

# Emerging Patterns を用いた都市の車両犯罪の発生に関する分析

## Applying Emerging Patterns to the Analysis of Relationship between Car-Related Crimes and Urban Space

瀧澤重志†, 川口史恵†, 加藤直樹†, 森 健治‡, 吉田和生‡

Atsushi Takizawa, Fumie Kawaguchi, Naoki Katoh, Kenji Mori and Kazuo Yoshida

### 1. はじめに

ひったくり、強盗、車両盗などの街頭犯罪は、全刑法犯の約半数を占め、われわれにも身近な危険がある。近年地理情報システム (GIS) を用いて、街頭犯罪を中心とした犯罪マップが多くの警察本部で作成され、インターネットを通じて広く公開されるようになった。犯罪マップは、犯罪発生地点の位置情報からカーネル密度推定法を用いて、観測地点における犯罪の発生密度を可視化したもので、特にホットスポットと呼ばれる犯罪多発地域の抽出を目的とし、その公開による犯罪抑制効果が期待されている<sup>1)</sup>。

日本においては、発生地点を含む犯罪データが情報化されたのは比較的最近のことであり、大量のデータを用いた多角的な分析が始まっている。都市は社会的・空間的属性など多種多様な属性で特徴付けられ、犯罪発生との関連も深い。しかしカーネル密度分析法はあくまで発生地点の情報を利用するだけなので、こうした属性を絡めた分析はできない。また、仮説検証型の一般的な統計分析手法では、大量の属性で特徴付けられるデータの中で、犯罪発生に関連する未知の重要なパターンを見逃す可能性がある。

筆者ら (瀧澤, 川口, 加藤) は建築学の立場から、特に空間属性と犯罪発生との関係に着目している。本研究では、京都市の中でも空間構成に変化があり、犯罪発生件数も比較的多い京都市西京区において、街頭犯罪の中で特に空間的影響が大きいと考えられる車両犯罪 (自動車盗, 部品ねらい, 車上ねらい) の発生データと都市の空間のデータから、データマイニングの中でも相関ルール分析に分類される Emerging Patterns (EP)<sup>2)</sup>を用いて、車両犯罪と関係が深い空間構成パターンを発見することを試み、併せて、EPの指標である GrowthRate を用いて、空間に着目した車両犯罪の発生のしやすさを地図化する。EPは、あるクラスにのみ顕著に現れる相関ルールであり、少数例でも重要な空間構成ルールを抽出することを目標としている。本研究でEPを用いる理由は、車両犯罪の多発地域としてよく知られている幹線道路沿いなどではなく、見落としやすい危険性が高い場所や空間構成を抽出することを主な目標としているからである。

### 2. 犯罪データと分析方法

本研究では、2002年10月から2005年6月の間に京都市西京区内で発生した車両犯罪の発生地点のデータを用いる。車両犯罪の内訳を表1に示す。車両犯罪の発生場所は、駐車場が6割で、その他が路上、住宅、私有地内部の車庫、空き地などである。駐車場には車止めやゲートなどが存在することがあり、防犯性が他の場所と比較して異なる。そ

† 京都大学工学研究科建築学専攻

‡ 京都府警察本部

こで分析を二つに分け、駐車場に限定して現地調査を行い分析するものと、駐車場とその他の場所を区別せず地域を細かいメッシュに分割し、より広範囲で分析する2通りの分析を行った。紙面の都合上本報では、後者の一般の空間に関する分析に限定して説明する。この分析では、平面を一辺  $w$  (m) のメッシュに区切り、犯罪発生地点を含むメッシュを犯罪発生メッシュ、含まないメッシュを犯罪非発生メッシュとする。このメッシュ一つ一つの中心部において、後で述べる空間特徴量を計測して説明属性とする。なお、メッシュ中心点が建物の内部にある場合はそのメッシュを分析から除外する。今回は  $w=30$  として実験を行い、全部で13,077個のメッシュをデータとして用いた。

表1 2002年10月~2005年6月の間の京都市西京区内の研究対象地域における車両犯罪の発生件数

自動車盗	部品ねらい	車上ねらい	計
167	648	933	1,748

(注: 発生件数は速報値である)

### 3. 空間特徴量

本研究では、表2に示す9個の空間特徴量をデータマイニングの説明属性とする。表2において、 $sd_p$ における主要道路は国道9号線と嵯峨野線の住宅地図が定義している県道とする。また、 $tl_p$ は半径  $ar$  (m) の円内の道路の長さとする。さらに、犯罪が空間の監視性と大きく関わることから<sup>3)</sup>、空間の可視性の指標  $v_p$ ,  $c_p^j$  を定義する。その詳細に関しては後述する。なおクラスは車両犯罪の罪種を一括して扱い、車両犯罪の発生の有無の2クラスとした。

表2 用いた説明変数とその値

名前: 変数	値	
周囲の土地利用: $lu_p$	1: 森林・荒地, 2: 農地, 3: 造成地, 4: 空地, 5: 工業地, 6: 一般低層住宅地, 7: 密集低層住宅地, 8: 中高層住宅地, 9: 商業・業務地域, 10: 道路, 11: 公園・緑地, 12: 公共施設用地, 13: 河川・湖, 14: その他	
主要道路までのユークリッド距離: $sd_p$	1: 0~34.1m, 2: 34.1~88.1, 3: 88.1~176.3, 4: 176.3~369.8, 5: 369.8~	
観測地点周囲の道路長: $tl_p$	1: 0~506.9m, 2: 506.9~688.9, 3: 688.9~849.9, 4: 849.9~1028.1, 5: 1028.1~	
視体積: $v_p$	1: 0~44.5%, 2: 44.5~65.4, 3: 65.4~82.5, 4: 82.5~93.2, 5: 93.2~	
建物用途別の視線の衝突本数: $c_p^j$	$j =$ 戸建て住宅 (DH)	1: 0~62本, 2: 63~125, 3: 126~188, 4: 189~251, 5: 252~315
	集合住宅 (MD)	1: 0~49本, 2: 50~100, 3: 101~151, 4: 152~212, 5: 213~315
	商業・業務用建物 (CB)	1: 0~23本, 2: 24~50, 3: 51~83, 4: 84~130, 5: 131~309
	公共建物 (PB)	1: 0~38本, 2: 39~77, 3: 78~121, 4: 122~165, 5: 170~315
	無壁舎 (NW)	1: 0~43本, 2: 44~87, 3: 88~134, 4: 135~191, 5: 192~315

なお、EPを適用するために、表2の「値」欄に示すように数値属性を離散化しているが、本研究では各水準に入るデータの数ができるだけ均等化するように値の幅を決めて離散化している。

次に、任意の地点における人間から見た空間の可視量を、図1のような視線集合の可視ドームによってモデル化する。可視ドームのパラメータは、半径  $r$  (m)、視線の水平/垂直方向の分割ピッチ  $iH/iV$  (度)、の3つである。任意の視点  $p$  の可視量を次のように計算する。 $a_p^i, i=1,2,\dots,N$  ( $N$ : 視線の数) を  $p$  から  $i$  番目の視線、 $l(a_p^i)$  を  $a_p^i$  の長さとする。建物の用途を  $j=1,2,\dots,M$  ( $M$ : 建物用途の種類数) として、 $a_p^i$  がぶつかる建物の用途を  $t(a_p^i)$  で表す。これらの基礎的な量から、視体積  $v_p = \sum_{i=1}^N l(a_p^i)/(r \cdot N)$  と、建物用途別に視線が建物への衝突する本数  $c_p^j = |\{a_p^i | t(a_p^i) = j\}|$  を求め、可視性の指標とする。

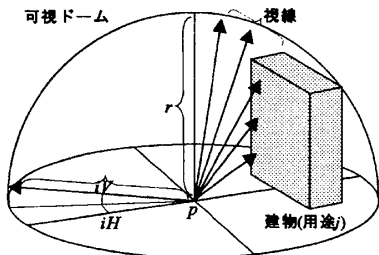


図1 可視ドーム

#### 4. Emerging Patterns (EP)<sup>2)</sup>

$T$  をデータベース全体、 $A$  を属性集合とする。ここでは2クラス問題に限定し、そのラベルをそれぞれ  $P, N$  とし、各クラスで  $T$  を2つのデータ集合  $D_P$  と  $D_N$  に分割する。トランザクションは、 $\{(A_i, v_i^t) | A_i \in A\}$  で定義される。ここで  $v_i^t$  は、トランザクション  $t \in T$  における属性  $A_i$  のデータの値である。また、それぞれの(属性、値)をアイテム、その集合をアイテム集合と呼ぶ。 $I = \{i_1, i_2, \dots, i_N\}$  を可能なアイテムの集合とすると、一つのトランザクションは、 $I$  の部分集合  $X$  で表現できる。 $k = |X|$  として、それを  $k$ -アイテム集合と呼ぶ。 $X \subseteq T$  のとき、トランザクション  $t$  がアイテム集合  $X$  を含むという。データ集合  $D$  における  $X$  のサポート  $supp_D(X)$  は、 $\#D(X)/|D|$  で求められる。ここで  $\#D(X)$  は、 $D$  内において  $X$  を含んでいるトランザクションの数である。 $\sigma$  を正の実数とし、 $supp_D(X) \geq \sigma$  のとき、 $X$  は  $D$  内において  $\sigma$ -large であるといい、そうでなければ  $\sigma$ -small であるという。 $Large_\sigma(X) / Small_\sigma(X)$  を、それぞれ、すべての  $\sigma$ -large /  $\sigma$ -small なアイテム集合全体とする。二つのデータ集合  $D_P, D_N$  において、 $X$  の  $D_N$  から  $D_P$  への  $GrowthRate(X)$  を下式のように定義する。 $GrowthRate(X)$  の値が大きい(通常1以上)  $X$  がEPとされる。なおEPの計算では、加藤がプロジェクトメンバーとなっているデータマイニングソフトのMUSASHI<sup>4)</sup>上で稼動するライブラリのプロトタイプを作成し、利用した。

$$Growthrate(X) = \begin{cases} 0, & \text{if } supp_{D_P}(X) = 0 \text{ and } supp_{D_N}(X) = 0 \\ \infty, & \text{if } supp_{D_P}(X) \neq 0 \text{ and } supp_{D_N}(X) = 0 \\ \frac{supp_{D_P}(X)}{supp_{D_N}(X)}, & \text{otherwise} \end{cases}$$

#### 5. 結果

既報<sup>5)</sup>の予備実験に基づき、 $ar = r=100, iH=iV=10$  として空間特徴量を計算した。車両犯罪は夜間発生が多く視界が限定されるが、 $v_p$  の分散を大きくするために  $r=100m$  としている。このデータから車両犯罪が発生しやすい空間構成のEPを求めた。表3に属性の組み合わせが4個のときのEPにおいて、 $GrowthRate(X)$  と  $\#D_P$  の順で降順にソートし、上位4位までとなる17個のEPを示す。発生件数を示す  $\#D_P$  は5~9件と、それなりの数がある。表3から、道路密集度が少なく周辺が無壁舎である地域(#1)、商業・業務地域で周辺に無壁舎が多い地域(#2-6)、集合住宅が多く可視領域が少ない地域(#10,17)、主要道路から少し離れた集合住宅の多い地域(#9,10,12-14,15,16)などの空間的特長が浮かび上がった。なお、得られたEPは12,761個と非常に多く、一つのメッシュに複数のEPが該当することも多いため、Strong EP<sup>6)</sup>と呼ばれるその部分集合がすべてEPであるようなEPに限定するなど、ルールの見通しをよくするために、結果を整理する必要があると考えている。

表3 犯罪発生に関する上位EP

#	$x_1$	$x_2$	$x_3$	$x_4$	$\#D_P$	$\#D_N$	$Growthrate(X)$
1	$tl_p\_1$	$c_p^{DH}\_1$	$c_p^{PB}\_1$	$c_p^{NW}\_5$	9	0	$\infty$
2	$tl_p\_9$	$tl_p\_1$	$c_p^{DH}\_1$	$c_p^{NW}\_5$	8	0	$\infty$
3	$tl_p\_9$	$tl_p\_1$	$v_p\_1$	$c_p^{NW}\_5$	8	0	$\infty$
4	$tl_p\_9$	$tl_p\_1$	$c_p^{PB}\_1$	$c_p^{NW}\_5$	8	0	$\infty$
5	$tl_p\_9$	$tl_p\_1$	$c_p^{CB}\_1$	$c_p^{NW}\_5$	8	0	$\infty$
6	$tl_p\_9$	$tl_p\_1$	$c_p^{MD}\_1$	$c_p^{NW}\_5$	7	0	$\infty$
7	$tl_p\_10$	$v_p\_1$	$c_p^{DH}\_1$	$c_p^{NW}\_4$	5	0	$\infty$
8	$tl_p\_9$	$sd_p\_3$	$c_p^{MD}\_5$	$c_p^{CB}\_1$	5	0	$\infty$
9	$sd_p\_4$	$tl_p\_1$	$c_p^{MD}\_5$	$c_p^{NW}\_1$	5	0	$\infty$
10	$sd_p\_4$	$tl_p\_1$	$v_p\_1$	$c_p^{MD}\_5$	5	0	$\infty$
11	$tl_p\_9$	$sd_p\_4$	$c_p^{MD}\_1$	$c_p^{NW}\_5$	5	0	$\infty$
12	$sd_p\_4$	$tl_p\_1$	$c_p^{DH}\_1$	$c_p^{MD}\_5$	5	0	$\infty$
13	$sd_p\_4$	$tl_p\_1$	$c_p^{MD}\_5$	$c_p^{CB}\_1$	5	0	$\infty$
14	$sd_p\_4$	$tl_p\_1$	$c_p^{MD}\_5$	$c_p^{PB}\_1$	5	0	$\infty$
15	$tl_p\_9$	$sd_p\_3$	$c_p^{MD}\_5$	$c_p^{NW}\_1$	5	0	$\infty$
16	$tl_p\_9$	$sd_p\_3$	$c_p^{MD}\_5$	$c_p^{PB}\_1$	5	0	$\infty$
17	$tl_p\_9$	$sd_p\_3$	$v_p\_1$	$c_p^{MD}\_5$	5	0	$\infty$

次に  $GrowthRate(X)$  の違いにより、各メッシュを色分けした地図を図2に示す。同じメッシュで複数のEPを有する場合があるが、この場合は  $GrowthRate(X)$  の値が大きいほうを採用している。比較のために図3に、犯罪マップで一般的に用いられているカーネル密度推定法により、犯罪発生密度の分布を色分けした地図を示す。

図2において濃く塗りつぶされているのは、国道9号線などの幹線道路沿いや洛西ニュータウンの高層住宅付近で、図3のホットスポットの分布とおおまかに対応する。しかし図3と比較して、値が局所的に大きく変動している。これは、本研究で提案した空間の可視性が局所的に大きく変動することに起因していると考えられる。カーネル法が発生地点の情報だけで犯罪発生密度分布を連続的に与えるのに対し、本研究では空間情報を用いることで、地域の空間特性を考慮してより細かく犯罪発生の危険性を抽出することができると思われる。ただし今回は、メッシュを一律に機械的に与えており、それらが微小変化した場合に結果がどのように変化するか、今後検証する必要がある。

なお図2は、各地点におけるEPの最大の  $GrowthRate(X)$  の大きさを示すものであり、各地点の車両犯罪発生総合的なリスクを直接示すものではない。送信するPDFファイルの容量から本報では掲載を見合わせた。研究ではロジスティック回帰などで車両犯罪の発生確率を求めるも行っている。この発生確率の地図を図2と比較すると、図2と大まかには類似しているものの、幹線道路沿いの発生確率が一律に上昇するといった違いが見られる。図2では幹線道路沿いの  $GrowthRate(X)$  は全体的に大きい、ところどころに細かな変動がある。これらの差からEPを用いた地図は、犯罪発生リスクと高い相関があるもの、見逃しがちな地域を優先的に評価しているのではないかと考えられる。なお、車両犯罪発生に関するより精密で総合的なリスクを計算するには、人口分布のデータなどが必要となるが、これは今後の課題とする。

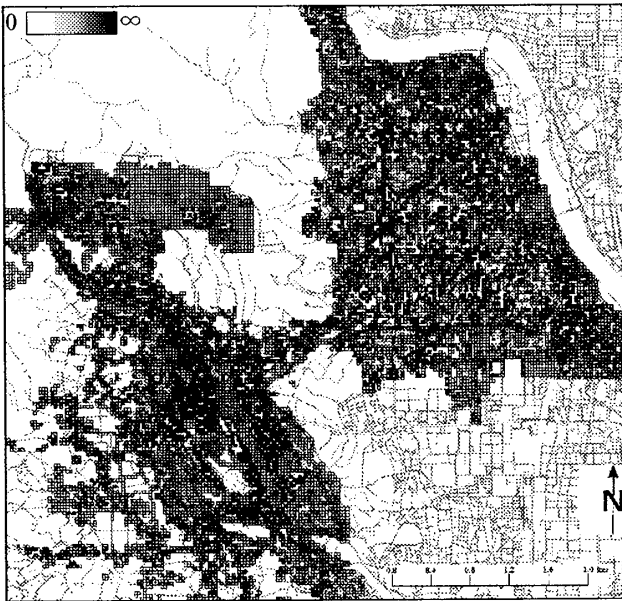


図2  $GrowthRate(X)$  の分布

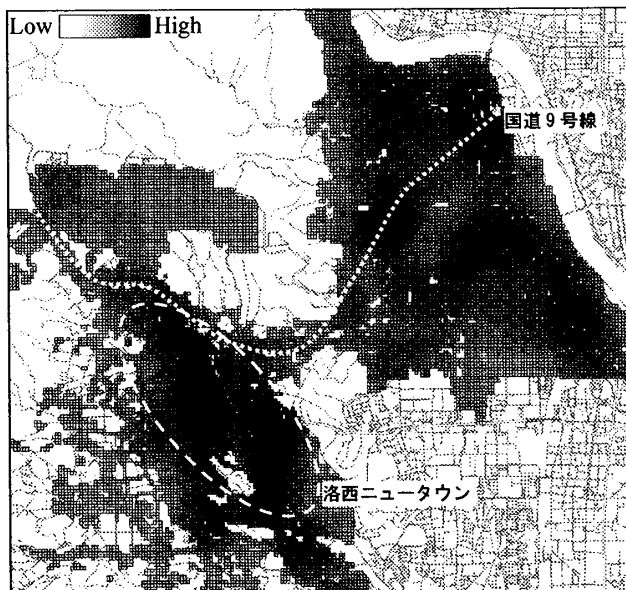


図3 カーネル密度推定法(バンド幅 300m)による車両犯罪マップ

## 6. まとめ

本研究では、京都市内の車両犯罪データと都市空間データから、EPを用いて車両犯罪の発生と関連のある都市の空間構成パターンを抽出した。さらにEPの  $GrowthRate$  により、従来よりもミクロなレベルの空間に着目した車両犯罪の起こりやすさを地図化した。

今後の課題としては、得られたEPの整理、可視領域計算の観測点に関する感度解析、リスクモデルとの結果の詳細な比較、デモグラフィック属性データの追加、などが挙げられる。

## 謝辞

本研究は、京都府警犯罪情勢分析室が主催するGIS活用研究会メンバーのご指導と、科学研究費補助金 基盤研究(C)(17500007)、若手研究(B)(18760460)の援助の下で行われました。

## 参考文献

- 1) 原田豊：都市の治安と犯罪地図，都市問題 第94巻9号，pp.3-19，2003.9.
- 2) Dong G., and Li J.: Efficient mining of emerging patterns: Discovering trends and differences, Proc. of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA: ACM Press, pp.43-52, 1999.
- 3) Oscar Newman 著，湯川利和，湯川聰子 訳：まもりやすい住空間-都市設計による犯罪防止，鹿島出版会，1976.
- 4) <http://musashi.sourceforge.jp/>
- 5) 瀧澤重志，川口史恵，加藤直樹：京都市西京区における都市空間の可視性分析，日本建築学会近畿支部研究報告集 計画系，pp.417-420，2006.
- 6) Li J. and Wong L.: Emerging Patterns and Gene Expression Data, <http://citeseer.ist.psu.edu/721728.html>.