

犯罪・交通事故の発生件数データを対象とした時系列解析モデル と分析ツールの開発

申請中 東京大学 *西颯人 NISHI Hayato
東京大学 樋野公宏 HINO Kimihiro

1. はじめに

犯罪や交通事故は人々の生活の安全を脅かす出来事であり、警察はこれらの防止に尽力している。これらは徐々に減少傾向にあるが、罪種によっては増加傾向にあるものもあり、また特定の時期に事件・事故が例年増加するような罪種・地域も見られる [1]。このような背景を踏まえると、効果的な防止活動を立案する上で、犯罪・交通事故のトレンドや季節性を考慮した時系列的な分析の必要性は大きいと考えられる。

しかし、犯罪・交通事故データの分析には以下のような困難がある。第一に、これらのデータはカウントデータであり、しかも特に我が国の犯罪は諸外国と比較してカウント数が少ない。時系列解析に広く用いられる自己回帰モデル・状態空間モデル [2] はガウス型のモデルであるため、これらの手法を直接的に適用するには難がある。特に時空間的に詳細な分析を行いたい場合には、集計単位を小さくする必要があるため、カウント数の少ないデータに適用可能な分析手法が求められる。第二に、分析ツールの現場での運用を考えると、いくつものハードルがある。例えば、特に次数の高い自己回帰モデルなどではモデルの解釈が容易ではない、分析に際してプログラミング技術が要求される場合がある、などのハードルが挙げられる。

これらを踏まえ、本研究では犯罪・交通事故の件数分析に適した時系列モデルとその推定法を提案するとともに、運用が容易な GUI を実装する。

2. 提案手法

2.1. モデル

犯罪・交通事故のトレンドと季節性を検証するため、これらに着目した状態空間モデルを拡張する形で統計モデルを構築する。状態空間モデルは潜在的なシステムモデルとシステムからデータが出力される観測モデルから成る。観測モデルとし

ては負の二項分布モデルを採用する。これは交通事故分析において外れ値に頑健なモデルとして利用されている [3]。時点 t における発生件数を y_t とすると、観測モデルは、

$$p(y_t | r_t, x_t) = \text{NB}(y_t | r_t, p_t) \quad (1)$$

$$p_t = \frac{\exp(x_t)}{1 + \exp(x_t)} \quad (2)$$

$$r_t = u_t r \quad (3)$$

となる¹。ここで p_t, r_t は負の二項分布のパラメータであり、前者はシステムモデルから生成される潜在変数 x_t をシグモイド関数で変換することで得られる。また後者は分散を調整するパラメータ r と時点 t における発生リスクの曝露 u_t の積によって表される。曝露は、例えば月別集計データであれば、その月の日数などが考えられる。 y_t の期待値は $\exp(x_t)r_t = \exp(x_t)u_t r$ であるから、 x_t (の指数変換) は時点 t での発生リスクと解釈できる。

次にシステムモデルについて、本研究ではトレンドと季節性に着目して発生リスクを以下のように分解する。

$$x_t = \mu_t + s_t + \varepsilon_{x,t} \quad (4)$$

$$\mu_t = 2\mu_{t-1} - \mu_{t-2} + \varepsilon_{\mu,t} \quad (5)$$

$$s_t = -\sum_{i=1}^{p-1} s_{t-i} + \varepsilon_{s,t} \quad (6)$$

$\varepsilon_{x,t}, \varepsilon_{\mu,t}, \varepsilon_{s,t}$ はそれぞれガウスノイズである。トレンド・季節成分はそれぞれ μ_t, s_t であり、それぞれ 2 次のトレンドモデルと周期 p の標準的季節調整モデルに従う [2]。これらの和に対して、式 (4) では分析をさらに頑健にするためのノイズ $\varepsilon_{x,t}$ を加えている。式 (4) から、リスク $\exp(x_t)$ は各成分の積となるから、このシステムモデルは発生リスクが割合で増減するモデルであると解釈できる。

¹NB は負の二項分布の確率質量分布である

2.2. 推定法

状態空間モデルの推定に広く使用される Kalman Filter は、ガウス型のモデルを仮定するため提案モデルに対して直接の適用はできない。一般の状態空間モデルの推定法としては Particle Filter が知られているが、提案モデルは指数変換を伴うため、モンテカルロ法の適用はオーバーフローのリスクを伴い、安定性に不安がある。

このため、本研究ではすべてのパラメータに共役事前分布を置いて、平均場近似による構造化変分ベイズ推論 [4, 5] を行う。負の二項分布モデルはそのままでは変分ベイズアルゴリズムを導出できないが、補助変数を導入することで解析的な更新式が導出できる [6]。

3. 分析結果

提案手法による分析例として、横浜市青葉区における月別の空き巣発生件数の分析結果を図 1 に示した。曝露は月当たり日数、周期は 12 とした。上段は実際の観測件数（黒色）と推定されたリスク $\exp(x_t)r_t$ の平均及び信頼区間（灰色）を示しており、中段はトレンド成分を、下段は季節成分を示す。図 1 から、空き巣発生リスクが解釈しやすい滑らかなトレンドと周期的な季節成分に分解されていることが読み取れる。モデルには 2010 年から 2016 年までの件数を入力しており、図の上段右端は予測と 2017 年の件数を比較している。これを見ると、実際の発生件数に近い予測がなされており、分析結果は妥当であると考えられる²。

4. GUI ツール

上記の手法を広く利用できるようにするため、提案手法の GUI を作成した (図 2)。これにより、データを csv 形式で用意するだけで先述の分析が可能となった。

5. おわりに

本研究では犯罪・交通事故を対象とした時系列解析モデルの提案とその GUI ツールの作成を行った。当該 GUI ツールは現在神奈川県警察に試供しており、今後フィードバックを経て改善・機能追加等を行ったうえで、ツールの公開を目指したい。

²上段のグラフは期待値の予測であるから、発生件数の予測信頼区間はさらに広がることになる。

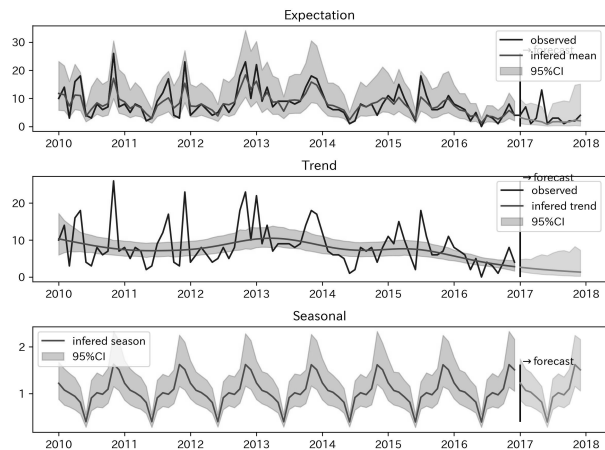


図 1: 横浜市青葉区の空き巣分析の例

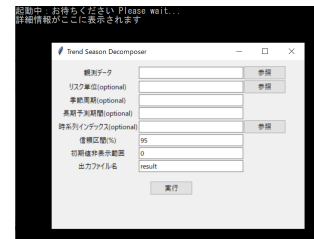


図 2: GUI 画面

謝辞

本研究は、神奈川県警察と締結した「犯罪・交通事故分析に関する共同研究協定」の枠組みに基づき提供されたデータで分析・評価を行いました。また本稿の執筆に当たり、薄井宏行助教 (東京大学) からご助言を頂きました。

参考文献

- [1] 西颯人, 樋野公宏: 交通事故発生件数のトレンドと季節性の類型化 神奈川県車両相互事故を対象として, 都市計画報告集 No.17 (2018).
- [2] 北川源四郎: 時系列解析入門, 岩波書店 (2005).
- [3] Dominique Lord, Fred Mannering: The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives, Transportation Research Part A: Policy and Practice(2010).
- [4] 須山敦志: ベイズ推論による機械学習入門, 講談社 (2017).
- [5] Matthew J. Beal: Variational algorithms for approximate bayesian inference, PhD Thesis(2003).
- [6] Mingyuan Zhou, et al.: Lognormal and gamma mixed negative binomial regression, 29th International Conference on Machine Learning(2012).