

卒業論文

長期的リスクと短期的リスクを考慮した
主成分回帰分析による地理的犯罪予測と
地域ごとの要因分析

English Title

富山県立大学 工学部 電子・情報工学科

1915043 島部達哉

指導教員 奥原 浩之 教授

提出年月: 2023年2月

目次

| | |
|--------------------------|-----|
| 図一覧 | ii |
| 表一覧 | iii |
| 記号一覧 | iv |
| 第1章 はじめに | 1 |
| § 1.1 本研究の背景 | 1 |
| § 1.2 本研究の目的 | 1 |
| § 1.3 本論文の概要 | 1 |
| 第2章 リスクと犯罪 | 1 |
| § 2.1 地理的犯罪予測 | 1 |
| § 2.2 長期的リスクと短期的リスク | 2 |
| § 2.3 オープンデータと GIS | 3 |
| 第3章 犯罪予測と要因分析に対する NN の活用 | 3 |
| § 3.1 主成分回帰分析を用いる GMDH | 3 |
| § 3.2 地理的犯罪予測と地域ごとの要因分析 | 3 |
| § 3.3 GMDH による予測と要因分析の例 | 3 |
| 第4章 提案手法 | 3 |
| § 4.1 | 3 |
| § 4.2 | 3 |
| § 4.3 | 3 |
| 第5章 数値実験並びに考察 | 4 |
| § 5.1 数値実験の概要 | 4 |
| § 5.2 実験結果と考察 | 4 |
| 第6章 おわりに | 5 |
| 謝辞 | 5 |
| 参考文献 | 5 |

圖一覽

表一覽

記号一覧

以下に本論文において用いられる用語と記号の対応表を示す.

| 用語 | 記号 |
|--|--------------|
| LiNGAM における i 番目の観測変数 | x_i |
| LiNGAM における j 番目の観測変数から i 番目の観測変数へのパス係数 | b_{ij} |
| LiNGAM における i 番目の観測変数に対する誤差 (非観測変数) | e_i |
| 主問題における各入力に対する重み | v^T |
| 主問題における各出力に対する重み | u^T |
| 主問題における対象 DMU の評価値 | z |
| CCR モデルにおける DMU _o の入力 | x_o |
| CCR モデルにおける DMU _o の出力 | y_o |
| CCR モデルにおける DMU の入力 | X |
| CCR モデルにおける DMU の出力 | Y |
| 双対問題における対象 DMU の評価値 | w |
| 入力指向モデルにおける対象 DMU の評価値 | θ |
| 入力指向モデルにおける各 DMU に対する重み | λ |
| 出力指向モデルにおける対象 DMU の評価値 | η |
| 出力指向モデルにおける各 DMU に対する重み | μ |
| 入力指向モデルにおける対象 DMU の i 番目の入力に対する改善案 | \hat{x}_i |
| 入力指向モデルにおける参照集合内の k 番目の DMU の i 番目の入力 | x_{ik} |
| 入力指向モデルにおける参照集合内の k 番目の DMU に対する重み | λ |
| 出力指向モデルにおける対象 DMU の j 番目の出力に対する改善案 | \hat{y}_j |
| 出力指向モデルにおける参照集合内の k 番目の DMU の j 番目の出力 | y_j |
| 出力指向モデルにおける参照集合内の k 番目の DMU に対する重み | μ |
| 提案手法における d 番目の市区町村の i 番目の入力 | x_{id} |
| 提案手法における d 番目の市区町村の i 番目の出力 | y_{id} |
| 提案手法における d 番目の市区町村に対する重み | λ_d |
| <i>robust Z-score</i> における正規化後の値 | ι |
| <i>robust Z-score</i> を用いて正規化するデータ集合内の値 | x |
| <i>robust Z-score</i> を用いて正規化するデータ集合 | X |
| <i>robust Z-score</i> を用いて正規化するデータ集合の中央値 | $median(x)$ |
| <i>robust Z-score</i> を用いて正規化するデータ集合の正規四分位範囲 | $NIQR$ |
| 0~1 変換の結果の値 | ι' |
| 0~1 変換を行うデータ集合内の値の最大値 | $max \iota $ |

はじめに

§ 1.1 本研究の背景

米国の心理学者スティーブン・ピンカーは、自らの著書『暴力の人類史』のなかで、現代は人類史上もっとも暴力の少ない時代だと述べている。国ごとに差はあるものの、

§ 1.2 本研究の目的

§ 1.3 本論文の概要

本論文は次のように構成される。

第1章 本研究の背景と目的について説明した。背景では、地理的犯罪予測の重要性について述べた。目的では、地理的犯罪予測の手法として、地域ごとの要因を可視化するために、主成分回帰分析を用いる GMDH を提案することを述べた。

第2章 地理的犯罪予測の概要と、その手法についてそれぞれ述べたあと、犯罪が発生するさまざまなリスクを説明変数とする機械学習の優位性を示す。また、犯罪が発生するリスクについて述べる。さらに、地理的犯罪予測には欠かせない GIS について、その概要を述べる。

第3章

第4章

第5章

第6章 本論文における前章までの内容をまとめつつ、本研究で実現できたことと今後の展望について述べる。

リスクと犯罪

§ 2.1 地理的犯罪予測

犯罪は、完全にランダムで発生するわけではなく、何らかの要因によって発生することが多い。こうした性質に着目し、過去に発生した犯罪のデータや、人口密度や学校までの距離、天候などといった、犯罪発生の変因となり得る間接的なデータを使って、そのエリアで犯罪が発生する確率を予測する研究が、欧米を中心に行われている。

地理的犯罪予測は、車上あらし、自動車や自転車などの盗難、空き巣など、犯罪を行いやすい時間や場所、状況を選んで行われるような罪種に対して行われる。その予測結果は、警察などに提供し、そのエリアに対して集中的にパトロールを行うなどといった判断を助けることを想定している。そのため、予測の範囲を、時間的には週や日、時単位、空間的には数十から数百メートルのグリッドセル単位としている研究が多い。

このような地理的犯罪予測に関する研究は、Cohen et al. (1980) が起源のひとつとして挙げられる。同研究では、1947年から1972年における、同罪種の犯罪発生率、失業率、15歳から24歳の人口比率を説明変数として、自己回帰モデルを用い、全米における強盗、侵入盗、自動車盗の、1年ごとの犯罪発生率を推定している。

同時期の研究は、全国や州における、1年ごとの予測といった、時間的や空間的に解像度が粗いものが多かった。さらに、時間的な作用と空間的な作用の交互を考慮しないものが多かった。これらが起源となり、時間的や空間的な解像度がより細かい予測や、それぞれの交互作用を考慮する予測をする研究へとつながっていった。

その結果、今日における地理的犯罪予測の手法は、大きく6つに分類できる。

i. 時空間クラスタの検出

(説明)

ii. 犯罪の時空間相互作用を考慮した犯罪発生強度推定

(説明)

iii. 犯罪生成・誘引要因からの犯罪発生リスクの予測

(説明)

iv. 多様な情報からの犯罪発生リスクの予測

(説明)

v. 地区単位での犯罪発生件数，または確率の時系列予測

(説明)

vi. 環境要因と犯罪の時空間相互作用を組み合わせた犯罪発生リスクの予測

(説明)

(諸外国における実務への応用例や，日本においても地理的犯罪予測の機運が高まっていることを説明)

(なぜさまざまな手法があるなかで，本研究では vi を採用するのか，なぜ機械学習，さらには GMDH を採用するのかを述べる.)

§ 2.2 長期的リスクと短期的リスク

2.1 節で述べたように，地理的犯罪予測では，過去に発生した犯罪のデータのほかに，犯罪発生の変因となり得る間接的なデータを用いる．Taylor et al. (2015) では，犯罪発生の変因となり得るものとして，「数年，あるいは数十年間固着しうる長期のリスク」と「むしろ数日から数週間にかけて増大する短期のリスク」の，大きく 2 つに分けている．以降，前者を「長期的リスク」，後者を「短期的リスク」と呼ぶ．

長期的リスク

i. 社会経済的変因

(説明)

ii. 犯罪発生・誘引変因

(説明)

iii. 犯罪可能変因

(説明)

短期的リスク

i. 近接反復被害

(説明)

ii. 前兆事案

(説明)

iii. 環境変因

(説明)

§ 2.3 オープンデータと GIS

2.1 節や 2.2 節で述べたように、地理的犯罪予測では、過去に発生した犯罪のデータや、犯罪発生の変因となり得る間接的なデータを用いる。特に、過去に発生した犯罪データや長期的リスクとするデータは、その多くが国や地方自治体などの公的機関が所有している。従来であれば、一般に公開していないか、公開していたとしても利用申請などが必要だったり、利用に制約があったりなど、特に本研究のような、多種多様なデータを用いる場合は、必ずしもそのハードルは低いと言えなかった。

しかし近年、公的機関が保有しているデータの「オープンデータ」化が進んでいる。すなわち、利用に制約がないデータの公開が進んでおり、

オープンデータ

(オープンデータの概要と、本研究で使用するオープンデータの例など)

GIS

(GIS の概要と、地理的犯罪予測における応用法)

(長瀬先輩の論文を参考にする)

犯罪予測と要因分析に対するNNの活用

§ 3.1 主成分回帰分析を用いる GMDH

本研究では、犯罪を予測する手法として、主成分回帰分析を用いた GMDH を適用する。この主成分回帰分析を用いた GMDH では、一般的に使用されることが多いニューラルネットワークと違い数法則式を導出することが可能である。法則についてブラックボックスになるニューラルネットワークでは特徴量の重要度や選択など不透明となるケースが多い。そのためモデルを構築する際に法則が見えづらく考察することが困難となる。一方、この GMDH では特徴量を自己選択し、重みによって特徴量の重要度を明確にすることによってモデルの構築についてクリアにすることが可能である。この手法は、評価基準を用いて変数の逐次選択を行い最適な部分表現式を自己選択する GMDH である。システムの完全表現式として、Kolmogorov-Gabor の多項式

§ 3.2 地理的犯罪予測と地域ごとの要因分析

§ 3.3 GMDH による予測と要因分析の例

提案手法

§ 4.1

§ 4.2

§ 4.3

数値実験並びに考察

§ 5.1 数値実験の概要

§ 5.2 実験結果と考察

おわりに

謝辞

本研究を遂行するにあたり，多大なご指導と終始懇切丁寧なご鞭撻を賜った富山県立大学工学部電子・情報工学科情報基盤工学講座の奥原浩之教授，António Oliveira Nzinga René 講師に深甚な謝意を表します．最後になりましたが，多大な協力をしていただいた研究室の同輩諸氏に感謝致します．

2023 年 2 月

島部 達哉

参考文献