積性が存在すると判断できる(杉浦;2003)。

3 空間回帰モデル

この節では通常の説明変数 x と目的変数の y の回帰モデル

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i (i = 1, 2, \dots, n) \quad (2)$$

*岡山大学大学院環境学研究所

**岡山大学アドミッションセンター

に対して、 y の空間的な従属性を考慮した 2-3 の空間回帰モデルを示す。

記号の表記は

X : デザイン行列

ϵ_i : 地区 i の誤差

$$\epsilon = [\epsilon_1 \ \epsilon_2 \ \dots \ \epsilon_i]^T$$

3.1 空間自己回帰モデル

空間自己回帰モデルとは当該地区に接している地区の平均値 $\sum_{j=1}^n w_{ij}y_j$ を、 y の空間的な従属性を表す説明変数として組み込んだモデルである。

$$y_i = \beta_0 + \beta_1 x_i + \rho \sum_{j=1}^n w_{ij}y_j + \epsilon_i \quad (3)$$

ρ : 空間自己回帰係数

w_{ij} : 一次の連結性指標

3.2 地理的加重回帰モデル

地理的加重回帰モデルとは、 x と y の関係自体が空間的に変動していると考えたモデルであり、式 (2) の回帰係数 β_0, β_1 が地区ごとに変動する。

地区 i のパラメータ β_{0i}, β_{1i} を推定するとき

$$W_i y = W_i X \beta_i + \epsilon \quad (4)$$

ここに $W_i = \text{diag}(h_{i1}, h_{i2}, \dots, h_{ij})$ 、 $\beta_i = [\beta_{0i} \ \beta_{1i}]^T$ 。

h_{ij} は地理的な重みとして距離減衰関数であるガウス型カーネル関数を用いる。

$$h_{ij} = \exp\left(-\frac{d_{ij}^2}{\theta^2}\right) \quad (5)$$

ここに、 d_{ij} は地区 i と地区 j 間との間の距離である。本研究では地区 i と地区 j の面積重心間の距離としている。

最適バンド幅 θ はクロスバリデーションスコアを最小にする θ とする。

3.3 回帰モデルと残差の空間的自己相関

モデルの指標として PSS(Prediction Sum of Square)

$$PSS = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{(i)i})^2} \quad (6)$$

を用いる。

$\hat{y}_{(i)i}$: 地区 i を除いて推定したパラメータによる、地区 i の予測値

モデルの適合度が優れていれば、PSS は小さくなる。

通常の回帰モデルでは、誤差は空間的に独立であり、互いに無相関であることを前提としている。その仮定が満たされているかどうかの確認に空間的自己相関を利用する。次のように残差の Moran'I 統計量を求める。

$$I_e = \frac{1}{W} \cdot \frac{\sum_i \sum_j w_{ij} e_i e_j}{\hat{\sigma}_e^2} \quad (7)$$

この I_e により誤差の空間的自己相関を比較する。

モデル比較の際は残差二乗和、PSS の比較によりモデルの当てはまりの良さ、残差の Moran'I 統計量の比較により誤差項の独立性の比較を行う。

4 埼玉県川口市の犯罪データの分析

本研究で用いるデータは(埼玉県川口市)で公開されている町丁字別犯罪(窃盗5手口)発生件数である。

平成18年度の合計データと平成19年度の各月の合計データが公開されており、本研究では一年間の合計数である平成18年度合計データを用いることにした。犯罪データの一部を表1に示す。

表 1 平成18年度川口市町丁別犯罪発生件数

町丁字	自転車盗	オートバイ盗	ひったくり
東川口5丁目	14	1	0
東川口4丁目	10	4	2
東川口1丁目	19	10	1
東川口3丁目	21	4	1
東川口6丁目	9	3	0

車上荒らし	空き巣
6	6
11	7
0	3
12	5
6	8

それぞれの犯罪の発生率(1,000世帯あたり)の要約数を表2に示す。

表 2 平成 18 年度川口市町丁別犯罪発生率要約

統計量	自転車盗	オートバイ盗	ひったくり
最小値	0	0	0
第一 4 分位値	5.305	1.340	0
中央値	8.016	3.058	0
平均値	13.073	4.464	1.069
第三 4 分位値	11.850	5.728	1.566
最大値	321.472	66.667	9.804

車上荒らし	空き巣
0	0
2.947	1.340
5.014	36
6.909	4.073
8.071	5.488
47.619	34.884

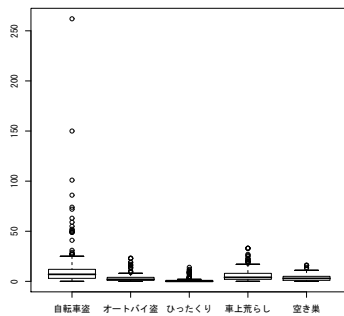


図 1 箱ひげ図

4.1 Moran'I 統計量

犯罪発生率の集積性を示す Moran'I 統計量を表 3 に示す。*は有意水準 5%, **は有意水準 1% で有意に集積性があることを意味する。車上荒らし以外には集積性が有意水準 5% で示唆された。

	Moran'I
自転車盗	0.11*
オートバイ盗	0.07*
ひったくり	0.11**
車上荒らし	0.01
空き巣	0.23**

では、この中でも特に集積性が高い空き巣に焦点をおいて解析を進めることにする。

4.2 空き巣発生率と地域指標の関係

空き巣発生率と地域指標データとの関係を探ることにする。地域指標として平成 17 年度国勢調査より町丁別住宅の建て方のデータ(単位は世帯)と、平成

18 年度川口市住民基本台帳のデータより町丁字年齢別人口数と世帯数のデータを入力した(埼玉県川口市)。地域指標の一部を表 4, 5 に示す。

表 4 平成 18 年度川口市町丁別人口、世帯データ

町丁字	人口	65 歳以上人口数	老人化率	世帯数
東川口 5 丁目	2512	197	0.08	917
東川口 4 丁目	2620	223	0.09	1075
東川口 1 丁目	1563	135	0.09	638
東川口 3 丁目	1042	66	0.06	463
東川口 6 丁目	2067	131	0.06	783

表 5 平成 17 年度住宅の建て方のデータ

町丁字	戸建	1.2 階建 ¹⁾	3-5 階建 ¹⁾
東川口 5 丁目	194	105	85
東川口 4 丁目	316	208	322
東川口 1 丁目	108	80	128
東川口 3 丁目	65	74	200
東川口 6 丁目	198	106	230
	6-10 階建 ¹⁾		11 階以上 ¹⁾
	387		0
	32		0
	112		45
	21		0
	97		0

1): 共同住宅の階数

上記の地域指標と空き巣発生率との散布図行列を図 2,3 に示す。なお、低層は 1.2 階建て、中層は 3-10 階建て、高層は 11 階建て以上とし、全世帯数で除することにより割合をデータとして用いる。図 2 の散布図行列はすべてのデータ、図 3 の散布図行列は空き巣が 0 件と 30 件以上の地区を除いている。図 2,3 の対角にデータ名、下三角にはおのこのデータの相関係数の絶対値、上三角には散布図を示している。

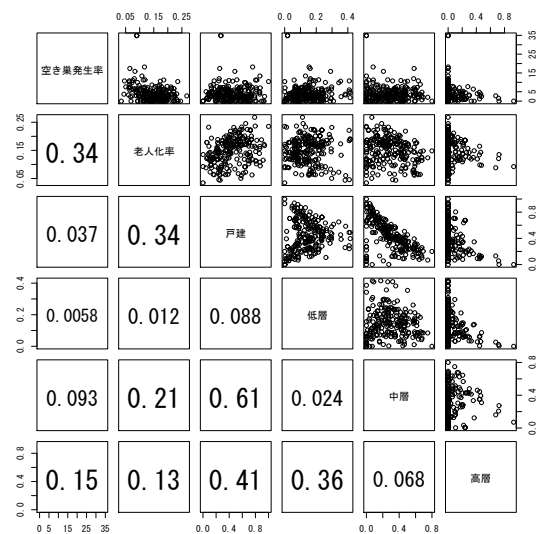


図 2 散布図行列

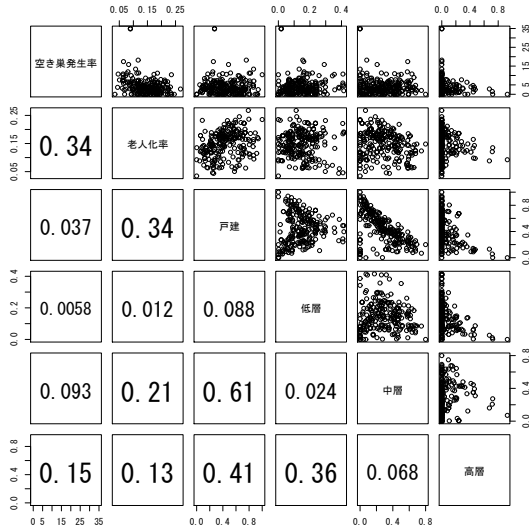


図 3 散布図行列

図 2,3 の散布図行列から、空き巣発生率と関係があり
 そうな地域指標は老人化率である。そこで空き巣発生
 率を被説明変数、老人化率を説明変数として以下のモ
 デルを構築しそれぞれの評価をする。

- 1 単回帰モデル
- 2 空間自己回帰モデル
- 3 地理的加重回帰モデル

各モデルの評価の基準として

- 1 残差二乗和
- 2 PSS
- 3 残差の Moran'I 統計量

各モデルの解析結果を次章で示す。

4.3 単回帰モデル

最小二乗法により、単回帰モデルの最小二乗推定量
 を求めた。

$$\hat{y}_i = 9.3 - 35.1x_i \tag{8}$$

係数はどちらも有意水準 1% で有意であった。

残差 e_i の Moran'I=0.13**。従って、単回帰モデルの残
 差に有意水準 1% で空間的な集積性があり独立性を仮
 定できない。決定係数が 0.11 であり、説明力の高いモ
 デルとは言えない。そこで次節以降の空間回帰モデル
 として空間自己回帰モデル、地理的加重回帰モデルの
 推定を行った。

表 6 単回帰モデル

残差二乗和	4838.6
PSS	12.06
残差の Moran'I	0.13**
決定係数	0.11

4.4 空間自己回帰モデル

空間自己回帰モデルを最尤法により推定した。

$$\hat{y}_i = 6.7 - 26.71x_i + 0.35 \sum_{j=1}^n w_{ij}y_j \tag{9}$$

係数はすべて有意水準 1% で有意であった。

表 7 空間自己回帰モデル

残差二乗和	4420.6
PSS	15.02
残差の Moran'I	-0.03
決定係数	0.19

$\rho=0.35$ という結果からも、川口市の空き巣発生率の
 集積性が伺える。

残差 e_i の Moran'I=-0.03。従って、空間自己回帰モ
 デルの誤差に独立性を仮定できる。しかし、PSS の値は
 悪化している。決定係数は 0.19 と単回帰よりも説明力
 は向上したものの、説明力が高いモデルとは言えない。

4.5 地理的加重回帰モデル

地理的加重回帰モデルを重み付き最小二乗法により
 推定した。

$$\hat{y}_i = \hat{\beta}_{0i} + \hat{\beta}_{1i}x_i \tag{10}$$

重み付き最小二乗法は

$$\min \left[\sum_{j=1}^n \{y_j - (\beta_{0i} + \beta_{1i}x_j)\}^2 h_{ij}^2 \right]$$

(5) の最適バンド幅は 1429.373 である。推定パラメー
 タは表 8 となった。となった。

表 8 地理的加重回帰モデルの推定パラメータ

	min	mean	max
$\hat{\beta}_{0i}$	-0.23	6.05	26.10
$\hat{\beta}_{1i}$	-128.14	-13.90	19.84

従って、地理的加重回帰モデルの誤差に独立性を仮
 定できる。また、残差二乗和、PSS とともに良化してい

表 9 地理的加重回帰モデル

残差二乗和	3197.92
PSS	4.1
残差の Moran'I	-0.03
決定係数	0.41

る。決定係数は 0.41 であり、単回帰、空間自己回帰モデルより説明力の高いモデルとなっている。

5 モデルの考察

誤差に集積性が示唆された単回帰モデルに対し、空間自己回帰モデルと地理的加重回帰モデルでは誤差の集積性がみられなかった。単回帰モデル、空間自己回帰モデルの老人化率の係数 $\hat{\beta}_{1i}$ は負である。これは高齢者の方々が監視役の役割を担うことによる結果であると思われる。

興味深いのは、GWR の $\hat{\beta}_{1i}$ が正になる地区が存在することである。正となっている地区は老人化率の抑制の効果があるとはいえない。GWR は老人化率の抑制力の空間的な変動を捉えたモデルとなっており、残差二乗和、PSS が一番小さいという結果になっており、決定係数の比較から説明力も一番優れていた。

図 4 に GWR のパラメータ $\hat{\beta}_{1i}$, 図 5 に残差の空間分布を示す。

正となっている地区は南西部の横曽根地区、北東部の戸塚地区である。

図 6 に正となった地区の $\hat{\beta}_{1i}$ の p 値のヒストグラムと、p 値が 0.05 よりも小さくなった地区を図 7 に示す。

ほとんどが有意水準 5% で有意であり、以後有意であった地区を β_{1i} が正である地区と解析を進める。

図 1 より正になった地区は老人化率が戸建て住宅割合からの影響を受けていると考えた。空き巣発生率と老人化率、戸建住宅割合との相関係数を表 10 に、 β_{1i} が正である地区と負になった地区とで層別し求めた偏相関係数を表 11 に示す。正となった地区の空き巣発生率と老人化率の偏相関係数（一戸建て割合の影響を除いた）は-0.05 であり相関係数の-0.34 とは異なればば無相関であり、空き巣発生率と戸建住宅との偏相関係数（老人化率の影響を除いた）は 0.26 と正の相関係数が見られる。正となった地区は戸建住宅の影響が強いことが分かる。

これに対して負となった地区の空き巣発生率と老人化率の偏相関係数は-0.36 であり相関係数とほぼ等しく、戸建住宅割合の影響は小さい。空き巣発生率と戸建住

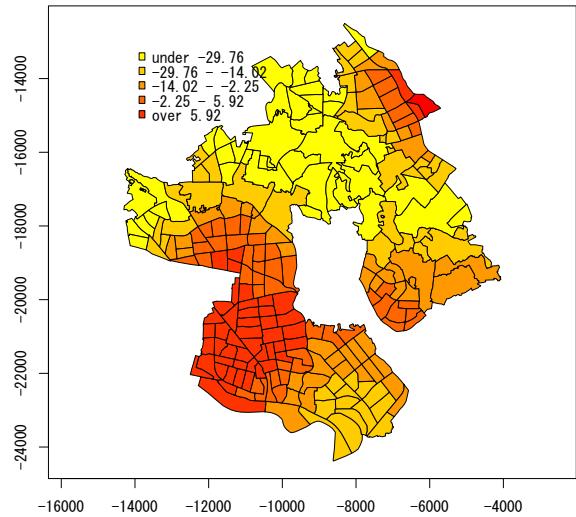


図 4 $\hat{\beta}_{1i}$

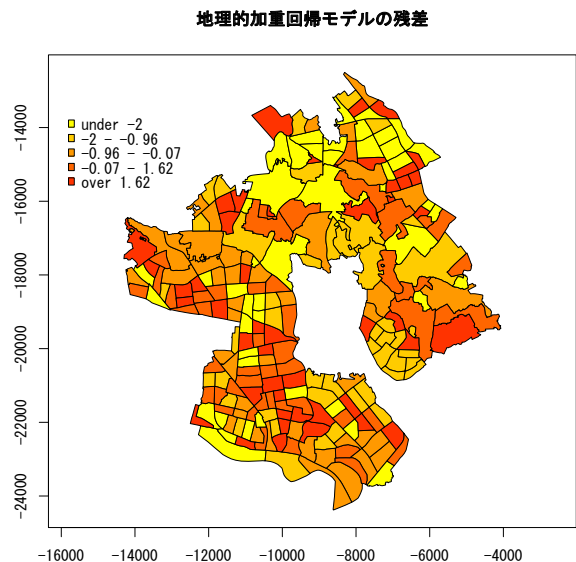


図 5 残差

宅の偏相関係数も 0.03 と無相関である。以上より、正となった地区 (ほぼ横曽根地区) は老人化率の抑制力がほぼ無いに等しく、戸建住宅割合による空き巣の誘発度が高い地区である。

表 10 相関係数

	老人化率	戸建住宅割合
空き巣発生率	-0.34	0.09

表 11 偏相関係数

	老人化率	戸建住宅割合
空き巣発生率 ($\beta_{1i} > 0$ の地区)	-0.05	0.26
空き巣発生率 ($\beta_{1i} < 0$ の地区)	-0.36	0.03

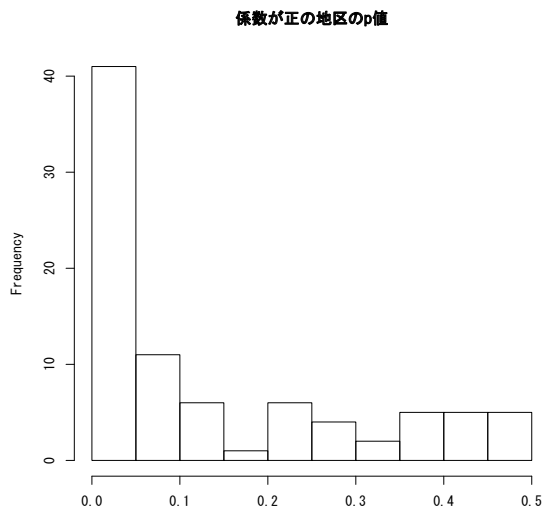


図 6 正となった地区の $\hat{\beta}_{1i}$ の p 値

参考文献

埼玉県川口市市民生活部
 防犯対策室犯罪情報ホームページ
 (<http://www.city.kawaguchi.saitama.jp/kbn/16050008/16050008.html>)
 埼玉県川口市市政情報ホームページ
 (<http://www.city.kawaguchi.saitama.jp/ctg/04013002/04013002.html>)
 杉浦芳夫編 (2002): 地理空間分析, 朝倉書店

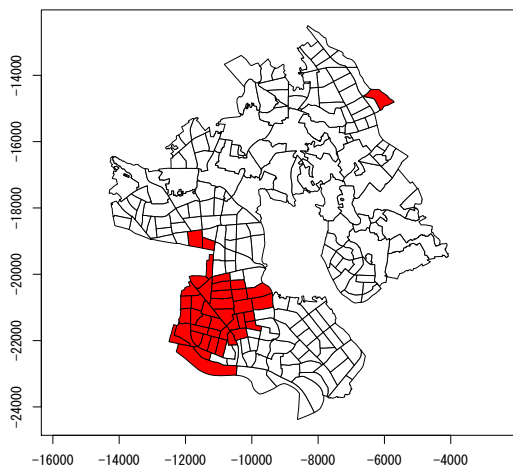


図 7 有意水準 5% で β_{1i} が正である地区