

様式（2a） 学修総まとめ科目 成果の要旨

学校名	久留米工業高等専門学校	専攻名	機械・電気システム工学専攻
専攻分野名称	工学	専攻の区分	情報工学
氏 名	平川智也	学籍番号	A3519
テーマ名	畳み込み辞書学習を用いたクラス分類の検討		
指導教員名	黒木祥光	指導補助教員名	

（1）背景

近年、畳み込みニューラルネットワーク（CNN）は、パターン・画像認識分野の様々なタスクにおいて強力な手法として知られている。CNNを画像分類問題に適用する場合、画像とラベルデータを用いて誤差逆伝搬法によってネットワークを学習する。これに対し、畳み込み辞書学習と呼ばれる手法が提案されている[1]。畳み込み辞書学習はフィルタと係数マップへの信号分離の最適化問題として定式化され、ラベルデータを用いずに画像のみからデータ駆動的に特徴を抽出する。畳み込み辞書学習は最適化ベースの特徴抽出法であるため、制約や目的関数の変更が柔軟にできる。本研究では畳み込み辞書学習問題における係数マップを特徴として、分類問題に適用する。

（2）目的

本研究では、係数マップのパワースペクトルを使用することにより、データセット内の各データ間に存在する位置ずれに対する頑健性を高める。CNNでは、新たなクラスが加わった際に、はじめから再学習する必要があるのに対して、部分空間法では、新たなクラスの張る部分空間を設計すればよいのでコストが低いことなどを考慮し、本研究では部分空間法を基にした分類器を適用する。具体的には、パワースペクトルは非負値をとるため、特徴群は錐の形状をなすと仮定し、部分空間法に錐制約を課した錐制約部分空間法を適用する。

（3）手法・手段

畳み込み辞書学習では、与えられたサイズが $H \times W$ の入力画像 $\mathbf{y} \in \mathbb{R}^{H \times W}$ を、サイズが $h \times w$ の K 個のフィルタ $\mathbf{d}_i \in \mathbb{R}^{h \times w}$ ($i = 1, \dots, K$)と、入力画像と同じサイズの K 個の疎な係数マップ $\mathbf{x}_i \in \mathbb{R}^{H \times W}$ ($i = 1, \dots, K$)の畳み込み和で表現する問題であり、以下の最適化問題で記述される。

$$D_{\text{op}}, X_{\text{op}} = \arg \min_{D, X} \frac{1}{2} \left\| \mathbf{y} - \sum_{i=1}^K \mathbf{d}_i * \mathbf{x}_i \right\|_2^2 + \lambda \sum_{i=1}^K \|\mathbf{x}_i\|_1 \quad \text{s.t. } \mathbf{d}_i \in C_d, \mathbf{x}_i \in C_x$$

ここで、 $D = [\mathbf{d}_1, \dots, \mathbf{d}_K] \in \mathbb{R}^{h \times w \times K}$, $X = [\mathbf{x}_1, \dots, \mathbf{x}_K] \in \mathbb{R}^{H \times W \times K}$ であり、 $*$ は巡回畳み込みを示す。パラメータ λ はデータの再現精度と係数マップの疎性を調節するパラメータである。 C_d, C_x はそれぞれ、フィルタと係数マップに対する制約集合である。上記の問題は、2変数の最適化問題であるため、各変数に関する2つの凸最適化問題に分割され、両問題ともに凸計画問題の反復解法である交互方向乗数法（Alternating Direction Method of Multipliers：ADMM）によって解かれ、各問題を交互に繰り返し解くことで所望の解を得る[2]。

特徴ベクトル空間でサンプルベクトルが張る凸錐は $C := \{\mathbf{x} \in \mathbb{R}^d | \mathbf{x} = \sum_{i=1}^N c_i \boldsymbol{\xi}_i\}$ で定義される。ここで N は錐の基底ベクトル $\boldsymbol{\xi}_i \in \mathbb{R}^d$ の数、 c_i は非負の結合係数である。錐は部分空間における結合係数に非負制約を課したものである。錐制約部分空間法では、入力ベクトルと各クラスの錐とのなす角度によって分類を行う。角度 θ は入力特徴ベクトル \mathbf{y} とその錐 C への正射影ベクトルのなす角度で定義される。

$$\theta = \arcsin \left\{ \min_{\mathbf{x} \in C} \frac{\|\mathbf{y} - \mathbf{x}\|}{\|\mathbf{y}\|} \right\} = \arcsin \left(\frac{1}{\|\mathbf{y}\|} \sqrt{\min_{c_i \geq 0} \left\| \mathbf{y} - \sum_{i=1}^N c_i \boldsymbol{\xi}_i \right\|^2} \right)$$

ここで $0 \leq \theta \leq \pi/2$ である。これは非負値最小二乗法を適用することで得られる。錐の構成方法として様々な方法が提案されている[3]が、本研究では非負値行列因子分解（NMF）による手法と包括凸錐を構成する方法を採用する。

(4) 内容

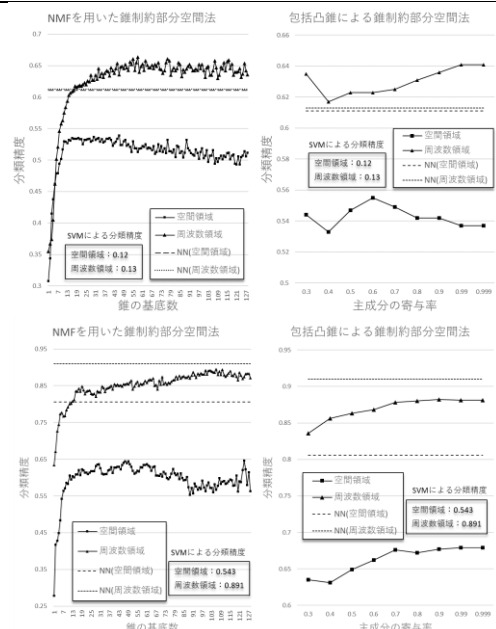
実験には 10 クラスの手書き数字文字画像からなるデータセットである MNIST を使用した。データセットは 32×32 ピクセルの 8 ビットグレースケール画像からなる。本実験では、データセットのうち 1000 枚をテスト用データに割り当てた。学習用データの画像枚数は 100, 200, 300, 400, 500, 1000 と変化させた。本研究では、パワースペクトルの使用が分類精度にどう影響するか、また錐制約部分空間法の分類問題への有効性、に焦点を当てる。したがって、分類に使用する特徴は、畳み込み辞書学習における係数マップおよび畳み込み辞書学習における係数マップのパワースペクトルの 2 種類で比較する。また、分類器については、NMF で構成された錐での錐制約部分空間法、包括凸錐による錐制約部分空間法に加えて、3 層ニューラルネットワーク、Support Vector Machine (SVM) を分類器として使用した。

以下に実験の手順を示す。まず、学習データについて畳み込み辞書学習を行い、フィルタを設計及び、学習データの係数マップを得る。本実験では 6 枚の 5×5 ピクセルのフィルタを学習させた。ここで、係数マップのパワースペクトルを特徴として使用する場合は、パワースペクトルを算出後にベクトル化する（係数マップを特徴として用いる場合はそのままベクトル化する）。これらを NMF により 256 次元まで次元削減し、それらを特徴ベクトルとする。各クラスの特徴ベクトルを用いて、錐制約部分空間法及び 3 層ニューラルネットワーク、SVM を学習させる。次に、テストデータについて係数を算出する。具体的には係数推定問題を解くのだが、フィルタは学習データで行った畳み込み辞書学習で得られたものを使用する。また、学習データについてと同様の操作で 256 次元の特徴ベクトルを得る。最後に、学習済の各分類器で分類を行い分類精度を算出する。なお、畳み込み辞書学習の計算には Python と、Python 用のライブラリ SPORCO[4] を用いた。

(5) 得られた結果と考察・将来展望

実験結果を右に示す。上図は学習画像が 200 枚、下図は学習画像が 1000 枚での実験結果である。結果からどの分類器に対してもパワースペクトルの使用が精度の向上に寄与していると考えられる。また、比較的小規模な学習用画像に対しては錐制約部分空間法による手法が他の手法に対して有効であることがわかる。学習用画像の枚数が増えてくるにつれて 3 層ニューラルネットワークでの分類精度は上がっていることから、十分に学習用画像が得られる状況下ではニューラルネットワークが有効になると考えられる。

今後の展望として、大規模データセットへの適用、多層の畳み込み辞書学習による特徴量に対してのパワースペクトルの適用が挙げられる。



[1] Bristow, Hilton, Anders Eriksson, and Simon Lucey. "Fast convolutional sparse coding," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013.

[2] Bristow, Hilton, and Simon Lucey. "Optimization methods for convolutional sparse Coding," arXiv preprint arXiv:1406.2407, 2014.

[3] 小林匠, and 大津展之. "パターン識別のための錐制約部分空間法." 電子情報通信学会論文誌 D 92.1, 104-111, 2009

[4] Wohlberg, Brendt. "SPORCO: A Python package for standard and convolutional sparse representations." Proceedings of the 15th Python in Science Conference,