

第II部門 集中豪雨の発生予測に対する畳み込みニューラルネットワークの応用

京都大学工学部 学生会員 ○鈴木紹晟 京都大学大学院工学研究科 正会員 Kim Sunmin
 京都大学大学院地球環境学堂 正会員 田中智大 京都大学大学院工学研究科 正会員 萬 和明
 京都大学大学院工学研究科 正会員 市川 温 京都大学大学院工学研究科 正会員 立川康人

1 序論 集中豪雨など局地的な豪雨の発生が増加しており、特にゲリラ豪雨とも呼ばれる局地的大雨は、短時間に大量の降水をもたらすことから洪水被害の原因となっている。この豪雨現象はメカニズムが完全には解明されていないことから予測が困難とされているが、本研究では従来の解析的な手法では把握できなかった豪雨の特徴を取り上げるために機械学習による発生予測を試みた。本研究では集中豪雨と局地的大雨をまとめて集中豪雨と設定し、これらの発生に関連すると思われる気象データに対し、画像認識の分野で活躍している畳み込みニューラルネットワーク(CNN)の応用を考え、集中豪雨予測モデルを考案した。

2 CNNの概要 本研究で用いるCNNは、画像データなど2次元または3次元の入力データに対して、フィルタと呼ばれる2次元配列を用いることで、入力データ中から特徴と呼ばれる特定の配列を取り上げることが可能とする。複数のフィルタを用いることで、各フィルタにより取り上げられた特徴の組み合わせから、入力データの分類を行うことができる。入力データ中の特定の配列に注目することから、画像データなどの配列が意味を持つデータに対して効果的に適用されている。

3 集中豪雨のメカニズムとCNNの応用 集中豪雨や局地的大雨は積乱雲の急激な発達により引き起こされ、下層に暖かい空気、上層に冷たい空気が存在する大気が不安定な状態における上昇気流の発生を原因とする。単独の積乱雲が発達した場合は局地的大雨となり、低気圧や前線の影響、地形効果が加わった場合は積乱雲の発生と発達が繰り返され集中豪雨となる。

この基本的なメカニズムから、集中豪雨の発生予測には発生前における複数地点の気象因子の時間変化が重要だといえる。ここで、観測地点を関係性を考慮して一列に配置し、もう一辺を時系列とする事で、2次元データの作成が考えられる。この2次元データは画像データ同様、配列が意味を持つデータであり、CNN

が適用可能と考えられる。さらに、画像認識においてカラー画像がRGB3色の画像データの重ね合わせであることから、気象変数ごとの2次元データを重ね合わせて3次元の入力データとして扱うことを考えた。

4 集中豪雨予測モデルの構築

4.1 使用データと対象地域 実験に使用するデータとして、複数の気象変数の観測があるAMeDAS気象観測データから1時間ごとの観測値を用いて、降水量、気温、日射量、風速、風向の5つの気象変数を取り出した。予測の対象地域として京都を設定し、さらに図1のように京都周辺一定の範囲から上記の5つの気象変数の観測のある地点を選び出し、計8つの地点に対し任意で順番を決定した。本研究では実験的に5つの気象変数全てと、この任意で決定した8つの観測地点を用いて実験を行った。

本研究では1998年～2017年の20年間の観測データの内、集中豪雨の発生が多い6月～8月のデータを用いた。ここで1998年～2013年の16年分を訓練データ、2014年～2015年の2年分を検証データ(validation)、2016年～2017年の2年分をテストデータ(test)とした。

4.2 データの整備 気象変数の中で、風速の情報は、16方位に静穏を含めた17パターンの表記であったため、静穏を0として北から北北西にかけて1～16の数値を当てはめた。また、最大値を1とする日射量以外の変数については、正規化を行い0～1の値に変換した。

図2のように横の一辺を観測地点、縦の一辺を時系

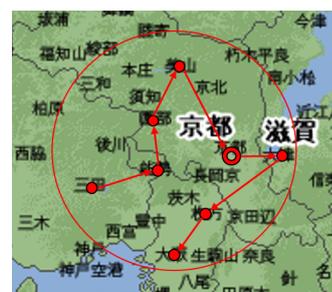


図1 対象地域および周辺の観測地点の選択

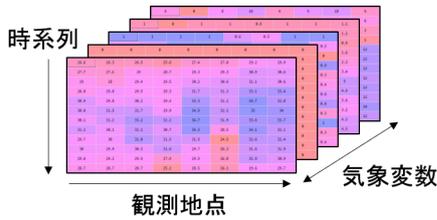


図2 本研究で用いた入力データの形式

表1 パラメータとその数値

パラメータ		数値
入力データ	時間数	6, 12, 24
畳み込み層	フィルタサイズ	3 × 3, 5 × 3, 5 × 5
プーリング層	ウィンドウサイズ	2 × 2
	ドロップアウト率	0.4
学習設定	バッチサイズ	10, 50, 100
	学習率	0.1, 0.01, 0.001
	減衰率	1.0
	エポック数	100

列とした2次元データを気象変数ごと5つ作成し、重ねて入力データとした。10mm/h以上の降水を集中豪雨として1時間後予測を考え、入力データの1時間後における対象地域での集中豪雨発生の有無を示すラベルを入力データに添付し、まとめてCNNへの入力とした。また、集中豪雨発生のデータが発生無しのデータに対し極端に少ないため、発生したデータのみ別途で100回訓練データに追加し、データの個数を調整した。

4.3 CNNのパラメータ設定 モデル構築に関しては一般的な画像認識におけるCNNのコードを参考に作成し、表1のようにパラメータの値を設定した。この中から4つのパラメータを選び出し、それぞれ3つの値を設定した。これらを組み合わせた際のモデルの挙動の違いから、それぞれ最適値を決定した。

5 結果と考察

5.1 評価方法 実験結果は、入力を実況、出力を予測として表2のような2 × 2の4ケースに分類される。本研究ではこの表より、適中率と捕捉率の2つの指標を用いて実験結果を評価した。交差エントロピーを用いたCNNの学習では、適中率が1になるように学習が行われるため、適中率を学習過程の確認に用いた。一方、予測指標には、集中豪雨が発生したデータ中の予測に成功したデータの割合を示す捕捉率を用いた。

5.2 学習パラメータの決定 4.3で述べた4つのパラメータについて調整を行った結果、入力データの時間

表2 実験結果の分類と評価指標

集中豪雨発生		実況	
		あり	なし
予測	あり	FO	FX
	なし	XO	XX

$$\text{適中率} = \frac{FO + XX}{FO + FX + XO + XX}$$

$$\text{捕捉率} = \frac{FO}{FO + XO}$$

表3 各テストデータにおける豪雨ケースごとの捕捉率（予測できた事例/全ての事例）

ケース	A	B	C
validation	67%(16/24)	33%(4/12)	17%(2/12)
test	67%(2/3)	43%(3/7)	0%(0/9)

数は6時間、フィルタサイズは3 × 3、バッチサイズは10、学習率は0.01が最適値だと分かった。さらに実験結果より、他のパラメータについてもドロップアウト率を0、減衰率を0.9、エポック数を50に変更した。

5.3 豪雨ケースごとの予測結果 決定した学習パラメータを用いて、豪雨ケースごとの予測精度を確認した。10mm/h以上の降水を豪雨として、入力データにおける集中豪雨発生の1時間前の降水状況から、対象地点で豪雨があった場合（ケースA）、対象地点で豪雨は無いが他の地点で豪雨があった場合（ケースB）、全地点で豪雨がなかった場合（ケースC）の3つに分類した。時間的・空間的スケールからケースAとBが集中豪雨を、ケースCが局地的大雨を表すとした。

validationとtest各テストデータにおける予測結果の豪雨ケースごとの捕捉率は表3となった。ケースAとBでは33%から67%までのデータを予測できたが、ケースCは予測できて17%程度であった。本研究の予測モデルは集中豪雨の発生には反応を示したが、局地的大雨は捉えられなかったといえる。

6 結論 本研究では、CNNを用いて気象データを任意の形で扱い、反応を得ることができた。集中豪雨に対しては高い精度ではないものの予測に成功したが、一方で、より短時間の内の現象である局地的大雨については予測できなかった。本研究では実験的にAMeDAS気象観測データより1時間ごとの観測値を用いたが、局地的大雨の予測のためには、より短い10分ごとの観測値や、局地的な気象変化を捉えるためのレーダー観測を用いた実験が考えられる。