

推薦論文

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

中川 淳子^{1,a)} 西村 祥治^{1,b)} 宮野 博義^{1,c)}

受付日 2019年8月5日, 採録日 2020年4月8日

概要: 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。本論文は、離散的な犯罪発生履歴データに時空間カーネル密度推定を適用して、発生件数密度分布の高密度エリアを犯罪発生場所として予測する、犯罪予測手法を提案する。カーネル関数のパラメータであるバンド幅は、パトロール実施可能な面積割合を制約条件として与え、その面積分の場所を予測した場合の予測件数的中率が最大になるように、発生履歴データを機械学習して推定する。これによりパトロールに割り当てるリソース量の多寡を最適バンド幅に反映する。既存手法ではこのようなパラメータ設定は犯罪学における知見に基づいて行う必要があったが、提案手法はそれを不要にする。米国立司法省研究所の公開データを用いた実験で、提案手法で得られた罪種ごとの最適バンド幅の傾向は、犯罪学における既存の知見と合致した。さらに、比較手法として犯罪発生集中場所を意味するホットスポットを分析する既存手法を改良して用いた実験で、罪種ごとの平均予測件数的中率を評価し、提案手法の犯罪予測への有効性を確認した。

キーワード: 犯罪予測, カーネル密度推定, 時空間分析, 機械学習, 地理情報システム (GIS)

Machine Learning Based Parameter Optimization of Spatiotemporal Kernel Density Estimation for Crime Prediction

JUNKO NAKAGAWA^{1,a)} SHOJI NISHIMURA^{1,b)} HIROYOSHI MIYANO^{1,c)}

Received: August 5, 2019, Accepted: April 8, 2020

Abstract: Crime prediction is a technique to predict places where crimes will occur in near future from crime records, and it is getting widely applied to patrol planning in police. In this paper, we propose a crime prediction method that estimates the density distribution on number of crimes from crime records with spatiotemporal kernel functions, and predicts high-crime rate area based on high-density area. Our method applies machine learning to optimize the bandwidth parameters of kernel functions that maximize the crime rate within the predicted area under the given condition of patrol area size constrained by patrol resource availability. This optimization enables our method to select the best bandwidth parameters, while traditional approaches required criminological knowledge to select them. We conducted experiments with a crime dataset provided by the National Institute of Justice and found that our method estimated reasonable values of the bandwidth parameters for each crime type from the point of criminology. We also evaluated the average precision of crime prediction and found that our method outperformed against traditional hot spot analysis. We thus confirmed the effectiveness of our method for crime prediction.

Keywords: crime prediction, kernel density estimation, spatiotemporal analysis, machine learning, geographic information system (GIS)

1. はじめに

犯罪の発生年月日と場所からなる履歴データ等を用い

本論文の内容は2018年9月のFIT2018第17回情報科学技術フォーラムにて報告され、同プログラム委員長により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

¹ NEC バイオメトリクス研究所
NEC Biometrics Research Laboratories, Kawasaki 211-8666,
Japan

a) jnakagawa.ay@nec.com

b) nishimura@nec.com

c) h-miyano@nec.com

て、近い将来における犯罪発生場所を予測する技術が、警察が行うパトロール業務に利用されはじめている。このような技術は犯罪予測と呼ばれ、欧米で商用サービスが開始し、日本でも警察の運用事例がある [1].

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。2016年、米国立司法省研究所 (National Institute of Justice : NIJ と略) は犯罪予測手法の比較を目的としたコンテストを主催し、米ポートランド警察における犯罪発生位置の緯度経度情報を含む履歴データを分析用に公開した [2]. コンテストのタスクは「予測対象エリアをセルに分割し、所定の予測期間に犯罪発生が集中しそうな場所とそうでない場所を 1/0 で区別せよ」であり、主な条件は、予測は罪種ごと、予測期間は特定の年月日から 1 週間/2 週間/1 カ月間/2 カ月間/3 カ月間、等である。NIJ はコンテストの目的の 1 つに、従来とは異なる分野の研究者による取り組み促進をあげていることから、今後多方面における研究活性化が期待できる。

警察には、犯罪被害を未然に防ぐという発想が重要で、そのためには発生が多い地点や時間を見極めようというリソース投入が必要という考え方がある [3]. 警察や地域の防犯ボランティアに対し、犯罪が起こりそうな場所や時間の予測情報を提供できれば、パトロールする場所を決める際の参考として有用である。そこで本論文は、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) は、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 [4]. 犯罪発生履歴データにも適用可能だが、パラメータとしてカーネル関数のバンド幅を与える必要がある。本論文は、犯罪発生履歴データだけから犯罪予測の処理を実現するため、バンド幅を発生履歴データの機械学習により推定する。そして機械学習の評価関数に、パトロール実施可能な面積割合を制約条件として入れることで、パトロール業務の運用状況を予測結果に反映できるようにする。

本論文は既発表論文 [5] に比較実験を追加したものである。まず 2 章で既存研究を紹介し、3 章で提案手法を説明する。4 章で提案手法を NIJ の公開データに適用し、機械学習によって得られた罪種ごとの最適バンド幅の傾向を、犯罪学における既存の知見を参考にして考察する。さらに既存の犯罪分析手法であるグリッドマッピングを改良した手法による実験結果と比較して、提案手法の有効性を示す。

2. 犯罪予測の既存研究

2.1 警察関係者や犯罪学専門家による分類と研究

犯罪予測に関する研究は、主に警察関係者や犯罪学専門

家により行われてきた。この理由として、まとまった件数の詳細情報を関係者以外が扱うことはなかったことがあげられる。米国警察の実務者向け報告書 [6] は、警察活動における予測技術を以下のように分類し、その 1 つとして本論文が対象にする犯罪予測を定義している。

- 犯罪予測：犯罪リスクが増加する場所と時間を予測
- 犯罪者予測：将来犯罪を行うリスクのある個人やグループを識別
- 犯罪被疑者の本人属性予測：特定の犯罪の被疑者プロフィールを高精度に作成
- 犯罪被害者予測：犯罪被害者になるリスクのある個人やグループを識別

犯罪予測手法の分類について、調査研究 [7] は、文献 [6] を含めた論文レビューに基づき、1) 時空間クラスタの検出、2) 犯罪の時空間的相互作用を考慮した犯罪発生の強度推定、3) 環境要因からの犯罪発生リスクの予測、4) 回帰分析による犯罪発生件数または確率の予測、としている。

分類 1) は犯罪学の知見で適切な統計的手法を選択して多発クラスタを計算する手法である。分類 2) は近接反復被害という犯罪学における知見を用いて予測を行う手法である。これは、ある場所で犯罪が起こると、その近隣において、短い期間のうちに再び犯罪被害が起こるという知見である [8]. まず、同一の人や場所に対して犯罪被害が繰り返し起こる現象が反復被害として知られた [9]. その後、住宅侵入盗の発生履歴を用いた研究 [10] では、被害発生後 2 週間程度の間、被害者宅から 200m 程度の範囲内で、住宅侵入盗発生リスクが高かった。この現象を近接反復被害と呼び、犯罪学の理論に基づいて考察された [8], [9]. 欧米警察における運用事例がある製品 (ProMap [11], PredPol [12]) の犯罪予測手法は、この知見をモデルに取り入れている。

分類 3) は犯罪発生リスクに影響する環境要因を犯罪学の知見に基づいて選定し、これらを説明変数として回帰分析等でモデル式を作成する手法である。この代表は米国警察の運用事例がある RTM (Risk Terrain Modeling : リスク起伏分析) [13] である。地理的・時間的な環境要因については、社会学や土木・建築学等、様々な学問領域における検討がある。警察庁による全国規模の研究報告 [14] や、居住者属性を用いて地域をセグメンテーションしたマーケティング用途向け市販データを用いる研究 [15] があり、一般的には既存研究で得た知見を参考に、環境要因を選ぶことがされる [16], [17], [18]. 分類 4) は環境要因データと犯罪発生履歴データを用いて回帰分析を基に犯罪発生確率を予測する手法である。

以上から、警察関係者や犯罪学専門家による犯罪予測研究における分析モデル作成と環境要因データ選択は、犯罪に関する知見に基づいてされてきたといえる。

2.2 異分野の研究者による分類-犯罪発生場所の予測

近年、犯罪予測をビッグデータ解析の問題として扱う研究が出ている。NIJ 犯罪予測コンテスト [2] は5年3カ月分の犯罪発生履歴データを公開して、犯罪学専門家以外の研究者にも参加を促した。文献 [19] はビッグデータ解析の視点で、犯罪予測で実現する内容に着目した既存研究の分類を、2-1) 犯罪発生時間中心、2-2) 犯罪発生場所中心、2-3) 犯罪者個人やコミュニティ中心、としている。本論文は近い将来における犯罪発生リスクが高い場所の情報提供を目指すため、2-2) の立場をとり、以降は2-2) に沿って既存研究を説明する。

犯罪発生場所の分析は、伝統的に、犯罪発生履歴からホットスポットと呼ばれる場所を求めることでされてきた [6]。ホットスポットとは、分析対象エリアのなかで過去の犯罪発生が相対的に集中するエリアである。犯罪はすでに起こった場所で起こりやすいと仮定すると、ホットスポットにおける犯罪発生リスクは高いと考えられ、この分析結果は将来の犯罪発生場所の予測と見なせる。

ホットスポット分析の具体的手法として、文献 [20] はグリッドマッピングをあげている。これは分析対象エリアにグリッドを生成してグリッドセルごとに犯罪発生件数を集計し、集計値が閾値以上のセルをホットスポットと見なす手法である。場所の解像度はセルの大きさで決まり、分析結果はセルの大きさや閾値設定で大きく異なる。文献 [20] は、セルの大きさ設定のアドバイスとして実験して経験を積むことと述べ、さらに、分析対象エリアを囲む矩形の長辺を一定値で割った値を初期値にする、というヒューリスティックな方法を提案している。

ホットスポット分析はまた、犯罪発生件数の空間分布をカーネル密度推定 [4] で求めることでも実現できる。カーネル密度推定ではカーネル関数のバンド幅をパラメータとして与える必要があり、文献 [20] はこれを経験的に定めている。カーネル関数は4次関数を選択している。これはNIJの資金で開発された犯罪データ分析ツール CrimeStat [21] のカーネル密度推定機能が提供するカーネル関数の1つで、犯罪分析の分野でよく利用される。時間方向に拡張し、時空間カーネル密度推定を用いた、犯罪発生の時空間パターンに関する研究もある [22]。

文献 [20] は、さらに、ホットスポット分析で求めたエリアが犯罪発生場所をどの程度予測できているかを、分析手法間で比較するための評価指標として、PAI (Prediction Accuracy Index) を提案している。これは、予測対象エリアに対する予測セルの面積割合と、発生件数に対する予測できた件数割合の比で、NIJ 犯罪予測コンテスト [2] で用いられる等、犯罪予測精度の一般的指標となっている。

2.3 既存研究の課題

犯罪予測研究において、データ選択やパラメータ設定は、

経験・知見に基づいてされてきた。時空間カーネル密度推定を用いたアプローチ [22] では、カーネル関数のバンド幅を、既存研究を参考にしたモデル式に別の研究によるパラメータを適用して求めている。しかし、このような経験・知見に基づくバンド幅設定は、従来の専門家以外にとっては難しい。一般的に、カーネル密度推定のバンド幅は尤度を評価指標にした交差検証法で最適化する [23], [24]。しかし犯罪予測への適用は十分に進んでいない。

一方、ビッグデータ解析のアプローチでは、予測する単位面積を、文献 [10] で得た近接反復被害の範囲である200m程度と比較し大きく設定するものが多い。犯罪発生データと人口等のデモグラフィック属性に、店舗や施設の位置等のPOI (Point of Interest) データとタクシーのトリップデータ [19]、行政サービス問合せ [25] 等の環境要因を組み合わせ、回帰分析やディープラーニングで、警察活動エリア [19], [25] や同一郵便番号エリア [26] ごとに、発生を予測する。しかし、犯罪発生の未然防止に向けてパトロール業務を支援する目的の予測情報としては、予測対象エリアの解像度が粗い。これらの研究は結果的に予測単位面積あたりのデータ件数を増やしており、解像度を上げた分析への適用は、単純には難しい。

また、予測精度の評価指標であるPAIについては、パトロール実施に割り当てるリソース量をなんらかの方法で制約したうえで予測精度評価を行うことが望ましい。しかし、従来このような検討はされていない。

そこで本論文は、経験・知見に基づく部分を極力なくしてデータ解析のみに基づくアプローチの有効性を示すために、カーネル密度推定を用いて犯罪発生場所を予測し、その最適バンド幅を、犯罪発生履歴データの機械学習により推定する手法を提案する。そして、バンド幅推定の評価関数に、パトロールリソース量の制約条件を導入する。

3. 提案手法 (最適化 KDE 法)

罪種ごとの犯罪発生件数の密度分布を、時間成分と空間成分の積からなる時空間カーネル密度推定でモデル化する。一般的にカーネル密度推定を適用する際に、データ定義領域をセルに分割する必要はないが、既存研究の手法と予測結果を合わせるために、データ定義領域内に設定した予測対象エリアを均一の形状のセルに分割して、セル単位で予測場所を出力する。パトロールリソースの制約条件を導入するために、パトロール実施可能なセルの、予測対象エリアに対する面積割合をセルカバー率、その範囲内で選択した予測セルにおける予測的中率をパトロールカバー率と定義する。そしてセルカバー率の制約下で、パトロールカバー率をカーネル関数のバンド幅に関して最大化する。

提案手法のブロック図を図1に示す。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

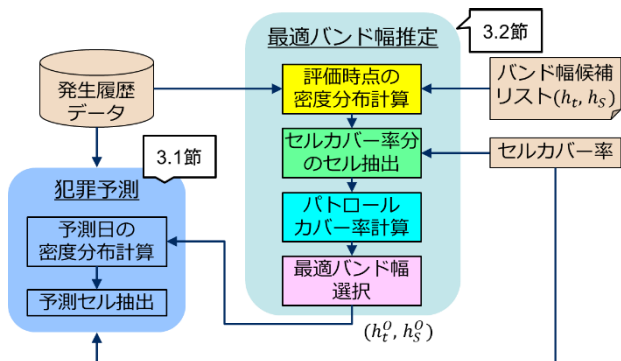


図 1 最適化 KDE 法のブロック図
Fig. 1 Block diagram of optimized KDE.

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 節). 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 節).

3.1 時空間カーネル密度推定による犯罪予測モデル

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.

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

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.

カーネル関数は, 犯罪データ分析ツール CrimeStat [21] 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 は文献 [20] 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 は式 (2) で表すことができる.

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

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

K_t, K_S の分布は図 2 に示すように, 時間成分はバンド幅 h_t の一定値, 空間成分はバンド幅 h_S 内の緩やかな減衰となる.

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

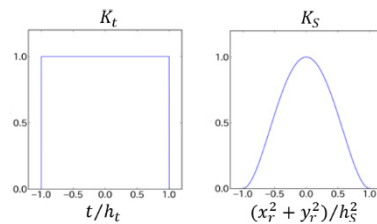


図 2 カーネル関数
Fig. 2 Kernel density functions.

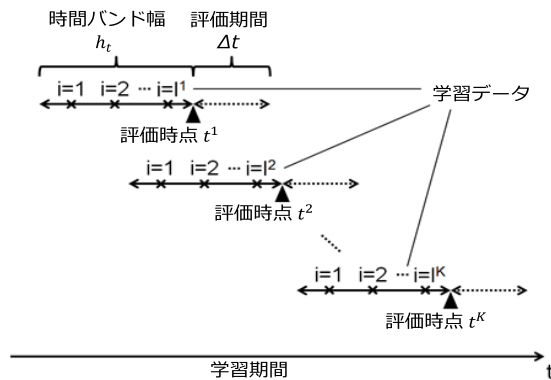


図 3 学習データの設定方法
Fig. 3 Arrangement of training datasets.

カバー率未達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.

$$\begin{aligned} &\text{セルカバー率 } \beta \\ &\equiv (\text{パトロール可能なセル数}) \\ &\quad / (\text{予測対象エリア内全セル数}) \quad (0 \leq \beta \leq 1) \quad (3) \end{aligned}$$

パトロール可能なセル数は, パトロール実行に割り当てるリソース量, すなわち, 警察官人員数や車両数に基づいて設定する.

3.2 機械学習による最適バンド幅推定

カーネル密度推定を用いる既存研究では, カーネル関数のバンド幅は経験と知見に基づいて設定した [20], [22]. これに対し提案手法は, バンド幅の候補リストを与え, 犯罪発生履歴データに含まれる年月日内に, 学習期間を設定して機械学習を行い, 候補リスト中から最適な組合せを選ぶ. ある学習期間における学習データ設定方法を図 3 に示す.

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. 評価時点は, 学習に用いるためには数多く設定することが望ましい. 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_t 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.

バンド幅候補の評価は、 $G^{k(\beta)}$ と t^k における犯罪発生セルの一致度合いで行う。まず、 t^k における発生と見なす期間として評価期間 Δt を設定し、 t^k の評価期間 Δt 内に犯罪が発生したセルのうち、 t^k における予測セル集合 $G^{k(\beta)}$ に含まれるセル数とそのセルにおける発生件数を集計する。そして学習期間内の全評価時点にわたる集計結果を、以下で定義するパトロールカバー率を用いて評価する。

パトロールカバー率

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

これは、セルカバー率 β 分の面積を予測した場合の、各評価時点の評価期間における予測対象エリア内発生件数に対する、予測セルにおける発生件数割合、すなわち予測的中率である。

バンド幅候補リスト中の組から、学習期間内のパトロールカバー率を最高にする組を最適バンド幅 (h_t^O, h_S^O) として選び、3.1 節の密度推定値の計算式 (1) に用いる。

3.3 評価関数

最適化 KDE 法のアルゴリズムを表 1 に整理する。セルカバー率とパトロールカバー率の関係は、模式的に図 4 に示すように、原点を通り、セルカバー率の増加とともに予測的中した発生件数が増えることでパトロールカバー率が増加するラインとなり、すべてのセルを予測するとパトロールカバー率は 100% になる。このラインは式 (1) で計算する密度推定値で変化し、カーネル関数のバンド幅組ごとに異なるラインとなる。さらに、一般的にあるセルカバー率に対して最高のパトロールカバー率となるバンド幅組は異なる。図 4 は、バンド幅組 A, B に対して異なるラインが得られ、パトロールカバー率の高いバンド幅組は、セルカバー率 10% 以下まではバンド幅組 B、それ以上ではバンド幅組 A に変化する様子を示す。

一方、既存研究では、犯罪予測精度の評価指標として、予測するセルの面積割合と予測できた件数割合の比である PAI を用いる [2], [17], [18], [20]。すなわち

$$n: \text{予測できた発生件数} \quad N: \text{予測期間内全発生件数}$$

$$a: \text{予測したセル数} \quad A: \text{予測対象エリア全セル数}$$

として PAI は次式で計算する。

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

PAI は犯罪予測の効率を示す指標であり、予測セルをす

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。

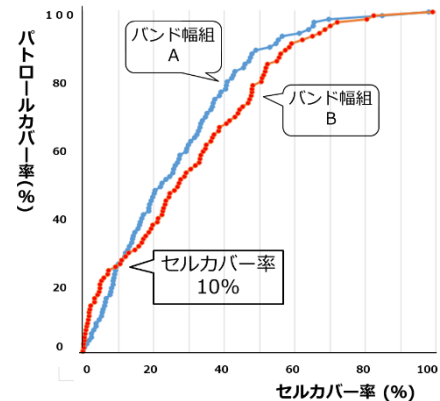


図 4 セルカバー率に対するパトロールカバー率 (イメージ)
Fig. 4 Image plot of patrol coverage versus cell coverage.

べてパトロールする場合は、セルカバー率とパトロールカバー率を用いて以下と書くことができる。

$$a/A = \text{セルカバー率}$$

$$PAI = (\text{パトロールカバー率})/(\text{セルカバー率}) \quad (6)$$

PAI を評価指標として最適化する場合、結果として得られる PAI に対応するセルカバー率は、必ずしもパトロールリソースの制約条件に合致しない。一方、提案手法は実際にパトロール実行に割り当てるリソース量を、予測面積の制約条件として学習手法に取り入れる。警察が行うパトロールは一般的に一地点にとどまるものではないため、セルカバー率の条件下でパトロールカバー率分の犯罪に遭遇するわけではない。それでも提案手法は、パトロール業務運用時の意思決定支援を考慮に入れた手法といえる。

4. 実験

最適化 KDE 法を NIJ 犯罪予測コンテスト [2] 公開データに適用して、自動車盗、路上犯罪、侵入盗の罪種ごとに予測を行う。次に、文献 [20] がホットスポット分析手法としてあげたグリッドマッピングを改良して同様の分析を行い、結果を比較・考察する。グリッドマッピングは主なパラメータはセルの大きさのみのため、実装の違いによる影響が少ない手法と考え、比較手法として選択した。

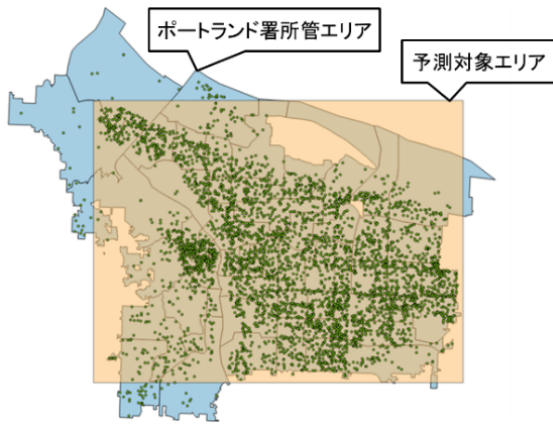


図 5 ポートランド署所管エリアと侵入盗発生地点
 Fig. 5 Portland police district and geographic plots of burglary.

表 2 罪種別犯罪発生データ数

Table 2 Number of crime events by type.

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.

4.1 条件設定

公開データ [27] は米ポートランド警察の 5 年 3 カ月分の犯罪発生履歴で、罪種、発生日月日、発生場所の緯度経度が付与されている。図 5 は侵入盗のデータ全期間分とポートランド署所管エリア [27] の、地理情報システム QGIS [28] による表示である。予測対象エリアとして発生が密集する地域を囲む矩形(東西約 23km, 南北約 17km)を設定し、NIJ コンテストの条件を参考にしてセルに分割する。その条件は、セルの形状は均一で最小面積は 62,500 平方フィートつまり 5,806 m² である。一辺の長さ 75m の正方形セルを設定すると、セル面積は 5,625 m² となり最小面積に近い。本論文はこのセルを用い、75m をセル幅と呼ぶ。表 2 に全期間分罪種別発生データ数を、予測対象エリア内とセルあたりについて示す。

セルカバー率に関連する NIJ コンテスト条件に、予測面積の制限がある。これはポートランド署所管エリア面積 147.71 平方マイルに対し、0.25 平方マイルから 0.75 平方マイルの間で、面積割合 0.17% から 0.51% の間に相当する。一方、文献 [18] は、警察の防犯パトロール向けにさいたま市内のあるエリアの犯罪多発地点を特定したところ、予測対象エリアに対する多発地点の面積割合は、4% から 5% になったとしている。最適化 KDE 法のセルカバー率はパトロールリソースを制約する値のため、本論文は両方を参考にして、セルカバー率 1% と 5% を設定する。

最適化 KDE 法に与える時空間カーネルのバンド幅候補を表 3 に示す。時間バンド幅は、1 週間や月単位の周期性に対応する 7 日、28 日、3 カ月程度の季節変動に対応する

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

91 日、および半年、1, 2, 3, 4 年に対応する日数とする。空間バンド幅は、セル幅 75m から予測対象エリアの一辺の長さの 1/10 程度である 2,000m までとする。本論文ではセルカバー率の条件下で、バンド幅組に対するパトロールカバー率の変化傾向を考察するために、時間と空間バンド幅の全組合せを探索する。

最適化 KDE 法の学習期間は、季節変動がある場合を考慮して 1 ケース 3 カ月間とし、公開データ 5 年 3 カ月内で最長時間バンド幅 1,460 日 (4 年) が設定可能な範囲に、5 ケースを設定する。表 4 に各学習期間と、期間内の発生データ数を示す。

また、なるべく多くの評価時点で学習データを収集するために、評価時点 t^k の間隔と評価期間 Δt は 1 日とする。

4.2 最適化 KDE 法の実験

セルカバー率 1% における罪種ごとの実験結果として、バンド幅組ごとのパトロールカバー率を、各学習期間内で正規化した値を示す(図 6)。すなわち図中の赤い部分のバンド幅組では、他の組に対してパトロールカバー率が相対的に高く、各学習期間における最適バンド幅はこれらの組に含まれる。この結果から、米国立司法省 (United States Department of Justice) の資金で設立された NPO 法人が作成した、米警察向け問題別ガイド [29] を参考に、罪種ごとの最適バンド幅の傾向を定性的に考察する。

4.2.1 自動車盗

学習期間ごとの最適バンド幅は、図 6(a) から、時間バンド幅 2~4 年、空間バンド幅 75~100m の範囲に含まれ、全学習期間でほぼ変わらない。これは過去 2~4 年の発生場所から 75~100m という近距離で多く発生する傾向を意味する。

文献 [29] は、自動車盗は駐車場での発生が多く、発生が多い駐車場には照明が暗い等の条件があるとしている。一

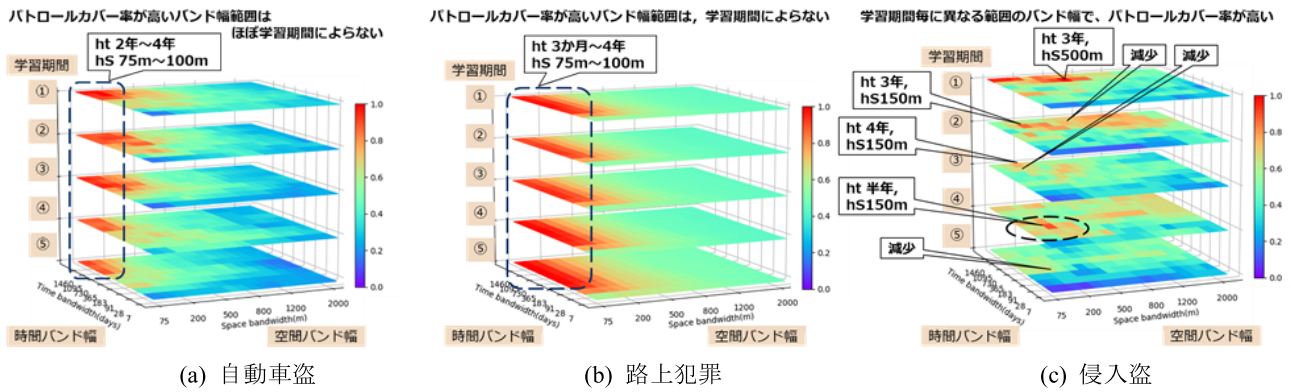


図 6 バンド幅候補に対する，学習期間ごとのパトロールカバー率（セルカバー率 1%）

Fig. 6 Patrol coverage versus candidate bandwidths for each training case (Cell coverage 1%).

一般的に駐車場の場所は数年間変わらないことを考え合わせると，実験結果は，過去数年間の発生場所と同じ駐車場で発生が続く状況に対応すると解釈することができる。

4.2.2 路上犯罪

図 6(b) から最適バンド幅は，時間バンド幅 3 か月～4 年，空間バンド幅 75～100 m の範囲に含まれ，過去の発生場所の近くで発生する傾向が安定している．時間バンド幅 3 か月以上であれば，パトロールカバー率の差は 5% 以内である。

路上犯罪のデータは複数の罪種を含み，暴行，強盗等の対人犯罪と騒乱等の秩序違反に大別される．文献 [29] は，暴行と飲酒店との関係，都市部におけるひったくり等の路上強盗は公共交通機関の駅のそばで多発，若者の秩序違反の種類と多発場所の対応付け，等の知見をあげている．これらをもとに，実験結果は，各々の罪種がそれぞれ特定の要因を持つ狭いエリアで頻発する状況に対応すると解釈することができる。

4.2.3 侵入盗

図 6(c) から，最適バンド幅や高パトロールカバー率のバンド幅範囲は，学習期間に対して変化し不安定である．学習期間ごとの最適バンド幅には，時間バンド幅 3 年以上と半年，空間バンド幅 150 m と 500 m の，各々 2 パターンがみられる．特に学習期間④では，最適バンド幅は，時間バンド幅半年，空間バンド幅 150 m，高パトロールカバー率の範囲は時間バンド幅 28 日～1 年，空間バンド幅 150 m であり，それ以前の学習期間よりも時間バンド幅が短い。

侵入盗には犯罪の近接反復被害の知見があり (2.1 節)，文献 [29] も被害の繰り返しを指摘している．学習期間④における高パトロールカバー率の範囲は，文献 [10] が示す近接反復被害の範囲，すなわち，発生後 2 週間程度の間，被害者宅から 200 m 程度という値にある程度近い値のため，近接反復被害の影響をとらえた結果と解釈することができる。

この解釈によると，最適バンド幅や高パトロールカバー率のバンド幅範囲の学習期間に対する不安定性は，近接反

復被害の有無や近接度合の変化をあらわすとも考えられる．提案手法は各学習期間のデータからそのつど最適バンド幅を求めるため，事前に設定した固定値を用いる既存研究よりも，犯罪発生状況が変化するケースの分析に適している．さらに時間経過とともに逐次得られる犯罪発生履歴データを継続的に再学習することで，状況変化への追従は一層容易になる。

学習期間④はまた，空間バンド幅 1 km 以上でも比較的高パトロールカバー率が高く，近接反復被害の影響と広いバンド幅の影響が混在する状況と考えられる．この状況は 1 つの時空間カーネルモデル (式 (1)) では表現できない．改善策として，複数の時空間カーネルモデルへの拡張が考えられる．文献 [8] は，近接反復被害の影響の大きさと，時間的空間的広がり範囲を，時空間分析分野の統計値を用いて計算する手法を提案している．この計算結果を 1 つのカーネル関数として最適化 KDE 法に直接的に組み込むことも考えられる。

4.2.4 罪種ごとの最適バンド幅組

セルカバー率 1% と 5% における，罪種ごとの最適バンド幅組を，学習期間ごとに表 5 に示す．最適空間バンド幅は自動車盗，路上犯罪で狭く，侵入盗では広い傾向がある．また，セルカバー率 5% における最適時間バンド幅は，セルカバー率 1% よりも長い傾向がある．これは 3.3 節で説明した，セルカバー率によってパトロールカバー率が高いバンド幅組が変化することの実例である。

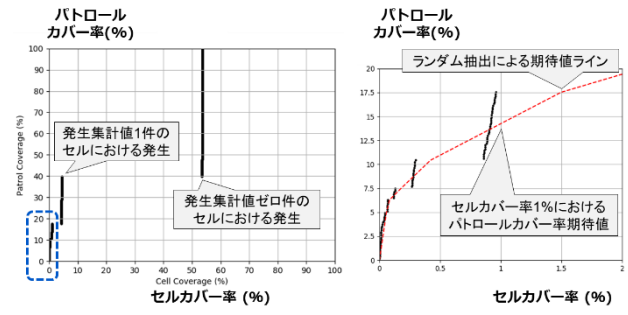
4.3 グリッドマッピングの実験

グリッドマッピングは，2.2 節で説明した既存のホットスポット分析手法で，図 7 右側に示す構成である．学習期間内に予測日 t_p を毎日設定し，予測日前日までのすべての発生をセルごとに集計して，件数が多い順に全セルを順位付けする．しかしこの手法は同一順位のセルが複数得られやすい．そこで本論文では最適化 KDE 法の実験と同様にセルカバー率を設定してパトロールカバー率を計算するため

表 5 最適バンド幅組 (時間バンド幅 (日), 空間バンド幅 (m))

Table 5 Optimal combinations of bandwidths (Temporal (days), Spatial (m)).

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)



(a) 全セルカバー率範囲 (b) 原点付近拡大図

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)

Fig. 8 Patrol coverage of grid mapping (Motor vehicle theft, training case ①).

の平均値と標準偏差を示す. 提案手法である最適化 KDE 法のパトロールカバー率は, 各学習期間の最適バンド幅を用いた値を示す. 比較手法である改良型グリッドマッピングは, 自動車盗と侵入盗で, セルカバー率付近の同一順位セル中から予測セルをランダム抽出した値である. パトロールカバー率平均値は, 自動車盗と侵入盗の合計 4 パターンで提案手法が, 路上犯罪 2 パターンで比較手法が高い.

表 6 にはさらに「ランダム」として, 予測対象エリア内の全 68,400 セルからセルカバー率分のセルを, 予測日ごとにランダムに選択し, 学習期間ごとのパトロールカバー率を計算した結果を示す. この値はほぼセルカバー率に等しく罪種によらない. また, 学習期間ごとの犯罪発生データ数の, 平均値と標準偏差を示す.

提案手法と比較手法のパトロールカバー率の差を検討するため, 2 手法間の, データに対応関係がある場合のノンパラメトリックな検定手法である, Wilcoxon の符号付順位検定 [30], [31] を適用する. 結果は表 6 に示すように, セルカバー率 1% の自動車盗以外の 5 パターンで, 15 ケースのパトロールカバー率の, 2 手法の差は有意であった. すなわち, 侵入盗 2 ($p < 0.0005$), 自動車盗 1 ($p < 0.0125$) の合計 3 パターンで, 提案手法のパトロールカバー率は比較手法に対して高く, 路上犯罪 2 パターン ($p < 0.0125$, $p < 0.0005$) で比較手法のパトロールカバー率が高かった.

侵入盗については, 4.2 節で, 提案手法の最適バンド幅は学習期間に対して不安定であり, 原因として学習期間ごとの近接反復被害の度合の変化が考えられるとした. 学習期間ごとに最適なバンド幅を選ぶことで, より高いパトロールカバー率を得る点は, 提案手法が優位である. 学習期間 15 ケースのパトロールカバー率の散布図 (図 9(c)) によると, 提案手法は比較手法よりも高いパトロールカバー率を維持し, 値の変動幅は比較手法よりも小さい. このことから提案手法の有効性が確認できる.

自動車盗についても, 最適空間バンド幅の値が大きく変化する期間は提案手法が有効である. 表 6 に示す検定結果

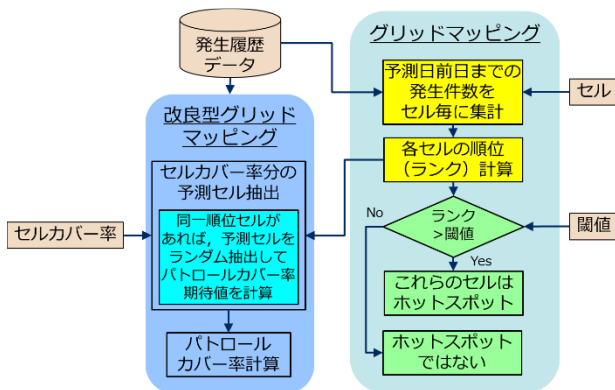


図 7 グリッドマッピングのブロック図

Fig. 7 Block diagram of grid mapping.

に, 図 7 左側に示す改良型グリッドマッピングを用いる.

図 8 (a) は, 集計結果が同一順位のセルに対して平均値の順位を割り当てた場合の, セルカバー率とパトロールカバー率の関係で, 同一順位セルに起因する離散的形状がみられる. 改良型グリッドマッピングは, 任意のセルカバー率 β でパトロールカバー率を計算できるように, 同一順位セルがある場合はそれらのなかからランダムにセルを抽出し, 合計 β 個のセルを予測セル集合 $G^{k(\beta)}$ とする. 図 8 (b) は図 8 (a) 左下破線内の拡大図である. 黒の実線で示す離散的な結果に対し, 同一順位セルから $G^{k(\beta)}$ をランダム抽出して得られるパトロールカバー率の期待値を赤の破線で示す.

実験データを多く得るために, 表 4 の学習期間 5 ケースを各々 1 カ月ごとに 3 分割して 15 ケースとし, 他の条件は同様の設定で罪種ごとに分析を行う.

4.4 最適化 KDE 法と改良型グリッドマッピングの比較

表 6 に, 3 罪種およびセルカバー率 1% と 5% の合計 6 パターンについて, 学習期間 15 ケースのパトロールカバー率

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

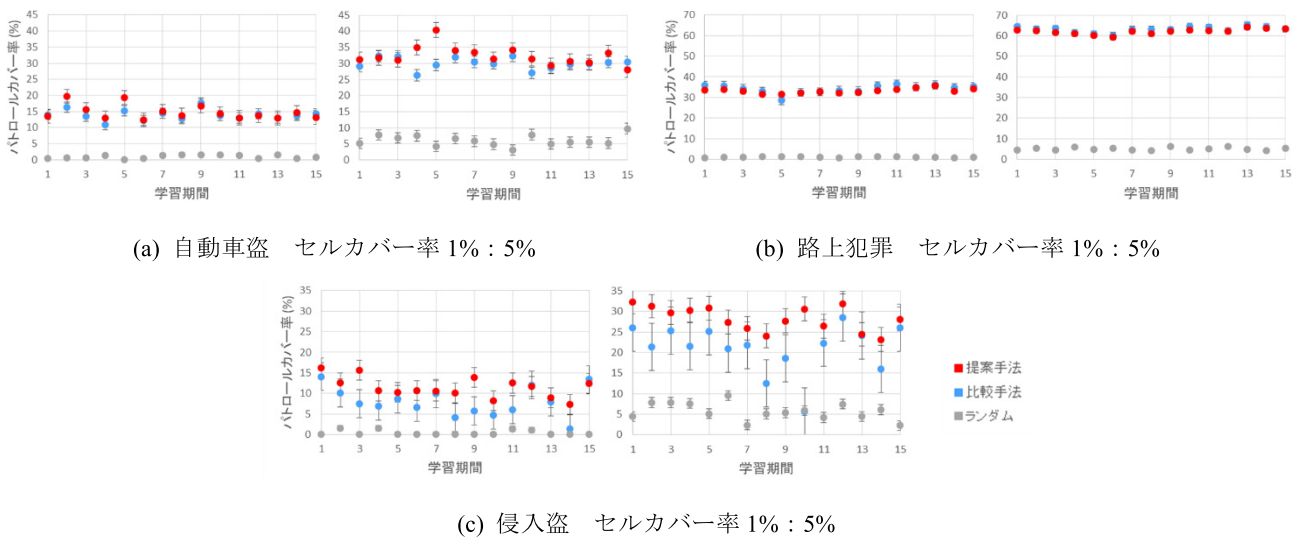


図 9 学習期間ごとのパトロールカバー率 (注) エラーバーは標準偏差を示す.

Fig. 9 Patrol coverage for each training case.

で有意差があったセルカバー率 5% の散布図 (図 9(a)) によると, 学習期間 4, 5 における提案手法のパトロールカバー率は, 比較手法よりも高い. これは表 5 の学習期間②に該当し, この期間の最適空間バンド幅は他の期間の 75 m から 150 m に変化する.

一方, 路上犯罪の最適空間バンド幅はセル幅 75 m に等しく学習期間に対して安定である (表 5). 4.2 節ではこれを特定の要因を持つ狭いエリアにおける頻発状況に対応すると解釈した. 表 6 に示す検定結果から, 路上犯罪では比較手法のパトロールカバー率が高く有意差がある. これは, 提案手法が発生事象によるリスク値を隣接セルに波及させるために, 相対的に発生セルでのリスク値が低く抑えられるのに対し, 比較手法はピンポイントの発生セル内にリスク値を集中する点が優位なためと考えられる.

このような発生状況を提案手法に包含するには, 提案手法の式 (1) における空間カーネル関数をディラックの δ 関数 (式 (7)) に置き換え, 発生場所をピンポイントで表現する. これにより提案手法は比較手法と同等と見なせ, 比較

手法と同様の結果が得られると考えられる.

$$K_S \propto \delta(x - x_i)\delta(y - y_i) \tag{7}$$

4.5 パトロールカバー率に関する考察

まず, 表 6 に「ランダム」として示した, 予測セルをランダム選択した場合は, セルカバー率程度のパトロールカバー率が得られる. 比較手法のパトロールカバー率はランダム選択の 3.7 倍以上で, さらに, 4.4 節で説明したように, セル幅が最適空間バンド幅程度の場合にパトロールカバー率が高い点で, 発生状況を反映した値が得られる. 比較手法である改良型グリッドマッピングは, 犯罪発生状況によっては, セルカバー率を制約したうえで犯罪発生場所を予測する有用な手法といえる.

次に, 本論文における犯罪発生件数とセルカバー率の関係から, 達成可能なパトロールカバー率の設定につき検討する. 予測対象エリア内の発生件数の, 学習期間ごとの平均値は, 表 6 に示すように路上犯罪が最多で, 1 日あたり

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。

90 件程度である。セルカバー率 1% の予測セル数 684 はこれより大きい。発生セルをすべて予測できたとすると毎日全発生をカバー可能である。パトロールカバー率は 100% となり、評価指標 PAI を式 (6) で計算すると 100 となるが、NIJ コンテストの結果、どの罪種でもこれに準ずる PAI は得られていない [2]。コンテスト結果には、PAI を評価指標として、予測面積以外にもセル形状、サイズや配置等を最適化した手法もある [32]。以上から、達成可能なパトロールカバー率の設定は現時点では現実的ではなく、これ自体が今後の課題といえる。

既存研究によると犯罪発生リスクは環境要因の影響を受ける (2.1 節)。すると国や地域、また罪種により犯罪予測精度の評価指標値は大きく異なると考えられる。予測精度の目標設定には、何らかの方法で環境要因を反映する必要があると考える。

試みに、本論文と同様に単純なセルを設定し狭い面積単位の予測を目的とした研究で得られた PAI との比較を示す。まず、本論文の予測結果である表 6 から式 (6) で計算した PAI を表 7 に示す。各罪種の最高 PAI 値は、三罪種とも、セルカバー率 1% における値となった。これは PAI がセルカバー率に依存する評価指標であることの一例である。文献 [17] は福岡県の一部地域における犯罪発生履歴データと環境要因データを用いて、2.1 節分類 3) の手法で 6 カ月間予測を行った。罪種は車上狙いで、得られた PAI の 6 カ月平均値は最高 2.11 であった。本論文で分析した罪種とは異なるため単純比較はできないが、本論文では相対的にこれより高い PAI を得ている。

なお本論文では、提案手法が学習期間ごとにバンド幅を変化させることの有効性を検討するために、最適バンド幅を推定した学習期間と同一期間における結果を示した。

5. おわりに

本論文は、犯罪発生履歴データのみを用いて、時空間カーネル密度推定で求めた発生件数密度分布の高密度エリアを近い将来における犯罪発生場所として予測する、最適化 KDE 法を提案した。時空間カーネル関数の最適バンド幅は、パトロールリソース量の制約条件としてパトロール実施可能な面積割合を与え、その面積分を予測した場合の予測的中率であるパトロールカバー率を最大化するよう

に、発生履歴データを機械学習して推定した。

NIJ の犯罪発生履歴公開データを用いた実験で、パトロール実施可能面積割合を予測対象エリアの 1%、5% として自動車盗、路上犯罪、侵入盗の最適バンド幅を求めた結果、これらは米警察向け問題別ガイドに記載する各罪種の知見と対応づけて解釈することができた。また、既存の犯罪発生集中エリア分析手法であるグリッドマッピングを改良し、パトロール実施可能面積割合を設定可能にした比較実験で、特に、最適バンド幅が時間的に変動する罪種である侵入盗における提案手法の有効性を確認した。

最適化 KDE 法は発生件数密度分布を推定するため、2.1 節で示した既存の犯罪予測手法分類における、近接反復被害モデル分類 2) に該当する。この手法は過去に発生件数密度分布が存在しなかったエリアを予測することはできない。一方、地理的・時間的な環境要因データを用いる手法分類 3) には、潜在的な発生リスクを評価できる可能性がある。また、犯罪が発生する時間・空間ではなく、被疑者や被害者を予測する既存研究にも同様の可能性がある。今後は、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や、継続的な学習方式の適用を検討したい。

参考文献

- [1] 国家公安委員会・警察庁編：平成 30 年版警察白書第 1 部 特集近年における犯罪情勢の推進と今後の展望, p.42 (2018), 入手先 (https://www.npa.go.jp/hakusyo/h30/pdf/03_tokushu.pdf) (参照 2019-07-02).
- [2] National Institute of Justice: Real-Time Crime Forecasting Challenge, available from (<https://nij.ojp.gov/funding/real-time-crime-forecasting-challenge>) (accessed 2019-12-16).
- [3] 渡辺昭一 (編著)：捜査心理学, 島田貴仁：犯罪情勢分析とクライムマッピング, pp.229-245, 北大路書房 (2004).
- [4] Bishop, C.M.: Pattern Recognition and Machine Learning, Springer-Verlag New York (2006). 元田 浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田 昇 (監訳)：パターン認識と機械学習上, シュプリンガー・ジャパン社 (2007).
- [5] 中川淳子, 小西勇介, 宮野博義：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, FIT2018 (第 17 回情報科学技術フォーラム) 講演論文集第 4 分冊, pp.97-100 (2018).
- [6] Perry, W.L., McInnis, B., Price, C.C., Smith, S.C. and Hollywood, J.S.: Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations, RAND Corporation (2013), available from (https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND-RR233.pdf) (accessed 2019-07-02).
- [7] 大山智也, 両宮 護, 島田貴仁, 中谷友樹：地理的犯罪予測研究の潮流, GIS-理論と応用, Vol.25, No.1, pp.33-43 (2017).
- [8] 菊池城治, 両宮 護, 島田貴仁, 齋藤知範, 原田 豊：近接反復被害の罪種間比較—時空間 K 関数の応用, GIS-理論と応用, Vol.18, No.2, pp.21-30 (2010).
- [9] Wortley, R. and Mazerolle, L.: Environmental Criminology and Crime Analysis (2008). 島田貴仁, 渡辺昭一 (監訳)：環境犯罪学と犯罪分析, ファレル, G., ピーズ, K.:

反復被害, pp.120-139, (財) 社会安全研究財団 (2010).

[10] Johnson, S.D., Bernasco, W., Bowers, K.J., Elffers, H., Ratcliffe, J., Rengert, G. and Townsley, M.: Space-Time Patterns of Risk: A Cross National Assessment of Residential Burglary Victimization, *Journal of Quantitative Criminology*, Vol.23, No.3, pp.201-219 (2007).

[11] ProMap, available from <https://www.promap.co.uk/> (accessed 2019-07-02).

[12] PredPol, available from <https://www.predpol.com/> (accessed 2019-07-02).

[13] RISK TERRAIN MODELING, available from <http://www.riskterrainmodeling.com/> (accessed 2019-07-02).

[14] 警察庁生活安全局：犯罪情勢分析手法の高度化に向けた調査研究報告書, 警察庁生活安全局生活安全企画課 (2009).

[15] 上杉昌也, 樋野公宏, 矢野桂司：ジオデモグラフィクスによる社会地区類型を活用した窃盗犯の発生要因に関する小地域分析, *E-journal GEO*, Vol.13, No.1, pp.11-23 (オンライン), DOI: 10.4157/ejgeo.13.11 (2018).

[16] 両宮 護：潜在成長曲線を用いた地域レベルでの犯罪の時系列変化と地区環境との関連の分析, *都市計画論文集*, Vol.48, No.3, pp.351-356 (2013).

[17] 大山智也, 両宮 護：地理的犯罪予測と人の流れデータ—犯罪抑止要因としての「人の目」と Risk Terrain Modeling, *電子情報通信学会 HPB 第 18 回研究会*, pp.38-41 (2018), 入手先 (http://ieicehpb.info/index.php?plugin=attach&refer=Kenkyukai18&openfile=IEICE_HPBO18_13.pdf) (参照 2019-12-12).

[18] 野 貴泰, 糸井川栄一：犯罪多発地点の予測に基づく防犯パトロール経路に関する提案, *地域安全学会論文集*, No.31, pp.195-204 (2017).

[19] Wang, H., Kifer, D., Graif, C. and Li, Z.: Crime Rate Inference with Big Data, *Proc. KDD'16*, pp.635-644, ACM (2016).

[20] Chainey, S., Tompson, L. and Uhlig, S.: The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime, *Security Journal*, Vol.21, pp.4-28 (2008).

[21] Levine, N.: CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 4.02), Ned Levine & Associates, and the National Institute of Justice (2015), available from <http://www.nij.gov/topics/technology/maps/pages/crimestat.aspx> (accessed 2019-06-04).

[22] Nakaya, T. and Yano, K.: Visualizing Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using space-time Kernel Density Estimation and Scan Statistics, *Trans. GIS*, Vol.14, No.3, pp.223-239 (2010).

[23] 杉山 将：密度比に基づく機械学習の新たなアプローチ, *統計数理*, Vol.58, No.2, pp.41-155 (2010).

[24] Härdle, W., Müller, M., Sperlich, S. and Werwatz, A.: Nonparametric and semiparametric models, Springer, Berlin, Germany (2004).

[25] Huang, C., Zhang, J., Zheng, Y. and Chawla, N.V.: DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction, *Proc. CIKM'18*, pp.1423-1432, ACM (2018).

[26] Wang, B., Luo, X., Zhang, F., Yuan, B., Bertozzi, A.L. and Brantingham, P.J.: Graph-Based Deep Modeling and Real Time Forecasting of Sparse Spatio-Temporal Data, arXiv:1804.00684v1, available from <https://arxiv.org/pdf/1804.00684.pdf> (accessed 2019-12-06).

[27] National Institute of Justice: Data for Download, available from <https://nij.gov/funding/Pages/fy16-crime-forecasting-challenge-document.aspx#data> (accessed 2019-12-25).

[28] QGIS, available from <https://qgis.org/ja/site/> (accessed 2019-07-02).

[29] Center for Problem-Oriented Policing, available from <https://www.popcenter.org/> (accessed 2019-12-25).

[30] 高木英行：使える！統計検定・機械学習—I—2 群間の有意検定, *システム/制御/情報*, Vol.58, No.8, pp.346-352 (2014).

[31] Zaitontz, C.: Real Statics Using Excel, available from <http://www.real-statistics.com/statistics-tables/wilcoxon-signed-ranks-table/> (accessed 2019-12-12).

[32] Mohler, G. and Porter, M.D.: Rotational grid, PAI-maximizing crime forecasts, available from <https://nij.ojp.gov/library/publications/rotational-grid-pai-maximizing-crime-forecasts> (accessed 2019-12-06).

推薦文

本論文は、専門家の知見に基づく統計量の選択やパラメータの設定を必要としない犯罪予測手法を目指したものである。本論文で提案された手法は、予測する犯罪発生場所を、犯罪発生履歴データから時空間カーネル密度推定で求める発生件数の密度分布の高密度エリアであると、パトロール可能な面積割合の場所を予測した場合に、予測した場所での発生件数が最大になるように、カーネル関数の最適バンド幅を機械学習で推定するものである。本論文では、米国立司法省 (NIJ) の犯罪予測コンテストの公開データを用いて提案手法を吟味し、自動車盗、路上犯罪については既存の知見と合致する結果を出しているが、侵入盗に関しては提案手法のモデルは改善の余地があることを報告している。本論文では、先行研究の基本的な確認結果が明確に提示されており、先行研究をベースラインとした提案手法の効果確認も明示的に行われている。結果の報告もよく整理されており、読者にとって、非常に分かりやすい。データの詳細やコンテストでの使用手法等を含む先行手法の知見等を加えることで、より、読者に有益な情報提示が可能になるとの見地から、本論文を推薦する。

(FIT2018 第 17 回情報科学技術フォーラムプログラム

委員長 河内谷 清久仁)



中川 淳子 (正会員)

1983 年筑波大学第一学群自然学類物理学主専攻卒業, NEC 入社。数値シミュレーション, テキストマイニング, 映像解析に関する研究開発に従事し, 現在, 犯罪予測の研究に従事。



西村 祥治 (正会員)

2001年京都大学情報学研究科修了，2017年東京工業大学情報理工学研究科博士課程修了。2001年NECに入社し現在まで，HPC，並列分散処理，大規模DB，映像分析に関する研究に従事。博士（工学）。



宮野 博義

1997年東京大学院工学系研究科情報工学専攻修士課程修了，NEC入社。パターン認識，動画像解析に関する研究開発に従事し，現在，バイオメトリクス研究所所長代理。2017年理化学研究所理研AIP-NEC連携センター副

連携センター長，2019年奈良先端科学技術大学院大学准教授。