

＜令和元年度専攻科学研究論文＞

畳み込み辞書学習を用いた

クラス分類の検討

指導教官: 黒木祥光 教授

機械・電気システム工学専攻 A3519 平川智也

目次

1	はじめに	2
2	畳み込み辞書学習	3
2.1	辞書学習問題	3
2.2	畳み込み辞書学習問題	4
3	錐制約部分空間法	6
3.1	錐の定義	6
3.2	錐への角度	6
3.3	錐の構成方法	7
3.3.1	NMF による錐の構成	7
3.3.2	包括凸錐の構成	7
4	提案法	9
5	実験	11
5.1	実験	11
5.2	実験結果	12
6	むすび	17
	謝辞	18
	参考文献	19

第1章

はじめに

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、単に特徴抽出だけでなく、画像の超解像度の研究 [2] などに応用されている。

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。その際に、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。

第2章

畳み込み辞書学習

2.1 辞書学習問題

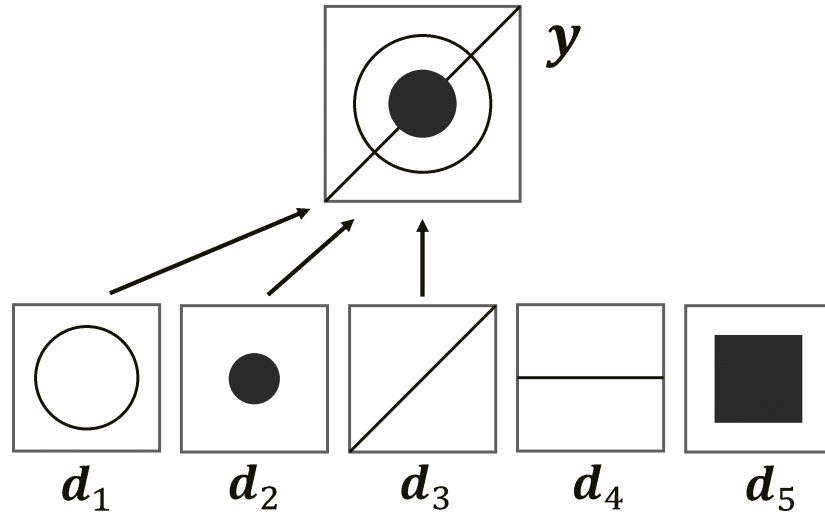


図 2.1: 辞書学習のイメージ

辞書学習問題では、与えられた d 次元の信号 $\mathbf{y} \in \mathbb{R}^d$ を、 K 個の辞書基底ベクトル $\mathbf{d}_i \in \mathbb{R}^d$ ($i = 1, \dots, K$) と、疎な係数 $\mathbf{x}_j \in \mathbb{R}^K$ の線形結合で表現する問題であり、以下の最適化問題で記述される。

$$D_{\text{opt}}, \mathbf{x}_{\text{opt}} = \arg \min_{D, \mathbf{x}} \frac{1}{2} \|\mathbf{y} - D\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{d}_i\|_2 \leq 1 \quad (2.1)$$

ここで、 $D = [\mathbf{d}_1, \dots, \mathbf{d}_K]$ であり、パラメータ λ はデータの再現精度と係数の疎性を調節するパラメータである。問題 (2.1) は、2 変数の最適化問題であるが一般に D の最

適化問題（辞書更新）、 \mathbf{x} の最適化問題（係数推定）に分割され、それらを繰り返し交互に解くことで解を得る．画像処理における辞書学習問題の応用では，入力画像を小領域のパッチに分割して処理することが多い．そのため，パッチ分割によるブロックノイズやエイリアジングが生じることがある．

2.2 畳み込み辞書学習問題

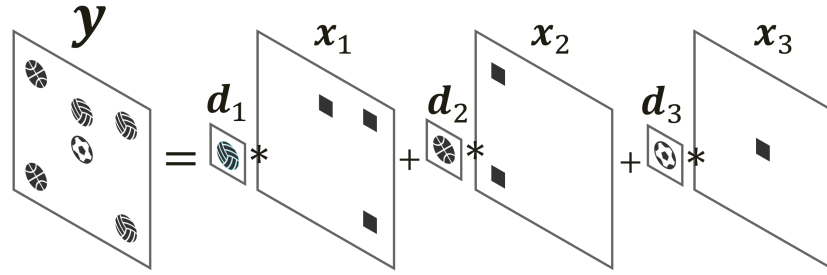


図 2.2: 畳み込み辞書学習のイメージ

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．

$$D_{\text{opt}}, X_{\text{opt}} = \arg \min_{\mathbf{d}_i, \mathbf{x}_i} \frac{1}{2} \left\| \mathbf{y} - \sum_{i=1}^K \mathbf{d}_i * \mathbf{x}_i \right\|_2^2 + \lambda \sum_{i=1}^K \|\mathbf{x}_i\|_1 \quad \text{s.t.} \quad \mathbf{d}_i \in C_d, \mathbf{x}_i \in C_x \quad (2.2)$$

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， $D_{\text{opt}}, X_{\text{opt}}$ は問題の最適値であることを示す．問題 (2.2) は，2 変数の最適化問題であるため，以下に示す各変数についての最適化問題に分割する．

$$D_{\text{opt}} = \arg \min_{\mathbf{d}_i} \frac{1}{2} \left\| \mathbf{y} - \sum_{i=1}^K \mathbf{d}_i * \mathbf{x}_i \right\|_2^2 \quad \text{s.t. } \mathbf{d}_i \in C_d \quad (2.3)$$

$$X_{\text{opt}} = \arg \min_{\mathbf{x}_i} \frac{1}{2} \left\| \mathbf{y} - \sum_{i=1}^K \mathbf{d}_i * \mathbf{x}_i \right\|_2^2 + \lambda \sum_{i=1}^K \|\mathbf{x}_i\|_1 \quad \text{s.t. } \mathbf{x}_i \in C_x \quad (2.4)$$

問題 (2.3) (2.4) を，辞書更新および係数推定の問題であり，どちらも凸最適化問題で定式化される．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）[7] 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．

第3章

錐制約部分空間法

3.1 錐の定義

特徴ベクトル空間でサンプルベクトルが張る凸錐は以下で定義される.

$$C := \left\{ \mathbf{x} \mid \mathbf{x} = \sum_{i=1}^N c_i \boldsymbol{\xi}_i = \Xi \mathbf{c}, \mathbf{c} \geq \mathbf{0} \right\} \quad (3.1)$$

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, $\Xi = [\boldsymbol{\xi}_1 \ \boldsymbol{\xi}_2 \ \dots \ \boldsymbol{\xi}_N]$, $\mathbf{c} = {}^t(c_1, c_2, \dots, c_N)$ 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り, 部分空間と同様にスケール変化・ベクトルの加法に対して閉じた表現となっている.

3.2 錐への角度

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.

$$\theta = \arcsin \left(\min_{\mathbf{x} \in C} \frac{\|\mathbf{y} - \mathbf{x}\|}{\|\mathbf{y}\|} \right) = \arcsin \left(\frac{1}{\|\mathbf{y}\|} \sqrt{\min_{\mathbf{c} \geq \mathbf{0}} \|\mathbf{y} - \Xi \mathbf{c}\|^2} \right) \quad (3.2)$$

ここで $0 \leq \theta \leq \frac{\pi}{2}$ 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.

3.3 錐の構成方法

凸錐の構成方法については様々な手法が提案されている [5] が, 本研究では非負値行列因子分解 (Nonnegative Matrix Factorization: NMF) [4] による手法と, 特徴ベクトルを包括する凸錐を構成する手法を採用した. 以下に概略を示す.

3.3.1 NMF による錐の構成

NMF は $n \times m$ の非負値行列 V を $n \times r$ の非負値行列 W と $r \times m$ の非負値行列 H の行列積で表現する手法である. $V = (v_{ij}), W = (w_{ij}), H = (h_{ij})$ とし, V, W, H の j 列目の列ベクトルをそれぞれ $\mathbf{v}_j = {}^t(v_{1j}, \dots, v_{nj}), \mathbf{w}_j = {}^t(w_{1j}, \dots, w_{rj}), \mathbf{h}_j = {}^t(h_{1j}, \dots, h_{rj})$ と表すと, \mathbf{v}_j については

$$\mathbf{v}_j \approx W\mathbf{h}_j = h_{1j}\mathbf{w}_1 + h_{2j}\mathbf{w}_2 + \dots + h_{rj}\mathbf{w}_r \quad (3.3)$$

と表され, V の列ベクトル \mathbf{v}_j を W の r 個の列ベクトルの線形結合で近似しており, その結合係数が \mathbf{h}_j となる. このため, 分解後の行列 W には錐の基底ベクトルが格納される.

3.3.2 包括凸錐の構成

錐のモデルとして [5] は厳密な凸錐, 包括凸錐, 円錐を提案しているが, 本研究では近似精度と計算量のバランスが良い包括凸錐を扱う. 包括凸錐は特徴ベクトル集合のなす凸錐を少数の基底ベクトルからなる包括的な凸錐で近似する手法である. 先ず, 特徴ベクトル $\mathbf{x}_i \in \mathbb{R}^d$ を単位超球面上に射影する.

$$\mathbf{z}_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|} \quad (3.4)$$

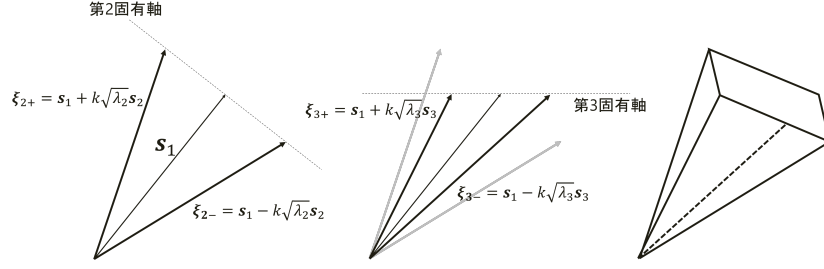


図 3.1: 包括凸錐の構成方法（3次元空間での例）

次に z_i の自己相関行列に主成分分析（PCA）を適用する．これにより得られた固有ベクトルを固有値の大きさに従い降順に並べたとき，第1固有ベクトルは原点からの凸錐の方向（中心），それに直交する第2以降の固有ベクトルは超球面上の特徴ベクトル集合の分布の広がりとなる．この分布を包括する凸包は第2以降の各固有軸上で分布を包括する（分布の端点などの）2点 (x_L, x_R) を選ぶことによって構成される．つまり，各固有軸 $(i \geq 2)$ に対して次の2つの錐の基底ベクトルが定まる．

$$\xi_{i-} = s_1 + x_L^i s_i = s_1 - k\sqrt{\lambda_i} s_i \quad (3.5)$$

$$\xi_{i+} = s_1 + x_R^i s_i = s_1 + k\sqrt{\lambda_i} s_i \quad (3.6)$$

ここで， s_i, λ_i はそれぞれ第 i 固有ベクトル，固有値となり， k はスケーリングのパラメータである．第1固有ベクトルが錐の中心ベクトルとなることから，第2以降の固有軸上での分布は原点が中心となり，その分散は固有値 λ_i で与えられる．ここでは外れ値の影響を考慮し，各軸上で標準偏差の k 倍の点を選んでいる．

第4章

提案法

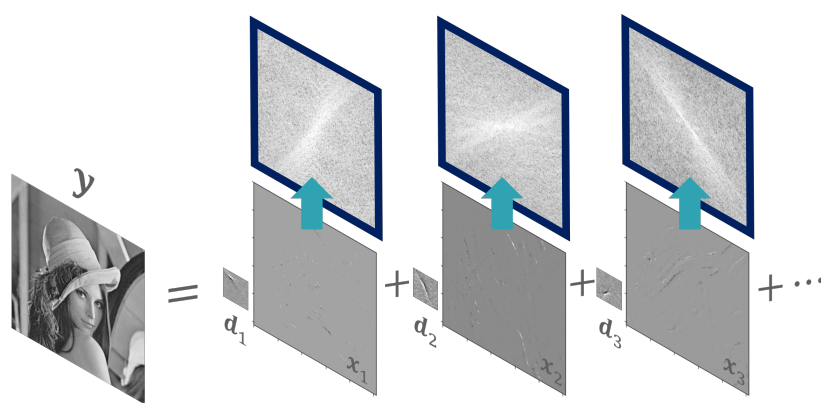


図 4.1: 提案法で使用する特徴（図中の太枠部分）

畳み込み辞書学習におけるフィルタは、エッジやテクスチャなどの局所的な特徴をとらえるのに対し、係数マップはフィルタの示す特徴が入力信号のどの部分に顕在しているのかを示す。したがって、入力信号がシフト変動を受けるとき、係数マップも近しい変動を受けると考えられる。そこで本研究では、疎な係数マップのパワースペクトル（図 4.2）を特徴として用いることで、シフト変動に頑健な特徴を設計する。パワースペクトルはその信号に含まれる各周波数成分の振幅を示すものである。これにより、データセット内の各データ間の位置ずれに対して頑健な特徴抽出が可能となる。

また、本研究ではパワースペクトルの非負値性を利用し、部分空間法ではなく、錐制約を課した部分空間法を分類問題に適用する。

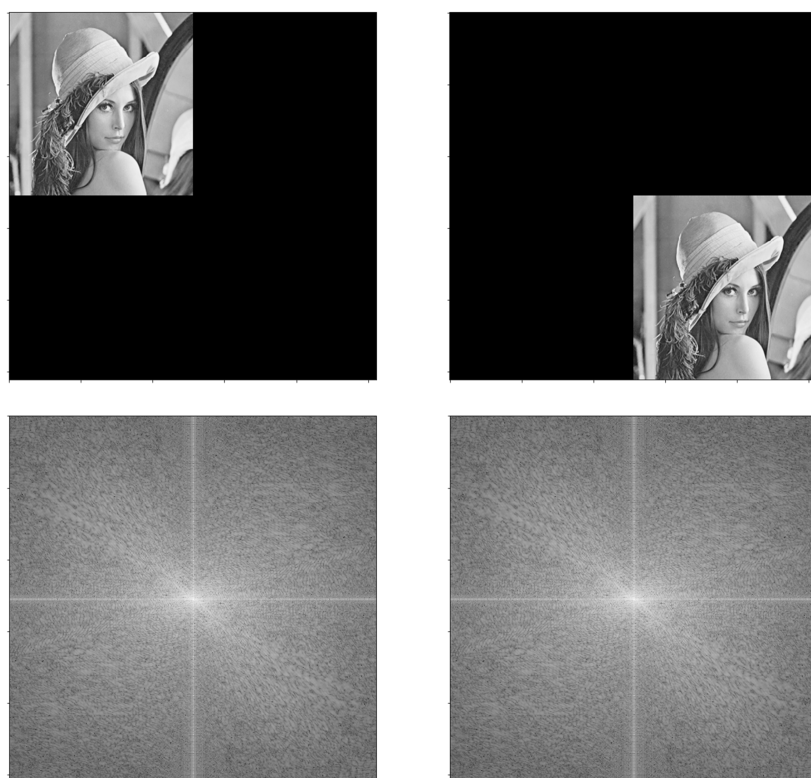


図 4.2: シフト変動させた信号に対するパワースペクトル

第5章

実験

5.1 実験

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は他の分類器と比較してどのような特徴を持つのか、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

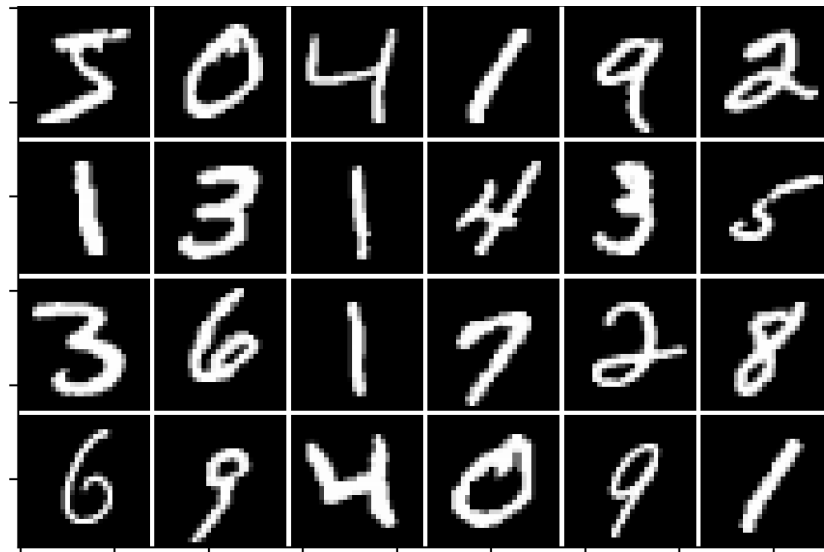


図 5.1: MNIST データセットの一部

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．具体的には問題 (2.4) 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，学習させた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[8] を用いた．

5.2 実験結果

各学習用画像の枚数に対する実験結果を図 5.2 から図 5.7 に示す．結果より，どの分類器を用いた実験においても，係数のパワースペクトルを特徴として使用すると認識精度が向上することが分かる．これは画像中の位置ずれを，周波数の使用により吸収

したためと考えられる。また、学習枚数が比較的少ない100枚、200枚、300枚の実験では錐制約部分空間法を用いた場合が最も良い精度であり、データの数が大きくなるにつれてニューラルネットやSVMを用いた方が精度が良くなることから、錐制約部分空間法は学習データが少ない状況下で有効な分類器といえる。

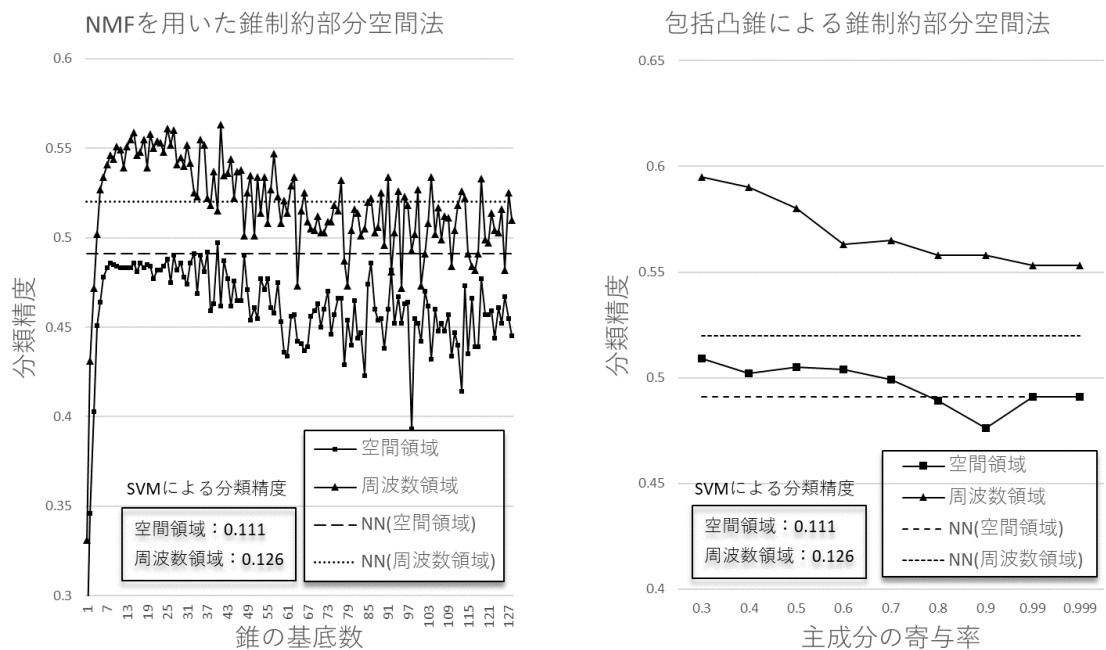


図 5.2: 学習枚数 100 枚における分類精度

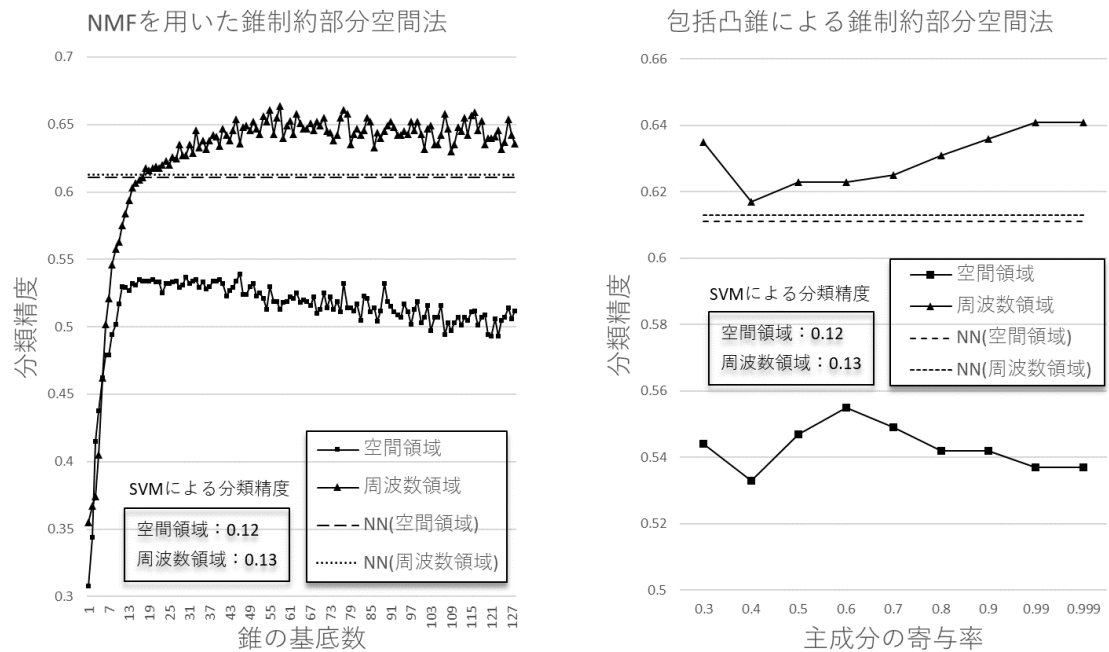


図 5.3: 学習枚数 200 枚における分類精度

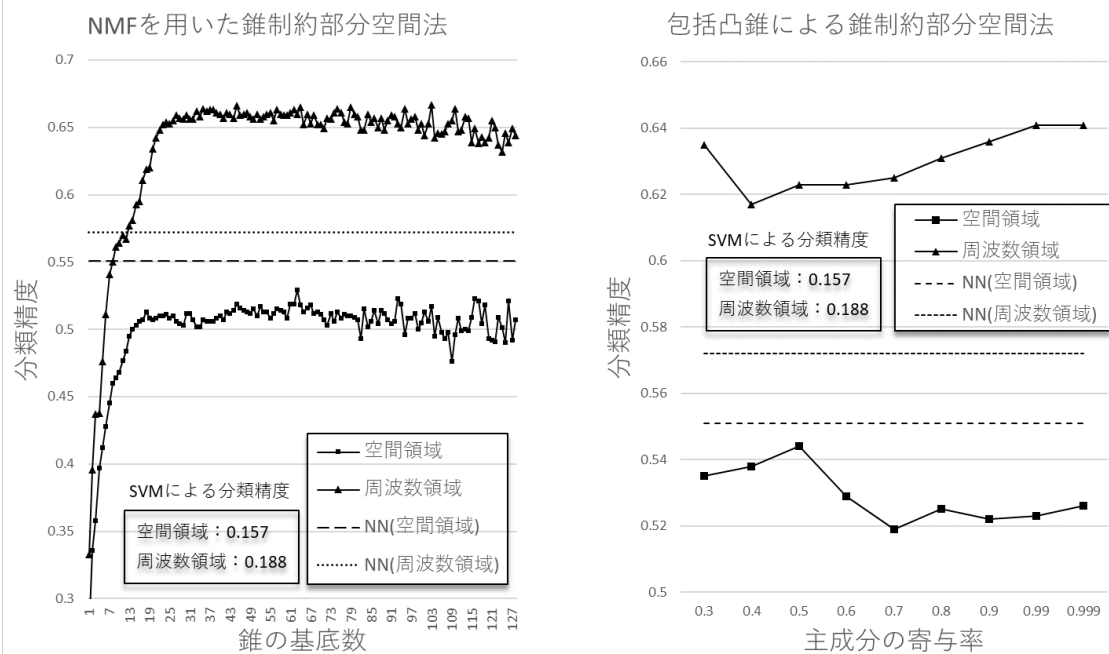


図 5.4: 学習枚数 300 枚における分類精度

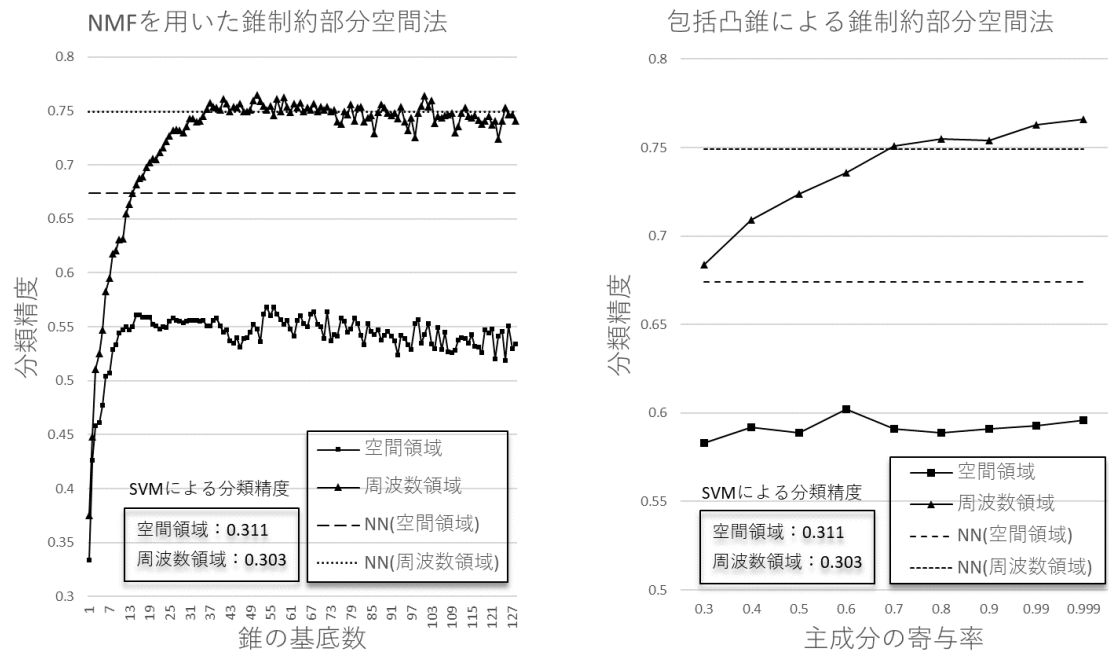


図 5.5: 学習枚数 400 枚における分類精度

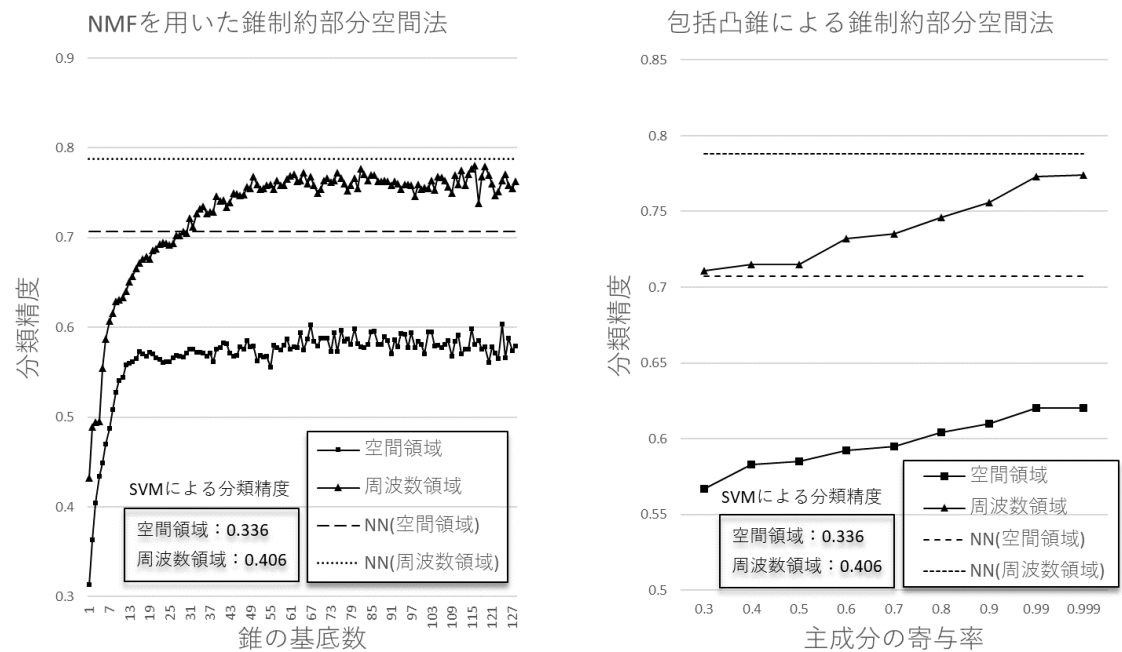


図 5.6: 学習枚数 500 枚における分類精度

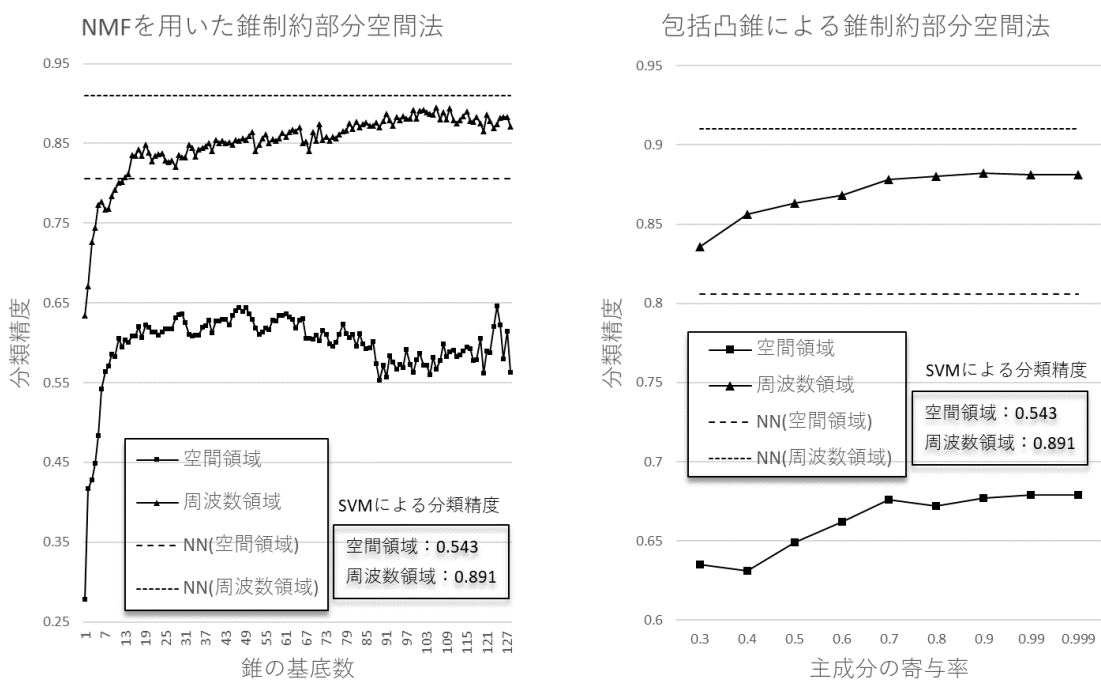


図 5.7: 学習枚数 1000 枚における分類精度

第6章

むすび

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を使用し錐制約部分空間法を提案した。パワースペクトルの使用は、分類器に依らず精度の向上が確認されたことから、シフト変動に頑健な特徴を示すことが分かった。また、錐制約部分空間法は他の手法と比較して、学習用データの枚数が少ないときに高い精度を示した。今後の課題としては、他のデータセットでの評価、多層化された畳み込み辞書学習による係数マップへの適用が考えられる。

謝辞

本研究及びその基礎の学習において，貴重な時間を割いてご指導して頂いた久留米工業高等専門学校制御情報工学科の黒木祥光教授，および日頃の研究活動において議論を行っていただいた同研究室の専攻科生に深く感謝致します．

参考文献

- [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] Gu, Shuhang, et al. "Convolutional sparse coding for image super-resolution," Proceedings of the IEEE International Conference on Computer Vision. 2015.
- [3] 福井和広, "部分空間法の今昔 (いまむかし)(下): 最近の技術動向: 相互部分空間法への拡張とその応用事例," 情報処理 49.6 (2008): 680-685.
- [4] 亀岡弘和, "非負値行列因子分解," 計測と制御 51.9 (2012): 835-844.
- [5] 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 (2009): 104-111.
- [6] Bristow, Hilton, and Simon Lucey, "Optimization methods for convolutional sparse coding," arXiv preprint arXiv:1406.2407 (2014).
- [7] Boyd, Stephen, et al, "Distributed optimization and statistical learning via the alternating direction method of multipliers," Foundations and Trends® in Machine learning 3.1 (2011): 1-122.
- [8] Wohlberg, Brendt, "SPORCO: A Python package for standard and convolutional sparse representations," Proceedings of the 15th Python in Science Conference, Austin, TX, USA. 2017.