

CAEPを用いた京都市郊外におけるひったくりの空間分析

京都大学工学研究科 瀧澤 重志
 南山大学数理情報研究センター 具 源龍
 京都大学工学研究科 加藤 直樹

1. はじめに

1970年代にアメリカにて防犯環境設計 (CPTED) の立場から、犯罪を起こしにくい環境を創り出す方法論が提案されて以来、日本でも防犯環境設計の在り方に関して研究がなされてきた。また、近年の地理情報や犯罪情報のデジタル化に伴い、蓄積されたデータを統計的に分析する事によって、犯罪マップの作成をはじめ、空間構造が犯罪に与える影響を客観的に分析する研究がなされてきた^{1,2)}。筆者らも京都市西京区において、街頭犯罪 (自動車盗難等) の空間分析をデータマイニングの手法により行い、洛西ニュータウンにおいて、空間の可視性などの属性の組み合わせが犯罪発生にかかわっていることなどを明らかにしてきた³⁾。続いて、京都市伏見区においてひったくりの研究を行ってきた^{4,5)}。ひったくりの研究では、空間的因子としてCPTEDにおいて重視される監視性や領域性の概念と関係する、可視領域に基づく街路の視覚的特長を加えて分析を行ったが、本研究ではこの研究を発展させ、分析方法としてCAEP⁶⁾と呼ばれる顕在パターンを基にしたクラス分類手法を用いて、ひったくりが起る場所と起らない場所を判別すると共に、発生場所の空間的特徴をパターンとして抽出することを旨とする。

2. データ

図1に分析対象地域を示す。分析対象地域は京都市伏見区の中心部の東西約1km、南北約2kmの範囲である。ここには約26,000人が住んでいる。また、京阪本線の丹波橋駅、伏見桃山駅、中書島駅と、近鉄京都線の近鉄丹波橋駅、桃山御陵前駅がある。この地域は、北は県道202号線、東と南は上記鉄道、西は新高瀬川で区切られており、地理的にまとまった単位となっている。

表1に使用するデータの出典を示す。使用した犯罪データは、京都府警察本部犯罪情勢分析室から提供を受けた、2004年1月～2005年12月に京都市伏見区で発生したひったくりのデータである。この期間に伏見区全体では343件のひったくりが発生した。このうち分析対象地域では96件のひったくりが発生したが、それらの中で9件はデータ#4の建物輪郭線の内部であったので、分析から除外した。

また#3の土地利用に関しては、分析するには詳細すぎると思われたので、図2のように概念階層を定義し、土地利用を一般化して用いた。概念階層はひったくりの発生を念頭において人の移動可能性を考慮して定義した。

さらに、2008年9月8日の20時から翌日3時にかけて、図1に示した対象道路の夜間照度測定を行った。GPSによ

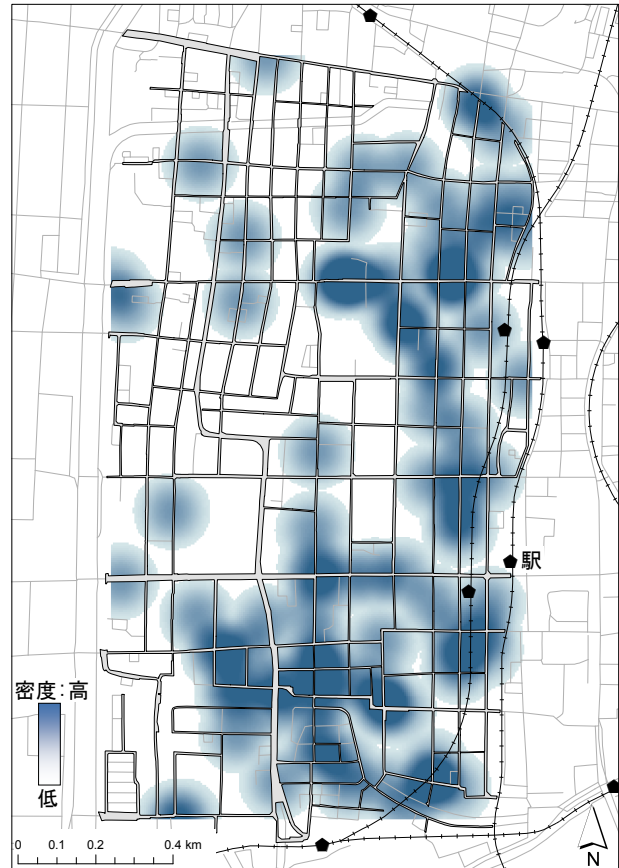


図-1 分析対象地域とひったくりのカーネル密度

表-1 使用したデータ

#	名称(対象地域, データ年など)	出典
1	ひったくりデータ (京都市伏見区, 2004年1月～2005年12月)	京都府警察本部犯罪情勢分析室
2	数値地図2500(空間データ基盤)(近畿-I, 2004年)	国土地理院
3	数値地図5000(土地利用)(近畿圏2001年)	国土地理院
4	Zmap Town II(京都市伏見区2005年)	㈱ゼンリン
5	平成12年国勢調査100mメッシュ推計データ(京都府)	㈱JPS

るナビゲーションソフトをインストールしたPDAを装着し、照度計をバイクのヘルメットに取り付けて、バイクにより当該道路を走行して1秒ごとの照度のログを記録した。照度計の高さは地上から約150cm、天候は晴れであった。測定後、GPSの軌跡と照度のログを時刻でマッチングさせ、

非建物用地 (os)
-- 移動可能空間 (ps)
---- 非道路 (nr)
----- 農地 (fi)
----- 空地 (va)
----- 公園 (pa)
---- 道路 (ro)
-- 移動不能空間 (=河川 (ri))
建物用地 (cs)
-- 住宅地 (ho)
---- 一般低層住宅地 (lh)
---- 密集低層住宅地 (dh)
---- 中高層住宅地 (mh)
-- 非住宅地 (nh)
---- 工業用地 (in)
---- 商業・業務用地 (bu)
---- 公共公益施設用地 (pu)

図-2 土地利用の概念階層 (斜体は著者が追加)

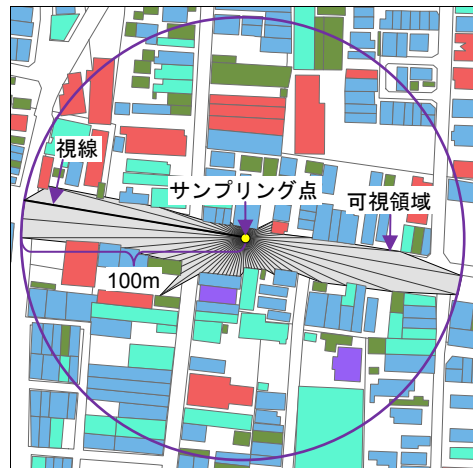


図-3 可視領域

照度に位置情報を付与した。

3. 説明属性

屋外の任意の点について、以下の説明属性を定義して計算する。

(1) 推定歩行者数

伏見区においては、駅やバス停から自宅への帰宅途中が最もひたたくりに合いやすい移動形態であることが、既往研究⁴⁾でわかっている。本報では既往研究と同様に、各住戸に割り当てられた人口が最寄駅に最短経路で歩行すると仮定し、その経路上に人口を加算していくことで、道路の推定歩行者数 (Pedestrian) を求める。考慮する駅は分析対象地域内の5つの駅と近隣の近鉄丹波橋駅、京阪宇治線観月橋駅、JR奈良線桃山町鍋島駅の3駅を加えた8駅とし、各駅を中心とする半径1km内を駅の利用圏と仮定した。計算方法を次に示す。データ#5の100mメッシュI内の住戸mの人口 P_m を、 $P_m = P_i S_m F_m / \sum_{i \in I} S_i F_i$ で与える。ここで P_i はIの夜間総人口、 S_m はmの面積、 F_m はmの階数である。つまり夜間人口をこの圏域の住居系建物の延べ床面積に比例させて、各住戸に配分する。

(2) 可視領域

可視領域の計算は指定された任意の点について行う。ここでは点を中心として半径100m範囲内で5°ピッチで72本の線分を発生させ、データ#4の建物輪郭線にぶつかる最短の線分をその方向の可視線分とする。それを基本データとして、可視線分の平均長さ (Visln_max), 平均 (Visln_mean), 最短 (Visln_min), 標準偏差 (Visln_std), 可視建物の棟数 (Visbsize), 可視建物の種類別の視線の総本数 (Visbtype1_A, Visbtype2_B), 可視線分の端点を順に結んでできる近似的な可視領域に含まれる土地利用の面積 (Visland_C) を求める。ここで、A, Bはそれぞれ異なる建物分類指標による建物種類を、Cは図2で示した土地利用の略号を示す。ここで、 $A = \{gb: 一般建物, ta: 目標物, nw: 無壁舎\}$, $B = \{pu: 各$

種公共施設, 神社, 仏閣など, ap : アパート, マンション, 団地など名称のある建物, ih : 個人の家屋, 事業: 事業所, bu : 事業所, その他}である。

(3) その他

その他の説明変数として、最寄り駅からの距離 (Distance), 点が属する100mメッシュ内の人口 (Population), サンプリング点の3つの最近傍の距離の重み付きの平均照度 (Illuminance), 点の半径20m以内の土地利用の面積 (Land_C) の合計49個の属性の値を計算する。

4. CAEP

CAEPはあるクラスについてのみ多頻度で、かつ他のクラスの事例においては相対的に頻度が少ない属性集合を示す顕在パターンを用いて、対象が所属するクラスを予測する新しいクラス分類手法である。CAEPの特徴は、分類精度の高さだけでなく、分類の過程で得られる顕在パターンにより、対象となるクラスがどのような属性の組合せで特徴付けられているかを、明らかにすることができる点である。CAEPの定義の準備として顕在パターンの説明をする。あるクラスCに属する事象データtの集合をDとして、属性値集合eを満たすtがDに含まれる頻度(支持度) $sup_c(e)$ は、次式で計算される。

$$sup_c(e) = \frac{|t \in D_c, e \subseteq t|}{|D_c|}$$

クラスCと他のクラスの集合 \bar{C} について、eの支持度を相対的に比較する値として、増加率 $growth_rate_c(e)$ を次式で定義する。

$$growth_rate_c(e) = \begin{cases} \frac{sup_c(e)}{sup_{\bar{c}}(e)} & (sup_{\bar{c}}(e) \neq 0) \\ \infty & (sup_{\bar{c}}(e) = 0) \end{cases}$$

$growth_rate_c(e) > 1$ となる e を顕在パターンと呼ぶが、その値が高いほど、 e が C のみに含まれる可能性が高くなる。以上が顕在パターンの定義である。CAEP では顕在パターンにより抽出された e の寄与度 $\alpha_c(e)$ (クラスのカテゴリ力) を、増加率と支持度の両者を考慮して以下のように定義する。

$$\alpha_c(e) = \frac{growth_rate_c(e)}{growth_rate_c(e) + 1} \cdot sup_c(e)$$

$growth_rate_c(e) = \infty$ のときは $\alpha_c(e) = \infty / (\infty + 1) \cdot sup_c(e) = sup_c(e)$ となる。トレーニングデータから抽出したクラス C に対する顕在パターンの集合を $E(C)$ とすると、インスタンス s が C に属する可能性を示す、集約スコア $score(s, C)$ は以下の式で計算される。

$$score(s, C) = \sum_{e \in s, e \in E(C)} \alpha_c(e)$$

集約スコアはクラスの件数に影響を受けるため、以下のよう
に基準化した値を最終的に用いる。

$$norm_score(s, C) = \frac{score(s, C)}{base_score(C)}$$

ここで $base_score(C)$ は、訓練データの集約スコアの中央値である。そして未知のインスタンス s について、 $norm_score(s, C)$ が最も大きいクラスに分類する。

5. 分析

(1) 準備

ひったくりの分析は道路ネットワークの単位で分析がなされることが多いが、本研究では空間的可視性を評価するために、点をサンプリングの単位とした。ひったくりは歩行者が日常歩行する道路の沿道で多く発生することを考慮し、図4のように、道路の境界から1mだけ内側に沿って歩行者動線を想定した線分を発生させた。なお十字型の交差点部分は、信号付きの交差点で横断歩道がある場所については横断歩道の中心を、信号が無い交差点では交差点に対して向かい合う道路の歩行者動線を最短で結ぶ線分で結んだ。その線分上10mごとにサンプル点を設け、7878点を得た。このようにして求めたサンプリング点7,878点と、ひったくり点87点を合わせて分析用データを構築する。各点にひったくりの発生：P、もしくは非発生：Nのクラスラベルを付与するが、既報3)に従い、ひったくり発生点の半径20m以内のサンプリング点もクラスPとみなした。なおサンプリング点が複数のひったくり点の円内に含まれる場合は、ひったくり多発地区とみなし、それらを新しくクラスPとしてデータをコピーした。結果としてPが978個、Nが7,071個からなる二つのデータセットを構築した。そしてCAEPを適用するために、各データセットのそれぞれの属

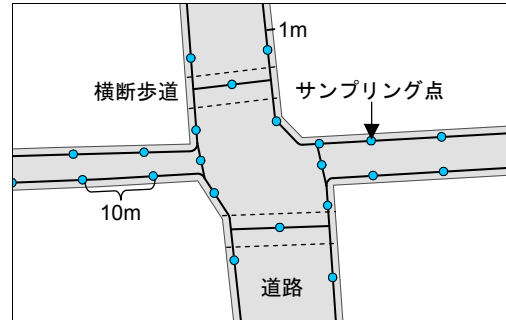


図-4 サンプリング点の設定

性を、その度数ができるだけ均等になるように3水準に離散化し、各水準を昇順に{L, M, H}とした。

(2) 結果

10回の交差検証によりCAEPの分類精度を求めた。その結果、クラスPで0.728、クラスNで0.807、全体で0.797と概ね良好な分類精度を得た。

表2はクラスPのデータセットにおいて、CAEPによって得られたサポートが上位10個の顕在パターンを示している。図5、6は表2の顕在パターンを構成する各属性の関係性を、可視化ソフト⁷⁾によりネットワーク表現したものである。太いエッジほどひったくりの発生に関して結びつき(サポートの和)が強いことを示しており、ここでは、 $Visland_nr=M$ 、 $Visland_ri=M$ 、 $Visland_pu=H$ という、見渡せる範囲内に人の気配があまり感じられない場所の特徴を示している。

図7、8は、どちらも駅周辺となる対象地域の北、南側での分類結果を示している。北側は住居系地域、南側は飲食店が多い商業・業務地域が大半を占める。分類が概ね正確な北側住居系地域では、空間的に比較的予測可能な場所でひったくりが発生する傾向が見られ、分類に不正確さが目立つ南側繁華街では、様々な場所でひったくりが起これることを示唆しているといえるかもしれない。

6. まとめ

本研究では、CAEPを用いて京都市伏見区中心部で発生したひったくりの空間分析を行い、ひったくりの発生に特徴的な空間的パターンと、住居系地域と繁華街での潜在的なひったくり発生場所の相違点について知見を得た。

謝辞

犯罪発生データを提供して頂いた京都府警察本部犯罪情勢分析室に感謝いたします。なお本研究は、科学研究費補助金若手研究(B)(20760405)の援助の下で行われました。

参考文献

- 1) 石川愛, 鈴木広隆: 道路ネットワークにおける見通し距離とひったくり発生との関係に関する研究: 大阪市住宅系地区を対象として, 日本建築学会環境系論文集, 73(623), pp.101-106, 2008.1.
- 2) 森口幸信, 吉川 眞, 田中一成: ひったくりのネットワーク空間分析, 地理情報システム学会講演論文集, 16, pp.343-346, 2007.

表-2 サポートが上位10位の顕在パターン

Class N				Class P					
Support	Total	Item 1	Item 2	Item 3	Support	Total	Item 1	Item 2	Item 3
0.153	1,084	Distance=H	Pedestrian=L	Population=L	0.072	70	Visln_mean=H	Visland_ri=M	Visland_nr=M
0.119	841	Distance=H	Visland_pu=L	Visland_nr=L	0.071	69	Visland_ri=M	Visland_nr=M	Visland_nh=H
0.117	826	Distance=H	Population=L	Visland_ri=L	0.067	66	Visln_std=H	Visland_ri=M	Visland_nr=M
0.116	823	Distance=H	Visland_bu=M	Visland_pu=L	0.067	66	Visland_ri=M	Visland_nr=M	Visland_cs=H
0.116	822	Distance=H	Visbtype2_pu=L	Visland_bu=M	0.064	63	Population=M	Visland_va=M	Visland_mh=H
0.114	805	Distance=H	Visbtype1_nw=L	Visland_va=L	0.064	63	Visbtype1_nw=L	Visland_pu=H	Visland_ri=M
0.111	784	Distance=H	Visland_mh=L	Visland_nr=L	0.064	63	Visbtype2_bu=H	Visland_mh=H	Visland_nr=M
0.109	769	Distance=H	Visbtype1_ta=L	Visland_bu=M	0.063	62	Distance=M	Visland_pu=H	Visland_ri=M
0.106	752	Distance=H	Visland_in=L	Visland_cs=L	0.063	62	Visland_pu=H	Visland_ri=M	Visland_nr=M
0.106	751	Distance=H	Population=L	Visland_va=L	0.062	61	Population=M	Visln_mean=H	Visland_ri=M

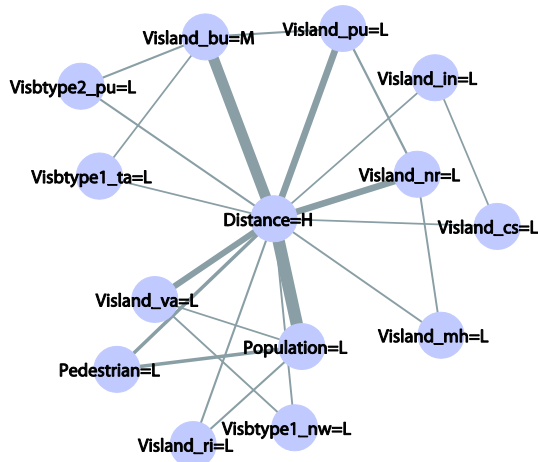


図-5 顕在パターン中のアイテムの関連: Class N

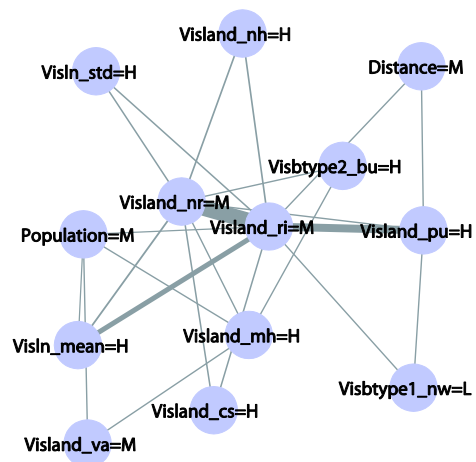


図-6 顕在パターン中のアイテムの関連: Class P

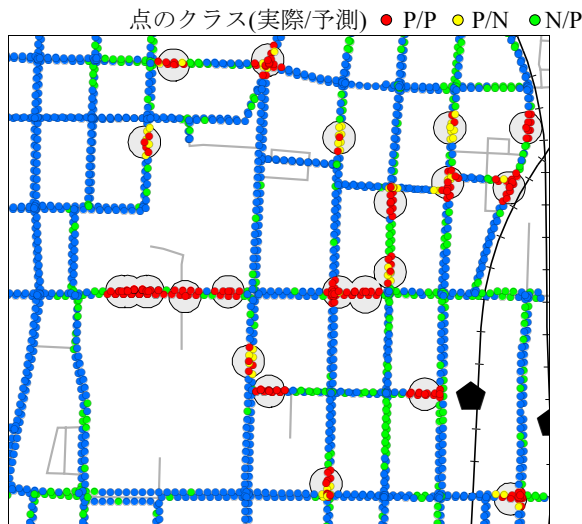


図-7 分類結果: 対象地域北側

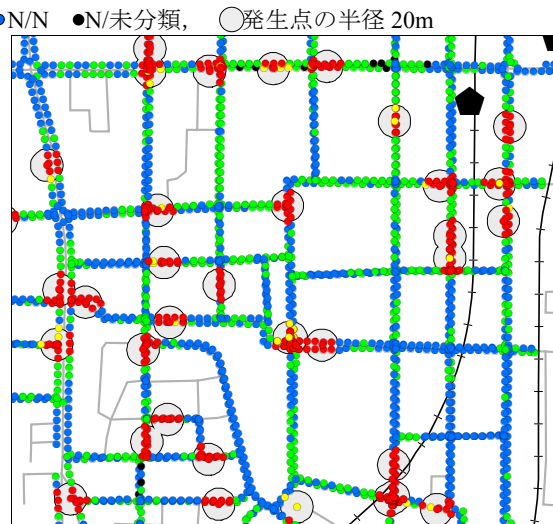


図-8 分類結果: 対象地域南側

- 3) A. Takizawa, et al.: Risk Discovery of Car-Related Crimes from Urban Spatial Attributes Using Emerging Patterns, Int. J. of Knowledge-based and Intelligent Engineering Systems (KES), 11(5), pp.301-311, 2007.12.
- 4) 瀧澤重志, 佐伯研, 加藤直樹: 京都市伏見区中心部におけるひたつくりの空間分析, 日本建築学会第30回情報・システム・利用・技術シンポジウム論文集, pp.165-168, 2007.
- 5) 瀧澤重志, 具源龍, 加藤直樹: 街路の視覚的情報を考慮した京都市伏見区中心部のひたつくりの空間分析, 日本建築学会第31回情報・システム・利用・技術シンポジウム論文集, pp.127-130, 2008.
- 6) G Dong, et al.: CAEP: Classification by Aggregating Emerging Patterns, Int'l Conference on Discovery Science, 30-42, 1999.
- 7) Cytoscape Consortium: Cytoscape, <http://www.cytoscape.org/>, 2009.6.1.