

犯罪予測を目的とした地図画像データセットの ズームレベルの比較

伊藤詩織[†] 福安真奈[‡] 向直人[§]

梶山女学園大学[†] 梶山女学園大学[‡] 梶山女学園大学[§]

1 はじめに

近年、過去に起きた犯罪の位置や時間などのデータから、将来の犯罪を予測する試みが注目を集めている。従来の犯罪予測手法には、過去に犯罪が起きた位置から犯罪を予測する「カーネル密度推定」[1]や、金融機関など犯罪発生リスクがある環境要因を用いて予測する「Risk Terrain Modeling(RTM)」[2]などがある。しかし、これらの手法は犯罪発生地域周辺の建物の配置や形状など地理的な特徴量が考慮されていない。そこで我々は、愛知県警察が公開している犯罪オープンデータを利用し、犯罪発生地域の地図画像を生成することで、地理的な特徴量を表現する。生成した地図画像を、畳み込みニューラルネットワークに学習させ、自動車盗や自転車盗などの犯罪種別の予測を試みる。また、生成する地図画像のズームレベルを 16 (一辺が約 468 メートル)、17 (一辺が約 234 メートル)、18 (一辺が約 117 メートル) に変化させ、異なるズームレベルが犯罪種別の分類精度に与える影響を明らかにする。

2 学習用データセットの構築

愛知県警察が公開している犯罪オープンデータには、2018 年から 2020 年に発生した犯罪の発生時刻や発生場所などのデータが含まれている。犯罪種別は 7 種類で構成されているが、本稿では地理的な要因に関連性が高いと思われる「自動車盗」「自転車盗」「自動販売機ねらい」の 3 種類に絞る。犯罪オープンデータから、各犯罪種別を 1,000 件、合計 3000 件のデータを無作為に抽出し、畳み込みニューラルネットワークの学習用データセットとした。地図画像の生成には、犯罪の位置が必要となるが、オープンデータには個人情報保護のため、緯度・経度ではな

く、「都道府県」「市区町村」「町丁目」の文字列しかない。そこで、これらを結合した住所文字列を基に、Mapbox Geocoding API を利用して、対応する緯度・経度を取得する(ジオコーディング)。しかし、取得される緯度・経度は必ずしも正確ではない。そこで、ジオコーディングの信頼度が 0.8 以上のレコードのみを採用した。取得した緯度・経度を基に Mapbox Static Tiles API で一辺が 256 ピクセルの地図画像を生成する。Mapbox のズームレベルは 16, 17, 18 に設定し、3 種類の地図画像のデータセットを構築する。図 1 が各ズームレベルの地図画像のサンプルである。地図画像は、表 1 に示すように、白、黒、緑、赤、青の 5 色で色分けされている。駐車場は点データとして記録されているため、同一半径の円で表す。また、Mapbox の仕様により、ズームレベル 16 においては、小さな建物や、一部の駐車場が表示されないことに留意されたい。

表 1 地図画像の配色

色	対象
白	道路領域
黒	背景領域
緑	建物領域
赤	駅
青	駐車場

3 実験結果と考察

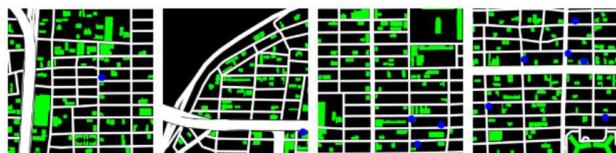
畳み込みニューラルネットワークとして AlexNet[3]を採用した。学習用のデータセットに含まれるサンプル(N=3000)の 8 割(2400 件)を訓練データ、2 割(600 件)を評価データとする。地図画像は 64×64 ピクセルに縮小し、n=256 のミニバッチ方式で、畳み込みニューラルネットワークを学習させる。

Comparison of zoom levels map image datasets for crime prediction purposes

[†] Shiori Ito, Sugiyama Jogakuen University

[‡] Mana Fukuyasu, Sugiyama Jogakuen University

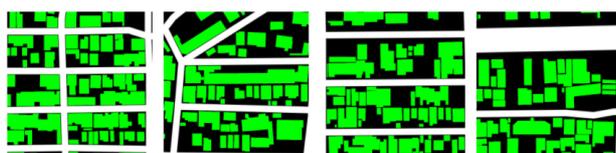
[§] Naoto Mukai, Sugiyama Jogakuen University



ズームレベル : 16



ズームレベル : 17



ズームレベル : 18

図1 各ズームレベルの地図画像

ズームレベル 16 の正解率を図 2, 17 の正解率を図 3, 18 の正解率を図 4 に示す。

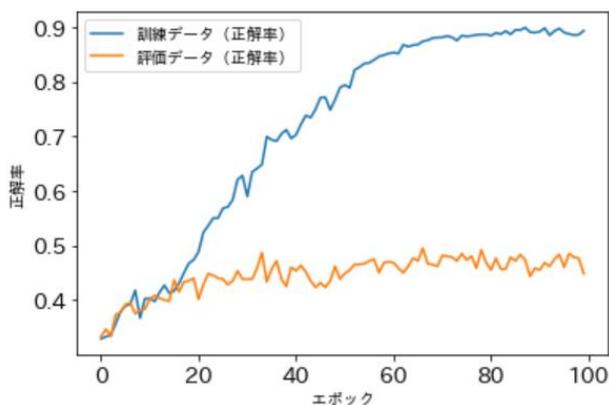


図2 ズームレベル 16 の正解率

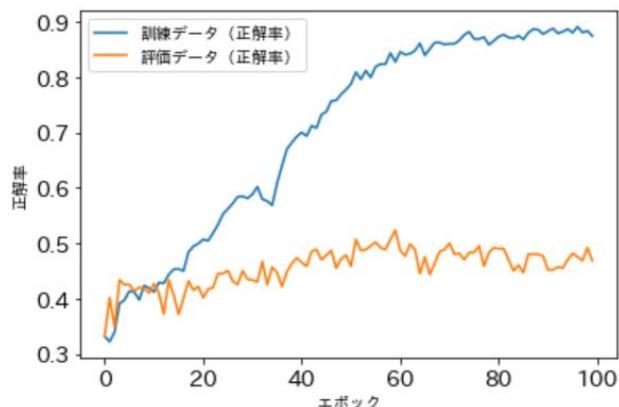


図3 ズームレベル 17 の正解率

正解率の最大値を比較すると、ズームレベル 16 が 49.5%, 17 が 52.3%, 18 が 50.3% であり、17

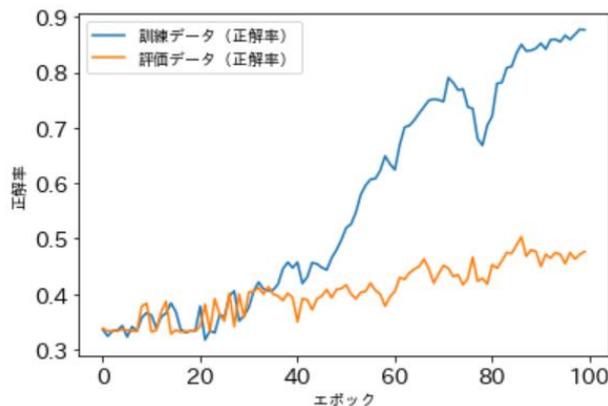


図 4 ズームレベル 18 の正解率が最も高い正解率を示した。このことから、ズームレベル 17 の地図画像が、犯罪発生地域の地理的特徴を最も適切に表現したといえる。一方で、ズームレベル 16 の正解率が最も低かった。小さい建物や一部の駐車場が地図画像に含まれていないことが原因と考えられる。過去研究から駅や駐車場の存在が、「自転車盗」に関与していることが示されている。また、ズームレベル 18 では生成した地図画像の範囲が狭すぎる可能性がある。上述したように住所文字列から地図画像を生成しており、対象の犯罪が地図画像の範囲外で発生した可能性が高くなってしまふ。

4 まとめ

異なるズームレベルの地図画像を学習に用い、犯罪種別の分類精度を比較した。実験の結果、ズームレベル 17 の地図画像が最も地理的特徴量を表現していることが分かった。今後は、分類精度を向上させるため、本稿では考慮しなかった街灯などの環境要因などを表現した地図画像を生成するなどの検討が必要である。

参考文献

- [1] 中川淳子, 西村祥治, 宮野博義. 犯罪発生履歴データの機械学習によ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. 7, pp. 1255-1266, 2020.
- [2] 野貴泰, 糸魚川栄一. 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, Vol. 31, pp. 195-204, 2017.
- [3] Alex Krizhevsky. One weird trick for parallelizing convolutional neural networks. CoRR, Vol. abs/1404.5997, 2014.