50m メッシュ・1 時間単位における犯罪リスク推定手法の提案

平元尚貴 [1]　梶田真実 [2]　村上大輔 [3]

Crime Risk Estimation Method at 50m Grid and Hourly Resolution

Naoki Hiramoto[1]　**Mami Kajita**[2]　**Daisuke Murakami**[3]

Abstract – This study proposes a crime risk estimation method at 50m grid resolution and hourly intervals to support crime prevention planning. While existing research enables prediction using spatial dependencies at 250m grid and monthly intervals, observations at 50m daily resolution are extremely sparse, making conventional methods difficult to apply. This research proposes (1) a daily risk estimation method with Poisson expected values in internal states that can be sequentially updated, and (2) estimation of mesh-specific 24-hour distributions by combining spatial eigenvectors and Fourier basis functions, integrating both approaches to achieve crime risk estimation at 50m grid and hourly intervals. The effectiveness of the proposed method was validated through empirical analysis using 2013-2020 data from Montevideo, Uruguay.

Keywords: 犯罪予測（Crime Prediction），空間統計（Spatial Statistics），逐次推定（Sequential Estimation），時空間解析（Spatiotemporal Analysis），高解像度予測（High-Resolution Prediction）

1　背景

過去 30 年にわたり地理的犯罪予測のための統計・機械学習手法が開発されてきた. 初期の研究では,Chainey et al.[1] によるカーネル密度推定や Caplan et al.[2] のリスク地形モデリングなど, 主に空間パターンの捕捉に焦点を当てた手法が提案されてきたが，それらは時間軸を明示的には考慮しておらず, 日々変化する犯罪パターンを柔軟に捉えることはできない.

一方で，時間軸を明示的に考慮した手法も考案されてきた. 例えば Mohler et al.[3] は時空間自己励起点過程モデルを開発し，50m メッシュ・日単位での犯罪予測を実施して，Flaxman et al.[4] は, シカゴを対象に約 1.6km メッシュ・週単位での長期の犯罪予測を実施した. 近年の機械学習・深層学習アプローチでは,Huang et al.[5] の階層的リカレントネットワークによる地域・日単位予測や,Lin et al.[6] による約 300–400m グリッド・月単位予測などがある. しかし，それらの既存手法には計算コストが大きいものも多く，例えば 50m メッシュ・日単位のような高解像度での犯罪予測やその実運用（日次更新等）には課題が残されている.

高解像度（空間・時間）での詳細な犯罪予測を実運用する際には次の 2 点が課題となる：(i) 高頻度で流入してくる犯罪データを用いて、犯罪リスクの空間分布を高速・安定的に逐次更新する方法の確立；(ii) 場所毎の用途差などによる空間異質性が強く、単純な空間連続性が仮定しづらい点への対処.

本研究は以上の 2 点に対処して，高い時間・空間解像度での犯罪予測を行うために, 犯罪の日別発生パターンと時間帯別発生パターンを分離する二段階分解アプローチを提案する。前者は逐次ベイズ更新で軽量に追跡し, 後者は長期平均として推定する. 後者には, 空間固有ベクトルとフーリエ基底を用いた加重リッジ推定を適用し, 疎なメッシュほど近傍・全体プロファイルへ収縮する部分プーリングを設計する. これにより,50m・1 時間単位という高解像度における犯罪リスクの逐次更新を高速・高精度に行うことを目指す. 著者の知る限り,50m・1 時間単位の推定を逐次更新で運用可能な形に落とし込む枠組みは, まだまだ限定的である.

本研究の主要な貢献は以下の通りである：(1) 疎な観測下での逐次的日別の犯罪リスク推定手法の開発,(2) 空間固有ベクトルと時間フーリエ基底を組み合わせた時間分布推定手法の提案,(3) 両手法を統合した 50m メッシュ・1 時間単位犯罪リスク推定システムの実現.

2　提案手法

2.1　日別犯罪件数のモデル化

犯罪件数 $n_{i,t}$（グリッド i, 日次 t）が犯罪リスク (ポアソン強度)$\lambda_{i,t}$ を期待値とするポアソン分布 $n_{i,t} \sim \text{Poisson}(\lambda_{i,t})$ に従うと仮定する.「ほとんどのグリッド・日次での犯罪件数がゼロ」という典型的なスパースデータ下での推定安定性を検証するため, 以下の 3 つの状態空間モデルを比較検討した.

2.1.1　ランダムウォークモデル（EMA）

内部状態 $\lambda_{i,t}$ がランダムウォーク過程に従うモデル：

$$\lambda_{i,t} = \lambda_{i,t-1} + \omega_t \tag{1}$$

$$n_{i,t} \sim \text{Poisson}(\lambda_{i,t}) \tag{2}$$

ここで $\omega_t \sim N(0, \sigma^2)$ はプロセスノイズである.

[1]総合研究大学院大学 統計科学コース,Singular Perturbations Email: hiramoto.naoki@ism.ac.jp
[2]東京大学 空間情報科学研究センター,Singular Perturbations
[3]統計数理研究所,Singular Perturbations

2.1.2 トレンド考慮モデル（T-MPF）

期待値の差分がランダムウォークに従う 2 次元状態空間モデル：

$$\lambda_{i,t} = \lambda_{i,t-1} + \mu_{i,t} + \omega_{1,t} \tag{3}$$
$$\mu_{i,t} = \mu_{i,t-1} + \omega_{2,t} \tag{4}$$
$$n_{i,t} \sim \text{Poisson}(\lambda_{i,t}) \tag{5}$$

ここで $\omega_{1,t}, \omega_{2,t}$ は独立なプロセスノイズである.

2.1.3 曜日効果モデル（W-MPF）

基本リスク $\lambda_{base,i,t}$ に曜日効果 $w_{i,t}$ を乗じるモデル：

$$\lambda_{base,i,t} = \lambda_{base,i,t-1} + \omega_{1,t} \tag{6}$$
$$w_{i,t} = w_{i,t-1} + \omega_{2,t} \tag{7}$$
$$\lambda_{i,t} = \lambda_{base,i,t} \times \begin{cases} w_{i,t} & \text{if weekend} \\ 1 & \text{otherwise} \end{cases} \tag{8}$$
$$n_{i,t} \sim \text{Poisson}(\lambda_{i,t}) \tag{9}$$

各モデルに対し, 融合粒子フィルタ (MPF) による推定を行った [7]. 基本モデルについては, 計算効率の観点から指数重み移動平均による近似解も併用した：

$$\hat{\lambda}_{i,t} = \frac{\sum_{d=1}^{t} n_{i,t-d+1} \cdot \rho^{d-1}}{\sum_{d=1}^{t} \rho^{d-1}}, \quad \rho = \frac{1}{1+r} \tag{10}$$

2.2 時間帯別犯罪件数のモデル化

50m メッシュ毎・日次では犯罪件数が極めて疎で, 空間相関パターンは観測されなかった. 一方, 各メッシュにおける犯罪発生の 24 時間分布の形状は (商業地・駅前・繁華街などの) 土地利用や人流のリズムに規定され空間相関パターンを持つことが期待できる. そこで本研究では, 空間相関パターンを捉えるための空間固有ベクトル [8] と時間変化パターンを捉えるためのフーリエ基底を組み合わせた加重リッジ回帰により, 各メッシュ固有の 24 時間分布を推定した.

空間相関構造を捉えるため,k-means 法により基底点を設定し, 最小全域木から得られるスケールパラメータを用いてガウシアンカーネル行列を構成後, 二重中心化・固有分解により空間固有ベクトル $\mathbf{E}$ を構成した. 時間パターンには,M 次までの調和波を含むフーリエ基底 $\mathbf{B} \in \mathbb{R}^{24 \times P}$ を用いた：

$$B_{h,p} = \begin{cases} 1 & p = 1 \\ \cos \frac{2\pi m h}{24} & p = 2m \\ \sin \frac{2\pi m h}{24} & p = 2m+1 \end{cases} \tag{11}$$

観測データ $\mathbf{Y} \in \mathbb{R}^{N \times 24}$ に対し, メッシュ別総発生件数 w_i による加重付きで以下の正則化問題を解いた：

$$\min_{\mathbf{\Theta}} \|\mathbf{W}^{1/2}(\mathbf{Y} - \mathbf{E}_{aug}\mathbf{\Theta}\mathbf{B}^{\top})\|_F^2 + \lambda_s \|\mathbf{\Theta}\|_F^2 + \lambda_t \text{tr}(\mathbf{\Theta}^{\top}\mathbf{W}_t\mathbf{\Theta}) \tag{12}$$

ここで $\mathbf{W} = \text{diag}(w_1, \ldots, w_N)$ は加重行列,$\mathbf{E}_{aug} = [\mathbf{1}, \mathbf{E}]$ は切片項を含む拡張空間行列,$\mathbf{W}_t$ は高調波次数の 2 乗に比例するペナルティ行列である. 最適化後, 非負化・正規化により時間分布 $\hat{\pi}_{i,h}$ を得た.

2.3 統合リスク推定

最終的な時間別リスクは、日別予測値（2.1 章）と時間帯別予測値（2.2）を掛け合わせた次式により推計する：

$$\hat{\lambda}_{i,t,h} = \hat{\lambda}_{i,t} \cdot \hat{\pi}_{i,h} \tag{13}$$

この方法により,「量」(日別リスクの大きさ) と「配分」(時間帯の偏り) を独立に推定し, 疎なデータ下でも安定した予測を実現する.

3 実際のデータでの推定

3.1 データセット

ウルグアイ・モンテビデオ市の 2013–2020 年犯罪記録 (緯度経度付き) を使用した。50m メッシュに集計後、最多発生犯罪である窃盗 (HURTOS) を中心に分析した。分析対象は犯罪発生頻度の高いエリアに限定し、総メッシュ数は約 14,015、総観測期間は 2,892 日である。

3.2 データの特徴

50m メッシュ・日単位の観測は極めて疎であり, 平均して 1 メッシュ1 日あたり 0.004 件の発生である. ホットスポットであっても, 日別の発生件数は 0 または 1 の不規則なパターンを示す. 一方で, 累積発生件数として観察すると多くのメッシュで一定の傾きを持つ直線的な増加パターンが確認され, 短期的には不規則に見える犯罪発生も中長期的には安定した発生率 (ポアソン強度) を持つことが示唆される (図 1).

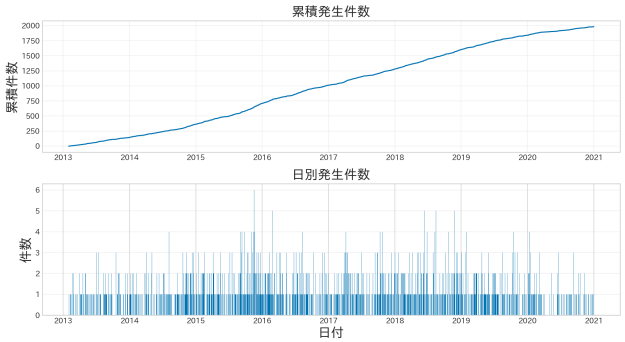


Figure 1: ホットスポットにおける日別発生件数の時系列例

評価指標としては, 上位の高リスクメッシュが実際の犯罪をどの程度捕捉できているかを示すヒットレートを採用した. 本研究では, 重点的な防犯資源配置の実務的要請を踏まえ, 上位 1%, 5%, 10%, 20%におけるヒットレートの値を評価対象とした. ベースライン手法としては, 過去 30 日間の平均に基づく単純移動平均のみを用いた. これは最も基本的かつ解釈容易な手法であり, 提案手法の性能を検証する上で適切な基準となる.

3.3 結果

提案手法と既存のベースライン手法について,2020 年のデータを用いてヒットレートを比較した. 提案手法として, ランダムウォークモデル (EMA), トレンド考慮モデル (T-MPF), 曜日効果モデル (W-MPF) を用い, ベースライン手法として単純移動平均 (SMA) を比較対象とした. 表 1 に各手法の結果を示す.

Table 1: 手法別・パーセンタイル別ヒットレート (HURTOS, 2020)

	SMA	T-MPF	W-MPF	EMA
1%	0.205	0.170	0.172	0.212
5%	0.413	0.348	0.351	0.458
10%	0.554	0.463	0.482	0.625
20%	0.717	0.607	0.633	0.811

各手法については, 指数移動平均の割引率や移動平均の日数など複数のハイパーパラメータを試行し, 最も良好な結果を示した設定を採用した. 移動平均および指数移動平均の詳細を表 2 に示す.

Table 2: 移動平均および指数移動平均モデルにおける平均ヒットレート (HURTOS, 2020)

手法	1%	5%	10%	20%
SMA(7 日)	0.124	0.252	0.301	0.385
SMA(30 日)	0.178	0.354	0.492	0.557
SMA(90 日)	0.205	0.413	0.554	0.717
SMA(全期間)	0.209	0.449	0.619	0.810
EMA $r = 0.0001$	0.209	0.450	0.619	0.810
EMA $r = 0.0005$	0.211	0.456	0.624	0.811
EMA $r = 0.001$	0.213	0.458	0.626	0.811
EMA $r = 0.003$	0.216	0.458	0.623	0.804
EMA $r = 0.005$	0.216	0.453	0.617	0.793
EMA $r = 0.02$	0.208	0.426	0.561	0.737
EMA $r = 0.08$	0.179	0.358	0.512	0.719

表 1 の結果から,MPF 手法 (T-MPF,W-MPF) は単純なベースライン手法 (EMA,SMA) に対して劣る性能を示した. この結果は,50m メッシュ・日単位という極めて疎なデータ環境における複雑なモデルの限界を示している.

T-MPF と W-MPF はそれぞれ 2 次元および条件付き状態遷移を持つ複雑な状態空間モデルであるのに対し,EMA は単一パラメータの指数重み付け平均という単純な構造である.50m メッシュでは平均して 1 日あたり 0.004 件という極めて疎な観測データとなるため, 複雑なモデルのパラメータを安定的に推定することが困難となる. 特に, トレンド成分や曜日効果といった追加パラメータは, 十分な観測データがない状況では粒子の退化や推定の不安定化を引き起こしやすい.

一方で,EMA のような単純な手法は, 少ない観測データでも比較的安定した推定が可能であり, 疎なデータ環境においてはより頑健性を示す. この結果は, 高解像度犯罪予測においては, モデルの複雑さよりもデータの疎性に対する頑健性がより重要であることを示唆している.

EMA 手法の中でも, 割引率の設定によって性能が大きく変化することが確認された. 全体として, 割引率が 0 から 0.001 程度まではヒットレートが向上し, その後は漸減する傾向を示した. 上位 1%,5%のホットスポットを対象とする場合は高めの割引率 ($r \approx 0.001$ 前後) が有効であり, 上位 20%のように広い範囲を対象とする場合には低割引率 ($r \to 0$) が安定して高い性能を示した. これは, 少数の高リスクメッシュでは短期的なトレンドを捉えることが重要である一方, 広域をカバーする場合は長期的な安定性が重視されることを反映している.

図 2 に示すように, 全てのエリアで時間分布は単峰性を示したが, ピークとなる時刻は昼から夜までエリアごとに異なっていた. 提案手法はこれらの変動に対して実測値をよく追随しており, ピーク位置や全体的な形状を精度高く再現できている.

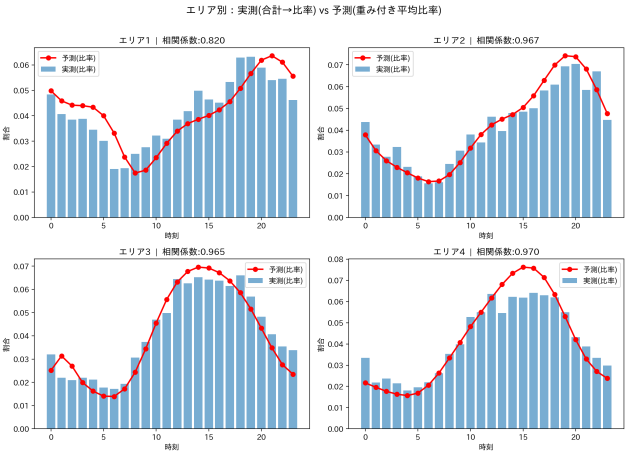


Figure 2: 地理的エリア別の実測時間分布 vs 予測時間分布の相関.4 つのエリアで 0.82-0.97 の高い相関を示した.

さらに, 図 3 では対象期間において 3 件以上の犯罪が発生したメッシュにおける予測と実測のピーク時刻差の分布を示している. 分析対象となったメッシュにおいて, 予測ピーク時刻が実測ピーク時刻よりも平均で+0.6 時間遅れる傾向が観察された. この結果は, 提案手法が実測データの時間的パターンを概ね正確に捉えているものの, わずかな系統的なずれが存在することを示している. ヒストグラムの形状は 0 時間差を中心とした対称に近い分布を示しており, 多くのメッシュで予測と実測のピーク時刻が一致またはわずかな差に収まっていることが確認できる.

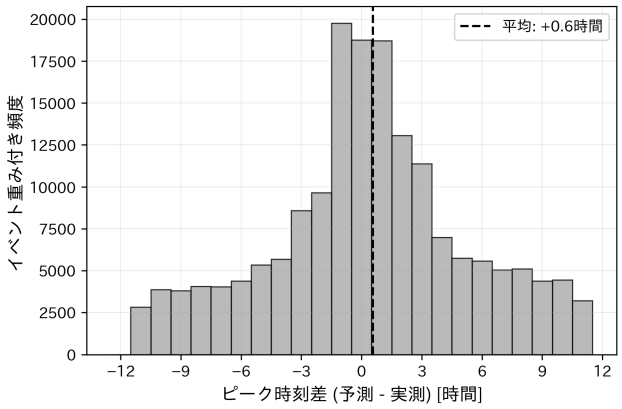


Figure 3: 予測と実測のピーク時刻差の分布. 平均+0.6時間のわずかな遅れが観察されるが, 多くのメッシュで良好な一致を示している.

これらの結果から, 提案手法は時間分布推定において高い精度を達成していることが確認された. 特に, 異なる地理的特性を持つエリア間での時間パターンの多様性を適切にモデル化できており, 商業地域, 住宅地域, 交通拠点など, 各エリアの機能的特徴に応じた犯罪発生の時間的集中を正確に予測していると考えられる.

4 結論と今後の課題

本研究では, 逐次更新による日別リスク推定と空間固有ベクトル・時間フーリエ基底を組み合わせた時間分布推定を統合することで, 従来困難とされた 50m・1 時間単位での犯罪リスク推定を実現した. 提案手法は大規模都市でもリアルタイム運用が可能なほどに軽量である. また, 新しいデータが得られ次第の逐次更新機能と, 空間・時間パターンを明示的に把握できる解釈可能性を併せ持つ.

ウルグアイ・モンテビデオ市の実データを用いた検証により, 多くの地理的エリアで 90%以上の高精度時間分布推定を達成し, 防犯実務に有用な高解像度リスク推定システムを構築した. この高解像度予測により, 時間帯別パトロール計画の最適化, 警備員配置の動的調整, 商業施設の防犯対策強化時間の特定, 住民への時間帯別注意喚起といった実務応用が期待される.

一方で, 本手法にはいくつかの制約も存在する. 極めて稀な犯罪 (年間数件) では推定精度が低下し, 突発的イベント (デモ, 災害等) への対応は困難である. また, 社会経済変数 (失業率, 人口動態等) の考慮が不十分であることも課題として挙げられる.

今後の研究では, 複数犯罪種別の同時推定と相互作用モデリング, 曜日・季節性・祝日効果の明示的導入, 社会経済・環境変数の統合が必要である. さらに, 他都市でのロバスト性検証と実際の警察業務での有効性評価を通じて, 手法の汎用性と実用性を検証する必要がある. 特に, 実務現場での運用評価と継続的な性能改善が重要な次のステップである.

謝辞

本研究は, モンテビデオ市警察からのデータ提供により実現された. 関係者各位に深く感謝申し上げる. なお，本研究は JSPS 科研費 25H00543,25K01339 の助成を受けたものである.

References

[1] Spencer Chainey, Lisa Tompson, and Sebastian Uhlig. The utility of hotspot mapping for predicting spatial patterns of crime. *Security journal*, Vol. 21, No. 1, pp. 4–28, 2008.

[2] Joel M Caplan, Leslie W Kennedy, and Joel Miller. Risk terrain modeling: Brokering criminological theory and gis methods for crime forecasting. *Justice quarterly*, Vol. 28, No. 2, pp. 360–381, 2011.

[3] George O Mohler, Martin B Short, P Jeffrey Brantingham, Frederic Paik Schoenberg, and George E Tita. Self-exciting point process modeling of crime. *Journal of the american statistical association*, Vol. 106, No. 493, pp. 100–108, 2011.

[4] Seth Flaxman, Andrew Wilson, Daniel Neill, Hannes Nickisch, and Alex Smola. Fast kronecker inference in gaussian processes with non-gaussian likelihoods. In *International conference on machine learning*, pp. 607–616. PMLR, 2015.

[5] Chao Huang, Junbo Zhang, Yu Zheng, and Nitesh V Chawla. Deepcrime: Attentive hierarchical recurrent networks for crime prediction. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pp. 1423–1432, 2018.

[6] Ying-Lung Lin, Meng-Feng Yen, and Liang-Chih Yu. Grid-based crime prediction using geographical features. *ISPRS International Journal of Geo-Information*, Vol. 7, No. 8, p. 298, 2018.

[7] Shinya Nakano, Genta Ueno, and Tomoyuki Higuchi. Merging particle filter for sequential data assimilation. *Nonlinear Processes in Geophysics*, Vol. 14, No. 4, pp. 395–408, 2007.

[8] Daniel A Griffith. Spatial filtering. In *Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization*, pp. 91–130. Springer, 2003.