

論文紹介

犯罪発生履歴データの機械学習による 時空間カーネル密度推定型犯罪予測の最適化

中市新太

富山県立大学
`u020025@st.pu-toyama.ac.jp`

May 12, 2023

自分の研究についての課題

未達成：GMDH について勉強し，自研究に応用できる点を探る．
未達成：引継ぎ研究をグリッドごとから町字ごとに変えてみる．
未達成：SHAP を勉強する．

今回紹介する論文で生かせそうなこと

パトロールリソースを制約条件に加え予測していた．
提案手法，比較方法のほかランダムを加えて予測精度を確認していた．
既存の理論を用いて犯罪が起こる犯罪別で適したモデルがあること．

1 本研究の背景と目的

3/20

背景

犯罪予測とは、犯罪の発生履歴データから近い将来における犯罪発生場所を予測する技術で、警察のパトロールにおける利用が始まっている。

犯罪予測に関する研究は、実際の発生位置を含む履歴データが広く公開されてこなかったため、これまで主に警察関係者や犯罪学専門家により進められてきた。

しかし、近年この状況が変わりつつある。

目的

本論文は、パトロール業務支援を目的に、犯罪発生履歴データから近い将来における犯罪発生リスクが高い場所を予測する手法として、最適化 KDE 法を提案する。

カーネル密度推定 (Kernel Density Estimation : KDE) :

離散的データから密度分布を推定する手法として知られる。

$$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^n K \frac{x - X_i}{h}$$

h をバンド幅と呼び、本研究では機械学習により推定する。

2 研究概要

4/20

概要

犯罪発生履歴データを入力し、予測日において、予測対象エリア内のセルにおける発生件数の密度推定値を時空間カーネル密度推定で計算する。この値が高いセルから、セルカバー率分のセル数の集合を抽出して、予測日において発生を予測するセルとする (3.1)

カーネル関数のバンド幅は、バンド幅候補リストを与えて、セルカバー率を制約してパトロールカバー率を評価関数にした犯罪発生履歴データの機械学習により、候補リストのなかから最適組を選択する。(3.2)

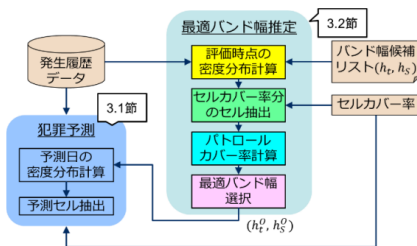


図 1 最適化 KDE 法のブロック図

図 1: ブロック図

3.1.1 提案手法

5/20

1

犯罪発生位置と発生年月日からなる，罪種ごとの犯罪発生履歴データ (x_i, y_i, t_i) , $(i = 1, 2, \dots, I)$ を用いて，予測日における発生場所を罪種ごとに予測する．

予測対象エリア内セルの中心点 (x, y) ，年月日 t における犯罪発生件数の密度推定値を，時間成分と空間成分のカーネル関数を各々 K_t , K_s として次式で定義する．

$$f(x, y, t) = \frac{1}{h_s^2 h_t} \sum_{i=1}^I K_t\left[\frac{t-t_i}{h_t}\right] K_s\left[\frac{x-x_i}{h_s}, \frac{y-y_i}{h_s}\right] (1)$$

h_t はカーネル関数 K_t のパラメータであるバンド幅， h_s はカーネル関数 K_s のバンド幅である．

3.1.2 提案手法

6/20

2

h_t はカーネル関数 K_t のパラメータであるバンド幅, h_s はカーネル関数 K_s のバンド幅である.

カーネル関数は, 犯罪データ分析ツール CrimeStat が提供するカーネル関数のなかから, K_t は最も単純な一様分布を, K_s は先行研究 が用いた 4 次関数を選択する.

セル中心点 (x, y) からある犯罪発生位置 (x_i, y_i) への相対座標値を (x_r, y_r) , すなわち, $x_r = x_i - x$, $y_r = y_i - y$ とすると, カーネル関数 K_t , K_s は下式 で表すことができる.

$$K_t(t) \propto 1 \quad \text{if } |t| < h_t$$

$$K_S(x_r, y_r) \propto \left(1 - \frac{x_r^2 + y_r^2}{h_S^2}\right)^2 \quad \text{if } x_r^2 + y_r^2 < h_S^2$$

図 2: 式

3.1.3 提案手法

7/20

はじめに
提案手法
まとめ

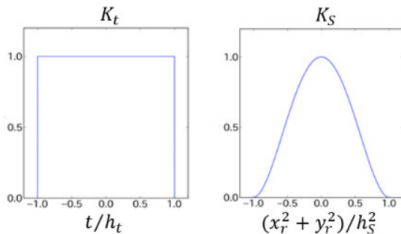


図 2 カーネル関数

図 3: 分布

3

年月日 t における発生予測場所は、予測対象エリア内の全セルから式 (1) の密度推定値が高い順に抽出した、セルカバー率未満の割合のセルとする。セルカバー率は式 (3) で定義する面積割合である。

3.1.4 提案手法

8/20

セルカバー率 β

\equiv (パトロール可能なセル数)

/(予測対象エリア内全セル数) ($0 \leq \beta \leq 1$) (3)

図 4: セルカバー率

3.2.1 機械学習

9/20

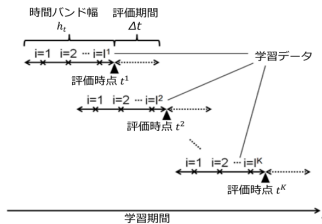


図 3 学習データの設定方法

図 5: 分布

1

まず、学習期間内の時間軸に対し、学習データ収集年月日として複数の評価時点 t^k ($k = 1, 2, \dots, K$) を設定する。

次にバンド幅候補リスト中のある組 (h_t, h_s) を用いて、評価時点 t^k における密度推定値 $f(x, y, t^k)$ を、 t^k から過去時間バンド幅 h^k 内の I^k 件の発生履歴データを学習データとして式 (1) で計算する。そして予測対象エリア内の全セルから、 $f(x, y, t^k)$ が高い順に、セルカバー率 β 以内の割合のセルを抽出して、 t^k における発生を予測するセル集合 $G_k(\beta)$ とする。

パトロールカバー率

$$\begin{aligned} &\equiv \sum_{k=1}^K (t^k \text{ の評価期間内発生のうち } G^{k(\beta)} \text{ における} \\ &\quad \text{発生数}) \\ &\quad / \sum_{k=1}^K (t^k \text{ の評価期間内発生数}) \end{aligned} \quad (4)$$

図 6: パトロールカバー率

n : 予測できた発生件数 N : 予測期間内全発生件数

a : 予測したセル数 A : 予測対象エリア全セル数

として PAI は次式で計算する.

$$PAI \equiv (n/N)/(a/A) \quad (5)$$

図 7: PAI

3.2.3 機械学習

11/20

表 1 最適化 KDE 法のアルゴリズム

Table 1 Algorithm of optimized KDE.

最適バンド幅推定 (3.2 節)

- Step 1: パトロールに割り当てるリソース量を考慮して、セルカバー率を設定。
 Step 2: バンド幅組ごとにパトロールカバー率を計算。
 Step 3: パトロールカバー率が最高になるバンド幅組を、最適バンド幅として選択。

犯罪予測 (3.1 節)

- Step 1: 最適バンド幅を使い、予測日の犯罪発生件数密度を推定。
 Step 2: 件数密度推定値が高いセルから順に、セルカバー率分のセルを、予測日における発生予測場所として出力。

図 8: アルゴリズム

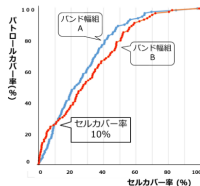


図 4 セルカバー率に対するパトロールカバー率 (イメージ)

図 9: イメージ

3.2.4 機械学習

12/20

2

予測したセルをすべてパトロールする場合以下の式となる.

予測したセル数/予測対象エリア全セル=セルカバー数

$PAI = \text{パトロールセルカバー数} / \text{セルカバー数}$

3

米ポートランド警察の 5 年 3 月分の犯罪発生履歴を使う.

東西 23 m, 南北 17 m の長方形セルに分ける.

NIJ コンテスト条件を基準にする.

3.2.5 機械学習

13/20

表 1 最適化 KDE 法のアルゴリズム

Table 1 Algorithm of optimized KDE.

最適バンド幅推定 (3.2 節)

- Step 1: パトロールに割り当てるリソース量を考慮して、セルカパー率を設定。
 Step 2: バンド幅組ごとにパトロールカパー率を計算。
 Step 3: パトロールカパー率が最高になるバンド幅組を、最適バンド幅として選択。

犯罪予測 (3.1 節)

- Step 1: 最適バンド幅を使い、予測日の犯罪発生件数密度を推定。
 Step 2: 件数密度推定値が高いセルから順に、セルカパー率分のセルを、予測日における発生予測場所として出力。

図 10: アルゴリズム

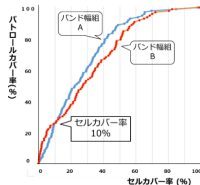


図 4 セルカパー率に対するパトロールカパー率 (イメージ)

図 11: イメージ

4.1 実験

14/20

はじめに
提案手法
まとめ

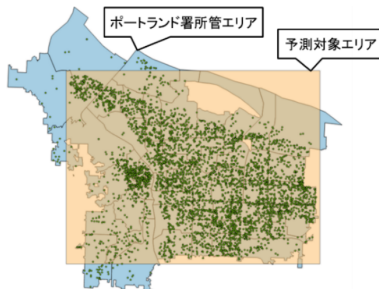


図 5 ポートランド署所管エリアと侵入盗発生地点

図 12: 実データ

罪種	予測対象エリア内件数	セルあたり件数
自動車盗	9,658 件	0.141 件
路上犯罪	163,785 件	2.395 件
侵入盗	5,277 件	0.077 件

(注) 予測対象エリア内セル数 68,400.

図 13: 犯罪数

4.2 実験

15/20

表 3 バンド幅候補リスト

Table 3 List of candidate bandwidths.

種類	値	組数
時間 バンド幅(日)	7, 28, 91, 183, 365, 730, 1095, 1460	8 組
空間 バンド幅(m)	75, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000	17 組
組合せ合計		136 組

表 4 学習期間と、予測対象エリア内の発生データ数

Table 4 Number of crime events for each training case.

No.	学習期間	自動車盗	路上犯罪	侵入盗
①	2016 年 3 月-5 月	508 件	8,278 件	196 件
②	2016 年 6 月-8 月	506 件	9,709 件	228 件
③	2016 年 9 月-11 月	710 件	8,157 件	258 件
④	2016 年 12 月 -2017 年 1 月	821 件	6,739 件	251 件
⑤	2017 年 3 月-5 月	792 件	8,358 件	261 件

図 14: 表

4.3 実験

16/20

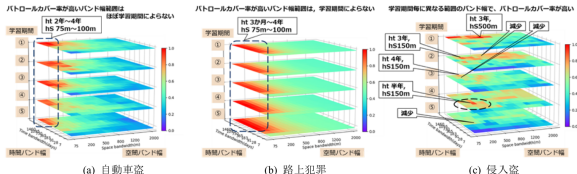


図 6 バンド幅候補に対する、学習期間ごとのパトロールカバー率（セルカバー率 1%）

図 15: 幅

罪種	学習期間	セルカバー率	
		1%	5%
自動車盗	①	(1095, 100)	(1460, 75)
	②	(1095, 100)	(1460, 150)
	③	(730, 75)	(1460, 75)
	④	(730, 75)	(1095, 75)
	⑤	(1460, 75)	(1460, 75)
路上犯罪	①	(365, 75)	(730, 75)
	②	(183, 75)	(730, 75)
	③	(365, 75)	(730, 75)
	④	(730, 75)	(730, 75)
	⑤	(730, 75)	(365, 75)
侵入盗	①	(1095, 500)	(1460, 300)
	②	(1095, 150)	(1460, 150)
	③	(1460, 150)	(1460, 300)
	④	(183, 150)	(1460, 400)
	⑤	(183, 150)	(1460, 300)

図 16: 最適バンド幅組

4.4 実験

17/20

はじめに
提案手法
まとめ

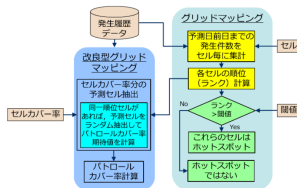


図 7 グリッドマッピングのブロック図

図 17: グリッド

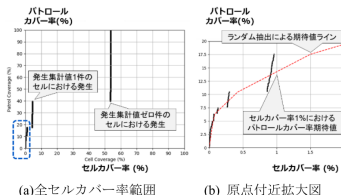


図 8 グリッドマッピング結果のパトロールカバー率 (自動車盗, 学習期間①)

図 18: カバー

4.5 実験

18/20

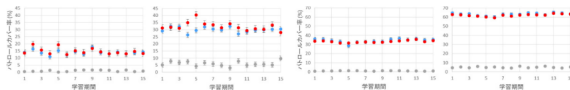
表 6 全学習期間の平均パトロールカバー率

Table 6 Average patrol coverage of all training cases.

罪種	予測手法	セルカバー率		学習期間ごとの 発生件数
		1%	5%	
自動車盗	提案手法	14.72 ± 2.16	32.35* ± 2.82	222.47 ± 48.26
	比較手法	13.95 ± 1.58	29.97 ± 1.72	
	ランダム	0.95 ± 0.53	6.07 ± 1.62	
路上犯罪	提案手法	33.22 ± 1.13	62.04 ± 1.26	2749.40 ± 360.11
	比較手法	34.17* ± 1.98	63.18** ± 1.38	
	ランダム	1.07 ± 0.19	5.08 ± 0.64	
侵入盗	提案手法	11.43** ± 2.42	28.26** ± 2.89	79.60 ± 10.20
	比較手法	7.94 ± 3.40	21.05 ± 5.74	
	ランダム	0.37 ± 0.62	5.69 ± 1.20	

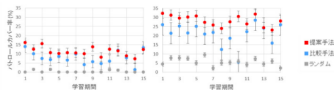
(注) 平均値 ± 標準偏差, 太字は罪種ごとの最高平均パトロールカバー率を示す,

Wilcoxon の符号付順位検定 (片側) * : $p < 0.0125$, ** : $p < 0.0005$



(a) 自動車盗 セルカバー率 1% : 5%

(b) 路上犯罪 セルカバー率 1% : 5%



(c) 侵入盗 セルカバー率 1% : 5%

図 9 学習期間ごとのパトロールカバー率 (注) エラーバーは標準偏差を示す,

図 19: カバー率

表 7 全学習期間の平均 PAI

Table 7 Average PAI of all training cases.

罪種	セルカバー率	提案手法	比較手法
自動車盗	1%	14.72	13.95
	5%	6.47	5.99
路上犯罪	1%	33.22	34.17
	5%	12.41	12.64
侵入盗	1%	11.43	7.94
	5%	5.65	4.21

(注) 太字は罪種ごとの最高 PAI を示す。

図 20: 平均

結果

既存研究の PAI2.11 よりも高い精度で予測できた。

今後の展開

最適化 KDE 法と環境要因データに基づく予測との統合的手法，複数の時空間カーネル関数を用いた犯罪予測モデルへの拡張．継続的な学習方式の適応を検討する．