

カーネル密度推定と回帰分析を用いた 盗難事件発生予測システムの開発

久野竜弥[†] 堀部典子[†] 樋口直哉[†]

崇城大学 情報学部 情報学科[†]

1 はじめに

近年、犯罪予測システムが普及しつつあり、システムが犯人検挙に役立った実績もある。しかし、これらのシステムは一般の人々向けに公開されていないため、一般の人々が未然に犯罪を防ぐことに利用することができない状況にある。本研究では、オープンデータを用いて、一般の人々向けにシステムを開発することで、犯罪に対しての意識を高めるとともに、犯罪発生件数を減らすことを目的としている。過去の犯罪データと気象データを基に、回帰分析を用いて犯罪が起きる可能性がある時間や件数を予測し、カーネル密度推定を用いて予測地点の犯罪発生リスクを予測する。予測結果を Web マップ上に表示することにより、一般の人々が自発的に犯罪への関りを防ぐための情報源として活用できるシステムを提案する。予測結果にはシステム使用者の現在地も表示する。本研究は犯罪の種別を、身近に発生している自転車盗難事件に限定して予測を行う。

2 予測に用いるデータ

「犯罪履歴データ」は、熊本県の公式 Web サイトで公開されているオープンデータから、公開されている平成 30 年から令和 3 年までの 4 年間の窃盗罪のデータを活用する[1]。データの内容は、罪名や手口など全 14 項目ありその中から必要な項目を抽出し、予測に使用する。

気象データとして、気象庁が Web サイトで公開されているオープンデータから犯罪履歴データと同じ期間のデータを取得する[2]。データの内容は天気と気温のデータを用いる。

予測する日の気象情報は、気象庁の API で公開されているアメダスのデータから取得する。気象情報は 1 時間ごとに新しい情報を取得できるようにする。

システム使用者の現在地の取得には、Python の pyicloud モジュールを用いる。このモジュールを使用することによって、システム使用者の

iPhone に紐づけられた AppleID とパスワードから端末の位置情報を取得できる。

3 犯罪予測のための手法

本研究で用いるカーネル密度推定と回帰分析について説明する。

カーネル密度推定とは、有限の標本点から全体の分布を予測する手法である[3]。予測を行うときは、バンド幅とカーネル関数を設定する必要がある。バンド幅とは分布の広がり幅のことで、値を小さくすると幅が小さくなり、一つの点の影響が強くなる。カーネル関数は広がり方を決める関数で、正規分布と三角比と均等化がある。

回帰分析には多項式回帰を用いる。多項式回帰とは、2 次以上の項を持つ関数を仮説として設定するもので、曲線でデータの関係を表すことができる手法である。本研究では Python のライブラリ scikit-learn に実装されている Polynomial-Features クラスを用いて予測を行う。degree の値を変えることで曲線の次元数を変更することができる。値を大きくすることで決定係数の値は 1 に近づくが過学習が起きる。決定係数は 1 に近づくほど予測精度が良く、0.9 以上あれば良いとされている。過学習とは一つ一つの要素に適合しすぎていると発生しやすく、予測するとき想定と大きく外れた値が出ることもある。過学習が発生した場合は、次元数を減らすか、データの要素数を減らして適切に適合させる必要がある。

4 犯罪予測の実験

多項式回帰を用いて犯罪発生件数の予測実験を行う。予測範囲を熊本市中央区に制限し、犯罪発生時の気温を説明変数にして予測を行った結果について図 1 に示す。青い点が実際の犯罪発生件数で、赤い線が予測結果である。25 度の場合の一番件数が多く、30 度を超えると激減している。予測値は 25 度の場合約 67 件で、35 度の場合 8 件となった。次に、犯罪発生時間を説明変数にした場合の結果について図 2 に示す。青い点が実際の犯罪発生件数で、赤い線が予測結果である。18 時の発生件数が一番多く、4 時の発生件数が一番少ない。予測値は 18 時の場合は約 139 件で、4 時の場合約 11 件となった。結

Development of a Theft Incident Prediction System by
Kernel Density Estimation and Regression Analysis

[†]Ryuya Hisano, Noriko Horibe, Naoya Higuchi,
Sojo University

果から、多項式回帰による予測ができていると考えられる。

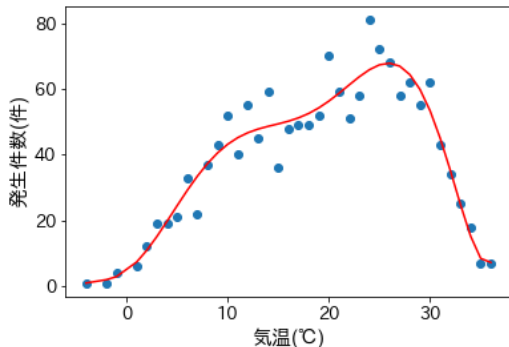


図 1. 発生時の気温による予測

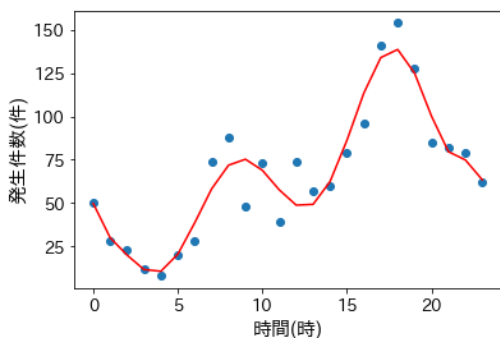


図 2. 発生時間による予測

多項式回帰の予測範囲を〇丁目ごとにした場合、データ数が少ないため予測を行うことはできなかった。そのため、多項式回帰の結果は市区町村ごとの発生リスクを示すために使用する。

次に、過去の犯罪データを用いてカーネル密度推定の実験を行った。また、システム使用者の位置情報をカーネル密度推定の結果に反映し、現在地周辺の犯罪発生リスクを認識できるようにする。最後に結果を地図と合わせて表示する。地図は、国土地理院の「地理院地図」を用いた[4]。予測範囲を熊本市南区に制限し、図 3 に過去の犯罪発生場所の散布図を示す。点が集まっている場所は駅や商業施設などがあり、その周辺の犯罪件数は多い傾向にある。

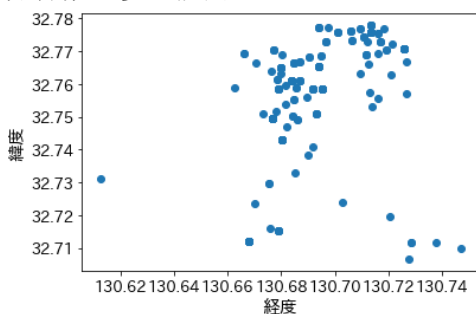


図 3. 熊本市南区の犯罪発生場所の散布図

次に、図 3 の散布図を基にカーネル密度推定を行い、システム使用者の現在地を表示する。カーネル密度推定を行うことで、点が多く集ま

っているエリアは一番色が濃く表示され、点が少ないエリアは色が薄く表示される。システム使用者の現在地は赤白の点で表示する。最後に地図を合わせて表示することでシステム使用者の位置や、犯罪リスクが高いエリアをより分かりやすく表示する。図 4 に実験結果を示す。図 4 から、現在地周辺の犯罪リスクが極めて低いことが分かる。



図 4. 地図を背景にした実験結果

5 まとめと今後の課題

本研究では、身近に発生する盗難事件の発生件数を減らすことを目的に、どこで犯罪が起りやすいかを可視化する犯罪発生予測システムの開発を行った。過去の犯罪履歴データから、犯罪が発生しやすい気温や時間を予測することに成功し、システム使用者周辺の犯罪発生リスクを表示することができた。

本システムは、一般の人々が利用することを想定しているため、AppleID やパスワード入力のプライバシー保護の検討や、新たに犯罪が発生したときのデータの追加の検討が必要である。警察や団体などからリアルタイムで情報を追加することや、システム使用者が自ら犯罪発生情報を入力できるようにする機能を追加することで、直近の犯罪発生情報を通知して警戒心を強めることができると考える。

参考文献

- [1] 熊本県, 熊本県オープンデータサイト-熊本県ホームページ, <https://www.pref.kumamoto.jp/soshiki/211/82808.html> (参照 2023-5-9).
- [2] 国土交通省 気象庁, 気象庁 | 過去の気象データ・ダウンロード, <https://www.data.jma.go.jp/gmd/risk/obsdl/> (参照 2023-10-12).
- [3] 中村 淳子, 西村 祥治, 宮野 博義, 「犯罪発生履歴データの機械学習における時空間カーネル密度推定型犯罪予測の最適化」, 情報処理学会論文誌, Vol. 61, No. 77, 1255-1266, July 2020.
- [4] 国土地理院, 地理院地図 / GSI Maps | 国土地理院, <https://maps.gsi.go.jp/>, (参照 2023-12-26).