

グラウンド識別問題と人検出問題におけるサポートベクターマシンと畳み込みニューラルネットワークの性能検証

○和田壮史 片田喜章(摂南大学)

概要 機械学習による画像識別において、問題によって有効な手法や必要なデータ数について検証するため、SVMとCNNを2種の識別問題に適用して学習実験を行った。その結果、データ数の多い方が過学習が抑制される傾向にあることがわかった。また、問題ごとに適切なデータ数が存在すること、CNNは問題に依らない頑健性の高い手法であること、SVMは画像内の識別対象の位置や色の変化が少ない場合に有効な手法であることを示唆する結果が得られた。

キーワード: 画像識別, サポートベクターマシン, 畳み込みニューラルネットワーク

1 研究背景および目的

画像識別の技術は深層学習の発展により進歩し続けているが、計算コストの高さなどの問題から深層学習以外の機械学習手法にも需要はある。また、学習に使用するデータの量や質がどの程度求められるかについては人の知識が必要である。本研究では、機械学習による画像識別において、識別問題によって有効な手法や必要なデータ数について検証することを目的とする。このために、グラウンド識別問題と人検出問題の2つの識別問題に対し、SVMとCNNの手法を用いて学習実験を行い、その識別性能を比較する。

2 機械学習手法

2.1 Support Vector Machine

Support Vector Machine (SVM)¹⁾ は、 n 次元の特徴量を持つ2クラスの学習データに対し、 $n-1$ 次元で境界を定めて識別器を得る手法である。SVMには線形な境界を定める線形SVMと非線形な境界を定める非線形SVMがある。SVMの識別関数を(1)式に、目的関数と制約条件を(2)式に示す。

$$f(x) = \text{sign} \left(\sum_{i=1}^N \lambda_i y_i K(x_i, x_j) + b \right) \quad (1)$$

$$\max_{\lambda} \sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i,j=1}^N \lambda_i \lambda_j y_i y_j K(x_i, x_j) \quad (2)$$

$$\text{s. t. } \sum_{i=1}^N \lambda_i y_i = 0, 0 \leq \lambda_i \leq C$$

ここで、 N は学習データ数、 x_i は学習データ i 番目の特徴量、 x_j は x_i を非線形空間に投影したもの、 y_i は学習データ i 番目の正解ラベルである。 C は誤識別をどの程度許容するかのパラメータであり、大きくなるほど誤識別を許容しなくなる。また、 $K(x_i, x_j)$ はカーネル関数であり、SVMの境界を定めるために用いる。本研究では、線形SVMのカーネル関数として線形カーネル((3)式)を用い、非線形SVMのカーネル関数としてRBFカーネル((4)式)を用いた。

$$K(x_i, x_j) = x_i^T x_j \quad (3)$$

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (4)$$

ここで、 γ はRBFカーネルが持つパラメータである。

C や γ は設定値によって識別性能に大きな影響を及ぼすため、本研究ではこれらのパラメータ調整を行い、最も良い結果を示したパラメータで識別性能の比較を行う。

2.2 Convolutional Neural Network

Neural Network (NN)²⁾ をFig. 1に示す。NNは人の神経細胞を模した人工ニューロン(Fig. 1内の丸の1つ1つ)を複数結合したものである。ニューロンはFig. 1の矢印で示す接続間に重みを持っており、重みを更新することによって学習を進め識別精度を高めていく。出力層のニューロンの数はクラス数と同数になっており、1つのニューロンが1つのクラスに対応している。学習には、入力層側から順方向にデータを入力したあと、出力結果を利用して出力層側から順に重みを更新する誤差逆伝搬法が多く用いられている。

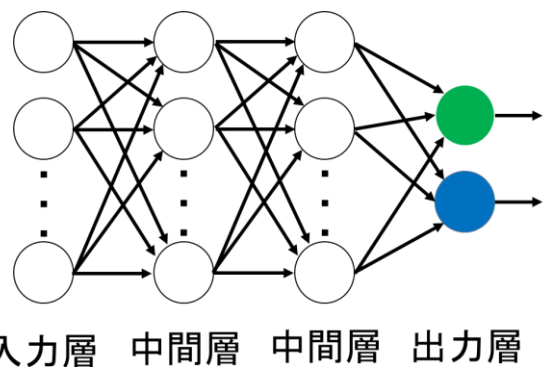


Fig. 1: Neural Network

Convolutional Neural Network (CNN)³⁾ は、このNNの中間層を畳み込み層やプーリング層と呼ばれる層にしたもので、深層学習の一種である。畳み込み層では、重みを持つ小さなフィルタを画像全体に渡ってかけることで出力を得る。フィルタは画像内にフィルタと一致する特徴がどの程度現れているかを抽出することを可能とし、学習により妥当な特徴を抽出できるように変化していく。プーリング層では画像の重要な特徴だけを抜き出し、画像の圧縮を行う。プーリング層は識

別対象の微小な位置のズレに対して頑健になる、計算コストを下げるといった効果を持っている。CNNはこのような構造から位置不変性を獲得しており、画像認識によく用いられる手法である。

3 識別問題

本研究では、グラウンド識別問題と人検出問題の2種類の識別問題を扱う。摂南大学寝屋川キャンパス第1グラウンドの全体図をFig. 2に示す。このグラウンドは球技を行うフィールド部分と、その周囲を囲う青色のトラック部分に大きく分けられる。グラウンド識別問題では、このグラウンドのフィールド画像とトラック画像の識別を行う。Fig. 3にこれらの画像の例を示す。本識別問題は、機械学習により得られた識別器をクローラ型移動ロボットに搭載し、トラックであれば回避を行いフィールドへ戻る、という動作を実現することを目的として設定されている。



Fig. 2: Artificial turf (sports ground) at Setsunan University



(a) Field image (b) Track image

Fig. 3: Field detection problem

人検出問題では、人検出研究分野のベンチマークであるINRIA Person Dataset⁴⁾から人が映り込んでいる画像と映り込んでいない画像の識別を行う。Fig. 4に画像の例を示す。



(a) image with human object (b) image with non-human object

Fig. 4: Human detection problem

4 実験設定

4.1 実験用データセット

識別器の入力特徴量として画像のRGB値を用いた。画像を32px×32pxに縮小した後、画像をランダムに抜き出してデータセットL {学習データ:3500枚, テストデータ500枚}およびデータセットS {学習データ:1000枚, テストデータ500枚}をそれぞれ20セットずつ作成した。SVMとCNNの各々に関して、学習を1セットずつ行い、学習後の20セットの識別率の平均を求めた。

4.2 SVM の設定

線形SVMと非線形SVMにより学習を行った。2.1節で述べたSVMのパラメータはC={0.001, 0.01, 0.1, 1, 10, 100}, γ ={0.0001, 0.001, 0.01, 0.1}とし、各組合せに関して実験を行った。

4.3 CNN の設定

実験に用いた5層CNNをFig. 5に示す。パラメータには活性化関数として出力層にソフトマックス関数、畳み込み層とプーリング層にReLU関数、損失関数として交差エントロピー誤差、重みの更新手法としてAdam⁵⁾を用い、バッチサイズを50、学習回数を50エポックとした。また、各プーリング層の出力にDropout⁶⁾を50%適用した。最終エポック時の識別率を実験結果とする。

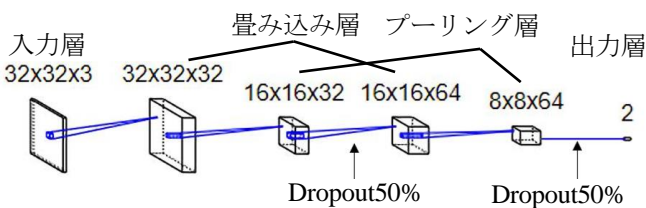


Fig. 5: CNN

5 実験結果

グラウンド識別問題における実験結果をTable 1~Table 3に示し、人検出問題における実験結果をTable 4~Table 6に示す。SVMについては、テストデータに対する識別率が最も高かったものを灰色で示している。

データの数による違いを見ると、全体の傾向として、データ数の多い方が過学習が抑制されていることがわかる。ここで、この差が有意なものかを確かめるために検定を行った。各識別問題の各手法において、データセットLとSの20セットずつに対するテストデータ識別率の間に有意差があるかを検定対象とした。SVMの場合は最も良い結果が得られたパラメータにおけるテストデータ識別率を用い、最良のパラメータが複数ある場合はそれら全ての組合せに対して検定を行った。検定手法にはWelchのt検定を用い、有意水準は1%とし、両側検定を行った。検定結果をTable 7に示す。なお、有意差があるものを灰色で示している。この結果より、グラウンド識別問題の非線形SVMとCNNにおいて、学習データが1000枚と3500枚の場合の識別率には有意差があるとはいえず、それ以外の場合には有意差があることがわかった。

Table 1: Performance of the Linear SVM on field detection problem (training set / test set) [%]

Dataset L						
C	0.001	0.01	0.1	1	10	100
Detection rate	97.80/97.73	99.25/98.59	10.00/99.12	10.00/99.08	10.00/99.08	10.00/99.08
Dataset S						
C	0.001	0.01	0.1	1	10	100
Detection rate	97.23/96.99	99.10/98.11	100/98.42	100/98.42	100/98.42	100/98.42

Table 2: Performance of the Non-linear SVM on field detection problem (training set / test set) [%]

Dataset L						
$\gamma \backslash C$	0.001	0.01	0.1	1	10	100
0.0001	67.31/67.63	67.31/67.63	65.45/65.55	96.76/96.70	98.31/98.03	99.71/99.24
0.001	65.74/65.94	64.27/64.36	96.68/96.74	98.84/98.60	99.93/99.63	100/99.65
0.01	75.05/75.09	93.65/93.65	99.85/99.77	99.97/99.85	100/99.85	100/99.85
0.1	73.52/73.52	73.48/73.40	83.59/81.81	100/94.73	100/95.16	100/95.16
Dataset S						
$\gamma \backslash C$	0.001	0.01	0.1	1	10	100
0.0001	64.28/64.73	64.28/64.73	64.28/64.73	80.55/80.44	97.75/97.56	99.62/98.49
0.001	64.23/64.59	64.23/64.59	78.12/77.97	98.11/97.80	99.86/99.03	100.0/99.28
0.01	75.25/74.88	75.25/74.88	99.23/98.63	99.96/99.74	100/99.78	100/99.78
0.1	75.97/74.46	75.97/74.46	79.87/76.34	100.0/91.02	100/92.06	100/92.06

Table 3: Performance of the CNN on field detection problem (training set / test set) [%]

Dataset	Dataset L	Dataset S
Detection rate	99.83/99.78	99.84/99.62

Table 4: Performance of the Linear SVM on human detection problem (training set / test set) [%]

Dataset L						
C	0.001	0.01	0.1	1	10	100
Detection rate	88.80/87.84	94.51/90.29	98.50/89.53	100/87.41	100/87.42	100/87.42
Dataset S						
C	0.001	0.01	0.1	1	10	100
Detection rate	87.92/85.63	95.24/88.41	99.89/87.10	100/87.06	100/87.06	100/87.06

Table 5: Performance of the Non-linear SVM on human detection problem (training set / test set) [%]

Dataset L						
$\gamma \backslash C$	0.001	0.01	0.1	1	10	100
0.0001	76.84/76.96	76.84/76.96	80.61/80.85	85.56/85.41	91.26/89.67	96.53/91.27
0.001	77.45/77.48	79.57/79.64	86.00/85.72	91.44/89.77	98.03/92.16	100/91.58
0.01	80.71/79.78	80.84/79.83	87.58/85.59	99.03/91.52	100/92.70	100/92.70
0.1	100/51.09	100/51.09	100/51.09	100/58.55	100/60.32	100/60.32
Dataset S						
$\gamma \backslash C$	0.001	0.01	0.1	1	10	100
0.0001	77.37/77.25	77.37/77.25	77.14/76.86	84.05/83.21	90.19/86.99	97.71/89.06
0.001	77.68/77.24	77.68/77.24	84.15/82.77	90.39/86.97	98.73/89.40	100/89.14
0.01	81.78/78.94	81.78/78.94	86.51/82.15	99.07/87.87	100/88.95	100/88.95
0.1	100/51.17	100/51.17	100/51.17	100/57.20	100/58.27	100/58.27

Table 6: Performance of the CNN on human detection problem (training set / test set) [%]

Dataset	Dataset L	Dataset S
Detection rate	98.92/98.07	98.18/96.00

つぎに手法による違いを見ると、グラウンド識別問題では非線形SVM, CNN, 線形SVMの順に高い識別率が得られ、人検出問題ではCNN, 非線形SVM, 線形SVMの順に高い識別率が得られた。この場合も、各識別問題の各データセットに対して、3手法により得られた識別率の間に有意差があるかを多重比較により検定した。検定手法にはHolm法(2群間の検定にはWelchのt検定)を用い、有意水準は1%とし、両側検定を行った。検定結果をTable 8に示す。Table 7と同様に、有意差があるものを灰色で示している。この結果より、グラウンド識別問題のデータセットLとSにおいて、非線形SVMとCNNの識別率に有意差があるとはいえず、それ以外の場合は有意差があることがわかった。

6 考察

データ数による差を検証したとき、グラウンド識別問題において、非線形SVMとCNNでは学習データ数を1000枚から3500枚に増やしても識別率に有意な差があるとはいえなかった。また、人検出問題ではどの場合もデータ数を増やすと識別率に有意な差が得られた。このことから、識別問題によって適切なデータ数が存在する可能性があると考えられる。

手法による差を検証したとき、グラウンド識別問題では非線形SVM, CNNで高い識別率が得られ、人検出問題ではCNNで高い識別率が得られた。このことから、CNNは問題に依らない頑健性の高い手法である可能性があると考えられる。また、2つの識別問題の相違点を考えると、人検出問題の画像はグラウンド識別問題の画像と比べ、識別対象の画像内における位置や色の変化が大きいことが挙げられる。このことが、非線形SVMを用いたときに、グラウンド識別問題ではCNNと同程度の識別率が得られ、人検出問題ではCNNより低い識別率となった要因ではないかと推測する。

7 まとめ

本研究の実験結果により、データ数の多い方が過学習が抑制される傾向にあることがわかった。また、識別問題ごとに適切なデータ数が存在する可能性があること、CNNは問題に依らない頑健性の高い手法である可能性があることや、SVMは画像内の識別対象の位置や色の変化が少ない場合に高い識別率が得られる可能性があることがわかった。

参考文献

- 1) B. Boser, I. Guyon, V. Vapnik: A training algorithm for optimal margin classifiers, Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 144/152, ACM, (1992)
- 2) G. Hinton: Learning multiple layers of representation, Trends in Cognitive Sciences, Vol. 11, No. 10, 428/434 (2007)
- 3) A. Krizhevsky, I. Sutskever, G. Hinton: ImageNet classification with deep convolutional neural networks, Neural Information Processing Systems, NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, Vol. 1, 1097/1105(2012)
- 4) <http://pascal.inrialpes.fr/data/human/>
- 5) D. Kingma, J. Ba, Adam: A method for stochastic optimization, International Conference for Learning Representations, San Diego, CA, USA (2015)
- 6) N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov: Dropout: A simple way to prevent neural networks from overfitting, The Journal of Machine Learning Research, Vol. 15, 1929/1958 (2014)

Table 7: Result of Welch's t-test

Detection problem	Method	Parameters (Dataset L)	Parameters (Dataset S)	p-value
Field detection problem	Linear SVM	C=0.1	C=0.1	1.08×10^{-4}
		C=0.1	C=1	1.33×10^{-4}
		C=0.1	C=10	1.33×10^{-4}
		C=0.1	C=100	1.33×10^{-4}
	Non-linear SVM	C=1, $\gamma=0.01$	C=10, $\gamma=0.01$	2.32×10^{-1}
		C=1, $\gamma=0.01$	C=100, $\gamma=0.01$	2.32×10^{-1}
		C=10, $\gamma=0.01$	C=10, $\gamma=0.01$	2.46×10^{-1}
		C=10, $\gamma=0.01$	C=100, $\gamma=0.01$	2.46×10^{-1}
		C=100, $\gamma=0.01$	C=10, $\gamma=0.01$	2.46×10^{-1}
		C=100, $\gamma=0.01$	C=100, $\gamma=0.01$	2.46×10^{-1}
CNN			1.69×10^{-1}	
Human detection problem	Linear SVM	C=0.01	C=0.01	4.79×10^{-6}
	Non-linear SVM	C=10, $\gamma=0.01$	C=10, $\gamma=0.001$	4.18×10^{-11}
		C=100, $\gamma=0.01$	C=10, $\gamma=0.001$	4.18×10^{-11}
	CNN			2.13×10^{-7}

Table 8: Result of Holm's sequentially rejective Bonferroni test

Detection problem	Dataset	Method	A parameter of Linear SVM	Parameters of Non-linear SVM	Correction p-value	
Field detection problem	L	Linear SVM	C=0.1	C=1, $\gamma=0.01$	5.84×10^{-7}	
			Non-linear SVM	C=0.1	C=10, $\gamma=0.01$	6.26×10^{-7}
				C=0.1	C=100, $\gamma=0.01$	6.26×10^{-7}
		Linear SVM	C=0.1		8.06×10^{-5}	
			Non-linear SVM		C=1, $\gamma=0.01$	2.41×10^{-2}
					C=10, $\gamma=0.01$	2.62×10^{-2}
				C=100, $\gamma=0.01$	2.62×10^{-2}	
		S	Linear SVM	C=0.1	C=10, $\gamma=0.01$	1.64×10^{-9}
				C=0.1	C=100, $\gamma=0.01$	1.64×10^{-9}
				C=1	C=10, $\gamma=0.01$	3.83×10^{-9}
	C=1			C=100, $\gamma=0.01$	3.83×10^{-9}	
	C=10			C=10, $\gamma=0.01$	3.83×10^{-9}	
	C=10			C=100, $\gamma=0.01$	3.83×10^{-9}	
	C=100			C=10, $\gamma=0.01$	3.83×10^{-9}	
	C=100			C=100, $\gamma=0.01$	3.83×10^{-9}	
	Linear SVM		C=0.1		1.29×10^{-7}	
			C=1		1.79×10^{-7}	
	CNN	C=10		1.79×10^{-7}		
		C=100		1.79×10^{-7}		
	Non-linear SVM		C=10, $\gamma=0.01$	1.40×10^{-2}		
		C=100, $\gamma=0.01$	1.40×10^{-2}			
Human detection problem	L	Linear SVM	C=0.01	C=10, $\gamma=0.01$	2.66×10^{-8}	
		Non-linear SVM	C=0.01	C=100, $\gamma=0.01$	2.66×10^{-8}	
		Linear SVM	C=0.01		5.23×10^{-22}	
			Non-linear SVM		C=10, $\gamma=0.01$	7.92×10^{-21}
		CNN		C=100, $\gamma=0.01$	7.92×10^{-21}	
	S	Linear SVM	C=0.01	C=10, $\gamma=0.001$	9.52×10^{-3}	
		Linear SVM	C=0.01		2.50×10^{-22}	
			Non-linear SVM		C=10, $\gamma=0.001$	3.92×10^{-19}
		CNN				