

正則化に基づく計量距離学習手法と
自動分類への応用に関する研究

Distance Metric Learning Based on Regularization
and Its Application for Automatic Classification

2016年2月

三 川 健 太
Kenta MIKAWA

正則化に基づく計量距離学習手法と
自動分類への応用に関する研究

Distance Metric Learning Based on Regularization
and Its Application for Automatic Classification

2016年2月

早稲田大学 大学院 創造理工学研究科

経営システム工学専攻

情報数理応用研究

三 川 健 太

Kenta MIKAWA

目次

第 1 章	序論	1
1.1	研究の背景	1
1.2	研究の目的	5
1.3	本論文の構成	8
第 2 章	メトリックラーニング及びその周辺研究における研究動向	11
2.1	問題設定	11
2.2	データの自動分類に関する従来研究	13
2.2.1	データの自動分類に対する従来研究	13
2.2.2	データのベクトル表現と距離尺度	15
2.3	メトリックラーニング	17
2.3.1	メトリックラーニングの概要	17
2.3.2	繰り返し処理を用いたメトリックラーニング	21
2.3.3	繰り返し処理を用いないメトリックラーニング	24
2.3.4	メトリックラーニングの問題点	26
2.4	周辺分野における研究動向	28
2.4.1	正則化を用いた機械学習手法との関連	28
2.4.2	多変量解析（重回帰分析，判別分析）手法との関連	29
第 3 章	代表元を用いた計量行列の正則化学習 [55]	31
3.1	本章の目的	31
3.2	代表元を用いた計量行列の正則化法	32
3.3	実験	35
3.3.1	実験条件	36

3.3.2	実験結果	38
3.4	考察	43
3.5	定理, 補題, 系の証明	45
3.5.1	定理 3.1 の証明	45
3.5.2	補題 3.1 の証明	47
3.5.3	補題 3.2 の証明	47
3.5.4	定理 3.2 の証明	48
3.5.5	系 3.1 の証明	49
3.5.6	系 3.2 の証明	50
3.6	最適化問題が最小値を持つことの例証	50
3.7	本章の結論	53
第 4 章	l_1 正則化を用いたスパースな計量行列の学習 [58]	55
4.1	本章の目的	55
4.2	ADMM を用いた最適化	57
4.3	l_1 正則化を用いたスパースな計量行列の導出法	58
4.3.1	l_1 正則化を用いた最適化問題の定式化	58
4.3.2	ADMM を用いた最適な計量行列の導出	59
4.3.3	M-update の導出	59
4.4	提案手法と Sparse Inverse Covariance Selection との関係性	60
4.5	実験	61
4.5.1	新聞記事データを用いた実験	61
4.5.2	UCI 機械学習リポジトリを用いた実験	62
4.6	考察	64
4.7	本章の結論	67
第 5 章	複数の計量行列を用いた距離計量の学習 [56]	69
5.1	本章の目的	69
5.2	複数の計量行列を用いた距離計量の学習とそれを用いた分類手法	70
5.2.1	カテゴリ毎に計量行列を仮定した計量距離学習	70

5.2.2	複数の計量行列を用いたデータの分類手法	73
5.3	実験	74
5.3.1	UCI 機械学習リポジトリを用いた実験	75
5.3.2	新聞記事データを用いた実験	76
5.4	考察	79
5.5	定理, 補題, 系の証明	82
5.5.1	定理 5.1 の証明	82
5.5.2	定理 5.2 の証明	84
5.5.3	系 5.1 の証明	85
5.5.4	系 5.2 の証明	85
5.6	本章の結論	86
第 6 章	考察	87
6.1	本研究の考察	87
6.2	提案手法の実問題への適用について	89
第 7 章	結論	91
7.1	本研究の結論	91
7.2	今後の課題と展望	95
付録	複数の計量行列を用いたスパースな正則化法	97
謝辞		99
参考文献		101
研究業績		109

第1章 序論

1.1 研究の背景

近年の情報技術の発展に伴い、人々の生活や企業活動で扱われる情報や記録（ログ）として、大規模かつ多様なデータが蓄積されるようになった。これらの大規模データのすべてを人手により分析することは物理的に不可能となっているため、その自動分析方法が広く研究されている。得られた電子データから電算機を用いて知識発見を行う、もしくは有用な情報を抽出するための手法はデータマイニング、テキストマイニングなどと呼ばれ、広く普及している [31].

他方、これらの知識発見手法の多くは機械学習に基づく手法がベースとなっている。機械学習は教師あり学習、教師なし学習、半教師あり学習の3種類に大別することができるが [30]、本研究ではあらかじめデータに付与された正解（カテゴリ）を基に、新規データのカテゴリを推定するための基準を作成する教師あり学習に着目する。

教師あり学習には確率モデルによるアプローチ [11], [33]、ベクトル空間モデルによるアプローチ [54]、ブーリアンモデルによるアプローチ [80] などが存在している。この中でも本研究では、入力データ間の距離構造に着目し、その関係性を検討するベクトル空間モデルによるアプローチに着目する。ベクトル空間モデルは、入力データ間の距離尺度のみを用いて所望の分析を行う手法であり、特徴空間上に何らかのデータ生成モデルを仮定する必要がない、直感的にデータ間の類似性が理解しやすい、実装が比較的容易である、などの特徴を持つため、広く用いられている。

このようなベクトル空間モデルを用いた手法には k -Nearest Neighbor (k -NN) 法 [20] やテンプレートマッチング法 [25] などの手法が存在しており、用いる距離尺度によりその性能が変化することが知られている。これらの性能を向上させるため、機械学習の分野ではメトリックラーニングと呼ばれる手法が提案されている [9], [28], [88].

メトリックラーニングは、距離尺度としてマハラノビス距離 [53] を仮定し、任意の制約条件のもと所望の分析手法の性能を向上させるような距離構造（マハラノビス距離における計量行列）を学習するための方法であり、分類問題、クラスタリング、情報検索や画像認識などのパターン認識

を行うための様々な手法が提案されている [9]. メトリックラーニングは主に教師あり学習に基づくメトリックラーニングと教師なし学習に基づくメトリックラーニングに分けることができ, さらに学習データ全体を用いて大域的な距離構造の学習を行うか, 局所的な関係性のみに着目して局所的な距離構造の学習を行うかにより細分化される. 本研究ではこのうち, 学習データの情報を基にして新規入力データが所属するであろうカテゴリを予測する, 教師あり学習に基づくメトリックラーニング手法に着目する.

一般的に, 教師ありメトリックラーニングではすべての学習データのペア間の類似, 非類似の情報が事前に与えられているという問題設定のもと, それらの情報, ならびに所望の制約条件を用いて最適な計量行列を学習する [88]. その際, 計量行列を用いたマハラノビス距離が距離の公理を満たすことを保証するため, 計量行列の半正定値性を制約条件とし, 半正定値計画 [14], [34], [35] などに問題を落とし込み, 繰り返し処理を用いることでその学習を行う.

これらのメトリックラーニング手法を用いることにより, 学習データの統計的特徴を考慮した距離尺度が学習できるものの, 一般に以下の問題点が存在する.

問題点 1 学習データ数, ならびにその次元数の増加に伴う計算量の増加.

問題点 2 学習データの次元数増加に伴うパラメータ数の増加.

問題点 3 大域的な距離計量のみを学習することによる局所的構造への不適合.

上記の問題点 1 は, 多くのメトリックラーニング手法は繰り返し処理を用いて最適な計量行列を導出し, 毎回の更新時に固有値分解 [29] を行うアルゴリズムを採用しているために生じる. これは, 導出した計量行列が距離の公理を満たすために必要な処理であり, 学習データの次元数の増加は収束までの繰り返し回数を増加させ, 計算量の大幅な増加を引き起こす.

問題点 2 は計量行列の要素数が学習データの次元数の 2 乗に比例していることに起因する. これにより, 高次元なデータに対しメトリックラーニング手法を使用する場合, パラメータ数, ならびにその学習に必要なデータ数の大幅な増加を引き起こす [32]. 加えて, 十分な量の学習データを用意することができない場合には過学習 [30] が発生する可能性が高まり, 識別精度の低下を引き起こすなど分析へ悪影響を及ぼす.

さらに, 問題点 3 は, メトリックラーニング手法のほとんどが唯一の計量行列のみを学習しようとしていることに起因する. 一般に, 学習データは複数のカテゴリから構成されるが, これに

対し、唯一の計量行列を学習しようとする、カテゴリ間の統計的特徴を打ち消し合ってしまう可能性が生じる。すなわち、学習データの統計的特徴を適切に表現できていない計量行列を学習してしまい、正確なパターン認識ができなくなってしまうと考えられる。

上記の議論より、メトリックラーニング手法を比較的高次元なデータに適用しようとした場合、これらの問題点を解決するための施策が必要不可欠であると言える。

一方、上記の問題点1の解決のため、最適な計量行列を解析的に得ることができる手法として、持橋らの手法 [59] が存在する。この手法では、カテゴリの代表元を用いることで、繰り返し処理を用いずに最適な計量行列を導出することが可能となる。このため、持橋らの手法は計算量の面で他の繰り返し処理を用いた手法より優れている。

問題点2を解決するためのメトリックラーニング手法も存在するものの、前述の通り、メトリックラーニング手法のほとんどは固有値分解を含む繰り返し処理を用いたアルゴリズムを採用しているため、問題点1の解決を同時に行うことは難しい。

問題点3に対しては、複数の計量行列の存在を仮定した最適化問題を定式化することは可能であり、このような観点から局所的な計量行列を学習する手法が提案されている [63], [83]。これらの手法では、学習データの統計的特徴を考慮可能な計量行列を複数個学習しているものの、最終的に学習する計量行列はその線形結合により表現される。このため、計量行列それぞれが局所的な統計的特徴を表現できていたとしても、結果的にはそれらが打ち消し合ってしまう可能性がある。したがって、これらの局所的な統計的特徴を直接的に表現可能な手法を構築することは局所的なデータ間の関係性を表現可能となるという点で有用であると言える。

他方、今日の情報技術の発展に伴い、取り扱うデータも多様化している。特に、データの高次元化、大規模化は著しく、従来のメトリックラーニング手法を直接的に用いては現実的な時間で最適な計量行列を学習することは難しい場合が存在する。このような状況を考えた場合、上記の改善は重要な課題であると考えられる。

上記の議論より、本論文では計算量の面で他のメトリックラーニング手法と比較し、優位性のある持橋らの手法に着目する。しかしながら、この手法には扱うデータの特性により、以下に示す問題点が存在する。

持橋らの手法の問題点1

高次元なデータを対象とした場合、もしくは次元数に対し学習データ数が著しく少ない場合、最適な計量行列の存在が保証されていない。

持橋らの手法の問題点 2

学習データの構成要素間の一部に統計的な関係性がないデータを対象とした場合、導出された最適な計量行列ではその構造を適切に表現できない。

持橋らの手法の問題点 3

学習データのカテゴリ毎に統計的特徴が異なるデータを対象とした場合、その局所的構造を表現できない。

持橋らの手法における問題点 1 は、学習データの次元数が学習データ数よりも多い時¹に、分散共分散行列の階数が落ちてしまうことに起因する。これにより、特に高次元なデータを対象とした場合、もしくは学習データの次元数に対し著しくデータ数が少ない場合には最適な計量行列の存在が保証されない。

他方、計量行列の各要素は学習に使用したデータの成分間の関係性を表現していると解釈することができる。持橋らの手法で導出する計量行列は、学習データの成分間に関係性がないものが存在している場合、これらに対しても微小な値を付与してしまう。すなわち、学習データの成分間の関係性に過度に適合してしまうことになり、これが識別精度の低下を引き起こしてしまう可能性がある。したがって、持橋らの手法の問題点 2 の通り、上記のような構造を持つデータ²の関係性を適切に表現できない。

持橋らの手法の問題点 3 は、一般のメトリックラーニング手法と同様の理由により生じる。例えば、学習データ間に見せかけの相関関係が存在するような場合、すべての学習データに対して大域的な計量行列を学習してしまうと、その局所的な構造を見落としてしまう可能性がある。また、計量行列を複数個導出した場合、それらを統合的に用いたデータ間の距離尺度の測定手法についてはこれまで提案されていない。したがって、仮に学習データの局所的な関係性を考慮した複数の計量行列を導出できたとしても、それをを用いた識別規則の検討は必要不可欠であると言える。

¹例えば、文書分類問題を対象とした場合、学習データのベクトルは文書集合中に出現した総異なり単語の出現頻度で構成される。一般にこれらの分析ではその次元数は数千から数万となり、これだけの量の学習データを事前に準備することは多大なコストを要する。

²例えば、電子商取引 (EC) サイト上の顧客の購買履歴などがこれに相当する。購買履歴は顧客の特性により、工学系の研究者は自身の研究に関係する分野の書籍を頻繁に購入する一方で、他分野の書籍は購入しないなど、取り扱う商品の一部の間にのみ関係性が考えられる。

1.2 研究の目的

前述の通り、持橋らの手法における計量行列導出の計算量は、繰り返し処理を用いた手法と比較し相対的に少ない。これは、データの大規模化や多様化が進んでいる今日では手法の適用面で有効であるということが出来るものの、前述の通り対象とするデータの特性によりいくつかの問題点が存在している。

本研究では、前述のデータを対象とした場合に生じる従来手法の問題点の改善を行うことにより、分類精度を向上させることを目的とし、以下の3種類の手法を提案する。これらの手法は、それぞれに異なる性質を有しているため、データの特性により使い分けるのみならず、それぞれの手法が持つ特徴の違いにより使い分けることが可能となる。

提案手法 1 正則化項を付与した最適化問題に基づく計量行列の学習方法 [55].

提案手法 2 スパースな計量行列の学習方法 [58].

提案手法 3 複数の計量行列の学習方法、ならびに複数の計量行列を用いた識別法 [56].

提案手法 1 は、持橋らの手法を正則化法を用いて拡張することで、その分類精度を向上させることを目的としている。具体的には、前述の持橋らの手法の問題点 1 にある、学習データ数がその次元数よりも少ない場合、すなわち、最適な計量行列が求められない場合を想定し、このような場合でも正則化項を導入した最適化問題を定式化することで、最適な計量行列が解析的に得られることを示す。これにより、最適解を解析的に得られるという従来の持橋らの手法の利点を活かしたまま、高次元なデータセットに対しても最適な計量行列が導出可能とし、分類精度の向上を図る。この手法は、他の提案手法と比較した場合、学習にかかる計算量が相対的に低いことが特徴であり、学習に要する時間をできるだけ短時間にしたい場合には有効な手法であると言える。

提案手法 2 の目的は、他のメトリックラーニング手法と比較し、分類精度を低下させずにスパースな構造のもとで計量行列を学習することにある。これにより、パラメータ数の削減に加え、新規入力データの分類時に必要な距離計算にかかる計算量を削減することが可能となる。提案手法 2 では、前述の持橋らの手法の問題点 2 の改善のため、学習データの成分間の関係性を表現している計量行列の各要素のうち、関係性が薄い成分ペアに対応する要素を 0 に縮退させることでスパースな計量行列の導出を行う。このため、持橋らの手法で定義した最適化問題に対し、計量行列の l_1 ノルムを正則化項として追加する。合わせて、提案手法 2 が統計学で用いられている Sparse

Inverse Covariance Selection (スパース共分散推定) [6] の特殊な場合であることを示す。この手法では、新規入力データを分類する際の計算量を削減することが可能となるため、短時間に多くの分類を行う必要がある適用場面では非常に有効な手法となる。

最後に、提案手法3では、計量行列を用いた分類の精度向上を目的として、各カテゴリに対して異なる計量行列の存在を仮定し、それらの学習方法を提案する。加えて、複数の計量行列を用いることによる新規入力データの分類手法についても提案を行う。本手法は、特にカテゴリ毎にその統計的特徴が大きく異なるようなデータを対象とした場合に分類精度の向上に有効な手法であるということができる。一方で、複数の計量行列を学習した場合、その学習時にかかる計算コストが大きな問題となる可能性が高い。本手法では、従来の持橋らの手法をベースとすることで提案手法1と同様に解析的に複数の計量行列がそれぞれ導出可能となることを示す。この手法では、複数の計量行列の導出を行うため、学習、分類それぞれの計算量は相対的に大きいものの、その精度は他の手法よりも高いため、計算量は犠牲にしても高精度な識別を行いたい場合には有効な手法であるということができる。

これらの提案手法と各提案手法が対象とするデータの関係性、各手法の特色について図1.1に示す。本研究で提案する各手法はデータや得られる計量行列の特性に合わせて使い分けることで、その特性に合わせた距離計量の学習が可能となる。これにより、多様なデータに対しその適用が可能となる。

各提案手法が対象とするデータの具体例について以下に示す。

例 1.1 (提案手法1 (高次元かつ学習データ数が少ないデータ))

デジタル形式で保存されている画像データなどが該当する。これらの画像データは各画素(ピクセル)が一つの特徴量を示し、一つの画像は膨大な量の特徴量から構成されている。このように特徴量が非常に大きなデータはパラメータ数が非常に多くなってしまうため、学習に必要なデータ数が多く必要になる。結果的に、学習に十分な量の学習データを集めることが難しくなり、限られた量の学習データで十分な精度を達成するためには何らかの工夫が必要になる。

例 1.2 (提案手法2 (多くの成分間で統計的関係性が薄いデータ))

文書分類などに用いる文書データが該当する。これらのデータをベクトル表現する際には、学習データに出現する全ての異なり単語をその特徴量とする。これにより、学習データのカテゴリが異なる場合、そこに出現する単語の傾向も異なり、同時に生起することがない単語の組が存在する可能性が高い。これらの不要な単語ペア間の特徴を適切に表現することができれば、不要なパ

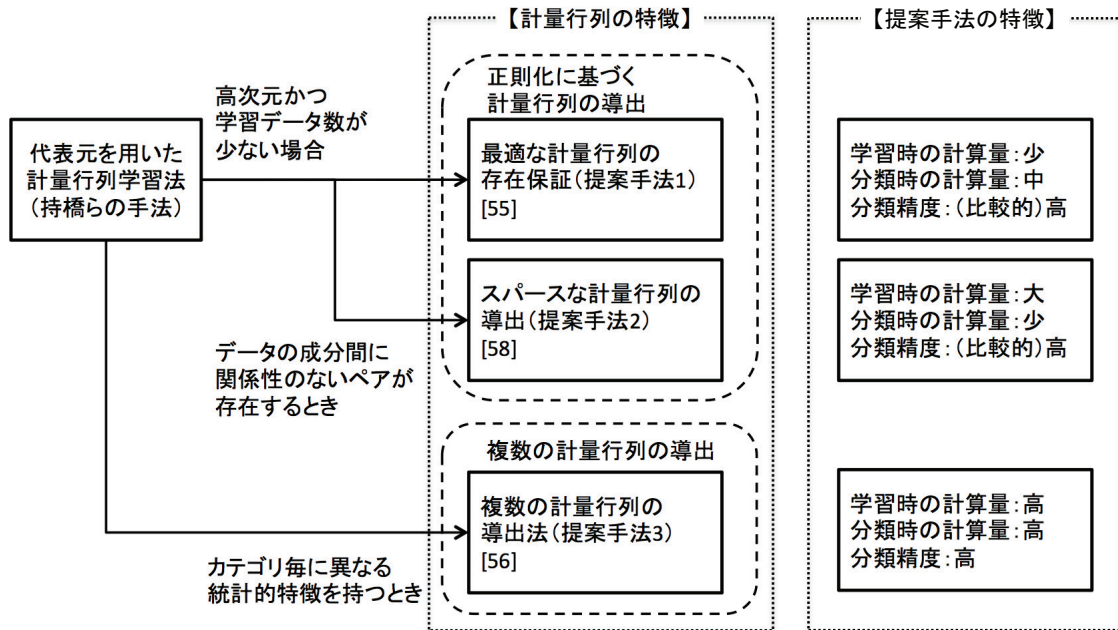


図 1.1: 各提案手法とそれが対象とするデータの関係性

ラメータの削減，ならびに分類精度の向上が期待できる。

例 1.3 (提案手法 3 (カテゴリ毎に異なる統計的特徴を持つデータ))

ファックスの通信で用いられる白黒の手書き文字のデータなどが該当する。手書き文字のデータも画像データと同様に各画素が一つの特徴量を示しており，対象領域において，黒画素の部分が1，白画素の部分が0であるようなベクトルとして表現される。具体例として，数字のデータを対象として考えた場合，各々の数字ではその形状，書き方などが大きく異なると考えられる。したがって，それぞれの数値ごとにその統計的特徴は大きく異なると考えられると共に，学習データ中に存在している特徴を適切に使用することができれば，それを用いた分類の精度も向上すると想定される。

前述の通り，本研究はメトリックラーニングに関する各種手法のうち教師あり学習に基づく手法を対象とする。一般にメトリックラーニングにおける教師あり学習には強教師あり学習，弱教師あり学習の二種類が存在する [9]。さらに，計量行列を学習データ全体の特徴を考慮し，唯一のものを学習する大域的メトリックラーニングと局所的なデータ間の関係性のみに着目してそれを学習する局所的メトリックラーニングに分けることができる。これらを踏まえ，本研究で提案す

る各手法のメトリックラーニング手法における位置づけについて、広く用いられている他手法との比較を表 1.1 に示す。

表 1.1: 本研究のメトリックラーニング手法における位置づけ

	教師あり学習		教師なし学習
	強教師あり学習	弱教師あり学習	
大域的 メトリックラーニング	LMNN [85], [84] 持橋らの手法 [59] 提案手法 1 [55] 提案手法 2 [58]	Xing et al. [86] ITML [22]	主成分分析 [62] 多次元尺度構成法 [78]
局所的 メトリックラーニング	提案手法 3 [56]	近傍成分分析 [27] 関連成分分析 [73]	ISOMAP [1] 局所線形埋め込み法 [67]

詳細については各章にて述べるが、本研究では代表元を用いたメトリックラーニング手法に対し、そのトレースを正則化項として付与したもの、 l_1 ノルムを正則化項として付与したもの、複数の計量行列の学習を行うものの3手法を提案することで、従来のメトリックラーニング手法の問題点の改善を図る。表 1.2 に、計量行列に対する正則化の観点から本研究の枠組みについて示す。

表 1.2: 正則化法に基づく本研究の枠組み

正則化	なし	$\text{tr}(M)$	l_1 ノルム
大域的計量行列	持橋らの手法 [59]	提案手法 1 [55] (3章)	提案手法 2 [58] (4章)
局所的計量行列	提案手法 3 [56] (5章)	提案手法 3 [56] (5章)	付録

表 1.2 における“付録”の内容に関しては、その定式化を付録に示す。

1.3 本論文の構成

本論文は7章から構成されている。次章以降の内容は以下の通りである。

まず、第2章では本研究で対象とする問題設定について説明する。さらに、マハラノビス距離 [53] を始めとする距離尺度や用いる変数の定義を与える。その後、本研究で対象とするメトリックラーニ

ングについて説明を行うと共に、広くベンチマーク手法として用いられている ITML (Information-Theoretic Metric Learning) [22], LMNN (Large Margin Nearest Neighbor) [85], [84], 持橋らの手法 [59], Xing らの手法 [86] のそれぞれについて説明を行う。

第3章では、データの次元数が学習データ数よりも大きな場合を想定した手法について検討を行う。このような高次元データを対象とした場合、従来の持橋らの手法では最適な計量行列の存在は保証されていない。これに対し、第3章で提案する手法では持橋らの手法に対し正則化を行うことで、計量行列が必ず一意に定まることを示す。加えて、正則化パラメータと導出された計量行列の間にある関係性や、最適な計量行列の持つ特性について解析を行う。提案手法1の有効性を新聞記事データを用いた分類実験を行うことにより示す。

第4章では、学習データの多くの成分間に統計的な関係性がないと想定されるデータを対象とした分析手法として、計量行列内の不必要な要素を0とするスパースな計量行列を学習する手法について検討を行う。このような計量行列の導出の際には、一般のメトリックラーニング手法と同様に繰り返し処理を用いた計算が必要であるが、ADMM (Alternating Direction Method for Multiplier) [13] と呼ばれる最適化手法を用いることで並列計算が可能となることを示す。加えて、このスパースな計量行列の導出が統計学の一手法である Sparse Inverse Covariance Selection [7] の特殊な場合であることを示す。提案手法2の有効性について、比較的低次元、密な（スパースではない）データセットとして、機械学習のベンチマークデータである UCI 機械学習リポジトリ [5] を、高次元、スパースなデータセットとして新聞記事データを用いた実験をそれぞれ行うことにより示す。

第5章では、用いる学習データのカテゴリ間でその統計的特徴が異なる場合を想定した分析手法について検討を行う。このようなデータに対して、提案手法3では複数の計量行列の存在を仮定し、カテゴリ毎の統計的特徴を表現可能な計量行列の導出を行う。この際に、従来の持橋らの手法同様に最適な計量行列がそれぞれ解析的に得られることを示す。また、第3章同様に各計量行列が持つ特性についても解析を行う。加えて、学習した複数の計量行列を用いた新規入力データと学習データとの距離を測定するための手法が構築可能であることを示す。提案手法3の有効性を第4章同様、UCI 機械学習リポジトリを用いた場合、新聞記事データを用いた場合の両実験により示す。

第6章では本研究で得られた考察を述べると共に、本研究にて提案を行った3つの手法の使い分けについて述べる。

最後に、第7章は結論であり、本研究の成果をまとめ、今後の展望について述べる。

第2章 メトリックラーニング及びその周辺研究における研究動向

2.1 問題設定

得られたデータに対し、何らかの規則のもとでそのデータが所属するであろうカテゴリを推定することをパターン認識と呼ぶ [4], [25], [28], [36], [39]. このようなパターン認識は様々な処理を包含しており, Web 上の類似したページの検索, 文書からの著者推定, EC サイト使用ユーザへの商品推薦など, 様々な使用用途が考えられる.

これらの作業のほとんどは人手で行うことにより高精度な認識が可能であるものの, 近年の高度情報化により, 大量のデータが蓄積されている今日では, 全てのデータに対し人手によるパターン認識を行うことは物理的に不可能である.

他方, このような膨大な量のデータに対し, 電算機による正確な識別が可能となれば, これらを高速に処理することができ, 人手で行う以上に大量のデータに対する処理, 延いてはそこからの知識発見が可能となる. 本研究で対象とするパターン認識は電算機による上記処理の性能を向上させ, より正確なデータの認識を可能とするための方法論を指し, 一般的には図 2.1 の手順により行われる [39].

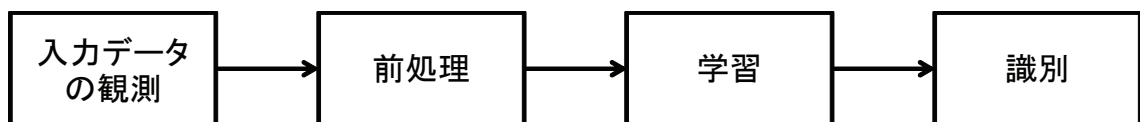


図 2.1: パターン認識の処理概要 [39]

図 2.1 における「前処理」は, 識別を行う入力データをコンピュータで処理しやすい形式に加工する処理を指す. 前処理では入力データのベクトル化に加え, 識別を容易にするために有効であると想定される特徴の抽出, すなわち用いる特徴を選択することもしばしば行われる.

同様に、「学習」は前処理で作成した入力データベクトルを用い、データの識別を容易にするような規則を何らかの基準のもとで抽出する処理を指し、識別は所属カテゴリが未知の新規入力データがどのカテゴリに所属するかを推定するための処理を指す。学習、識別それぞれの処理は何らかのアルゴリズムに基づいて行われ、これらの手法を総じて（統計的）機械学習と呼ぶ [30], [71], [75], [76].

パターン分類の一連の流れは、以下のような例が考えられる。

例 2.1 (文書データを対象とし、新規入力文書を分類する場合)

電子的に作成されている文書データを対象として、新規入力文書を既に与えられたカテゴリへ分類する問題を考える。前処理では、得られた文書に対し形態素解析 [47], [61] と呼ばれる処理を行うことで文書を形態素へ分割し、異なり単語の出現頻度をその成分として持つベクトルで表現し処理を行う。

特に文書データを対象とした場合には、前処理として入力データの形態素によるベクトル化に加え、識別を容易にするために有効であると想定される特徴の選択が行われる、例えば、ベクトルの成分のうち分類に有効な形態素を品詞によって選択することもしばしば行われる。文書データの場合では、助詞、助動詞、接続詞などは識別に影響を与えない“不要語”として削除し、それら以外の品詞を用いて文書ベクトルを構成することも多い [48].

これらのデータに対し、何らかの基準のもとで新規入力データを正しいカテゴリに分類するための規則を学習し、学習した規則を用いてその識別を行う。

例 2.2 (EC サイト上に存在する顧客の購買履歴を対象に商品の推薦を行う場合)

顧客の購買履歴を基にして、商品の推薦を行う場合を考える。この場合、購入している商品が類似しているユーザ同士を比較し、類似ユーザが購入しているもののうち推薦対象ユーザが未購入の商品を推薦する方法が一般的である。

前処理として、購買履歴データに対し、顧客が購買した商品の個数を成分とするベクトルで表現する [65].

学習では、これらのデータに対し、顧客間の類似度を測定し、購買行動の類似性が高いと考えられる顧客の抽出を行う。さらに、識別では類似性が高い顧客が既に購入している商品のうち、対象とする顧客が未購入であるような商品の推薦を行う。

機械学習は一般的に、予めデータの所属カテゴリが既知の学習データが与えられたもとで、その統計的特徴から新規入力データの識別規則を学習、新規入力データの予測を行う教師あり学習 [2],

[60] と、正解データを用いず、所属カテゴリが未知のデータのみを使用してそれらの特徴を抽出する教師なし学習 [37] に大別される。

一般に、教師あり学習に関する手法は、あらかじめ学習データに付与されているラベルが連続か離散かにより、回帰、分類の二種類に大別される。主な回帰問題に対するアプローチとして、重回帰分析 [62] やロジスティック回帰分析 [49], Support Vector Regression (SVR) [19], ニューラルネットワーク [10] などが用いられる。一方で、分類問題に対するアプローチとしては、確率モデルを用いた分析 [8], [11], [62] や特徴空間の分割に基づく手法 [2], [71], 入力データ間の類似度に基づく手法 [20] などが用いられる。

他方、教師なし学習に関する手法としては、入力データを任意の複数個のクラスタへ分割するクラスタリング、その次元を縮約しデータの構造把握を容易にする次元圧縮、入力データ間の関係性を発見する相関ルールなどがあげられる。

このうち、クラスタリングでは、 k -means 法 [25] と呼ばれる手法や、クラスタ分析 [62] などがしばしば用いられる。加えて、近年では各クラスタが多次元正規分布に従うと仮定し、そのパラメータを推定することでクラスタリングを行う凸クラスタリング [51], や距離尺度として Bregman Divergence [15] を用いたクラスタリング手法 [6] などが提案されている。

また、次元圧縮に関しては、多変量解析などで用いられている主成分分析 [62] や Latent Semantic Indexing (LSI) [24], LSI と同等の処理を確率モデルで表現した Probabilistic Latent Semantic Indexing (PLSI) [33], 多次元尺度構成法 [78] などが存在する。

他方、大規模なデータから有用な知識発見を試みるための方法論としてデータマイニングが用いられているが、このデータマイニングで用いられる手法のうち、基本的かつ代表的な手法が相関ルールに基づく手法 [38] と言える。

以下、本研究ではこれらの手法のうち、教師あり学習における分類問題を対象とする。

2.2 データの自動分類に関する従来研究

2.2.1 データの自動分類に対する従来研究

前述の通り、データの自動分類は、あらかじめデータの所属カテゴリが付与された学習データの情報を基に、何らかのアルゴリズムによりその識別規則を学習し、新規入力データの所属するであろうカテゴリを予測する問題とすることができる。

一般に、機械学習における分類問題は以下に示すアプローチに大別することができる。

- 確率モデルに基づく手法
- 識別関数に基づく手法
- 類似度（距離尺度）に基づく手法
- 特徴空間の分割に基づく手法

このうち、確率モデルに基づく手法では、「学習データがあるカテゴリ特有の何らかの確率分布に従い生起する」、という仮定のもと、その確率分布のパラメータを推定し、識別規則を学習する。確率モデルに基づくアプローチとしては、迷惑メールのフィルタリングなどで広く用いられているナイーブベイズ法 [8], [11] や優れた二値判別を行うことが可能な RVM (Relevance Vector Machine) [79] など、多くの分類手法が存在している。

加えて、学習データと所属カテゴリの間には、観測されない潜在的な何らかの関係性があるという前提のもと、その潜在的な関係を表現した確率モデルとして潜在クラスモデルが存在している。潜在クラスモデルは一般に、表現能力が高く、学習データのカテゴリ毎の潜在的な傾向を推定可能なモデルとして近年注目されている。これらには、潜在ディリクレ配分法 (LDA) [12], 確率的潜在意味解析 (PLSA) [33], 混合 polya 分布に基づく手法 [3], [68] など、様々な手法が提案され、広く用いられている。

識別関数に基づく手法では、新規入力データを正しくカテゴリに分類するための識別関数を定義し、そのパラメータを学習データを用いることで推定するための手法である。識別関数に基づく手法としては、線形識別モデル [25], SVM (Support Vector Machine) [19], [74], [82], ニューラルネットワーク [10] などが存在する。

類似度に基づく手法では、ベクトル空間上で表現された各入力データに対し、それらの間の類似度（距離）尺度を用いた分類を行う。これらの手法としては、新規入力データと、その周辺に存在する学習データのカテゴリ毎の多数決により所属カテゴリを推定する k -NN 法 [20], 各カテゴリの代表元と新規入力データとの距離を測定し、最も近いカテゴリへ分類するテンプレートマッチング法 [25] などが存在する。

他方、特徴空間の分割に基づく手法は、類似した学習データ同士を集められるような分割の規則を学習するための手法であると言える。これらの手法としては、決定木による手法 [25], 決定木

の欠点である過学習や分類精度の低さをアンサンブル手法 [46] を用いて改善したランダムフォレスト [16] などがあげられる。

上記の手法の持つ特徴をまとめると表 2.1 のようになる。

表 2.1: 機械学習における分類問題に対するアプローチ

	手法	実装の 容易さ	解釈の 容易さ	大規模データ への適応性	学習 時間	分類 時間	分類 精度
確率モデルに基 づく手法	ナイーブベイズ法	○	◎	○	○	○	△
	RVM	×	◎	△	△	○	◎
識別関数に基 づく手法	ニューラルネットワーク	△	△	△	△	△	△
	SVM	×	△	○	○	○	○
類似度に基づく 手法	k -NN 法	◎	◎	○	◎	△	△
	テンプレートマッチング	◎	◎	○	◎	△	△
特徴空間の分割 に基づく手法	決定木	△	○	×	○	○	△
	ランダムフォレスト	△	○	○	○	○	◎

本論文では、このように様々な手法が存在している分類手法のうち、類似度（距離尺度）に基づく手法を採用する。

類似度に基づく手法はベクトル空間モデルを用いた手法であり、入力データ間の距離尺度のみを用いて所望の分析を行うことが可能となる。このため、入力データ間に何らかのモデルを仮定することなく、どのようなデータが対象であってもパターン認識を行うことができる。そのため、新しい学習データが追加された際に、最初から識別規則を学習しなおすことなく、データベースにそのデータを追加しておくだけで、新たな入力データの分類に活用できるという実用上の利点を持つ。また、類似度に基づく手法を採用する理由として、表 2.1 からわかる通り、実装、結果の解釈が比較的容易であること、学習した計量行列を使用したデータの分類がしやすいことがあげられる¹。

2.2.2 データのベクトル表現と距離尺度

本節では本研究で用いる変数についてその定義を行う。併せて、ベクトル空間モデルに基づくアプローチで用いられる距離尺度についての説明を述べる。

¹従来用いられているメトリックラーニング手法のほとんどは、 k -NN 法をはじめとする類似度に基づく手法の使用を前提として距離計量の学習が行われているため、この点からも本研究ではこれを使用することとする。

いま、ベクトル表現された W 次元の学習データを $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iW})^T$ とし、 $i = 1, 2, \dots, D$ とする。ただし T はベクトルの転置を、 W は総異なり特徴数を表し、 D は総学習データ数を表すものとする²。また、各学習データにはそれぞれ一つのカテゴリ情報が与えられているものとし、カテゴリ集合を $\mathcal{C} = \{C_1, C_2, \dots, C_N\}$ とする。ただし、 N は総カテゴリ数を表すものとする。このように入力データをベクトルとして表現することにより、ベクトル空間上でデータ間の距離を考えることが可能となる。ただし、ここでの距離尺度は、入力データ $\mathbf{x}_i, \mathbf{x}_j$ に対し、以下に示す距離の公理をすべて満たすような距離関数 $d(\mathbf{x}_i, \mathbf{x}_j)$ とする。

1. 非負性 $d(\mathbf{x}_i, \mathbf{x}_j) \geq 0$
2. 対称性 $d(\mathbf{x}_i, \mathbf{x}_j) = d(\mathbf{x}_j, \mathbf{x}_i)$
3. 非退化性 $d(\mathbf{x}_i, \mathbf{x}_j) = 0 \Leftrightarrow \mathbf{x}_i = \mathbf{x}_j$
4. 三角不等式 $d(\mathbf{x}_i, \mathbf{x}_k) + d(\mathbf{x}_k, \mathbf{x}_j) \geq d(\mathbf{x}_i, \mathbf{x}_j)$

k -NN 法、テンプレートマッチング法に代表されるベクトル空間モデルでは、用いる距離尺度によってその性能が異なることが知られているものの、計算の簡便さなどからユークリッド距離やコサイン類似度が頻繁に用いられている³。

いま、2つの入力データをそれぞれ $\mathbf{x} = (x_1, \dots, x_W)^T, \mathbf{y} = (y_1, \dots, y_W)^T$ とし、これらの間のユークリッド距離、コサイン類似度をそれぞれ $d_E(\mathbf{x}, \mathbf{y})$, $\text{sim}(\mathbf{x}, \mathbf{y})$ と記述することになると、 $d_E(\mathbf{x}, \mathbf{y})$, $\text{sim}(\mathbf{x}, \mathbf{y})$ は以下のように定義される。

$$d_E(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{w=1}^W (x_w - y_w)^2}, \quad (2.1)$$

$$\text{sim}(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} = \frac{\sum_{w=1}^W x_w y_w}{\sqrt{\sum_{w=1}^W x_w^2} \sqrt{\sum_{w=1}^W y_w^2}}. \quad (2.2)$$

ただし、 $\|\cdot\|_2$ はベクトルの l_2 ノルムを示すものとする。

前述の通り、これらの尺度は容易に計算可能であるため広く用いられているが、共に入力データの要素ごとの独立した類似性のみを考慮しており、要素間の統計的な関係性については考慮す

²このような特徴は、例えば文書データを対象とした場合には、学習データの文書集合中に出現する各異なり単語を、EC サイトを対象とした場合は、ユーザが購入したアイテム数などで与えられる。

³ただし、コサイン類似度は距離尺度ではなく、類似度尺度であるため、上記の距離の公理を満たさない。

ることができない。これに対し、要素間の関係性を考慮したマハラノビス行列（以下計量行列） $M = [M_{ij}] \in \mathcal{R}^{W \times W}$ を用いたマハラノビス距離 [53] を考えることで、データ間の関係性を適切に表現できることが知られている。いま、入力データ \mathbf{x} と \mathbf{y} 間のマハラノビス距離 $d_M(\mathbf{x}, \mathbf{y})$ は以下で定義される⁴。

$$d_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T M (\mathbf{x} - \mathbf{y})}. \quad (2.3)$$

式 (2.3) で定義したマハラノビス距離はその計量行列が半正定値対称行列である場合、距離の公理を満たすことが知られている。また、以降ではある行列 A が半正定値対称行列である場合、 $A \succeq 0$ を用いる。いま、式 (2.3) における計量行列 M が半正定値対称行列であるとした場合、ある行列 $L \in \mathcal{R}^{W \times W}$ が存在して、 $M = L^T L$ と分解することができる [53]。これを用いて式 (2.3) を表現すれば、

$$\begin{aligned} d_M(\mathbf{x}, \mathbf{y}) &= \sqrt{(\mathbf{x} - \mathbf{y})^T L^T L (\mathbf{x} - \mathbf{y})} \\ &= \|L\mathbf{x} - L\mathbf{y}\|_2 \\ &= d_E(L\mathbf{x}, L\mathbf{y}), \end{aligned} \quad (2.4)$$

となる。式 (2.4) は入力データを行列 L により射影した空間上で \mathbf{x}, \mathbf{y} 間のユークリッド距離を測定していることと等価であることが分かる。このため、 $d_E(L\mathbf{x}, L\mathbf{y})$ を正規化ユークリッド距離と呼ぶこともある。

マハラノビス距離を用いることで、入力データの成分間関係性を考慮した距離尺度を導入することが可能となるが、ここで用いる計量行列は何らかの方法を用いて適切に推定する必要がある。

2.3 メトリックラーニング

2.3.1 メトリックラーニングの概要

前述の通り、ベクトル空間モデルを用いたパターン認識を行う際には、用いる距離尺度によってその性能が大きく依存する。特に入力データとして用いるベクトルの成分間に何らかの関係性

⁴式 (2.3) は統計学で用いられているマハラノビス距離とは計量行列がデータから得られる精度行列ではないということから厳密には一致しない。しかし、一般のメトリックラーニング手法では式 (2.3) をマハラノビス距離と呼ぶことが一般的であり、本研究でもこれに則り、式 (2.3) をマハラノビス距離と呼ぶものとする。

があるような場合、通常のユークリッド距離やコサイン類似度を用いると、これらの統計的関係性を考慮していないこととなり、高精度な分類器を構成できない可能性がある。このような場合、距離尺度として式(2.3)で定義したマハラノビス距離を用いることで、統計的関係を考慮した距離構造を学習することが有効であるものの、計量行列の学習が必要となる。

メトリックラーニングは、用いる分類手法の性能を向上させるような計量行列を距離の公理を満たしながら行う手法である [17], [41], [52], [64], [88]。すなわち、新規入力データへの分類性能を向上させると想定される制約条件のもとで、計量行列 M を学習によって得る手法であると言える。メトリックラーニングにおける処理の流れを以下の図 2.2 に示す。

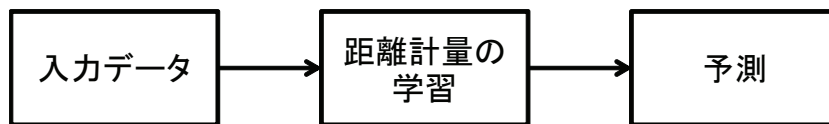


図 2.2: メトリックラーニング手法の処理

上記の図 2.2 における「距離計量の学習」が、メトリックラーニング手法により計量行列を学習することに相当する。すなわち、一般のメトリックラーニング手法は、図 2.1 に示した一連のパターン認識の処理のうち、「学習」までのプロセスを指しており、「識別」における方法はこれらの手法とは独立に構成されることが多い。したがって、図 2.2 における「予測」が一般のパターン認識における「識別」部分に対応することとなる。

一般の機械学習手法同様、メトリックラーニングもあらかじめ学習データが与えられたもとでその計量行列の学習を行う教師あり学習に基づく手法、ならびに次元削減を行いデータの低次元表現を得ようとする教師なし学習に基づく手法の二つに大別される。

教師なし学習に基づくメトリックラーニングでは、学習データが与えられない状況下でデータの低次元表現を入力データから抽出するための手法であると言える。すなわち、入力データの特徴を可視化し、そのデータセットが持っているであろう特徴を削減した次元で確認するための方法論と言える。この教師なし学習に基づくメトリックラーニングには、主成分分析 [62]、多次元尺度構成法 [78]、ISOMAP [1] などが存在する。

他方、教師あり学習に基づくメトリックラーニングは、その最適な計量行列を半正定値計画や凸計画問題 [45] を用いて学習する繰り返し処理を用いたメトリックラーニング手法と、最適解を解析的に得ることのできる（繰り返し処理を用いない）メトリックラーニング手法に分類するこ

とができる。

この教師あり学習に基づくメトリックラーニングは学習データのカテゴリ特性により以下の二つに分けることができる [9].

強教師あり学習：

学習データに対し，その所属カテゴリがラベルとして一意に与えられているという（強い）条件のもとで計量行列を学習する手法。

例 2.3 (新聞記事データの分類を対象とした場合)

新聞記事データには，その所属カテゴリ（政治，経済，スポーツ，芸能など）が与えられていることが多い。このような場合，学習データの所属カテゴリとして，これらのカテゴリ情報を用い，適切な距離計量を学習する。学習した距離計量を用い，新規入力データと各学習データとの距離を測定することでその分類を行うことが可能となる。

弱教師あり学習：

上記の強教師あり学習とは異なり，学習データにカテゴリラベルが付与されていない状況下で計量行列を学習する手法。ただし，学習データ間の類似度や非類似度は何らかの基準によって与えられているため，この（弱い）条件を用いることで所望の計量行列を学習する。

例 2.4 (類似した特徴を持つ Web ページ群を抽出する場合)

Web ページのクラスタリングを Web ページ間の共起単語情報を用いて行う場合を考える。例えば，異なる二つの Web ページで共起する単語が一定数以上あった場合，両者を類似しているものと定義する。この情報を用いることで異なる Web ページ間での類似，非類似情報を事前に得ることが可能となり，これを用いて Web ページ間の距離計量を学習することが可能となる。

強教師あり学習，弱教師あり学習ともに，距離計量を学習する際には学習データ間の関係性を用いた距離計量を学習する。この際には“類似した学習データ間の距離をできるだけ小さく”，かつ“類似していない学習データ間の距離はできるだけ大きく”，という考え方にに基づき，その学習を行うことで分析性能を向上させるような距離計量を求める。図 2.3 に，教師あり学習に基づくメトリックラーニング手法の概要を示す。

上記の考え方のもとで所望の距離計量を得る際には，学習データとして x_1, x_2, \dots, x_D が与えられたとき，類似データ集合 S ，非類似データ集合 D をそれぞれ定義し，これらに所属するデータ

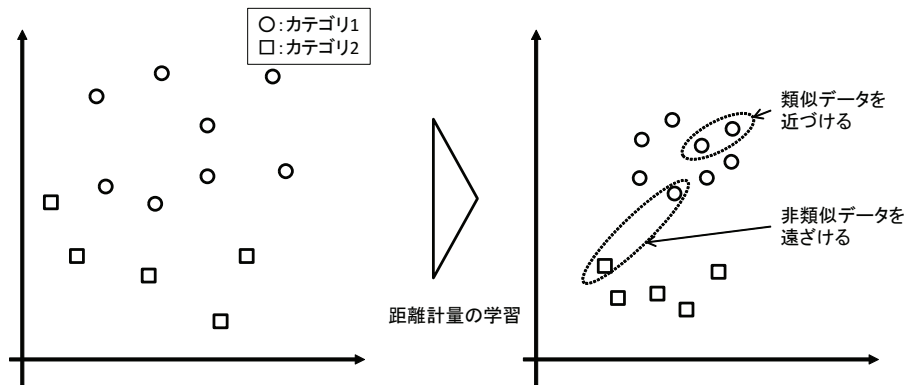


図 2.3: 教師あり学習に基づくメトリックラーニング手法の概要

間の関係性を用いる [86]. ただし, 集合 \mathcal{S} , \mathcal{D} はそれぞれ,

$$\begin{aligned} \mathcal{S} : (x_i, x_j) \in \mathcal{S} & \quad \text{if } x_i \text{ and } x_j \text{ are similar,} \\ \mathcal{D} : (x_i, x_j) \in \mathcal{D} & \quad \text{if } x_i \text{ and } x_j \text{ are dissimilar,} \end{aligned}$$

により与えられる.

これらの制約条件は強教師あり学習, 弱教師あり学習の双方に対して定義することができ, 例えば強教師あり学習の場合は同一カテゴリに所属している学習データのペアは \mathcal{S} に属し, 異なるカテゴリに所属している学習データのペアは \mathcal{D} に所属する, などによって最適な計量行列を学習する. 同様に弱教師あり学習では, 類似しているデータペアは \mathcal{S} に, 非類似データペアは \mathcal{D} に所属させることで計量行列の導出を行う.

この制約条件を用いることで, 最適な計量行列の学習は, 例えば \mathcal{S} に所属する学習データペア間のマハラノビス距離の総和を小さくしながら, \mathcal{D} に所属する学習データペア間のマハラノビス距離の総和をできるだけ大きくすることにより行われる.

他方, 学習データ x_i, x_j, x_k の組に対し, x_i と x_j の距離が x_i と x_k の距離よりも小さいという条件を制約として計量行列を学習する手法も提案されている [87]. いま, このような制約を満たすデータ組の集合を \mathcal{T} とし, ベクトル x, x' 間の距離関数を $f(x, x')$ とすれば, \mathcal{T} は,

$$\mathcal{T} = \{(x_i, x_j, x_k) \mid f(x_i, x_j) \leq f(x_i, x_k)\}, \quad (2.5)$$

で与えられる.

教師ありメトリックラーニング手法はこれらの制約条件に加え、データの分類精度向上に寄与すると想定される任意の制約条件を課したもとの、適切な距離計量を学習する。本研究では学習データが所属しているカテゴリが所与である強教師あり学習の問題設定のもと、学習データのカテゴリ情報を用いた距離計量を学習する。

以下では、本研究で対象とする教師あり学習に基づくメトリックラーニング手法のうち、繰り返し処理を用いた手法として、最も基本的な手法である Xing et al. の手法 [86]、比較的高い分類精度を達成でき、ベンチマーク手法として広く用いられている Information Theoretic Metric Learning (以下, ITML) [22], Large Margin Nearest Neighbor (以下, LMNN) [85] についてその説明を行う。加えて、本研究で対象とする代表元を用い、解析的に解を得ることができる繰り返し処理を用いないメトリックラーニング手法として持橋らによる手法 [59] の説明を与える。

2.3.2 繰り返し処理を用いたメトリックラーニング

Xing et al. による手法 [86]

計量行列の学習に関する基本的な考え方は、前述の通り、類似データ間の距離を近づけながら、非類似データ間の距離を離すことにより用いる分析手法の精度を向上させるという概念に基づく。このような考え方は2003年に Xing ら [86] によって導入され、以降様々な手法が提案されている。

Xing らは前述の制約条件を用い、求めるべき計量行列 \hat{M} を以下の最適化問題を解くことで学習する手法を提案した。

$$\begin{aligned} \hat{M} &= \arg \max_M \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}} d_M(\mathbf{x}_i, \mathbf{x}_j), \\ \text{subject to } & \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}} d_M^2(\mathbf{x}_i, \mathbf{x}_j) \leq 1, \\ & M \succeq 0. \end{aligned} \quad (2.6)$$

上記の最適化問題を解くため、制約条件を満たす集合 $\mathcal{Q}_1, \mathcal{Q}_2$ を

$$\mathcal{Q}_1 = \left\{ M : \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}} (\mathbf{x}_i - \mathbf{x}_j)^T M (\mathbf{x}_i - \mathbf{x}_j) \leq 1 \right\}, \quad (2.7)$$

$$\mathcal{Q}_2 = \{ M : M \succeq 0 \}, \quad (2.8)$$

として定義し、これらの集合に対し距離を最小とする射影を繰り返し、 M を収束させる方法を示している。

Xing et al. による最適化手法のアルゴリズムについて Algorithm 1 に示す.

Algorithm 1 Xing et al. による計量行列の学習法

```

1: repeat
2:   repeat
3:      $M := \arg \min_{M'} \{\|M' - M\|_F\} : M' \in \mathcal{Q}_1$ 
4:      $M := \arg \min_{M'} \{\|M' - M\|_F\} : M' \in \mathcal{Q}_2$ 
5:   until  $M$  converges
6:    $M := M + \varepsilon \nabla_M g(M) \perp_{\nabla_A f(M)}$ 
7: until  $M$  converges

```

ただし, Algorithm 1 における $g(M), f(M)$ はそれぞれ

$$g(M) = \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}} d_M(\mathbf{x}_i, \mathbf{x}_j), \quad (2.9)$$

$$f(M) = \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}} d_M^2(\mathbf{x}_i, \mathbf{x}_j), \quad (2.10)$$

で定義される関数である. また, ε は微小な正定数を表し, \perp は垂直方向への射影を表す演算子とする. 加えて, Algorithm 1 の4行目は学習する計量行列が半正定値対称行列となることを保証するための処理であり, その固有値 λ_i を全て0以上の値とする処理を示す. いま, 正規直交行列 $\Gamma \in \mathcal{R}^{W \times W}$, 計量行列 M の固有値をその対角要素として持つ対角行列を $\Lambda = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_W] \in \mathcal{R}^{W \times W}$ とする. このとき, 計量行列 M は $\Gamma^T \Lambda \Gamma$ と固有値分解できる. Λ の要素のうち, 負の値を取るものを0とした行列 Λ' を $\Lambda' = \text{diag}[\max\{0, \lambda_1\}, \max\{0, \lambda_2\}, \dots, \max\{0, \lambda_W\}]$ とし, $\Gamma^T \Lambda' \Gamma$ を計算することでその半正定値性を保証する.

Xing et al. による手法は, 上記の計算を繰り返すことで最適な計量行列を学習することができるものの, 上記アルゴリズムの3, 4行目にある通り, アルゴリズム中に計量行列 M の最適化がネストされた構造になっている. また, 更新毎に固有値分解を実施するため収束するまでの計算コストが非常に大きいという問題点を含んでいる. 加えて, 一般のメトリックラーニング手法同様, 学習データの次元数が増加した際には計算コストがさらに増大すると共に, 過学習を引き起こし分類精度が低下してしまう可能性が高いことが知られている.

Information Theoretic Metric Learning (ITML) [22]

情報理論などで用いられる divergence [21] の考え方を距離構造の学習に援用した手法として、ITML がある。ITML では Xing らの手法と同様に類似データペア間の距離を小さくしながら、同時に非類似データペア間の距離を大きくするような距離構造を学習する。

ITML は以下の最適化問題を解くことにより、最適な計量行列を学習する。

$$\begin{aligned}
 & \underset{M}{\text{minimize}} \quad d_{LD}(M, M_0), \\
 & \text{subject to} \quad d_M(\mathbf{x}_i, \mathbf{x}_j) \leq u, \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}, \\
 & \quad \quad \quad d_M(\mathbf{x}_i, \mathbf{x}_j) \geq l, \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}, \\
 & \quad \quad \quad M \succeq 0.
 \end{aligned} \tag{2.11}$$

ただし、 u, l はそれぞれ類似（非類似）データ間の距離を一定値以上（以下）に制約するための正の値をとる閾値であり、目的関数の d_{LD} は、

$$d_{LD}(M, M_0) = \text{tr}(MM_0^{-1}) - \log \det(MM_0^{-1}) - W, \tag{2.12}$$

を満たす log det divergence [42] である。ただし、 $\text{tr}(\cdot)$ は行列のトレースを表すものとする。したがって、式 (2.11) は類似/非類似データ間の距離を制約条件とし、計量行列 M と目標とする行列 M_0 との log det divergence をできるだけ小さくするような最適化問題であると言える。一般的に、 M_0 には単位行列や分散共分散行列の逆行列が用いられる。

Large Margin Nearest Neighbor (LMNN) [85], [84]

LMNN は k -NN 法をベースとし、データ間の局所的な関係性を考慮した距離計量を学習するための手法であり、学習時には SVM [19], [74] で用いられるマージン最大化の考え方をを用いる。この手法は他のメトリックラーニング手法と比較し、高性能であることが知られており、広く用いられている。

LMNN における最適な計量行列は以下の最適化問題を解くことにより導出する.

$$\begin{aligned}
& \underset{M}{\text{minimize}} \sum_{ij} \gamma_{ij} d_M(\mathbf{x}_i, \mathbf{x}_j) + c \sum_{ijl} (1 - \delta_{il}) \xi_{ijl}, \\
& \text{subject to } d_M(\mathbf{x}_i, \mathbf{x}_l) - d_M(\mathbf{x}_i, \mathbf{x}_j) \geq 1 - \xi_{ijl}, \\
& \xi_{ijl} \geq 0, \\
& M \succeq 0.
\end{aligned} \tag{2.13}$$

ただし, ξ_{ijl} はスラック変数とし, c は $[0, 1]$ 上の実数値をとる重みパラメータであるとする. また, γ_{ij} はデータ \mathbf{x}_j がデータ \mathbf{x}_i の近傍かつ類似データだった場合には 1 を, そうでない場合には 0 をとるようなインジケータ関数とし, δ_{il} はデータ \mathbf{x}_i と \mathbf{x}_l が類似データならば 1 を, そうでない場合には 0 を取るようなインジケータ関数とする.

上記の最適化問題は, 対象とするデータ \mathbf{x}_i に対し, その周辺にあり同一カテゴリに所属する学習データを近づけ, 別カテゴリに所属する学習データをマージンを考慮して離すような距離計量をすべての学習データに対して求めることに相当する.

2.3.3 繰り返し処理を用いないメトリックラーニング

持橋らの手法 [59]

繰り返し処理を用いず, 解析的に最適解を得ることができるメトリックラーニング手法として持橋らの手法 [59] がある. 持橋らの手法では, 最適な計量行列を各カテゴリの代表元と同一カテゴリに所属する学習データとのマハラノビス距離の 2 乗の和を最小とすることで導出する.

いま, μ_n を n 番目のカテゴリの代表元と定義すれば, μ_n は

$$\mu_n = \frac{1}{|C_n|} \sum_{\mathbf{x}_i \in C_n} \mathbf{x}_i, \tag{2.14}$$

で定義される. ただし, $|C_n|$ はカテゴリ C_n に所属している学習データ数を表すものとする. これらを用いて, 以下の通り最適化問題を定式化する.

$$\begin{aligned}
\hat{M} &= \arg \min_M \sum_{n=1}^N \sum_{\mathbf{x}_i \in C_n} (\mathbf{x}_i - \mu_n)^T M (\mathbf{x}_i - \mu_n), \\
& \text{subject to } \det M = 1.
\end{aligned} \tag{2.15}$$

ただし, $\det M$ は行列 M の行列式を指すものとし, 制約条件にある $\det M = 1$ は $\det M$ が 0 となる自明な解を持たないようにするための制約条件である.

いま, 簡素化のため, S_n を,

$$S_n = \sum_{\mathbf{x}_i \in \mathcal{C}_n} (\mathbf{x}_i - \boldsymbol{\mu}_n)(\mathbf{x}_i - \boldsymbol{\mu}_n)^T, \quad (2.16)$$

とすれば, 式 (2.15) は

$$\hat{M} = \arg \min_M \sum_{n=1}^N \text{tr}(S_n M), \quad (2.17)$$

として表現することができる. 式 (2.15), (2.17) で定式化される最適化問題は, ラグランジュの未定乗数法を用いることで最適な計量行列 \hat{M} を解析的に得ることができ, \hat{M} の数式解は

$$\hat{M} = \det \left(\sum_{n=1}^N S_n \right)^{\frac{1}{N}} \left(\sum_{n=1}^N S_n \right)^{-1}, \quad (2.18)$$

で与えられることが示されている [59].

持橋らの手法は, 前述の通り探索的な繰り返し処理を用いず, 最適な計量行列を解析的に得ることが可能な手法である. したがって, Xing et al. の手法を始めとする繰り返し処理を用いる手法と比較して相対的に計算量が少ないという利点が存在している.

また, この持橋らの手法はフィッシャーの判別分析 [4] を多カテゴリに拡張したものと解釈できる. 2 カテゴリの場合の両者の等価性について以下に示す.

持橋らの手法とフィッシャーの判別分析の等価性について (カテゴリ数が 2 の場合)

カテゴリ数が 2 の場合 ($\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2\}$) のフィッシャーの線形判別分析では, 以下の関数 $h(\mathbf{x})$ の正負によって入力データのカテゴリを識別する.

$$h(\mathbf{x}) = w_0 + \mathbf{w}^T \mathbf{x}. \quad (2.19)$$

ただし, 重みパラメータを w_0, \mathbf{w} とした. このとき, フィッシャーの方法による重みパラメータ \mathbf{w} は, $\mathbf{w} \propto S_w^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ で与えられる. ここで, S_w は $S_w = (S_1 + S_2)$ とし, S_1, S_2 はそれぞれ $S_1 = \sum_{\mathbf{x}_i \in \mathcal{C}_1} (\mathbf{x}_i - \boldsymbol{\mu}_1)(\mathbf{x}_i - \boldsymbol{\mu}_1)^T$, $S_2 = \sum_{\mathbf{x}_i \in \mathcal{C}_2} (\mathbf{x}_i - \boldsymbol{\mu}_2)(\mathbf{x}_i - \boldsymbol{\mu}_2)^T$ とする.

したがって、式(2.19)に上記のパラメータを代入することで識別関数は、

$$\begin{aligned} h(\boldsymbol{x}) &= w_0 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top (S_w)^{-1} \boldsymbol{x} \\ &= w_0 + \boldsymbol{\mu}_1^\top S_w^{-1} \boldsymbol{x} - \boldsymbol{\mu}_2^\top S_w^{-1} \boldsymbol{x} \\ &= w_0 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top S_w^{-1} \boldsymbol{x}, \end{aligned} \quad (2.20)$$

として得られる。

一方、持橋らの手法では式(2.18)を用いて、識別関数を、

$$h_m(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu}_n)^\top M (\boldsymbol{x} - \boldsymbol{\mu}_n), \quad (2.21)$$

とし、 $\hat{C} = \arg \min_{C_n} h_n(\boldsymbol{x})$ となるカテゴリ \hat{C} に入力データを分類することに相当する。いま、カテゴリ数は2であるので、識別は $g_1(\boldsymbol{x}) < g_2(\boldsymbol{x})$ となった場合カテゴリ C_1 に、そうでない場合にはカテゴリ C_2 に分類する。このとき、 $h_1(\boldsymbol{x}) - h_2(\boldsymbol{x})$ は

$$\begin{aligned} h_1(\boldsymbol{x}) - h_2(\boldsymbol{x}) &= \boldsymbol{x}^\top M \boldsymbol{x} - 2\boldsymbol{\mu}_1^\top M \boldsymbol{x} + \boldsymbol{\mu}_1^\top M \boldsymbol{\mu}_1 - \boldsymbol{x}^\top M \boldsymbol{x} + 2\boldsymbol{\mu}_2^\top M \boldsymbol{x} - \boldsymbol{\mu}_2^\top M \boldsymbol{\mu}_2 \\ &= -2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top M \boldsymbol{x} - C, \end{aligned} \quad (2.22)$$

となる。ここで、 C は \boldsymbol{x} によらない定数である。

いま、式(2.18)を(2.22)に代入すれば、式(2.22)は、

$$\begin{aligned} h_1(\boldsymbol{x}) - h_2(\boldsymbol{x}) &= -2(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \det(S_w)^{\frac{1}{W}} (S_w)^{-1} \boldsymbol{x} - C \\ &= -2 \det(S_w)^{\frac{1}{W}} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top (S_w)^{-1} \boldsymbol{x} - C \\ &\propto (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top (S_w)^{-1} \boldsymbol{x}, \end{aligned} \quad (2.23)$$

となる。

したがって、式(2.20), (2.23)より、カテゴリ数が2の場合に本質的に両者の内容は等価である。

2.3.4 メトリックラーニングの問題点

上記の繰り返し処理を用いるメトリックラーニング手法、用いないメトリックラーニング手法はそれぞれ異なる制約条件を用いて最適な計量行列を学習することが可能となる。しかしながら、第1章にて述べた通り、以下のような問題点が存在している。

問題点 1 学習データ数, ならびにその次元数の増加に伴う計算量の増加.

問題点 2 学習データの次元数増加に伴うパラメータ数の増加.

問題点 3 大域的な距離計量のみを学習することによる局所的構造への不適合.

前述の繰り返し処理を用いたメトリックラーニング手法では, 特に学習データの次元数が大きい場合に上記の問題点 1, 2, 3 のいずれにも対応することができない.

他方, 問題 2 に対して, 今日では繰り返し処理を用いるメトリックラーニング手法に対し, 正則化 [30] を用いることでロバストなパラメータ推定を行う手法も存在している [17], [41], [43], [44], [50], [64]. これらの手法を用いることで学習データ数がパラメータと比較し少ない状況でも計量行列の学習が可能となるものの, そのいずれも繰り返し処理を用いて計量行列を導出するため, 大規模なデータを対象とした場合, 上記の問題点の解決には至らない.

これに対し, 持橋らの手法では解析的に最適解を得ることができるため, 上記の問題点 1 の改善は可能であるものの, 第 1 章にて述べた通り, 以下の問題点が存在する.

持橋らの手法の問題点 1

高次元なデータを対象とした場合, もしくは次元数に対し, 著しく学習データが少ない場合, 最適な計量行列の存在が保証されていない.

持橋らの手法の問題点 2

学習データの構成要素間の一部に統計的関係性がないデータを対象とした場合, 導出された最適な計量行列ではこれを適切に表現できない.

持橋らの手法の問題点 3

学習データのカテゴリ毎に統計的特徴が異なるデータを対象とした場合, その局所的構造を表現できない.

これらの問題点は, 持橋らの手法により得られる最適な計量行列 \hat{M} が式 (2.18) により導出されることに起因している. 持橋らの手法では, 式 (2.18) における $\sum_{n=1}^N S_n$ が非正則になってしまった場合, \hat{M} の導出が不可能となるため, このような条件下では最適な計量行列を求めることができない. また, $\sum_{n=1}^N S_n$ はその要素に何らかの値を保持する密な行列になることから, その逆行列により算出される最適な計量行列もまた密な構造となる. このため, 学習データの成分間に統

計的關係性がない場合でも何らかの重みを付与してしまうと考えられ、このような構造を適切に表現することができないと考えられる。同様に、式 (2.18) により得られる計量行列は学習データ全体に対して唯一のものであるため、学習データの局所的な構造には対応することができない。

上記の議論より、持橋らの手法におけるこれらの問題点を改善することは、相対的に計算量が少ないという持橋らの手法の利点を保持しつつ、対応可能なデータの幅を広げ多様なデータに対する分析を可能にすると考えられる。

2.4 周辺分野における研究動向

本研究で提案する手法は、正則化に基づく方法、並びに複数の計量行列の学習に基づく方法の二つのアプローチにより構成されている。

以下本節では、3章以降で提案する各手法と、本研究の周辺分野で既に存在する、

- 正則化を用いた機械学習手法
- 多変量解析（重回帰分析、線形判別分析）における手法

との関連性について述べる。

2.4.1 正則化を用いた機械学習手法との関連

近年のデータの大規模化、多様化に伴い、例えば生物統計学で用いられるような遺伝子データなど、学習データ数と比較して極めて大きな次元数を持つようなデータが存在している。このようなデータに対して、既存の機械学習手法を直接適用してしまうと、(次元数と比較して相対的に数の少ない) 学習データに過度にパラメータが適合してしまう過学習の問題が発生してしまう。過学習が生じた場合、前述の通り、学習データに過度にフィットしたパラメータが推定されてしまうため、推定したパラメータを用いて新規入力データを分類、予測しようとした場合、その精度が大きく低下してしまう可能性が存在している。

このような問題点に対し、学習データのパラメータへの適合度合いに応じて何らかのペナルティを加えることにより、過学習を抑える正則化と呼ばれる方法が存在している [30]。ペナルティ項（以下、正則化項）としては、推定するパラメータの l_1 ノルムや l_2 ノルムが用いられることが比較的多い。

前者の l_1 ノルムを正則化項として用いた場合、学習データの特徴量のうち、不要な項を 0 とするようなパラメータが推定される。これにより、学習するパラメータ数の削減が可能となる。これは、多変量解析などで行われる変数選択を学習データの内容に応じて自動的にやっていることと等価であると言えるものの、 l_1 ノルムは原点において非連続であることが知られている。このため、後述する l_2 ノルムを用いた正則化とは異なり、解析的に最適解を得ることが難しく、計算には何らかの探索的な最適化アルゴリズムを適用することが必要となる。

他方、 l_2 ノルムを正則化項として用いた場合、推定するパラメータ値の大きさに対して罰則を与えるような推定を行うため、過度に大きな値となることを妨げる効果がある。 l_1 ノルムを正則化項として用いた場合と異なり、 l_2 ノルムは微分可能であるため、最適解を解析的に得られることが多い。

これらの他にも l_1, l_2 ノルム両者を正則化項として用いる Elastic Net[23] 型の正則化を用いた機械学習手法も存在している。

このような正則化手法は、ほとんどの機械学習手法への適用が可能であり、本研究で対象とするメトリックラーニング手法に対しても正則化を行った手法が提案されている [22], [44], [52], [64]。これらの手法と本研究との大きな違いは、最適化問題を定式化する際にカテゴリの代表元を使用している点にある。従来の手法では、類似、非類似データの集合 S, D に属する全てのデータペアを制約条件として用いているため、学習データ数に応じて制約条件の数が増加してしまう。一方で、本研究では代表元を用いることにより制約条件の個数を削減することが可能となっており、この点は従来のメトリックラーニング手法との大きな差異であると言える。

2.4.2 多変量解析（重回帰分析，判別分析）手法との関連

既に述べた通り、本研究で着目している持橋らの手法は、フィッシャーの線形判別分析 [4] を多値分類へ拡張したものであると考えることができ、両者には様々な類似点が存在している。また、得られたデータの統計的特徴を基にして線形モデルを用いて予測を行うという点では、多変量解析との関連性も存在することから、フィッシャーの線形判別分析と連続値の予測を行う重回帰分析との関連性について述べる。

重回帰分析は連続型の目的変数をよりよく予測するための説明変数の回帰パラメータを最小二乗推定する手法であり、本研究で対象とする分類の問題とは目的が異なる。また、重回帰分析における正則化手法は lasso[77] やリッジ回帰 [11] に代表されるように、本研究で用いる正則化との

表 2.2: 本提案手法と重回帰分析, フィッシャーの判別分析との関係性

	l_1 正則化	l_2 正則化	目的	学習の対象
重回帰分析	lasso	リッジ回帰	予測 (目的変数: 連続値)	ベクトル
判別分析	スパース判別分析	正則化判別分析	二値分類 (目的 変数: 離散値)	ベクトル
メトリックラーニング	提案手法 2	提案手法 1	多値分類	行列

共通点が多い。前者の lasso は、正則化項として回帰係数ベクトルの l_1 ノルムを付与したものであり、不要な回帰係数を 0 とするような推定量を与える手法である。後者のリッジ回帰は、回帰係数の l_2 ノルムを正則化項として付与したものであり、特に回帰分析を行う際に問題となる多重共線性の解決を可能とする手法である。

一方、識別関数を用いた手法として、フィッシャーの判別分析が知られている。これは、データの二値分類に用いられる手法であり、正例、負例を識別するための識別超平面を構成するための方法といえることができる。フィッシャーの判別分析に対しても、大規模なデータや学習データの変数間に関係性が薄いデータを対象とするために正則化を行う手法が存在している [18], [26]。

このように、重回帰分析、フィッシャーの判別分析共に、多様なデータへの適用を可能とするため、本研究同様に正則化を行っている。本提案手法と、これらの手法の正則化の観点から見た関係性は表 2.2 の通りとなる。

これらの手法と、本研究で対象とするメトリックラーニング手法との大きな違いは、学習するパラメータがベクトルか、行列かという点にある。すなわち、重回帰分析、フィッシャーの判別分析は共にデータを予測、識別するための超平面を学習しようとしているが、本研究で対象とするメトリックラーニングはデータを識別するための空間を学習していることに相当し、これらの手法との大きな差異であるといえることができる。

第3章 代表元を用いた計量行列の正則化学

習 [55]

3.1 本章の目的

既に述べた通り，学習データが与えられたもとで計量行列の学習は

- 繰り返し処理を用いた学習
- 繰り返し処理を用いない学習

に大別される。

両者に共通する問題点として，計量行列の要素数が学習データの次元数の2乗に比例するため，高次元なデータを対象とした場合，十分なパラメータの学習に必要なデータ数が膨大になってしまう点あげられる。

一方で，繰り返し処理を用いたメトリックラーニングの各種手法は得られた計量行列を用いたマハラノビス距離が距離の公理を満たす必要があるため，学習した計量行列が半正定値対称行列となることを保証する必要がある。このため，学習アルゴリズムにおける計量行列の更新の際には，固有値分解を使用し，計量行列の全固有値を0以上にするという処理を用いることが一般的である。固有値分解に要する計算量は $O(W^2) \sim O(W^3)$ であり，これを毎回の更新時に行うことは特に入力データが高次元である場合，最適な計量行列を現実的な時間で得ることが難しくなってしまう。

また，持橋らの手法では第1章で述べた通り，学習データが高次元でデータ数が少ない場合，最適な計量行列を解析的に得ることができない。このような問題は，扱うデータが多様化，高次元化している今日では実用上の大きな制約であると考えられ，その解決が望まれる。

本章では，上記の問題点を解決するため，繰り返し処理を用いずに解析的に最適な計量行列を得ることができる持橋らの手法をベースとし，最適な計量行列の存在が保証された手法の提案を行う。具体的には，機械学習の分野で盛んに研究されている正則化手法 [30] を適用することで，持橋

らの手法 [59] に対し、その利点を保持したまま、計量行列が必ず一意に存在することを示す。併せて、導出した計量行列を用いることで高い分類精度を達成することを示し、最適な計量行列を持つ特徴を述べる。

本章で提案する手法の有効性を示すため、高次元スパースな文書データセットを用いた実験を行い、分類精度の評価を行う。

3.2 代表元を用いた計量行列の正則化法

既に述べた通り、持橋らの手法 [59] は繰り返し処理を用いずに最適な計量行列を導出することが可能となるものの、式 (2.16) における行列 $\sum_{n=1}^N S_n$ が正則であることは保証されていない。

したがって、持橋らの手法を用いる際には、 $\sum_{n=1}^N S_n$ が正則であるという条件が必要となるため、学習データの総数とその次元数について注意を払う必要がある。いま、行列の階数を $\text{rank}(\cdot)$ とする。このとき、一般に $\sum_{n=1}^N S_n$ の階数 $\text{rank}(\sum_{n=1}^N S_n)$ は学習データが互いに独立であるという条件のもと、以下を満たすことが知られている。

$$\text{rank}\left(\sum_{n=1}^N S_n\right) = \min(D, W).$$

ただし、 D は学習データの総数とする。したがって、 $D < W$ となる場合、すなわち、学習データの総数よりもその次元数が大きい場合、 $\sum_{n=1}^N S_n$ は非正則な行列となってしまう、その逆行列は存在しない。このため、式 (2.15) における $\det M = 1$ という制約条件と矛盾してしまうことになる。

以下では、入力データが特に高次元な場合、 $(\sum_{n=1}^N S_n)$ が非正則な行列となりやすいことに着目する。この問題を解決し、逆行列の存在を保証するため、本章では持橋らの手法で定義された最適化問題に対し、適切な正則化項を追加することを考える。正則化項としては、例えば行列のフロベニウスノルムやトレースなど、様々なものが機械学習の分野では用いられている。しかしながら、その種類によっては最適な計量行列を解析的に求めることができず、繰り返し処理を用いる必要が生じてしまう。繰り返し処理を用いる最適化問題はその計算量の多さが一つの問題であり、上記のような $\sum_{n=1}^N S_n$ が非正則となるような場合、すなわち学習データの次元数とその総数よりも大きい場合には特にその問題は顕著であるといえる。既に述べている通り、持橋らの手法の利点の一つは解析的に最適な計量行列を導出できるというところにある。したがって、この

条件を満たした適切な正則化項を導入することは、分類精度を向上させながら上記の問題点を解決できる可能性がある。

上記の議論のもと、解析的に解を得ることができる正則化項を検討するために以下の最適化問題を考える。

$$\underset{M}{\text{minimize}} \text{tr}(M), \quad (3.1)$$

$$\text{subject to } \log(\det M) = 0. \quad (3.2)$$

ここで、計量行列 M の固有値をそれぞれ、 m_1, m_2, \dots, m_W とすると、 $\text{tr}(M) = \sum_{i=1}^W m_i$ 、 $\det M = \prod_{i=1}^W m_i$ が成立する。したがって、上記の最適化問題は計量行列の固有値を $m_1 = m_2 = \dots = m_W = 1$ に近づけるような計量行列を学習することと解釈できる。したがって、制約条件である式 (3.2) のもとで $\text{tr}(M)$ を式 (2.15) に正則化項として導入することは、直感的には計量行列 M の固有値をすべて 1 に近づける効果がある。ここから、計量行列が非正則になる、換言すればいずれかの固有値が 0 となる条件を回避できる可能性が向上し、最適な計量行列が必ず一意に定まることが想定される。

いま、正則化パラメータ η を $\eta \geq 0$ とし、持橋らの手法に対し、正則化項を $\text{tr}(M)$ として追加すれば、最適な計量行列 \hat{M} は以下の最適化問題を解くことで得ることができる。

$$\begin{aligned} \hat{M} &= \arg \min_M \left(\sum_{n=1}^N \text{tr}(S_n M) + \eta \cdot \text{tr}(M) \right), \\ &\text{subject to } \log(\det M) = 0, \\ &M \succeq 0. \end{aligned} \quad (3.3)$$

この最適化問題は持橋らの手法同様に解析解を導出でき、ここから以下の定理を得る。

定理 3.1 (最適な計量行列)

式 (3.3) で定義した最適化問題により得られる最適な計量行列は

$$\hat{M} = \det(A)^{\frac{1}{W}} A^{-1}, \quad (3.4)$$

により得られる。ただし、行列 $A \in \mathcal{R}^{W \times W}$ は以下の式を満たすものとする。

$$A = \sum_{n=1}^N S_n + \eta I. \quad (3.5)$$

ここで、 I は $W \times W$ の単位行列とする。 □

(証明) 3.5.1 節参照.

本章で導出した式 (3.4) の最適な計量行列と持橋らの手法で求められた式 (2.18) の計量行列の違いは, η が行列 $\sum_{n=1}^N S_n$ の対角要素に加算されているかどうかにある. $\sum_{n=1}^N S_n$ に ηI が加算されているため, A の逆行列の存在が保証されることとなり, 最適な計量行列を一意に算出することが可能となる.

上記の議論から, 式 (3.4) に対して以下の補題を得ることができる.

補題 3.1 (最適な計量行列の正定値対称性)

$\eta > 0$ のとき, 式 (3.4) で導出される最適な計量行列 \hat{M} は正定値対称行列となる. すなわち, 式 (3.4) は必ず一意に存在することが保証される. \square

(証明) 3.5.2 節参照.

さらに, 式 (3.3) で定義される最適化問題は M について下界を持つことがわかり, 以下の補題が成立する.

補題 3.2 (式 (3.3) で定義した最適化問題における下界の存在保証)

式 (3.3) で定義した最適化問題は下界を持つ. \square

(証明) 3.5.3 節参照.

また, 式 (3.3) により定義した最適化問題は計量行列の固有値の分散を抑える効果があることもまたわかる. ここから, 正則化項として $\text{tr}(M)$ を追加することは計量行列により射影される空間を過度に大きく変更させない作用があることがわかる.

上記の議論から, 正則化パラメータ η と最適な計量行列の固有値との関係性について以下の定理が成立する.

定理 3.2 (正則化パラメータ η と最適な計量行列 \hat{M} の固有値の関係性)

いま, $\hat{m}_k, k = 1, 2, \dots, W$ を式 (3.4) で与えられる最適な計量行列 \hat{M} の固有値とし, その平均 \bar{m} を $\bar{m} = \frac{1}{W} \sum_k \hat{m}_k$ とする. このとき, 以下の性質が成り立つ.

η が増加した場合,

$$\left\{ \begin{array}{l} \bar{m} < \hat{m}_k \text{ が成立するとき, } \hat{m}_k \text{ は減少し,} \\ \bar{m} > \hat{m}_k \text{ が成立するとき, } \hat{m}_k \text{ は増加する.} \end{array} \right.$$

ただし, $\hat{m}_1, \dots, \hat{m}_W, \bar{m}$ は η に関する関数であることに注意する. \square

(証明) 3.5.4 節参照.

上記の定理 3.2 から, 以下の系がただちに成立する.

系 3.1 (最適な計量行列の固有値の取り得る範囲と分散)

最適な計量行列 \hat{M} の固有値の取りえる範囲と分散は, すべての固有値が等しい場合を除き, η の増加に比例して単調に減少する. \square

(証明) 3.5.5 節参照.

系 3.2 (最適な計量行列の固有値と正則化パラメータ η との関係性)

$\eta \rightarrow \infty$ のとき, $\hat{m}_k \rightarrow 1$ がすべての k について成立する. \square

(証明) 3.5.6 節参照.

上記の系 3.1 より, 正則化パラメータ η が増加した場合, 最適な計量行列の各固有値のとり得る範囲は減少することが分かる. 加えて, 系 3.2 より, 式 (3.4) において正則化パラメータ η が増加した場合, 最適な計量行列は単位行列に漸近していくこともまた分かる. この場合, 導出した最適な計量行列を用いたマハラノビス距離はユークリッド距離と等価となる¹. したがって, 正則化パラメータ η を変化させることは, 用いる距離尺度を学習データの統計的特徴を最大限反映したマハラノビス距離とユークリッド距離の間で調整する効果があることがわかる. 加えて, 学習データ数 D が無限, すなわち $D \rightarrow \infty$ とすることが可能であると仮定した場合, 正則化項の役割は徐々に小さくなり, 本章で定義した最適化問題は持橋らの手法と等価となる.

3.3 実験

本章で提案した手法の有効性を確認するため, 新聞記事データを使用した分類実験を行った. 以下, 本節ではこれらの実験の条件について述べるとともに, その結果について示す.

¹いま, 式 (2.3) における計量行列 M を単位行列 I とすれば,

$$d_I(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T I (\mathbf{x} - \mathbf{y})} = \sqrt{\|\mathbf{x} - \mathbf{y}\|_2^2},$$

となる. このとき, マハラノビス距離とユークリッド距離は等価となる.

3.3.1 実験条件

本節では2000年、2005年に発行された毎日新聞記事を用いた文書分類実験を行う。文書分類実験は以下に示す2種類から構成されている。

- 正則化パラメータ η の効果を調べるための実験
- 他のメトリックラーニング手法との比較実験

両者の実験ともに、上記の新聞記事データから5カテゴリを抽出し、各カテゴリの学習データを100, 300, 500件と変化させるとともに、すべてのカテゴリに対し、500件のテストデータ（合計2,500件）を使用するものとした。評価尺度は分類精度とし、すべての実験は5回繰り返した平均とする。

なお、正則化パラメータ η に関しては、最適なその値を決定するための方法には一般的には存在しない。したがって、本研究では、事前に10分割交差検定 [30] を行うことにより、適切なパラメータを検討するものとした。

2000年、2005年の新聞記事に対する10分割交差検定の結果を図3.1, 3.2にそれぞれ示す。

図3.1, 3.2の凡例である，“Train”は距離計量の学習に使用したデータに対する分類精度を示し，“Test”は評価用データ²に対する分類精度を示すものとする。図3.1, 3.2から、評価データに対する分類精度は η が100～200近傍の場合が相対的に高いことが分かる。一方で、学習データに対する分類精度は正則化パラメータ η が相対的に小さい場合には比較的高く、その値が大きくなるにつれて低くなっていくことが分かる。これらより、 η を200程度に設定することにより適切な分類精度を達成できることが分かる。

10分割交差検定を実施することで決定した正則化パラメータを使用し、テストデータに対する分類精度を測定する実験を行った。本実験では、文書ベクトルの素性として、各異なり単語の出現頻度である *tf* measure [48] を用いたもの、ならびに各異なり単語の出現頻度に、ある単語の文書出現頻度の逆数の対数値である *idf* 値 [66] を乗じたものをそれぞれ用いた。ただし、 w 番目の素性 x_w に対する *idf* 値、 $idf(w)$ は以下の式で算出する。

$$idf(w) = \log \frac{D}{df(w)}. \quad (3.6)$$

²交差検定を行う際には、学習データを学習用のデータ、ならびに評価用のデータへ分割する。学習した分類器を用いて、学習用データを分類した場合には“Train”，評価用データを分類した場合には“Test”と記載することとする。

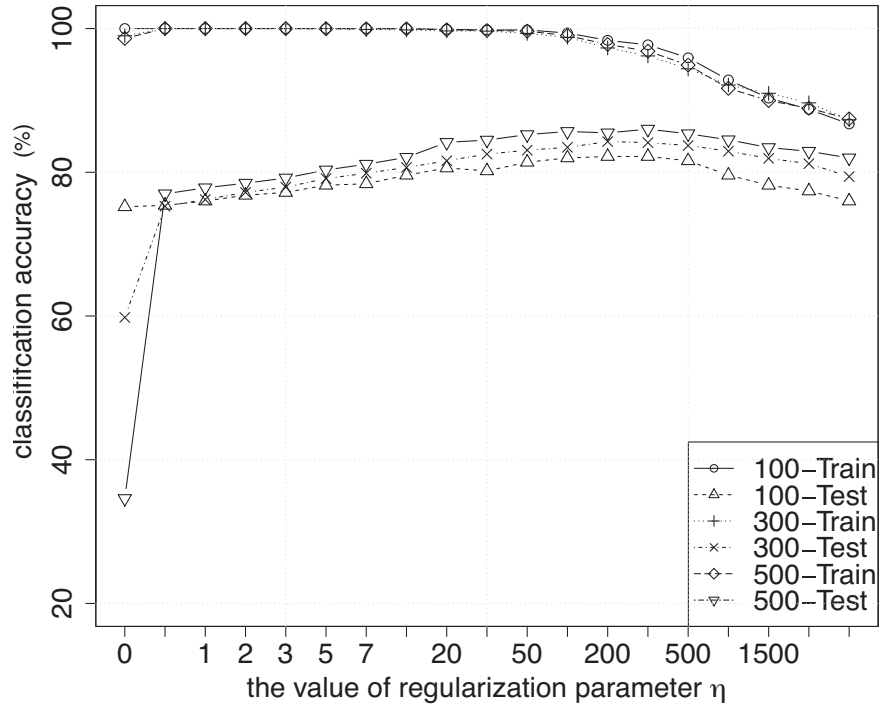


図 3.1: 10 分割交差検定の結果 (毎日新聞 2000 年) [55]

ここで, $df(w)$ は, w 番目の素性出現する学習データの数を示すものとする. また, 正則化パラメータ η は 0.5 から 3000 へと変化させるものとした.

また, 他のメトリックラーニング手法との比較実験では性能比較のため, 前述の ITML [22], LMNN [85], 持橋らの手法 [59] に加え, 通常ユークリッド距離, コサイン尺度とも分類精度の比較を行った. ただし, LMNN, ITML では $k = 3$ とした k -NN を用いるものとし, 繰り返し回数の最大値を 1,000 回とした. それ以外の手法における分類規則はテンプレートマッチング [25] を用いた. いま, テストデータを \mathbf{x} とすれば, テンプレートマッチングによる分類規則は,

$$\hat{C} = \arg \min_{C_n} d_{\hat{M}}(\mathbf{x}, \mu_n), \quad (3.7)$$

として与えられる.

実験の基本的な手順は以下の通りとなる.

Step 1) 各カテゴリの学習データを使用して, 代表元を算出する.

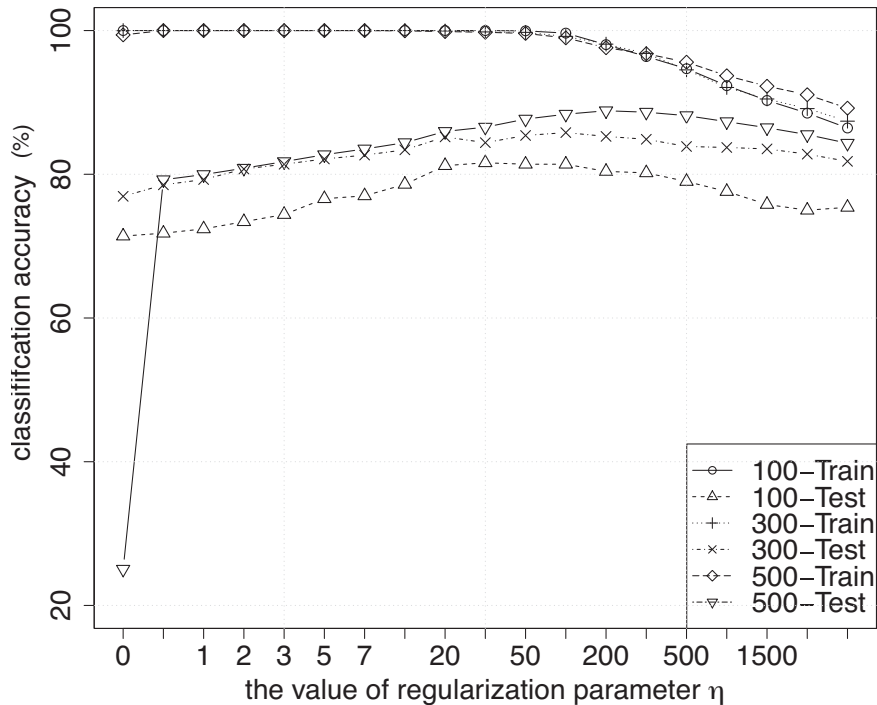


図 3.2: 10 分割交差検定の結果（毎日新聞 2005 年） [55]

Step 2) 式 (3.4) を用いて，最適な計量行列 \hat{M} を学習する．

Step 3) 各テストデータと各カテゴリの代表元とのマハラノビス 2 乗距離を測定する．

Step 4) テストデータを式 (3.7) により，最もマハラノビス 2 乗距離が小さい代表元が所属するカテゴリに分類する．

Step 5) すべてのテストデータの所属カテゴリを推定し，その精度を算出する．

3.3.2 実験結果

正則化パラメータの違いによる分類精度の違い

正則化パラメータの変化による分類精度の差を図 3.3 から図 3.6 に示す．

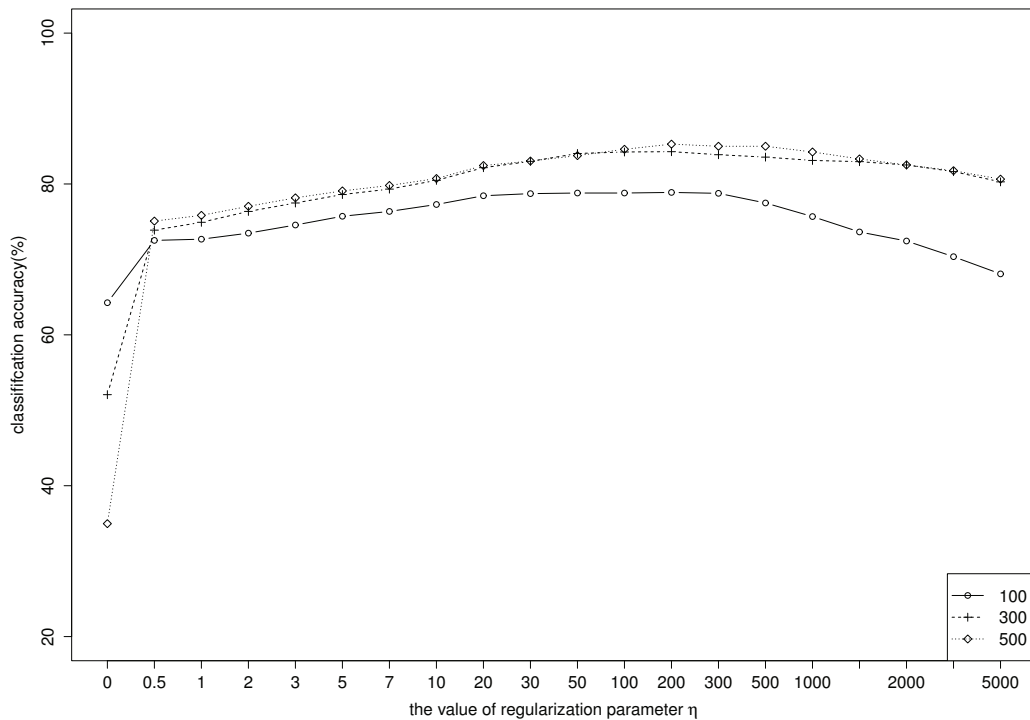


図 3.3: 正則化パラメータ η の変化に対する分類精度への影響 (毎日新聞 2000 年, tf measure) [55]

ただし, 図 3.3, 図 3.4 は 2000 年発行の毎日新聞記事を用いたものであり, 図 3.5, 図 3.6 は 2005 年毎日新聞記事を用いたものを示す. このうち, 図 3.3, 図 3.5 は tf measure を用いたもので, 図 3.4, 図 3.6 は $tf-idf$ measure を用いたものを示す. なお, 図 3.3, 図 3.5 において $\eta = 0$ とした場合は, 正則化項がないものとみなせるため, 提案手法は従来の持橋らの手法と等価となる.

これらの実験結果より, 本章で提案した手法は持橋らの手法と比べ, 高い分類精度を達成できていることが分かる. 加えて, 持橋らの手法では特に学習データの数が増加した場合, 分類精度が低下している. これは, 学習データの総量が増加した場合, より多くの単語が学習データ中に出現するため, その次元数が増加し過学習が発生したことに起因すると考えられる.

図 3.3, 3.5 にある通り, 素性として単語出現頻度 (tf measure) を使用した場合, 正則化パラメータの変化と分類精度の変化には大きな関係性はないことが見て取れる. 前述の 10 分割交差検定法の結果から, 適切な正則化パラメータは $\eta = 200$ 程度であったが, 本実験の結果からも同様の傾向

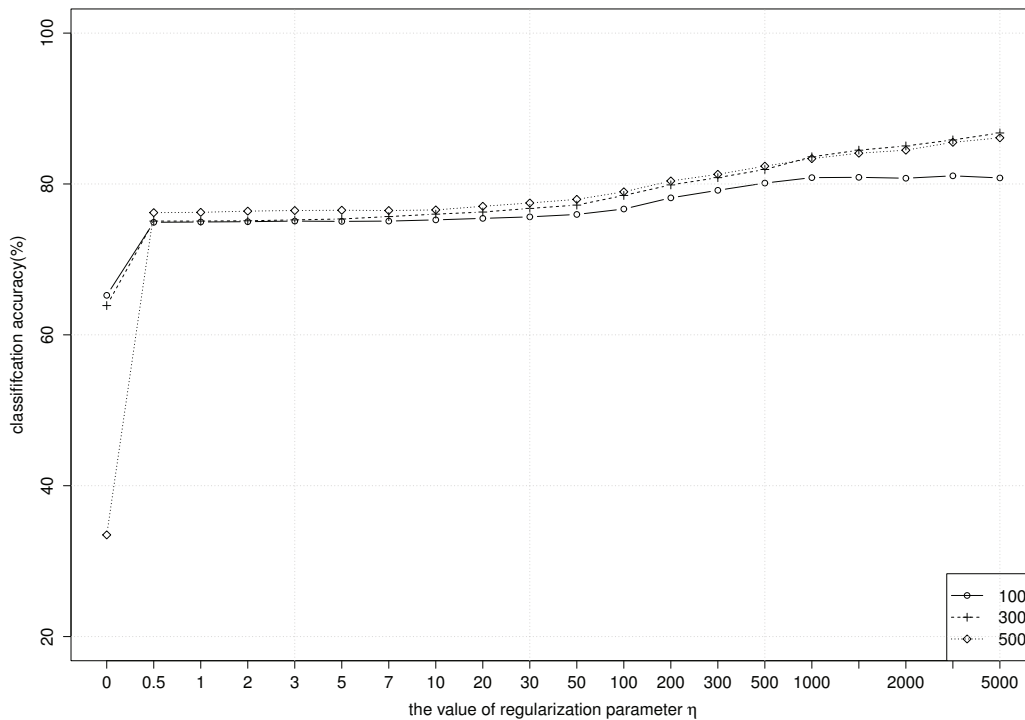


図 3.4: 正則化パラメータ η の変化に対する分類精度への影響 (毎日新聞 2000 年, $tf-idf$ measure) [55]

が見て取れ, 適切な正則化パラメータの設定が高い分類精度の達成につながることも分かる. したがって, 本手法を用いる場合には事前に何らかの方法で適切な正則化パラメータを決定する必要があるが, その一つの手法として, 交差検定が有効であることも示された. しかしながら, 交差検定を事前に実施することは, 最適な計量行列を導出するために式 (3.5) で定義された A の逆行列の計算を伴うことから, 特に本実験のように高次元なデータを対象としている場合, 計算量の面で大きな課題を残している. 他方, 正則化パラメータ η を $\eta > 0$ として用いた場合, その差異による分類精度への影響は過度に大きいものではなく, 交差検定の使用などが計算機資源の制約などで難しい場合には, $\eta = 50 \sim 200$ 程度の比較的大きい値を設定しておくことで, ある程度の分類精度を達成できることが示唆される.

一方で, 図 3.3, 図 3.4 より (図 3.5, 図 3.6 も同様の傾向を示す), 素性として tf measure を用いた場合と, $tf-idf$ measure を用いた場合では, 正則化パラメータ η を変化させた場合の分類精

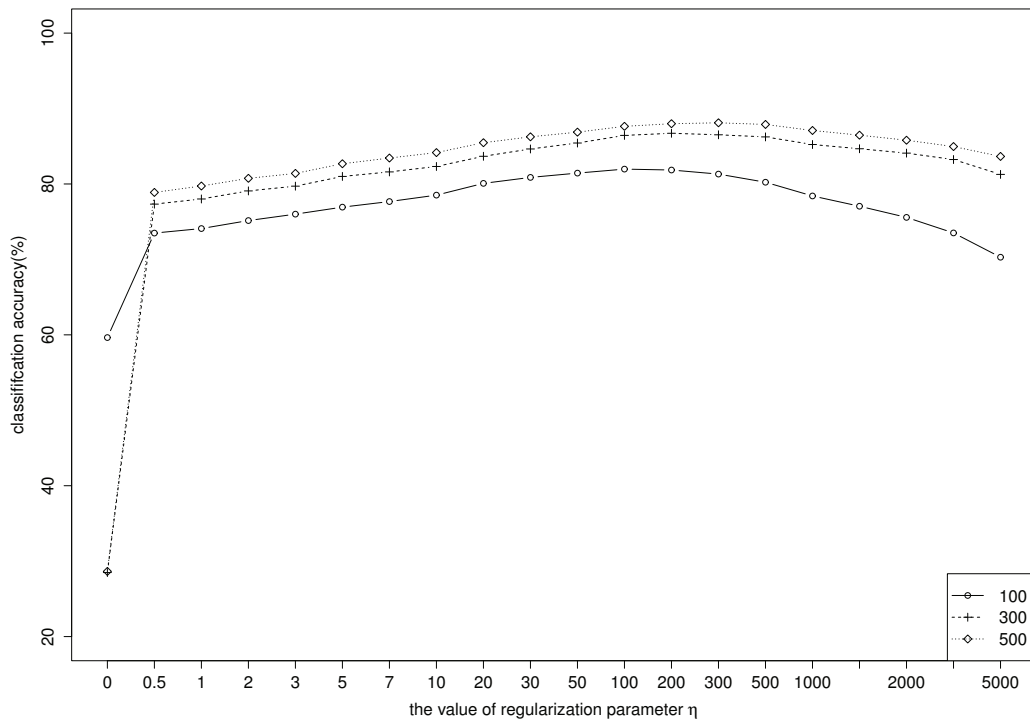


図 3.5: 正則化パラメータ η の変化に対する分類精度への影響 (毎日新聞 2005 年, tf measure) [55]

度の変化の傾向が大きく異なることがわかる。また、正則化パラメータが比較的小さい場合、素性として tf measure を用いた方が分類精度は高く、逆に比較的大きな場合は $tf-idf$ measure を用いた方が分類精度は高いことが分かる。 $tf-idf$ measure を用いた場合、 idf 値のように各単語の文書集合中の特徴を表現する素性を用いることは、計量行列の学習に悪影響を与えてしまう可能性がある。これは正則化パラメータを大きくした場合、すなわち、用いる距離尺度がユークリッド距離に近づく場合に、 $tf-idf$ measure における分類精度が向上していることから示唆される。また、 tf measure を用いた場合には全文書中での統計的特徴を一切考慮していないことから、マハラノビス距離を用いることで成分間の関係性を表現することができたと考えられる。

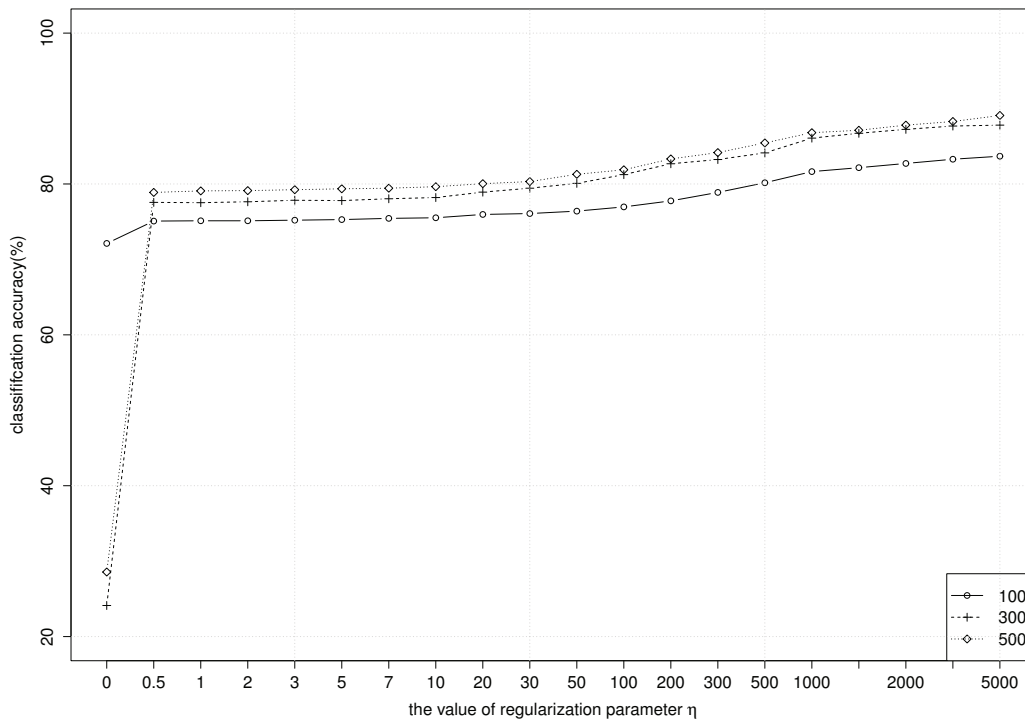


図 3.6: 正則化パラメータ η の変化に対する分類精度への影響 (毎日新聞 2005 年, $tf-idf$ measure) [55]

他手法との分類精度の比較

他の従来手法との分類精度の比較実験の結果を図 3.7, 図 3.8 に示す. ただし, 図 3.7 は 2000 年の毎日新聞記事の結果を示し, 図 3.8 は 2005 年の毎日新聞記事に対する結果を示すものとする. この実験では, 正則化パラメータ η を両者が比較的高い分類精度となる $\eta = 100$ とした. 図 3.7, 図 3.8 の横軸における “Pro_t” は素性として tf measure を用いた場合, “Pro_i” は $tf-idf$ measure を用いた場合を示し, “Euc”, “Cos” はそれぞれユークリッド距離, コサイン尺度を用いた場合の分類精度を示すものとする. これらの図から, 提案手法は他の従来手法と比較して高い分類精度を示していることが分かる.

以上より, 本章にて提案した手法は学習データの次元が大きな場合, 分類精度の面で有効であると共に, 繰り返し処理を用いないため, ITML, LMNN に代表される繰り返し処理を用いた手法と比較した場合, 計算コストの面でも有効であることが分かる.

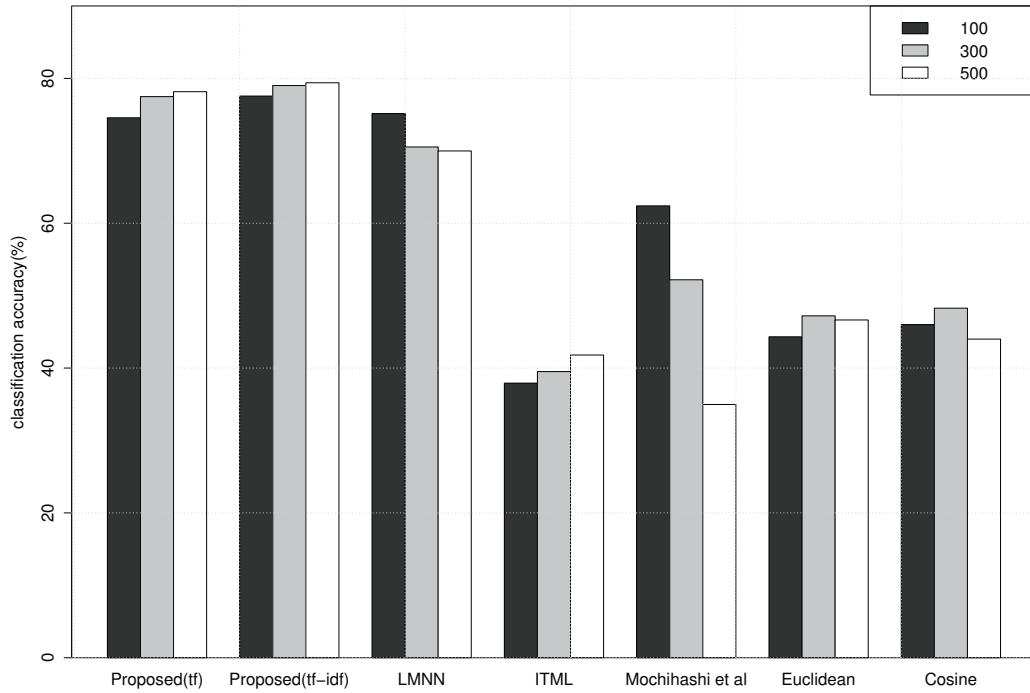


図 3.7: 他手法との分類精度の比較 (毎日新聞 2000 年) [55]

3.4 考察

本章では、持橋らの代表元を用いた手法に対し、正則化項として $\text{tr}(M)$ を用いた場合の最適化問題を定義し、式 (3.5) より、行列 A が正則となることが保証され、最適な計量行列が常に一意に求まることを示した。提案した手法を用いれば $\sum_{n=1}^N S_n$ が非正則な場合、例えば学習データ数 D がその次元数 W よりも少ない場合でも逆行列を導出可能となり、持橋らの手法と比較して分類精度の面からも有効であることが分かる。

また、他の繰り返し処理を用いる従来手法では、計量行列の半正定値性を保証するため、半正定値計画 [14] に基づくアルゴリズムを採用している。このため、毎回の更新時に固有値分解を行う必要があり、最適な計量行列を導出するためには多くの時間を要する。一方、本章にて提案した手法の計算量は、特に高次元なデータを使用した場合には相対的に低いことが分かる。

また、補題 3.1 より、提案手法では $\sum_{n=1}^N S_n + \eta I$ が正定値対称行列となることが保証され、常に逆行列を持つことが保証される。一方で、持橋らの手法では $\sum_{n=1}^N S_n$ が非正則となる可能性が

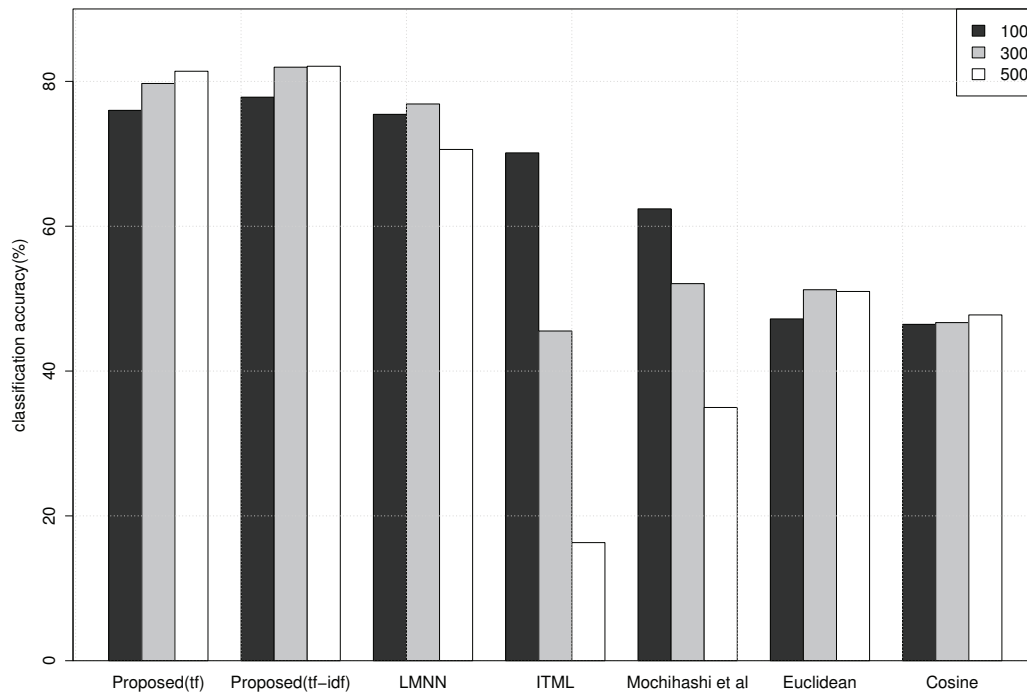


図 3.8: 他手法との分類精度の比較（毎日新聞 2005 年） [55]

存在しているため、これが保証されない。この点は本手法と持橋らの手法の大きな違いであるといえることができる。

計量行列の要素は各カテゴリの代表元とそこに所属する学習データとのマハラノビス距離を最小にするという意味での重みと解釈できる。したがって、以下では提案手法にて導出した最適な計量行列 \hat{M} の各要素の値について確認を行う。具体的には、その対角要素の取る値についてを確認し、さらにすべての要素がどのような値を取るのかを確認することとした。前者を図 3.9 に、後者を図 3.10 に示す。なお、両者の図共に、各カテゴリ 100 件、正則化パラメータ η を $\eta = 3$ とした場合の計量行列を使用した。

図 3.9 より、多くの対角要素は距離尺度としてユークリッド距離を仮定した場合（すなわち、すべての対角要素が 1 となる場合）と比較して高い値を示していることが分かる。さらに図 3.10 より、すべての計量行列の要素を対象とした場合、多くの要素が 0 付近の値を取り、一部の要素が局所的に高い値、あるいは低い値を取っていることも分かる。これらから、分類精度向上には一

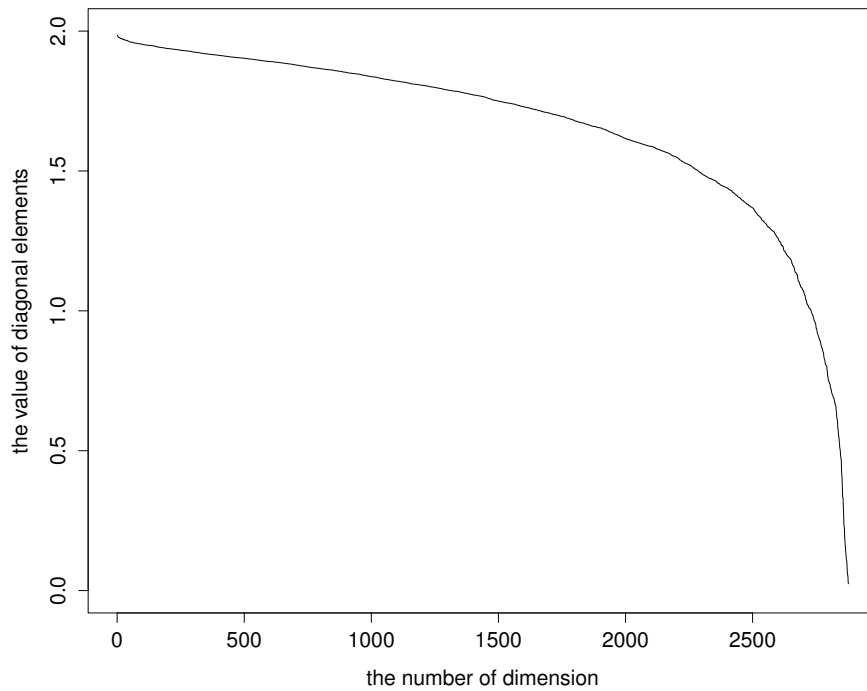


図 3.9: 最適な計量行列 \hat{M} の対角要素のランキング（降順） [55]

部の関係性の強い要素が大きく寄与しており，提案手法ではこれらの関係性を学習データから抽出することができたため，従来手法と比較してその精度を向上させることができたと考えられる．

3.5 定理，補題，系の証明

3.5.1 定理 3.1 の証明

式 (3.3) で定義された最適化問題を解くため，以下ではラグランジュの未定乗数法 [35] を用いる．

いま，ラグランジュ関数 $L(M, \psi)$ を以下の通り定義する．

$$L(M, \psi) = \sum_{n=1}^N \text{tr}(S_n M) + \eta \cdot \text{tr}(M) - \psi (\log \det M - \log 1). \quad (3.8)$$

ただし， ψ は未定乗数とする．いま，式 (3.8) の勾配を $\nabla_M L(M, \psi)$ とし，これを $\mathbf{0}$ とすれば，

$$\nabla_M L(M, \psi) = \sum_{n=1}^N S_n + \eta I - \psi M^{-1} = \mathbf{0}, \quad (3.9)$$

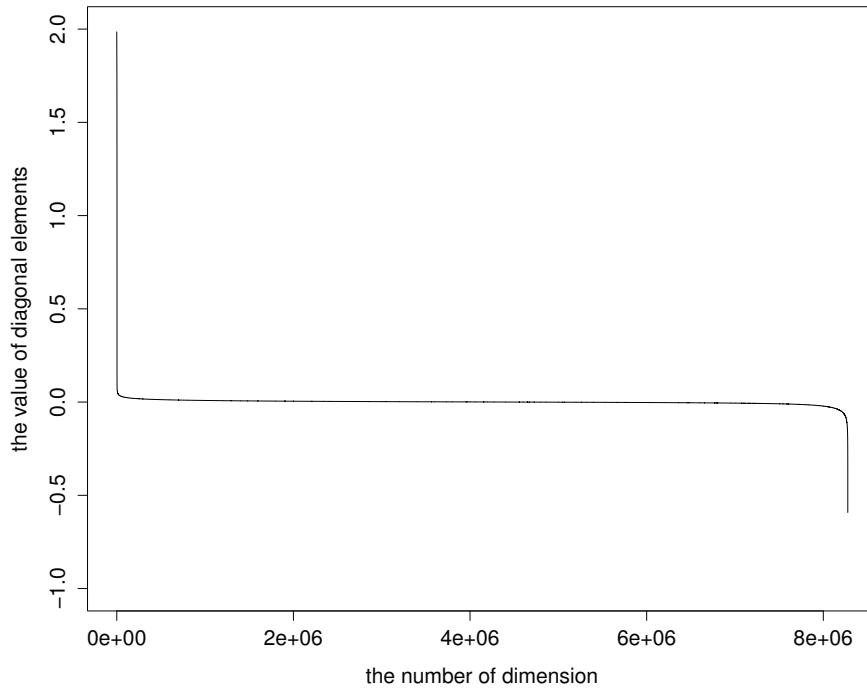


図 3.10: 最適な計量行列 \hat{M} の全要素のランキング（降順） [55]

を得る。ここから、

$$\psi M^{-1} = \sum_{n=1}^N S_n + \eta I, \quad (3.10)$$

がただちに導かれる。

いま、式 (3.10) の行列式を求めれば

$$\psi^W \det(M^{-1}) = \det \left(\sum_{n=1}^N S_n + \eta I \right), \quad (3.11)$$

となり、両辺の対数を取ることで

$$\log \psi^W = \log \det \left(\sum_{n=1}^N S_n + \eta I \right), \quad (3.12)$$

を得る。ただし、制約条件である $\log \det M = 0$ を用いた。

したがって、

$$\psi = \det \left(\sum_{n=1}^N S_n + \eta I \right)^{\frac{1}{W}}, \quad (3.13)$$

が導かれる.

これらの議論から, 最適な計量行列 \hat{M} は

$$\hat{M} = \det \left(\sum_{n=1}^N S_n + \eta I \right)^{\frac{1}{W}} \left(\sum_{n=1}^N S_n + \eta I \right)^{-1}, \quad (3.14)$$

で与えられる. □

3.5.2 補題 3.1 の証明

式 (2.16) より, $\sum_{n=1}^N S_n \in \mathcal{S}_+^W$ は明らかである. ただし, \mathcal{S}_+^W は半正定値対称行列からなる集合である. したがって, 半正定値対称行列の性質から, $\mathbf{0}$ ベクトル以外の 任意の W 次元ベクトル \mathbf{x} に対し, 以下の式が成立する.

$$\mathbf{x}^T \left(\sum_{n=1}^N S_n \right) \mathbf{x} \geq 0. \quad (3.15)$$

いま, $\eta > 0$ とし, 式 (3.15) における $\sum_{n=1}^N S_n$ に ηI を加えることで,

$$\mathbf{x}^T \left(\sum_{n=1}^N S_n + \eta I \right) \mathbf{x} = \mathbf{x}^T \left(\sum_{n=1}^N S_n \right) \mathbf{x} + \eta \mathbf{x}^T \mathbf{x}, \quad (3.16)$$

と変形できる. $\eta \mathbf{x}^T \mathbf{x} > 0$ より,

$$\mathbf{x}^T \left(\sum_{n=1}^N S_n + \eta I \right) \mathbf{x} > 0, \quad (3.17)$$

が成立する.

したがって, $\sum_{n=1}^N S_n + \eta I$ は正定値対称行列となり, 式 (3.4) における \hat{M} は一意に存在することが保証される. □

3.5.3 補題 3.2 の証明

提案手法で定義した最適化問題におけるラグランジュ関数は式 (3.8) で与えられる.

加えて、式 (3.13) より、 ψ は $\psi = \det \left(\sum_{n=1}^N S_n + \eta I \right)^{\frac{1}{W}}$ で与えられる。したがって、式 (3.8) にこれを代入すれば、

$$L(M) = \text{tr} \left(\sum_{n=1}^N S_n M \right) + \eta \text{tr}(M) - \det \left(\sum_{n=1}^N S_n + \eta I \right)^{\frac{1}{W}} (\log \det M - \log 1), \quad (3.18)$$

を得る。

式 (3.18) の右辺第一項、第二項はそれぞれ、 M に対して線形であり、補題 3.1 から、 $\det \left(\sum_{n=1}^N S_n + \eta I \right)$ は正となる。これらの議論と $\log \det(M)$ の凹性 [14] より、式 (3.18) は凸関数であることがわかる。

したがって、上記の議論から、本章で提案した最適化問題は正則化パラメータ η と正則化項 $\text{tr}(M)$ の効果により、下界を持つことがわかる。

3.5.4 定理 3.2 の証明

いま、 $\text{Eigen}[\cdot]$ をある行列の固有値を各成分に持つベクトルと定義する。このとき、

$$\text{Eigen} \left[\hat{M} \right] = (\hat{m}_1, \dots, \hat{m}_W)^T = \text{Eigen} \left[|A|^{\frac{1}{W}} A^{-1} \right], \quad (3.19)$$

が成立する。ただし、 \hat{m}_k は提案手法にて導出された最適な計量行列 \hat{M} における k 番目の固有値を指すものとする。式 (3.5) で定義した行列 A に対し、その固有値は $\lambda_k + \eta$ で与えられる³。

したがって、 $\text{Eigen} \left[\hat{M} \right]$ は

$$\text{Eigen} \left[|A|^{\frac{1}{W}} A^{-1} \right] = \prod_{i=1}^W (\lambda_i + \eta)^{\frac{1}{W}} \left(\frac{1}{\lambda_1 + \eta}, \dots, \frac{1}{\lambda_W + \eta} \right)^T, \quad (3.21)$$

として表現することができる。

³いま、 B を $B = \sum_{n=1}^N S_n$ とする。このとき、 B は対称行列であることから、固有値分解を用いることで $B = C^T \Lambda C$ が成立する。ここで、 C は正規直交行列であり、 Λ は B の固有値 λ_w を対角要素に持つような対角行列となる。すなわち、 $\Lambda = \text{diag}[\lambda_1, \dots, \lambda_W]$ となる。

一方で、式 (3.5) から $A = B + \eta I$ が成立する。ここから、 A の固有値分解は

$$A = C^T \Lambda C + C^T \eta I C, \quad (3.20)$$

として表現できる。これらの議論から、 $A = C^T (\Lambda + \eta I) C$ が成り立ち、 A の固有値は $\lambda_w + \eta$ ($w = 1, 2, \dots, W$) で与えられる。

いま、最適な計量行列における k 番目の固有値 \hat{m}_k は、

$$\hat{m}_k = \left\{ \prod_{w=1}^W (\lambda_w + \eta) \right\}^{\frac{1}{W}} \frac{1}{\lambda_k + \eta}, \quad (3.22)$$

として与えられる。ここから、

$$\log \hat{m}_k = \frac{1}{W} \sum_{w=1}^W \log (\lambda_w + \eta) - \log (\lambda_k + \eta), \quad (3.23)$$

が成立する。いま、 \hat{m}_k の増減は $\log \hat{m}_k$ の増減と一致するため、以下では $\frac{d \log \hat{m}_k}{d \eta}$ について検討を行う。 $\frac{d \log \hat{m}_k}{d \eta}$ は

$$\frac{d \log \hat{m}_k}{d \eta} = \frac{1}{W} \sum_{w=1}^W \frac{1}{\lambda_w + \eta} - \frac{1}{\lambda_k + \eta} = \frac{\bar{m} - \hat{m}_k}{\left\{ \prod_{w=1}^W (\lambda_w + \eta) \right\}^{\frac{1}{W}}}, \quad (3.24)$$

と表現できる。

したがって、式 (3.24) より、 $\bar{m} > \hat{m}_k$ が成立する場合には $\frac{d \log \hat{m}_k}{d \eta} > 0$ が成立し、 $\bar{m} < \hat{m}_k$ が成立する場合、 $\frac{d \log \hat{m}_k}{d \eta} < 0$ が成立する。したがって、 $\bar{m} > \hat{m}_k$ の場合、 \hat{m}_k は η に対し単調増加し、 $\bar{m} < \hat{m}_k$ の場合、 \hat{m}_k は η に対し単調減少する。□

3.5.5 系 3.1 の証明

いま、提案した最適化問題によって得られた計量行列 \hat{M} に対し、最大の固有値と最小の固有値、固有値の平均をそれぞれ、 \hat{m}_{\max} 、 \hat{m}_{\min} 、 \bar{m} と定義する。最大の固有値 \hat{m}_{\max} 、最小の固有値 \hat{m}_{\min} 、ならびにすべての固有値の平均 \bar{m} はそれぞれ異なる値を取るため、 $\hat{m}_{\max} > \bar{m}$ 、ならびに $\hat{m}_{\min} < \bar{m}$ が成立することは明らか。

また、定理 3.2 より、 η が増加した場合、 \hat{m}_{\max} は減少し、 \hat{m}_{\min} は増加する。したがって、 $\hat{m}_{\max} - \hat{m}_{\min}$ の取り得る値の範囲は η に対して単調減少することが分かる。同様にして、 \bar{m} と異なる値を持つ固有値 \hat{m}_k は η が増加するにしたがって、 \bar{m} に近づいていくことが分かる。これらの理由により、固有値の分散は η が増加するにしたがって、単調減少する。□

3.5.6 系 3.2 の証明

いま, \hat{m}_{\max} (最適な計量行列 \hat{M} における固有値の最大値) と \hat{m}_{\min} (最適な計量行列 \hat{M} における固有値の最小値) について検討する. これらは,

$$\hat{m}_{\max} = \left\{ \prod_{w=1}^W (\lambda_w + \eta) \right\}^{\frac{1}{W}} \frac{1}{\lambda_{\min} + \eta}, \quad (3.25)$$

$$\hat{m}_{\min} = \left\{ \prod_{w=1}^W (\lambda_w + \eta) \right\}^{\frac{1}{W}} \frac{1}{\lambda_{\max} + \eta} \quad (3.26)$$

とそれぞれ表現することができる. ただし, $\lambda_{\max}, \lambda_{\min}$ は行列 A の固有値のうち, それぞれ最大, 最小のものを示す. いま, 両者の差分を

$$\beta(\eta) = m_{\max} - m_{\min} \quad (3.27)$$

と定義すれば, $\beta(\eta)$ は

$$\beta(\eta) = \left\{ \prod_{w=1}^W (\lambda_w + \eta) \right\}^{\frac{1}{W}} \left(\frac{1}{\lambda_{\min} + \eta} - \frac{1}{\lambda_{\max} + \eta} \right), \quad (3.28)$$

として表現することができる.

いま, 式 (3.28) は

$$\begin{aligned} \beta(\eta) &= \eta \left\{ \prod_{w=1}^W \left(1 + \frac{\lambda_w}{\eta} \right) \right\}^{\frac{1}{W}} \left\{ \frac{\frac{1}{\eta}}{\frac{\lambda_{\min}}{\eta} + 1} - \frac{\frac{1}{\eta}}{\frac{\lambda_{\max}}{\eta} + 1} \right\} \\ &= \left\{ \prod_{w=1}^W \left(1 + \frac{\lambda_w}{\eta} \right) \right\}^{\frac{1}{W}} \left\{ \frac{1}{\frac{\lambda_{\min}}{\eta} + 1} - \frac{1}{\frac{\lambda_{\max}}{\eta} + 1} \right\}, \end{aligned} \quad (3.29)$$

のように変形することができる. ここから, $\beta(\eta)$ について極限をとれば, $\lim_{\eta \rightarrow \infty} \beta(\eta) = 0$. となることが分かる. この意味は, 固有値の最大値と最小値が $\eta \rightarrow \infty$ となった場合に等しくなることと等価である. したがって, このときにすべての固有値は互いに等しくなることが分かる. 加えて, $\log \det M = 0, \det M = 1$ の制約条件を考慮すれば, すべての w に対し, $m_k = 1$ が $\det M = \prod_{k=1}^W m_k$ という条件のもと, 成立する. \square

3.6 最適化問題が最小値を持つことの例証

本章で定義した最適化問題が最小値を持つことを以下の $W = 2$ の時の例を用いて説明する.

いま，行列 Γ, V をそれぞれ

$$\Gamma = \begin{pmatrix} a & b \\ b & d \end{pmatrix}, \quad (3.30)$$

$$V = \begin{pmatrix} x & y \\ y & z \end{pmatrix}, \quad (3.31)$$

として定義する．ただし， Γ は正定値対称行列であるとする．ここから， $a \geq 0, d \geq 0, ad - b^2 \geq 0$ がそれぞれ成立することが分かる．

いま，上記の行列 Γ, V を用いて従来の特橋らの手法で定義されている最適化問題を以下の通り定式化すると，

$$L(x, y, z) = \text{tr}(\Gamma V), \quad (3.32)$$

$$\text{subject to } V \succeq 0, \quad (3.33)$$

$$|V| = t^2, \quad (3.34)$$

となる．上記の最適化問題において，式 (3.32) は目的関数で式 (3.33), (3.34) はそれぞれ制約条件を表すものとする．さらに，式 (3.30), (3.31) を用いることで式 (3.32) は

$$L(x, y, z) = ax + 2by + dz, \quad (3.35)$$

と表現できる．加えて，上式の特関数は行列 V が半正定値対称行列であることから

$$\begin{aligned} f(\beta) &= \det |V - \beta I| \\ &= (x - \beta)(z - \beta) - y^2 \\ &= \beta^2 - (x + z)\beta + xz - y^2, \end{aligned} \quad (3.36)$$

と書くことができる．したがって，式 (3.36) と， V の固有値がすべて 0 以上であるという半正定値対称行列の条件より

$$x + z \geq 0, \quad (3.37)$$

$$xz - y^2 \geq 0, \quad (3.38)$$

$$\begin{aligned} D &= (x + z)^2 - 4(xz - y^2) \\ &= (x - z)^2 + 4y^2 \geq 0. \end{aligned} \quad (3.39)$$

を得る。ただし、すべての変数の組 (x, y, z) に対し、式 (3.39) が成立することに注意されたい。加えて、 $f(0) \geq 0$ より、式 (3.38) を得ることができる。さらに、式 (3.34) から、

$$xz - y^2 = l^2, \quad (3.40)$$

もまた成立する。

したがって、式 (3.32) – (3.34) は

$$L(x, y, z) = ax + 2by + dz, \quad (3.41)$$

$$\text{subject to } x + z \geq 0, \quad (3.42)$$

$$xz - y^2 = l^2, \quad (3.43)$$

$$|x| < \infty, \quad (3.44)$$

$$|y| < \infty, \quad (3.45)$$

$$|z| < \infty, \quad (3.46)$$

のように変形することが可能となる。もし、 $z = 0$ が成立した場合、式 (3.43) は不成立となる。このため、 $z \neq 0$ としても一般性を失わない。

したがって、

$$x = \frac{y^2 + l^2}{z}, \quad (3.47)$$

が成立するとともに、式 (3.41) は以下の通り変形することが可能となる。

$$L(y, z) = a \frac{y^2 + l^2}{z} + 2by + dz, \quad (3.48)$$

$$\text{subject to } \frac{y^2 + l^2}{z} + z \geq 0, \quad (3.49)$$

$$|y| < \infty, \quad (3.50)$$

$$|z| < \infty. \quad (3.51)$$

式 (3.49) から、 $z > 0$ が成立する。したがって、上記の最適化問題は、

$$L(y, z) = a \frac{y^2 + l^2}{z} + 2by + dz, \quad (3.52)$$

$$\text{subject to } |y| < \infty, \quad (3.53)$$

$$z > 0, \quad (3.54)$$

のように記述することが可能となる。いま、 y の定義域は実数であることから $\frac{\partial L(y,z)}{\partial y} = 0$ は一意に解が求まる。すなわち、

$$y = -\frac{b}{a}z, \quad (3.55)$$

が成立する。

式 (3.55) を式 (3.48) に代入すれば $L(z)$ は、

$$\begin{aligned} L(z) &= a \frac{\frac{b^2}{a^2}z^2 + l^2}{z} - \frac{2b^2}{a}z + dz \\ &= \frac{1}{a} (ad - b^2)z + \frac{al^2}{z}, \end{aligned} \quad (3.56)$$

$$\text{subject to } z > 0, \quad (3.57)$$

として表現できる。したがって、 $a \geq 0, d \geq 0, ad - b^2 \geq 0$ の条件から、式 (3.56) は $a > 0, d > 0$ の時に限り成立することが分かる。換言すれば、行列 Γ がフルランクのときにのみ、従来の持橋らの手法では定義した最適化問題は最小値を持つことが分かる。

一方、本章で提案している最適化問題においても、上記同様に $L(x, y, z)$ を表現すれば、

$$L(x, y, z) = (a + \eta)x + 2by + (d + \eta)z, \quad (3.58)$$

を得ることができる。同様の式展開を行うことで提案した最適化問題は η を正の正則化パラメータとして

$$\frac{1}{a + \eta} \left\{ (d + \eta)(a + \eta) - b^2 \right\} z + \frac{a + \eta l^2}{z}, \quad (3.59)$$

$$\text{subject to } z > 0, \quad (3.60)$$

のように表現することが可能となる。

したがって、 $a \geq 0, d \geq 0, ad - b^2 \geq 0$, ならびに $\eta > 0$ の条件から、式 (3.59) は z に関して常に一意な最小値を持つことが分かる。□

3.7 本章の結論

本章では、繰り返し処理を用いないメトリックラーニング手法 [59] に着目し、最適な計量行列の存在保証を行うため、正則化付き最適化問題の定式化を行った。正則化項を計量行列のトレー

スとし、最適化問題を解くことで、最適な計量行列が必ず存在すること、導出した最適な計量行列の持つ特性を明らかにした。

また、新聞記事データを用いた分類実験を行うことにより、本章で提案した手法が従来用いられている手法と比較し、高い分類精度を達成できることを示した。

本章で提案した手法は、繰り返し処理を用いない計量行列学習法と比較し、ほぼ同等の計算量で従来手法以上の分類精度を達成することができる。また、必ず最適な計量行列が導出できることから、多様なデータへの適用が可能となると共に、その拡大が期待できる。

第4章 l_1 正則化を用いたスパースな計量行列の学習 [58]

4.1 本章の目的

既に述べている通り、計量行列の各要素は学習データの成分間の関係性を示したものであると解釈できる。したがって、学習データの成分間に相関関係がない場合、推定される計量行列の解釈容易性や過学習回避のため、計量行列の要素のうち相関関係のない成分間の要素は0となることが望ましい。一方、近年の高度情報化により、高次元で多様な種類が存在しているデータには、多くの変数間には統計的な関係性がなく、ごく限られた変数間でのみ統計的な関係性があるような構造を持つものが多く存在している。すなわち、このような少数の変数間の統計的な関係性を適切に表現できる計量行列を学習することはデータ分析、知識発見の観点からも必要であると考えられる。このような構造を持つデータには、例えば以下のようなものがあげられる。

例 4.1 (文書データ)

文書データをベクトル表現する際には、学習データ集合中に出現する異なり単語をその成分とする。したがって、その次元数は膨大（一般に数千から数万次元）となることが容易に想定される。加えて、これらの文書ベクトルには、互いに相関の強い（同一の文書に共起しやすい）単語が存在するとともに、相関がないものも存在する。例えば、“国会”という単語に対し、“衆議院”という単語は同一文書中で生起しやすいと想定されるが、“ゴール”という単語はほとんど生起しないと想定される。このように、一般的に総異なり単語数が非常に大きい文書データの分析においては、学習データの成分間に相関が強いものがある反面、統計的な関係性がほとんどない成分の組み合わせも存在すると考えて差し支えない。

例 4.2 (顧客の購買履歴データ)

顧客の購買履歴も同様に、ある商品と関連付けて購入される商品や、商品間の相関関係が強い商品ペアが存在すると考えられる。例えば、機械学習の書籍を購入している顧客は、統計学の書籍や確率論の書籍を購入しやすいと想定されるのに対し、高度な医学の専門書を購入する可能性は

低いと想定される。一般に、EC サイト等で販売されている商品数は膨大であり、上記のような構造を保持していると考えられる。

これらの例にあるようなデータに対し、スパースな計量行列を自動的に学習することができれば、相関関係のない成分間の関係性が正しく表現され、学習データの成分間の関係性を適切に表現することが可能となる。

併せて、メトリックラーニングにおける計量行列の学習にはパラメータ数が過多となり、学習に必要なデータ数が膨大になってしまうという問題点が存在する。このような場合、各パラメータの推定に用いることのできるデータ数は相対的に少なくなり、パラメータ推定精度の低下、ならびに分類精度の低下を引き起こす。

パラメータ数の過多に関する問題を回避するための手法としては、学習に使用するデータ数を増加させる、モデル選択基準 [4] などを用いてデータの特徴空間の次元削減を行うなどが考えられる。しかしながら、学習データ数を増加させることは収集のためのコストが増加するため、パラメータ推定に必要となる十分な数のデータを収集することは難しいことが多い。また、学習の計算量の観点から、高次元データの学習に十分な量の学習データを学習されることが事実上困難になることも考えられる。このため、一般的に学習の性能は与えられたデータ数に対してどれほど精度を向上できるかで議論されることがほとんどである。他方、モデル選択基準を適用するためには何らかの確率モデルを仮定し、尤度を計算する必要がある。メトリックラーニングでは、事前に与えられたクラスの確率モデルを仮定していないため、そのままでは尤度を計算することができない。以上の理由により、どちらの方法も実問題に対し汎用的に用いることができるとは言い難い。

これに対し、 l_1 正則化を用いてスパースな計量行列を導出することにより、学習に必要な要素数は元の計量行列の要素数と比較して相対的に少なくなり、パラメータ推定に必要な学習データ数を低減させることにつながる。これにより、統計的関係性が互いに低い要素に対応するパラメータの値が0となり、推定に必要なパラメータ数が減少し、学習データ数が少ない場合でも安定したパラメータ推定が可能となる。加えて、計量行列の非ゼロ要素数が減少するため、新規データの分類時には計算量、コンピュータ上へ実装する際のメモリ量の低減にもつながる。

本章では従来手法と比較し、分類精度を落とさずにシンプルな構造を持つ計量行列を学習する方法を構成することを目的とする。これにより、不要なパラメータ（成分間の統計的関係性が弱いような成分ペア）を0に縮退させたスパースな計量行列を得るとともに、新規入力データの分類時

の計算量を低減させることが可能となる。計量行列の導出の際には Alternating Direction Method of Multiplier (ADMM) [13] に基づいたアルゴリズムを用いることで効率的に計量行列の計算が可能となることを示す。

また、本章で提案する手法は統計的モデル選択などの分野で用いられる Sparse Inverse Covariance Selection [7] の特殊な場合であることを示す。

4.2 ADMM を用いた最適化

本章で用いる最適化問題では、Alternating Direction Method for Multiplier (以下、ADMM) [13] を導入することにより、スパースな計量行列を学習する。

ADMM は以下で定義される最適化問題を解くための手法であり、特定の条件のもとで最適解を得ることができるアルゴリズムである。

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) + g(\mathbf{z}), \\ & \text{subject to} && A\mathbf{x} + B\mathbf{z} = \mathbf{c}. \end{aligned} \tag{4.1}$$

ただし、各変数は $\mathbf{x} \in \mathcal{R}^W$, $\mathbf{z} \in \mathcal{R}^K$, $A \in \mathcal{R}^{P \times W}$, $B \in \mathcal{R}^{P \times K}$, $\mathbf{c} \in \mathcal{R}^P$ とする。加えて、 $f(\mathbf{x})$, $g(\mathbf{z})$ は (広義の) 凸関数であると仮定する。いま、上記の最適化問題における拡張ラグランジュ関数 L_ρ は

$$L_\rho(\mathbf{x}, \mathbf{z}, \mathbf{u}) = f(\mathbf{x}) + g(\mathbf{z}) + \frac{\rho}{2} \|\mathbf{r} + \mathbf{u}\|_2^2, \tag{4.2}$$

で定義される¹。ただし、 \mathbf{r} は $\mathbf{r} = A\mathbf{x} + B\mathbf{z} - \mathbf{c}$, ρ は $\rho > 0$ を満たす正定数とし、 \mathbf{u} は未定乗数ベクトルとする。

ADMM は式 (4.2) の各変数に対し、任意の初期値を与えたもとで、以下のような更新を繰り返すことで最適解を得るアルゴリズムである。

$$\mathbf{x}^{k+1} := \arg \min_{\mathbf{x}} L_\rho(\mathbf{x}, \mathbf{z}^k, \mathbf{u}^k), \tag{4.3}$$

$$\mathbf{z}^{k+1} := \arg \min_{\mathbf{z}} L_\rho(\mathbf{x}^{k+1}, \mathbf{z}, \mathbf{u}^k), \tag{4.4}$$

$$\mathbf{u}^{k+1} := \mathbf{u}^k + \rho(A\mathbf{x}^{k+1} + B\mathbf{z}^{k+1} - \mathbf{c}). \tag{4.5}$$

ただし、 $\mathbf{x}^k, \mathbf{z}^k, \mathbf{u}^k$ は、 k 回目の繰り返しで得られる $\mathbf{x}, \mathbf{z}, \mathbf{u}$ をそれぞれ表すものとする。

¹簡単のため、拡張ラグランジュ関数を scaled form で表現した。scaled form に関しては、文献 [13] 参照。

4.3 l_1 正則化を用いたスパースな計量行列の導出法

既に述べた通り，学習する計量行列のパラメータ数は次元数の2乗に比例しており，高次元なデータを使用した場合には学習のために大量のデータを必要とする．そこで本研究では，推定パラメータ数を減少させ，その推定精度を向上させるための方法として l_1 正則化 [11] を用いる．

4.3.1 l_1 正則化を用いた最適化問題の定式化

以下では，代表元を用いたメトリックラーニング手法同様，教師あり学習の視点からスパースな計量行列を得るため，以下のように最適化問題を定義する．

$$\begin{aligned} & \underset{M}{\text{minimize}} \left\{ \sum_{n=1}^N \text{tr}(S_n M) - \log \det M + \eta \|M\|_1 \right\}, \\ & \text{subject to } M \succeq 0. \end{aligned} \quad (4.6)$$

ただし， $\|M\|_1$ は行列の各要素の絶対値の和とし， S_n は式 (2.16) で定義したものをを用いる．すなわち， $\|M\|_1 = \sum_{i,j} |M_{ij}|$ ， $S_n = \sum_{\mathbf{x}_i \in \mathcal{C}_n} (\mathbf{x}_i - \boldsymbol{\mu}_n)(\mathbf{x}_i - \boldsymbol{\mu}_n)^T$ である．また，上記の最適化問題における $\log \det M$ は $M = \mathbf{0}$ となる縮退した解となることを避けるための制約条件であり， η は正則化パラメータ ($\eta \geq 0$) とする．上記の最適化問題により，計量行列が半正定値であるという制約のもと，各カテゴリの代表元と，そのカテゴリに所属している学習データ間のマハラノビス距離を最小とし，かつスパースな計量行列を得ることが可能となる．

式 (4.6) に対し ADMM を用いるため，これを

$$\begin{aligned} & \underset{M}{\text{minimize}} \left\{ \sum_{n=1}^N \text{tr}(S_n M) - \log \det M + \eta \|Z\|_1 \right\}, \\ & \text{subject to } M = Z, \\ & M \succeq 0, \end{aligned} \quad (4.7)$$

と表現する．ただし， $Z = [Z_{ij}] \in \mathcal{R}^{W \times W}$ である．このとき，式 (4.7) に対する拡張ラグランジュ関数は次のように表現できる．

$$L_\rho(M, Z, U) = \sum_{n=1}^N \text{tr}(S_n M) - \log \det M + \eta \|Z\|_1 + \frac{\rho}{2} \|M - Z + U\|_F^2, \quad \rho > 0. \quad (4.8)$$

ここで， $\|A\|_F$ は行列 A のフロベニウスノルムを示し， $\|A\|_F = \sqrt{A^T A}$ で定義される．また， U は未定乗数を要素として持つ行列とする．式 (4.8) は式 (4.2) とは異なり，行列を対象として最適化

を行う。従って、拡張ラグランジュ関数のペナルティ項のノルムが l_2 ノルムではなくフロベニウスノルムとなっていることに注意されたい。

4.3.2 ADMM を用いた最適な計量行列の導出

上記の拡張ラグランジュ関数に対する ADMM Form は式 (4.3)–(4.5) と同様に、以下のように表現できる。

$$M^{k+1} := \arg \min_M L_\rho(M, Z^k, U^k), \quad (4.9)$$

$$Z^{k+1} := \arg \min_Z L_\rho(M^{k+1}, Z, U^k), \quad (4.10)$$

$$U^{k+1} := U^k + M^{k+1} - Z^{k+1}. \quad (4.11)$$

したがって、上式をそれぞれ解くことで、各変数の更新は

$$M^{k+1} := Q^k \tilde{M}^k Q^{kT}, \quad (4.12)$$

$$Z_{ij}^{k+1} := S_{\eta/\rho} \left(M_{ij}^{k+1} + U_{ij}^k \right), \quad (4.13)$$

$$U^{k+1} := U^k + M^{k+1} - Z^{k+1}. \quad (4.14)$$

となる。式 (4.12) の導出については 4.3.3 節を参照されたい。ここで、 Q は式 (4.18) で算出される正規直交行列を指す。また、 $\tilde{M} = Q^T M Q$, $\tilde{M} = \text{diag}[\tilde{M}_{11}, \tilde{M}_{22}, \dots, \tilde{M}_{WW}]$, $\tilde{M}_{ii} = \frac{\lambda_i + \sqrt{\lambda_i^2 + 4\rho}}{2\rho}$ を満たすものとする。また、 $S_\kappa(v)$ は成分ごとに以下の処理を行う soft thresholding operator で、

$$S_\kappa(v) = \begin{cases} v - \kappa & (v > \kappa), \\ 0 & (|v| \leq \kappa), \\ v + \kappa & (v < -\kappa), \end{cases}$$

を満たすものとする。上記の式 (4.12)–(4.14) を順に繰り返すことで最適解を得ることが可能となる。

4.3.3 M-update の導出

式 (4.6) で定義された最適化問題は、式 (4.8) から以下の更新式を解くことで最適解を得ることができる。

$$\hat{M} = \underset{M}{\text{minimize}} \left\{ \sum_{n=1}^N \text{tr}(S_n M) - \log \det M + \frac{\rho}{2} \|M - Z^k + U^k\|_F^2 \right\}. \quad (4.15)$$

いま、式 (4.9) の勾配を $\mathbf{0}$ とすれば、

$$\nabla_M M = \sum_{n=1}^N S_n - M^{-1} + \rho (M - Z^k + U^k) = \mathbf{0}, \quad (4.16)$$

を得る。したがって、

$$\rho M - M^{-1} = \rho Z^k - \sum_{n=1}^N S_n - U^k, \quad (4.17)$$

となる。いま、式 (4.17) の左辺は対称行列となることから、右辺の対称性が保証される。従って、右辺は $Q\Lambda Q^T$ と固有値分解することができて、

$$\rho Z^k - \sum_{n=1}^N S_n - U^k = Q\Lambda Q^T, \quad (4.18)$$

を得ることができる。ただし、 λ_w を $\rho Z^k - \sum_n S_n - U^k$ の w 番目の固有値、 $\Lambda = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_W]$ 、 $QQ^T = I$ を満たすものとする。 $\tilde{M} = Q^T M Q$ より、式 (4.17) は

$$\rho \tilde{M} - \tilde{M}^{-1} = \Lambda, \quad (4.19)$$

となる。 $\tilde{M} = \text{diag}(\tilde{M}_{11}, \dots, \tilde{M}_{WW})$ となることから、

$$\rho \tilde{M}_{ii} - \frac{1}{\tilde{M}_{ii}} = \lambda_i, \quad (4.20)$$

を得る。したがって、

$$\rho(\tilde{M}_{ii})^2 - \lambda_i \tilde{M}_{ii} - 1 = 0, \quad (4.21)$$

より

$$\tilde{M}_{ii} = \frac{\lambda_i + \sqrt{\lambda_i^2 + 4\rho}}{2\rho}, \quad (4.22)$$

が成立する。 \hat{M} に再び Q をかけることで M -update が実行できる。□

4.4 提案手法と Sparse Inverse Covariance Selection との関係性

式 (4.7) で定義された提案手法は、モデル選択などで用いられる sparse inverse covariance selection [7] の拡張とみなすことができる。sparse inverse covariance selection は多次元正規分布 [81]

における分散共分散行列の逆行列をスパースな解として得ようとする手法を指す。いま、あるデータ $\mathbf{a}_i \in \mathcal{R}^W, i = 1, 2, \dots, D$ が平均 $\boldsymbol{\mu}$ 、分散共分散行列 Σ の多次元正規分布に従うとすると、その対数尤度関数 R は

$$R = -\frac{NW}{2} \log 2\pi - \frac{N}{2} \log \det \Sigma + \sum_{i=1}^N \left\{ (\mathbf{a}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{a}_i - \boldsymbol{\mu}) \right\}, \quad (4.23)$$

として表現できる。 $S = \frac{1}{N} \sum_{i=1}^N (\mathbf{a}_i - \boldsymbol{\mu})(\mathbf{a}_i - \boldsymbol{\mu})^T$ とすれば、式 (4.23) は

$$R = -\frac{NW}{2} \log 2\pi - \frac{N}{2} \log \det \Sigma + \text{tr}(S\Sigma^{-1}), \quad (4.24)$$

として表現することが可能となる。いま、 $Y = \Sigma^{-1}$ とし、 $\boldsymbol{\mu}$ は既知であるとした上で、スパースな精度行列を求めるため、上記の対数尤度関数に Y の l_1 ノルムを正則化項として付与した最適化問題を考える。

$$\underset{Y}{\text{minimize}} \left\{ \text{tr}(SY) - \frac{N}{2} \log \det Y + \lambda \|Y\|_1 \right\}. \quad (4.25)$$

これを解くことで、前述の条件を満たすスパースな精度行列を求めることが可能となる。本章で定義した最適化問題 (4.7) と式 (4.25) との違いは、カテゴリが複数存在し、その平均が $\boldsymbol{\mu}_n$ に従い、精度行列 M が未知の場合であるということにある。

4.5 実験

本研究で提案した手法について高次元、スパースなベンチマークデータ、ならびに低次元密なベンチマークデータを用いた分類実験を行い、その有効性について検討を行う。以下では、高次元スパースなデータセットを用いた場合、比較的次元、密な（スパースではない）データセットを用いた場合の実験とその結果について示す。

4.5.1 新聞記事データを用いた実験

実験条件

新聞記事データを用いた実験では、2000年の毎日新聞記事（5カテゴリ）を用いた。各カテゴリに存在する学習データは100件（合計500件）、テストデータは500件（合計2500件）として

分類精度を求めた。テストデータ x の分類は、第3章同様に式 (3.7) を用いてカテゴリに分類し、その正解率を求めるものとした。

また、従来手法も第3章と同様、持橋らの手法 [59], LMNN (Large Margin Nearest Neighbor) [85], ITML (Information Theoretic Metric Learning) [22] を用い、比較のために $\text{tr}(M)$ を正則化項として導入した第3章での手法を用いるものとした。提案手法における正則化項 η については 0.01, 0.05, 0.07, 0.1, 0.2, 0.3, 0.4, 0.5 の8種類を用いた場合の分類精度を求めるものとし、他手法との分類精度の比較に際しては最も分類精度が高かった、 $\eta = 0.05$ を用いるものとした。合わせて、 η の変化に伴い、計量行列における0の要素数がどのように変化するかについても検討することとした。その他の実験条件は第3章と同様とする。

実験結果

η の変化による提案手法の分類精度の変化とスパース率の変化を図4.1に、他手法との分類精度の差異を比較した結果について図4.2に示す。ここで、スパース率とは得られた計量行列の要素数のうち、ゼロとなった要素の割合を示すものとする。

また、目的関数が収束していることを確認するため、繰り返し毎の行列 M , Z の間のフロベニウスノルムの差を測定した結果を図4.3に示す。

図4.1より、正則化パラメータ η が比較的小さい場合には分類精度は高く、大きくなるにつれて計量行列のゼロ要素の増加と共に徐々に分類精度が低下していくことがわかる。これは、計量行列のゼロ要素が多くなりすぎると、自由度が下がり、データ構造へのフィッティングが悪くなるためと考えられる。一方、図4.2より、提案手法はLMNN、第3章で提案した手法と同程度の分類精度を達成できていることが分かる。これは、同程度の分類精度でよりシンプルな表現のモデルが得られていることを意味する。アルゴリズムの収束に関しても、図4.3から、500から600回程度の繰り返しで収束していることも分かる。

4.5.2 UCI機械学習リポジトリを用いた実験

実験条件

上記の実験とは異なり、比較的次元、密なデータセットに対する分類精度を検討するため、UCI機械学習リポジトリ [5] を使用した分類実験を行った。UCI機械学習リポジトリのデータは以下の

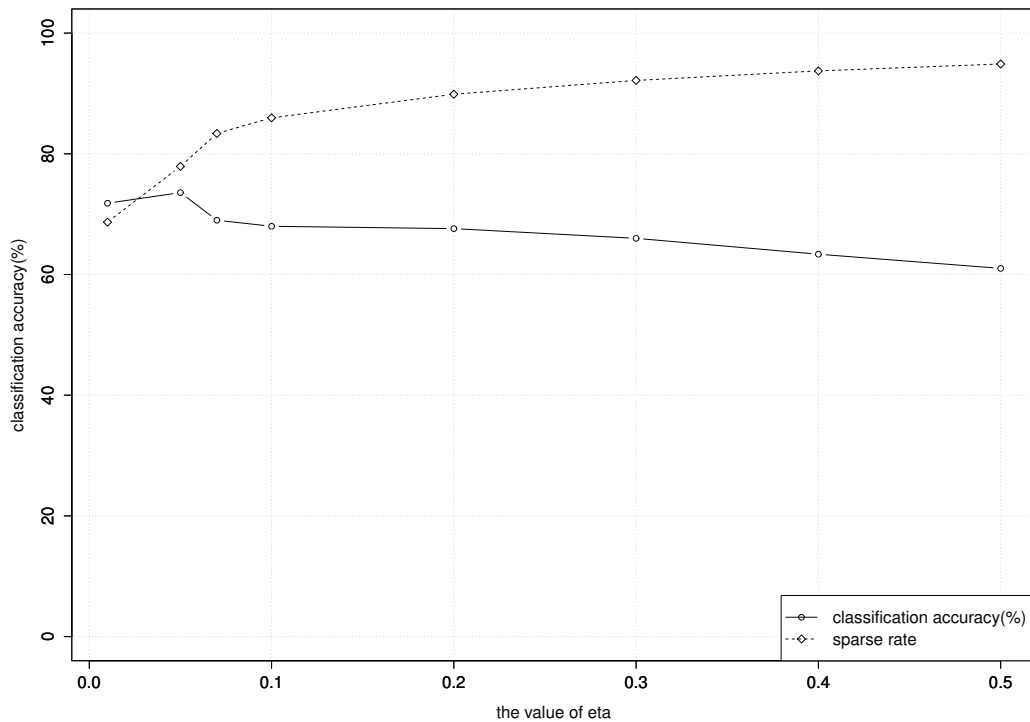


図 4.1: η の変化による分類精度と計量行列のゼロ要素の割合の変化 [58]

表 4.1 に記載の 5 種類のデータセットを用いた。

これらのデータについては、データをランダムに 10 分割し、そのうちの 9 つを学習データ、残りをテストデータとすることで分類精度を算出するものとし、新聞記事データを用いた実験同様、従来手法と分類精度の比較を行った。

実験結果

UCI 機械学習リポジトリを用いた場合の各種手法との分類精度の違いについて以下の表 4.2 に示す。すべての手法において、分類精度が最大となっているものについては太字としている。

また、正則化パラメータ η の変化に伴い、分類精度と計量行列のスパース率がどのように変化しているかについて確認するために、上記のデータセットを対象に、 η を 0 から 1.5 まで 0.1 ずつ変化させた際の分類精度とそのスパース率を算出した。これを図 4.4–4.7 に示す。

表 4.2 より、提案手法は比較的次元、密なデータセットを対象とした場合、従来手法とそれほど大きく結果が変わらないことが見て取れる。これは学習データが低次元、密であるため、パラ

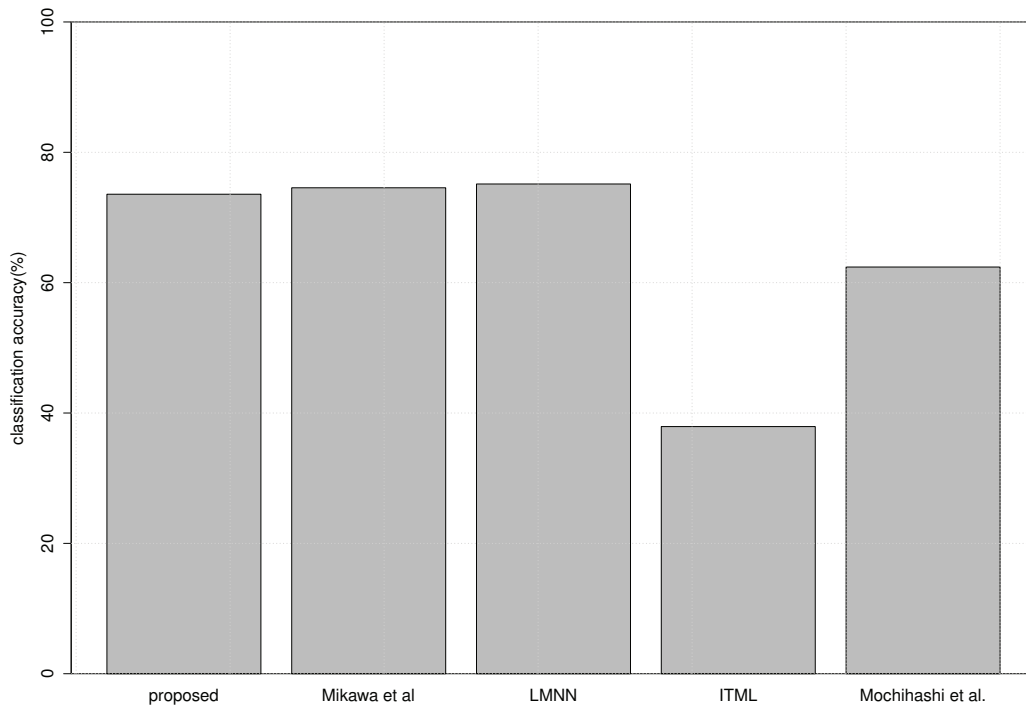


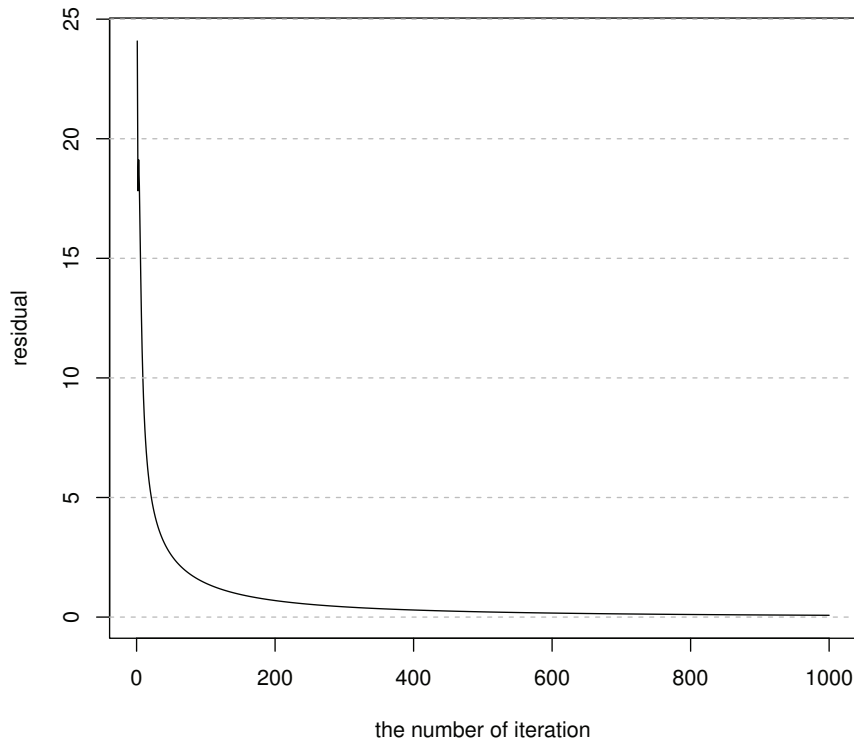
図 4.2: 他手法との分類精度の比較 [58]

メータ数に対し学習データ数が相対的に多く、 l_1 正則化を用いなくとも十分なパラメータ推定が可能となっているためであると想定される。このことは balance, ionosphere 以外のデータセットの実験結果にある通り、正則化パラメータ η を大きくしても分類精度はそれほど大きく変化していないことから確認できる。

一方で、balance, ionosphere データセットについては正則化パラメータを大きくした際に計量行列のスパース率が大きくなり、結果的に分類精度が低下してしまっている。これらから、適切な正則化パラメータを事前に検討することが必要であると言える。

4.6 考察

本研究では、 l_1 正則化を用いたスパースな計量行列の学習方法について、教師あり学習の問題設定のもと、提案を行った。式 (4.6) で定義した目的関数は行列 M の l_1 ノルムを含むため、原点で非連続となり、通常の勾配法などによる最適化アルゴリズムを用いることができない。これに

図 4.3: 繰り返し毎の計量行列 M と Z の差 (フロベニウスノルム) [58]

対し、本研究では ADMM を用いることにより、非連続な関数の最適化を可能としている。また、式 (4.13) より、行列 Z の各要素は並列に更新することができるため、効率的に計算を行うことが可能である。しかしながら、 M の更新には固有値分解を行う必要があるため、繰り返し処理毎の計算コストが大きくなってしまふ。これは高次元データを分析する際には依然として大きな問題であると考えられる。

他方、実験結果より、本提案手法は特に入力データが高次元な場合に有効であることが示された。これは前述の通り、不要なパラメータを 0 にすることで学習に必要なパラメータ数が低減され、各パラメータの推定精度が向上したことによるものであると考えられる。また、本実験で比較を行った他のメトリックラーニング手法と比べ、提案手法では不要なパラメータ (計量行列の要素) を 0 に縮退させることが可能となる。これにより、例えばスパース行列の圧縮表現に使用されるハーヴェル・ボーディング形式 [48] などを適用することが可能となり、実装時に必要となるメモリ量を大幅に節約することができる (いま、計量行列の非ゼロ要素数を ξ とすれば必要な

表 4.1: UCI 機械学習リポジトリの基本情報

	データ数	次元数	カテゴリ数
iris	150	4	3
wine	178	13	3
balance	625	4	3
ionosphere	351	34	2

表 4.2: UCI 機械学習リポジトリを用いた際の分類精度の比較 [58]

	LMNN	ITML	持橋ら	$\text{tr}(M)$	$ M _1$
iris	95.83	96.00	98.00	98.00	98.00
wine	91.96	72.47	98.32	98.86	96.67
balance	86.86	88.96	71.08	73.16	75.00
ionosphere	87.10	86.89	87.96	88.06	88.06

メモリ量は $2\xi + W + 1$ となる)。

加えて、計量行列の非ゼロ要素のみを用いてマハラノビス距離の算出を行うことにより、従来手法で得られる計量行列を用いた場合と比較して少ない回数で距離の計算が可能となる²。換言すれば、スパースな計量行列を用いることで分類時の計算コストを低減することが可能となる。図 4.8 に、正則化パラメータ η を変化させた際の新聞記事データに対する分類精度と分類時に必要とする計算回数の関係性を示す。

また、実験結果より高次元なデータについて計量行列をスパースに求めることの有効性は示されたものの、低次元で密なデータに対しては、本手法はそれほど有効ではないことも分かる。これらから、例えば、推定するパラメータ数に対して学習データが比較的少ない場合には提案手法を、そうでない場合には従来手法を用いるなど、分析対象のデータ特性に合わせて各手法を使い分けることが重要であるということが示唆される。

²例えば、 $W = 1,000$ の入力データ間のマハラノビス距離を計算する場合、計量行列の全要素が非ゼロであれば、その計算回数は $W^3 = 1.0 \times 10^9$ となるが、要素の 20% が非ゼロであればその計算回数は $W^2 \times 0.2W = 0.2 \times 10^9$ となり、必要な計算回数は $1/5$ となる。

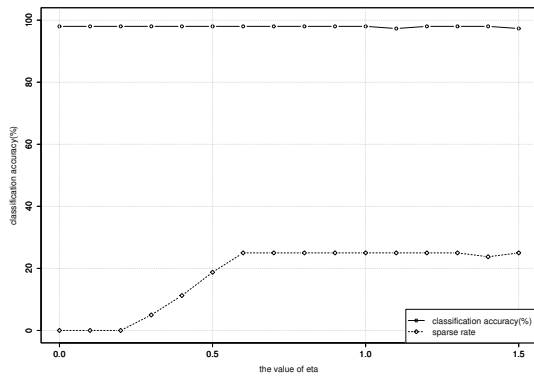


図 4.4: η の変化による分類精度と計量行列のゼロ要素の割合の変化 (iris) [58]

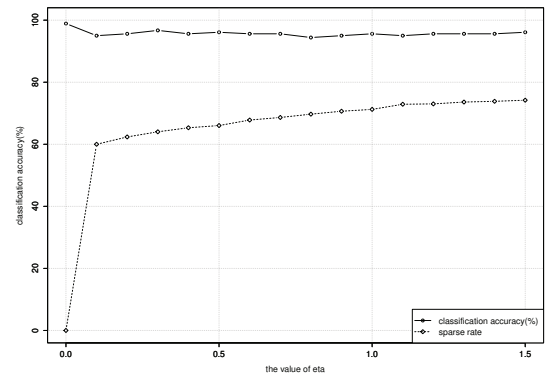


図 4.5: η の変化による分類精度と計量行列のゼロ要素の割合の変化 (wine) [58]

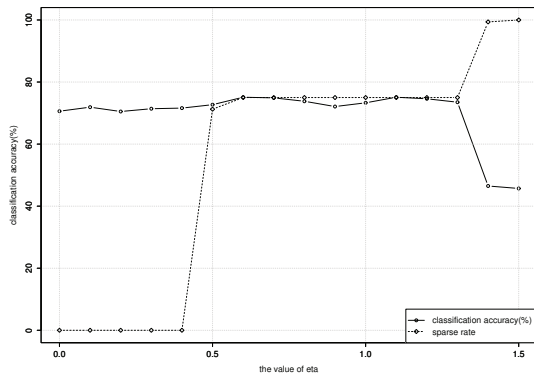


図 4.6: η の変化による分類精度と計量行列のゼロ要素の割合の変化 (balance) [58]

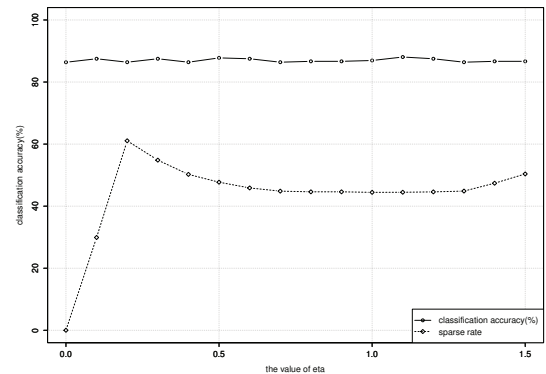


図 4.7: η の変化による分類精度と計量行列のゼロ要素の割合の変化 (ionosphere) [58]

4.7 本章の結論

本章では、従来のメトリックラーニング手法と比較し、分類精度を落とさずにシンプルな構造を持つ計量行列を学習するため、不要なパラメータを 0 に縮退させたスパースな計量行列の導出法の提案を行った。加えて、本章で提案した手法が Sparse Inverse Covariance Selection の特殊な場合であることを示した。

新聞記事データ、機械学習におけるベンチマークデータを用いた分類実験を行うことにより、スパースな構造を持つ計量行列を従来手法と同様の分類精度で構成可能となることを示した。

計量行列の学習には第 3 章で提案した手法と比較して多くの計算量を要するものの、新規入力

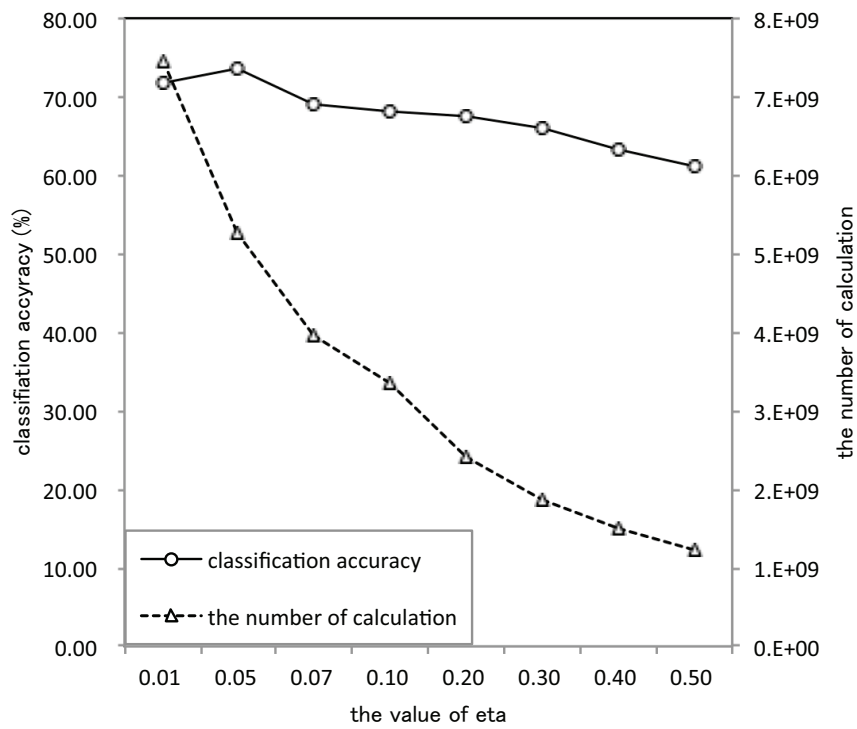


図 4.8: 正則化パラメータ η と分類精度, 計算回数の関係性

データを少ない計算量で分類することが可能となる。これにより、実応用を視野に入れた場合、その適用対象を拡大させることが可能となるため、より多くの実問題への適用もまた期待できる。

第5章 複数の計量行列を用いた距離計量の学 習 [56]

5.1 本章の目的

これまでに述べてきた計量行列の学習はすべて大域的な、すなわち学習データに対してただ一つの計量行列を学習するものであった。しかしながら、大域的な計量行列では、複数のカテゴリ間で統計的特徴に差異がある場合、これを考慮した距離尺度を得ることができない。このような場合、学習データで唯一の計量行列を学習するのではなく、カテゴリ毎に計量行列の存在を仮定することで各カテゴリの統計的特徴を考慮した距離構造を学習できる可能性がある。これらから、分類精度の向上にはあるカテゴリとそれ以外のカテゴリを識別しやすくすることができるような距離構造を学習することが有効であると考えられる。

したがって、本章では持橋らの手法 [59] をベースとした繰り返し演算を必要としない計量行列の学習において、カテゴリ毎に異なる計量行列の存在を仮定し、それらの最適解の導出を行う。その際、カテゴリ間の学習データの統計的な関係性を考慮し、分類に寄与すると考えられるような距離構造を学習するため、カテゴリ毎の学習データの統計的特徴の差異を考慮した計量行列を得るための最適化問題を定式化する。

既に述べた通り、一般の計量行列学習では繰り返し法を用いることで最適な計量行列の導出を行うが、データが大規模になった場合には計算量が多く必要となるという問題点がある。すなわち、学習データセットに対し、複数の計量行列を導出しようとした場合、その計算量は単純にカテゴリ数に比例して線形で増加してしまい、現実的な時間で解を得られなくなってしまう可能性がある。このため、本章でも第3章にて提案を行った手法をベースとし、解析的に最適な計量行列を求める方法について提案を行う。

持橋らの手法 [59] では計量行列の学習を各カテゴリの代表元を用いて行っており、導出した計量行列を用いた分類は、入力データと代表元の距離を用いたテンプレートマッチングを行うこと

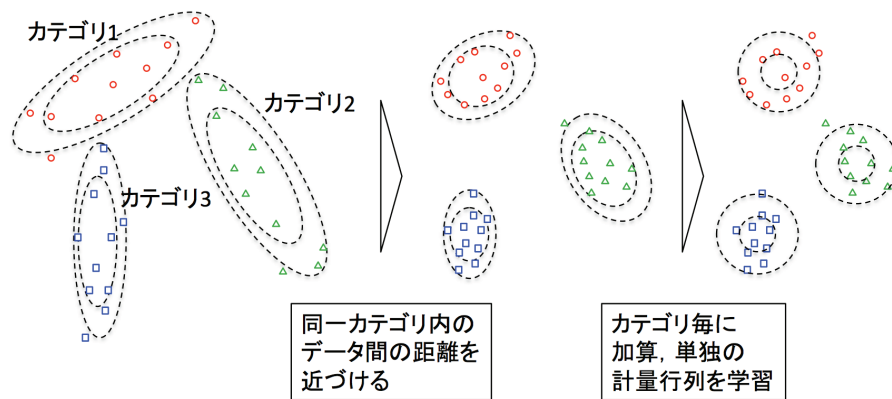


図 5.1: 大域的な計量行列の学習イメージ

は比較的容易である。しかし、この方法を複数の計量行列を用いた場合に単純に拡張しようとする
と、任意のデータ間の距離が一意に定義されていないため、 k -NN 法を適用することが困難になっ
てしまう。一方で、一般的なメトリックラーニング手法では k -NN 法による分類のための距離尺
度を学習することが多いこと、識別境界が非線形の問題では k -NN 法の性能が優れていると考え
られることから、 k -NN 法を適用できるように任意のデータ間の距離が定義される方が望ましい。
したがって、本研究ではさらに、複数の計量行列を用いて入力データと近傍のデータとの距離測
定を可能とするための手法について提案を行う。これにより複数の計量行列を用いた k -NN 法の
実行が可能となる。

5.2 複数の計量行列を用いた距離計量の学習とそれを用いた分類手法

5.2.1 カテゴリ毎に計量行列を仮定した計量距離学習

これまでに述べてきた手法は、学習データ全体から大域的な計量行列を導出するための手法で
あるが、学習データの統計的特徴がカテゴリ毎に大きく異なる問題も多いと考えられる。しかし
ながら、式 (2.18) における計量行列では各カテゴリの代表元とそこに所属する学習データとの関
係のみを考慮しているため、カテゴリ間の統計的な差異を考慮していない。加えて、カテゴリの
特徴として得られた情報を全てのカテゴリに渡って加算しているため、これらの特徴が互いに打
ち消しあってしまう可能性がある。これらの例について以下の図 5.1 に示す。

このような問題は大域的な計量行列の存在を仮定していることに起因するものであり、カテゴ

り単位に計量行列の存在を仮定することで緩和できる可能性がある。したがって、本研究ではカテゴリ毎に計量行列が存在するとし、これらの計量行列の学習が従来手法同様、解析的に行うことができることを示す。

いま、カテゴリ \mathcal{C}_n における計量行列を $M_n \in \mathcal{R}^{W \times W}$ とする。このとき、カテゴリ \mathcal{C}_n に対応する計量行列を以下の最適化問題を解くことで得るものとする。

$$\begin{aligned} \hat{M}_n = \underset{M_n}{\text{minimize}} \quad & \left\{ \alpha \frac{\text{tr}(M_n S_n)}{|\mathcal{C}_n|} - \frac{1-\alpha}{N-1} \sum_{n':n \neq n'} \frac{\text{tr}(M_n S_{n'})}{|\mathcal{C}_{n'}|} + \eta_n \text{tr}(M_n) \right\}, \\ \text{subject to} \quad & \forall n \in \{1, 2, \dots, N\}, \quad \log \det M_n = 0. \end{aligned} \quad (5.1)$$

ただし、第3章と同様に、最適な計量行列の存在を保証するため、式(5.1)における正則化項として $\text{tr}(M_n)$ を加えている。また、 η_n は正則化パラメータである。上記の最適化問題における制約条件の $\log \det M_n = 0$, ($n = 1, 2, \dots, N$) はそれぞれの計量行列により構成される距離空間が同様の尺度となっていることを保証し、各計量行列を用いたマハラノビス距離同士を比較可能にするためのものである¹。これにより、本研究で提案を行う複数の計量行列を用いた k -NN 法の実行が可能となる。

式(5.1)で定義された最適化問題は代表元と同一カテゴリの学習データ間の距離を可能な限り小さくしながら、その学習データと別カテゴリの代表元との距離をできるだけ離すように計量行列を学習する。これにより、各カテゴリの学習データの統計的特徴を表現した異なる計量行列をそれぞれ学習することが可能となる²。

上記の最適化問題をラグランジュの未定乗数法を用いて解くことで、以下の定理を得る。

定理 5.1 (最適な計量行列)

上記の最適化問題の解として得られる計量行列 \hat{M}_n は

$$\hat{M}_n = (\det A_n)^{\frac{1}{W}} A_n^{-1}, \quad (5.2)$$

で与えられる。ただし、 A_n は

$$A_n = \left(\frac{\alpha}{|\mathcal{C}_n|} S_n - \frac{1-\alpha}{N-1} \sum_{n':n \neq n'} \frac{1}{|\mathcal{C}_{n'}|} S_{n'} + \eta I \right), \quad (5.3)$$

¹この制約条件は、例えば $\log \det M_1 = \log \det M_2 = \dots = \log \det M_N$ としてもよいが、一般性を失わないため、上記のように定義することとした。

²上記の最適化問題を異なる等価な形で構成することも可能であるが、既に述べた通り、本提案手法では計算量をできるだけ少なくした形で最適な計量行列を導出するため、このような定式化を行うものとした。例えば、式(5.1)の目的関数は減算ではなく除算の形で定式化可能であるが、このような場合、繰り返し処理を用いる必要が生じる。したがって、これを避けるため減算の形で最適化問題を定式化している。

を満たすものとする. □

式 (5.2) の導出は 5.5.1 節に示す.

また, 式 (5.1) で定義した最適化問題における正則化項と最適な計量行列の関係について, 第3章同様に以下の定理が成立する.

定理 5.2 (最適な計量行列 \hat{M}_n の固有値と正則化パラメータ η_n の関係)

最適な計量行列 \hat{M}_n の固有値を \hat{m}_{nk} , ($k = 1, 2, \dots, W$), 固有値 m_{nk} の平均値を $\bar{m}_n = \frac{1}{W} \sum_k \hat{m}_{nk}$ とする. このとき, 以下の関係が成立する.

いま, η_n が増加した時

$$\left\{ \begin{array}{l} \bar{m}_n < \hat{m}_{nk} \text{ が成り立てば, } \hat{m}_{nk} \text{ は増加し,} \\ \bar{m}_n > \hat{m}_{nk} \text{ が成り立てば, } \hat{m}_{nk} \text{ は減少する.} \end{array} \right.$$

ただし, $\hat{m}_{n1}, \dots, \hat{m}_{nW}, \bar{m}_n$ はそれぞれ η_n に関する関数とする. □

(証明) 5.5.2 節参照.

上記の定理 5.2 より, ただちに以下の系が成立する.

系 5.1 (最適な計量行列の固有値のとり得る範囲と分散)

最適な計量行列 \hat{M}_n の固有値 m_{nk} の取り得る範囲と分散は, \hat{M}_n の固有値がすべて等しい場合を除き, η_n の増加に伴い単調に減少する. □

(証明) 5.5.3 節参照.

系 5.2 (正則化項 η_n と固有値の関係)

$\eta_n \rightarrow \infty$ とした場合, $\hat{m}_{nk} \rightarrow 1$ がすべての k に対して成立する. □

(証明) 5.5.4 節参照.

提案手法の最適化問題は, 式 (5.1) より, 各カテゴリの代表元と同一カテゴリに所属する学習データの距離を小さくしながら, 別カテゴリの代表元との距離を遠くすることに相当する. これにより, 新規データをより分類しやすい距離構造を学習できると想定される.

5.2.2 複数の計量行列を用いたデータの分類手法

テンプレートマッチングに基づく分類手法

式 (5.2) により得られたカテゴリ毎の計量行列を用いることで、学習データの統計的性質の差異を考慮した計量行列を学習することが可能となる。いま、カテゴリ情報が未知のテストデータ \boldsymbol{x} を得られた計量行列を用いて分類する際には以下の分類規則を用いる。

$$\hat{c} = \arg \min_{c_n} \left\{ \alpha d_{M_n}^2(\boldsymbol{x}, \boldsymbol{\mu}_n) - \frac{1-\alpha}{N-1} \sum_{n': n \neq n'} d_{M_n}^2(\boldsymbol{x}, \boldsymbol{\mu}_{n'}) \right\}. \quad (5.4)$$

式 (5.4) は、学習で得られた計量行列のうち、最もよくテストデータを分類することのできる計量行列のみを用いたテンプレートマッチング法であると解釈できる。

k -NN 法に基づく分類手法

式 (5.4) による分類規則では、学習によって得られた複数の計量行列のうち、ただ一つを選択的に用いて分類していることと等価である、この方法は、分類対象のデータがあるカテゴリの代表元に近い位置に存在しているときには有効であると考えられるが、複数のカテゴリの代表元から離れた位置に存在している場合には単独の計量行列によって代表元との距離を測ることは有効ではない可能性がある。このような場合、複数の計量行列による距離関係を統合して用いることができれば、複数の計量行列の保持している情報を有効活用した分類規則を規定することができると考えられる。加えて、式 (5.4) では新規入力の分類対象データ \boldsymbol{x} と学習データ \boldsymbol{x}_i 間の距離を測定することは難しい。すなわち、分類対象データの近傍に位置する学習データの情報を用いて分類を行う k -NN 法を適用することはできない。本研究ではこれらの問題点を解決するため、複数の計量行列を活用した k -NN 法に基づく分類規則を以下の通り定義する。いま、入力データを $\boldsymbol{x}_1, \boldsymbol{x}_2$ とし、これらの間の距離 $D_m(\boldsymbol{x}_1, \boldsymbol{x}_2)$ を

$$D_m(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{n=1}^N \frac{d_{M_n}^2(\boldsymbol{x}_1, \boldsymbol{x}_2)}{d_{M_n}^2(\boldsymbol{x}_1, \boldsymbol{\mu}_n) + d_{M_n}^2(\boldsymbol{x}_2, \boldsymbol{\mu}_n)}, \quad (5.5)$$

と定義する。これにより、学習した複数の計量行列を用いたデータ間の距離計算が可能となり、 k -NN 法に準ずる分類を行うことも可能となる。いま、カテゴリ情報が未知のテストデータ \boldsymbol{x} の (距離関数として式 (5.5) を用いた場合の) 近傍 k 個の学習データの集合を $\mathcal{N}_{D_m}(\boldsymbol{x}) = \{\boldsymbol{x}_1^*, \dots, \boldsymbol{x}_k^*\}$

とする。さらに、カテゴリ C_n に $\mathcal{N}_{D_m}(\mathbf{x})$ のデータが所属する数を $k_n(\mathbf{x})$ とすれば、式 (5.5) を用いた k -NN 法に基づくテストデータの分類は

$$\hat{C} = \arg \max_{C_n} k_n(\mathbf{x}), \quad (5.6)$$

とすることで可能となる。

式 (5.5) の距離のイメージを明確にするため、以下の図 5.2 にカテゴリ数が 2 の場合、入力データ $\mathbf{x}_1, \mathbf{x}_2$ 間の距離を測定する例を示す。

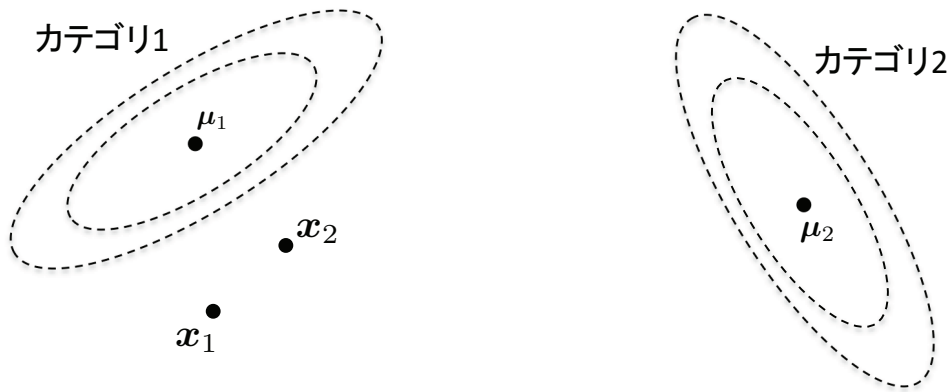


図 5.2: カテゴリ数 2 の場合の式 (5.5) による分類の例

いま、上記の図 5.2 において、学習したカテゴリ毎の計量行列を用いた各ベクトルと各代表元間の距離は $d_{M_1}(\mathbf{x}_1, \mu_1) = 3$, $d_{M_1}(\mathbf{x}_2, \mu_1) = 1$, $d_{M_1}(\mathbf{x}_1, \mathbf{x}_2) = 2$, $d_{M_2}(\mathbf{x}_1, \mu_2) = 10$, $d_{M_2}(\mathbf{x}_2, \mu_2) = 8$, $d_{M_2}(\mathbf{x}_1, \mathbf{x}_2) = 4$ で与えられているものとする。

このとき、式 (5.5) を用いた分類規則では、 $\mathbf{x}_1, \mathbf{x}_2$ 間の距離を $\frac{2}{3+1} + \frac{4}{10+8} = \frac{13}{18}$ として与える。これを用いることにより、相対的にカテゴリ 1 に近いデータ間の距離を測定する際、すなわち、カテゴリ 1 の代表元の近傍に存在するデータ間の距離を測定する場合には計量行列 M_1 により得られた距離計量を重要視し、両カテゴリの代表元からの距離が同程度の場合には両者を等しく用いることで、局所的な統計的特徴を考慮したデータ点間の距離を測定することが可能となる。

5.3 実験

提案手法の有効性を示すため、ベンチマークデータを用いた分類実験を行う。ベンチマークデータは低次元、密なものとして、UCI 機械学習リポジトリ [5] のうち iris, wine, balance, ionosphere

を、高次元、スパースなデータセットとして毎日新聞 2000 年、2005 年、2010 年発行の新聞記事データをそれぞれ用いるものとした。

両実験共に、計量行列の導出には式 (5.1) を用い、テストデータに対する分類は、上記の式 (5.4), (5.5) をそれぞれ用いることとし、比較手法として、他のメトリックラーニング手法（持橋らの手法, ITML, LMNN）, 通常の距離尺度（ユークリッド距離, 余弦尺度）を用いた。また, 式 (5.3) における正規化パラメータ η_m はカテゴリ毎に定義することが可能であるが, 本実験ではすべてのカテゴリに対して同一の値を用いるものとした。

5.3.1 UCI 機械学習リポジトリを用いた実験

実験条件

前述の通り, 以下では UCI 機械学習リポジトリ [5] を用いた実験を行う。これらのデータセットの基本的な情報は表 4.1 と同様である。

また, 提案手法において, 適切なパラメータ α を求めるため, 学習データを対象として 10 分割交差検定を実施し, パラメータを決定することとした。 α は 0 ~ 1.0 までを 0.1 刻みとし, 最も分類精度の高い α を用いて他手法との分類精度の比較を行った。

実験結果

以下の図 5.3–5.6 にそれぞれのデータセットを用いた場合のパラメータ α の違いによる 10 分割交差検定の結果を示す。

図 5.3–5.6 より, α を変化させた場合でも分類精度そのものは大きく変化していないことが分かる。また, ここで得られた α のうち, 分類精度が最も高いものを用いて, 他手法と比較を行った。表 5.1 にこの結果を示す。ただし, 各データセットで最も高い分類精度を達成した手法を太字としている。

表 5.1 から, UCI 機械学習リポジトリを用いた場合の分類精度は式 (5.4) を用いた場合には従来手法とほぼ同等の分類精度となっていることがわかる。しかしながら, 式 (5.5) を用いた場合には, 従来手法以上か同等の分類精度となっていることが分かる。

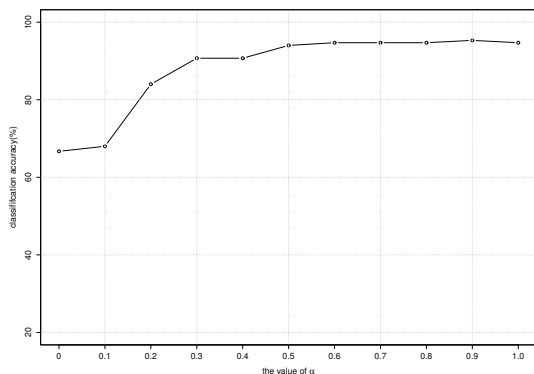


図 5.3: 各 α に対する 10 分割交差検定の分類精度 (iris) [56]

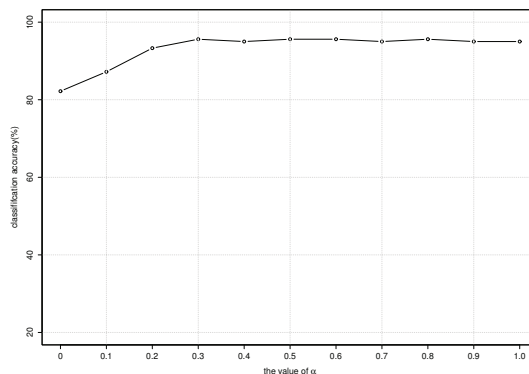


図 5.4: 各 α に対する 10 分割交差検定の分類精度 (wine) [56]

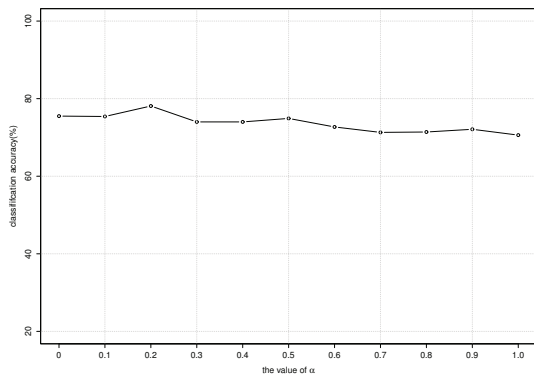


図 5.5: 各 α に対する 10 分割交差検定の分類精度 (balance) [56]

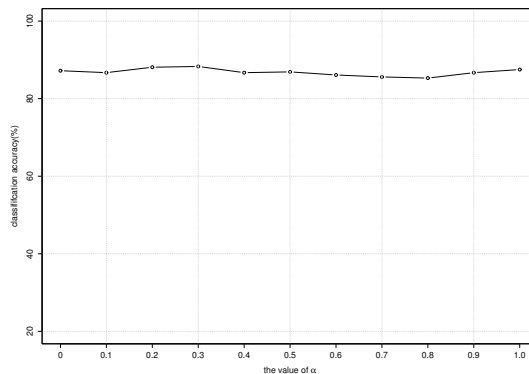


図 5.6: 各 α に対する 10 分割交差検定の分類精度 (ionosphere) [56]

5.3.2 新聞記事データを用いた実験

実験条件

新聞記事データを用いた実験では、国際、経済、芸能、スポーツ、社会の5カテゴリを使用することとした。学習データ数の違いによる分類精度の変化を確認するため、各カテゴリ100件（合計500件）、300件（合計1500件）、500件（合計2500件）とし、テストデータは各カテゴリ500件（合計2500件）として分類精度を算出した。入力データの構成要素は各形態素の出現頻度とし、テストデータに対する分類精度を式(5.4), (5.5)を用いて求めた。

加えて、実問題への適用可能性について検討を行うため、各カテゴリに存在する実データ数に

表 5.1: UCI 機械学習リポジトリを用いた際の分類精度の比較 [56]

	LMNN	ITML	持橋ら	Euclid	Cosine	提案手法 (式 (5.4))	提案手法 (式 (5.5))
iris	95.83	96.00	98.00	92.53	33.30	95.33 ($\alpha = 0.9$)	96.77($\alpha = 0.9$)
wine	91.96	72.47	98.32	72.35	32.90	95.56 ($\alpha = 0.6$)	98.89 ($\alpha = 0.6$)
balance	86.86	88.96	71.08	75.00	87.30	78.09 ($\alpha = 0.2$)	84.73($\alpha = 0.2$)
ionosphere	87.10	86.89	87.96	71.85	75.90	88.33 ($\alpha = 0.3$)	92.76 ($\alpha = 0.3$)

比例したデータ数で他手法との比較も行うものとした。この実験では、毎日新聞 2000 年のデータを対象に前述の 5 カテゴリに含まれるデータをカテゴリ毎のデータ比率を保ったまま全量のデータからランダムサンプリングし、学習データ、テストデータとして各カテゴリで 210, 165, 51, 319, 553 個、合計 1298 個のデータを用いた。

なお、新聞記事を用いた実験でも UCI 機械学習リポジトリを用いた実験同様、事前に 10 分割交差検定を行うことによって適切なパラメータ α を決定している。

実験結果

図 5.7 に 10 分割交差検定の結果を示す。

図 5.7 より、 $\alpha = 0.5$ までは分類精度は上昇した後に低下し、 $\alpha = 0.7$ を境に再度上昇していることが分かる。これより、以降の実験では $\alpha = 0.5$ を用いて他手法と分類精度の比較を行うものとする。

図 5.8 – 図 5.10 に 2000 年、2005 年、2010 年の新聞記事データを用いた他手法との比較結果を、図 5.11 に実データに比例したデータ数による他手法との比較結果を示す。ただし、図 5.8 – 図 5.9 における「Pro_T」は計量行列の学習を提案手法にて行い、テンプレートマッチング (式 (5.4)) で分類を行ったものを示し、「Pro_M」は計量行列の学習を提案手法にて行い、 k -NN 法 (式 (5.5)) で分類を行ったものを示す。

図 5.8 – 図 5.11 より、提案手法は既存手法と比較して式 (5.4), (5.5) のどちらを用いた場合でも従来手法と比較して同程度以上の分類精度を達成していることが分かる。特に式 (5.4) を用いた場合はいずれのデータセットについても従来手法以上の分類精度を達成できている。しかしながら、UCI 機械学習リポジトリを用いた場合とは異なり、式 (5.5) を用いた場合には式 (5.4) を用いた場

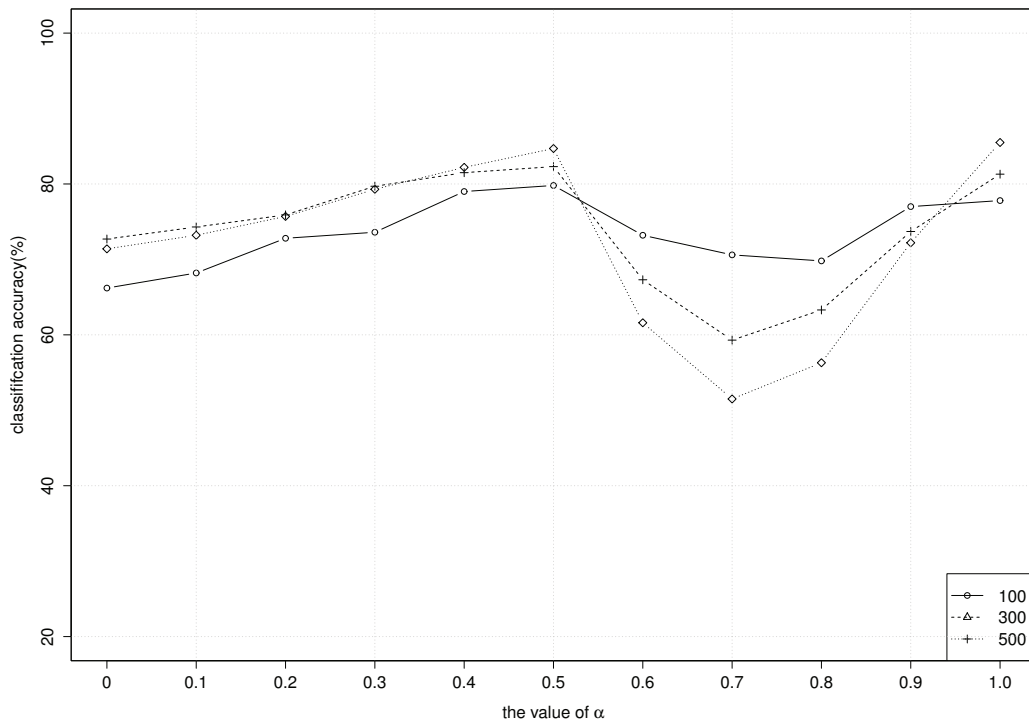


図 5.7: 各 α に対する 10 分割交差検定の分類精度 [56]

合と比較して、分類精度が若干低下していることもわかる。

一方で、既存のメトリックラーニング手法では、学習データの増加に伴い分類精度が低下している。これは、学習データが増加するにつれて、その文書集合中に出現する異なり総単語数が増加し、入力データの次元数が増加したことに起因していると考えられる。入力データの次元数の増加は学習に必要なパラメータ数の増加を意味すると共に、その数は入力データの次元数の二乗に比例する。したがって、従来手法では推定に用いられた学習データ数が、パラメータ数と比較して相対的に少なくなってしまう、分類精度の低下につながったと考えられる。

他方、図 5.11 から、実データと同様の割合のデータを使用した場合も提案手法の分類精度が他手法と比較して高いことが分かる。以上より、本提案手法を実際の文書分類タスクに適用した場合の有効性を示すことができた。

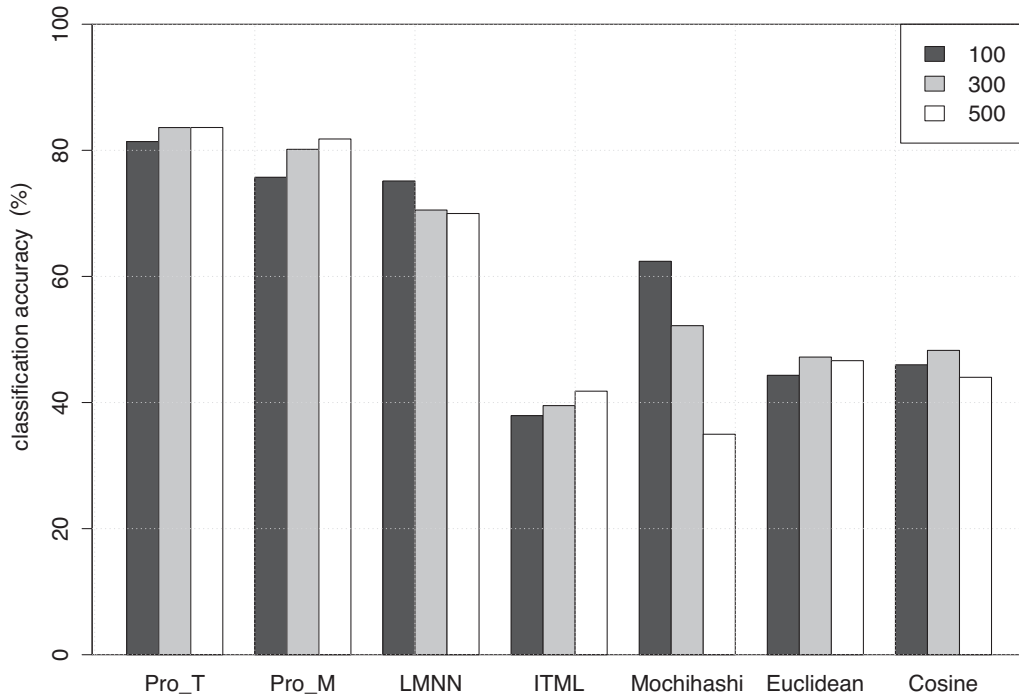


図 5.8: 他手法との比較結果 (毎日新聞 2000 年) [56]

5.4 考察

本研究では、学習データの局所的な統計的特徴を考慮した距離構造を学習するため、カテゴリ毎に計量行列が存在することを仮定し、その導出を行った。式 (5.2) の通り、各カテゴリの計量行列は解析的に導出できることを示した。式 (5.1) で定義される最適化問題は、前述の通り、カテゴリの代表元とそこに所属する学習データとの距離を近づけながら、他カテゴリの代表元とを遠ざける仕組みとなっていることが分かる。

また、定理 5.2, 系 5.1, 5.2 より、学習の結果得られた計量行列は正則化項 $\text{tr}(M_n)$ の効果により、 η_n の増加に伴いすべての固有値が同じ値に近づく。これはすなわち、学習の結果得られた計量行列は、学習データを各カテゴリの代表元を中心とした真球に近づけ、同時に別カテゴリの代表元との距離をできるだけ離す効果があると考えられることができる。

また、式 (5.5) を用いることによって、特に UCI 機械学習リポジトリを用いた場合には分類精度が上昇している。 k -NN 法を基にした手法を用いることでその分類精度は向上することも明らか

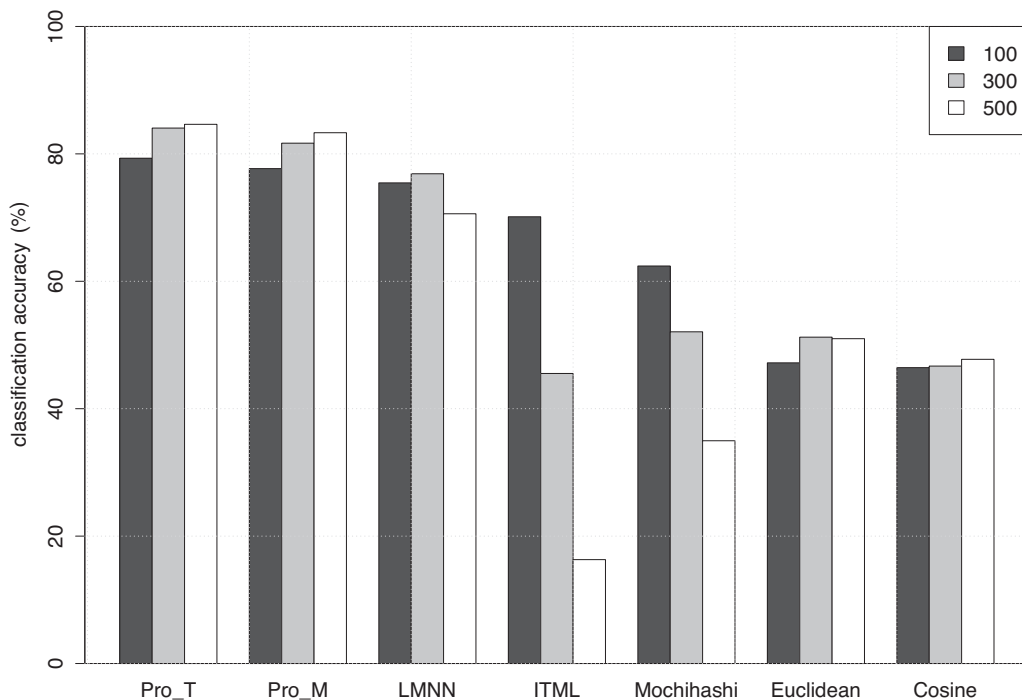


図 5.9: 他手法との比較結果 (毎日新聞 2005 年) [56]

となり、これが有効な手法の一つとなり得ることが判明した。UCI 機械学習リポジトリを用いた際に精度が向上した理由は高次元スパースなデータと比較して、学習データが低次元空間に密集して存在しており、基準となる学習データの近傍に位置する学習データは基準となるデータと類似しているためであると考えられる。一方で、高次元空間に存在しているデータは次元の呪いにより、基準データの近傍に存在している類似データを正確に抽出できなかった可能性があると思定される。したがって、従来手法と比較して k -NN 法に基づく分類手法は有効ではあるものの、テンプレートマッチングに基づく手法と比較すると、高次元なデータを用いた場合にはその性能が若干衰えてしまう可能性があることも示唆される。

一方、各種実験の結果から、本提案手法は特に高次元データに対して有効であることが見て取れる。特に他のメトリックラーニング手法では、学習データ数が増加した場合に分類精度が低下しているのに対し、提案手法は分類精度が上昇している。これは正則化項を追加したことによる影響であると考えられる。この点について検証を行うため、正則化項を変化させた際の

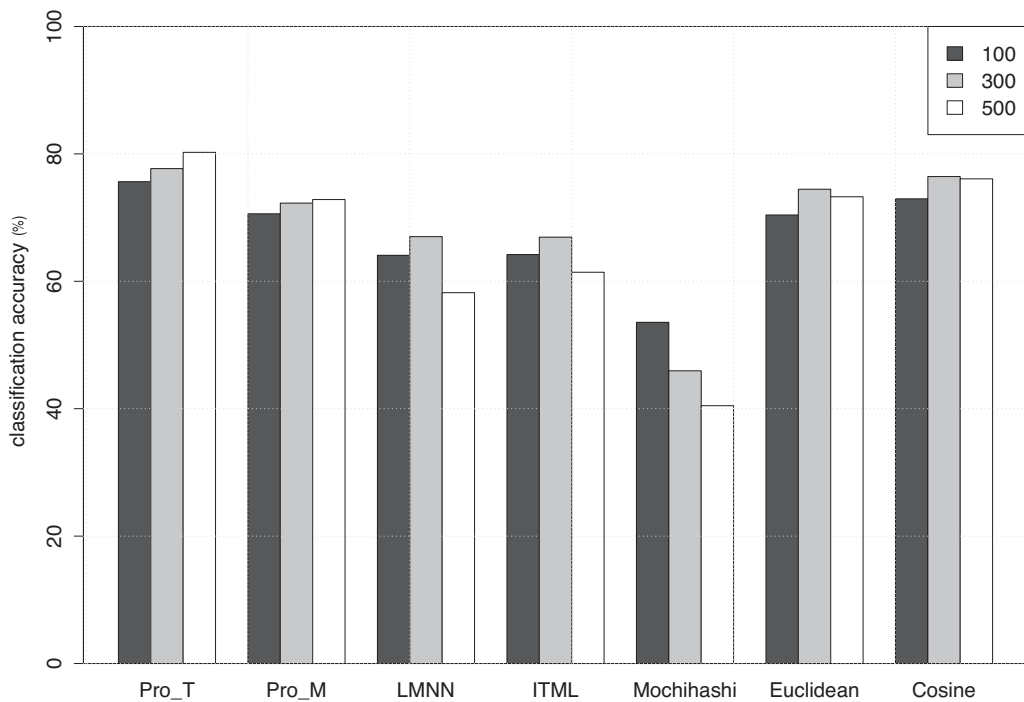


図 5.10: 他手法との比較結果 (毎日新聞 2010 年) [56]

分類精度の違いについて新聞記事データ (毎日新聞 2000 年) を用いた実験を行った。カテゴリ数はこれまでの実験と同様に 5 とし、学習データ数は各カテゴリ 100 件、正則化パラメータをすべてのカテゴリに対して等しく 0 から 100 へ変化させ、各パラメータでの分類精度を測定するものとした。実験の結果を図 5.12 に示す。

図 5.12 より、正則化項が存在しない場合と比較し、正則化パラメータ η_m が比較的小さい場合には分類精度が向上していることが分かる。ここから、適切な正則化パラメータの設定が必要であるものの、その追加は分類精度を向上させる効果があることが分かる。

しかし一方で、使用するデータが低次元で密なものであった場合、従来手法と比較し、分類精度はそれほど大きく変化していない。これは次元数に対し十分な量の学習データが存在している低次元かつ密なデータセットに対しては、正則化により逆に分類精度の低下を引き起こしてしまった可能性が考えられる。提案手法は従来手法と比較し、あらかじめ決定するパラメータ数 (α, η_m) が増加している。これらの影響により、どの程度分類精度に変化があるのか、最適な α, η_m の組み

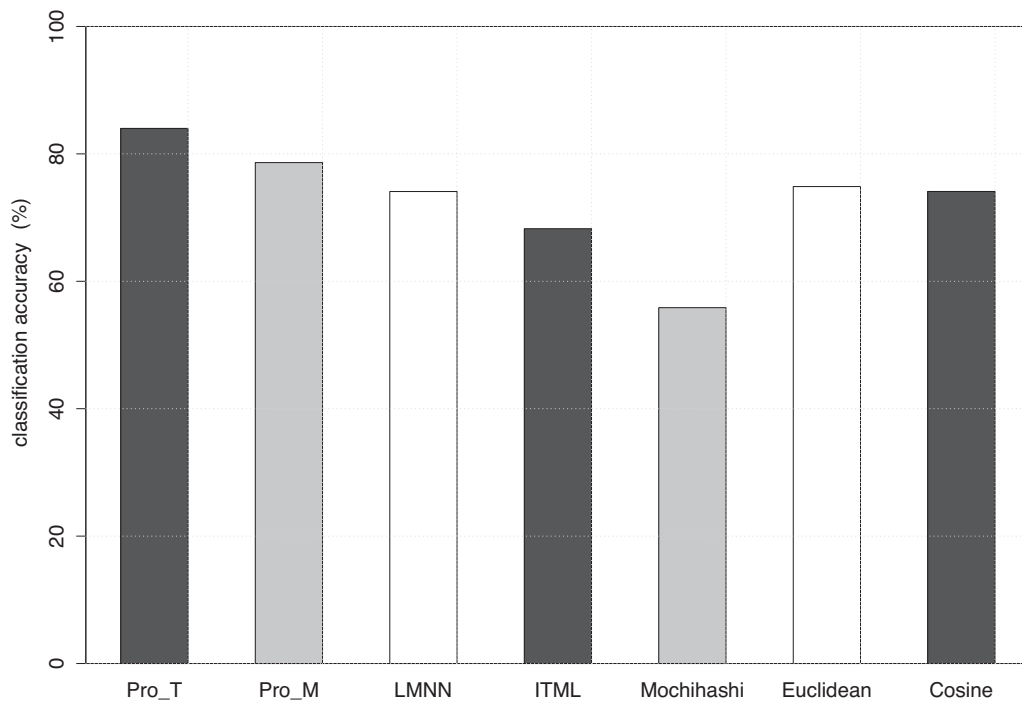


図 5.11: 実データに比例したデータ数による他手法との比較結果 (毎日新聞 2000 年) [56]

合わせについてはより詳細な検討が必要である。加えて、本研究で定義した制約条件はすべてのカテゴリに対し、計量行列の行列式の対数値が等しいというものであるが、例えば、各カテゴリでデータのばらつき具合が大きく異なる場合などデータの分布によっては、このような条件が適切ではない可能性がある。

5.5 定理, 補題, 系の証明

5.5.1 定理 5.1 の証明

式 (5.1) で定義された最適化問題を解くため、ラグランジュ関数 $L(M_n, \rho_n)$ を以下の通り定義する。

$$L(M_n, \rho_n) = \alpha \frac{\text{tr}(M_n S_n)}{|C_n|} - \frac{1 - \alpha}{N - 1} \sum_{n': n \neq n'} \frac{\text{tr}(M_n S_{n'})}{|C_{n'}|} + \eta_n \text{tr}(M) - \rho_n \log \det M_n. \quad (5.7)$$

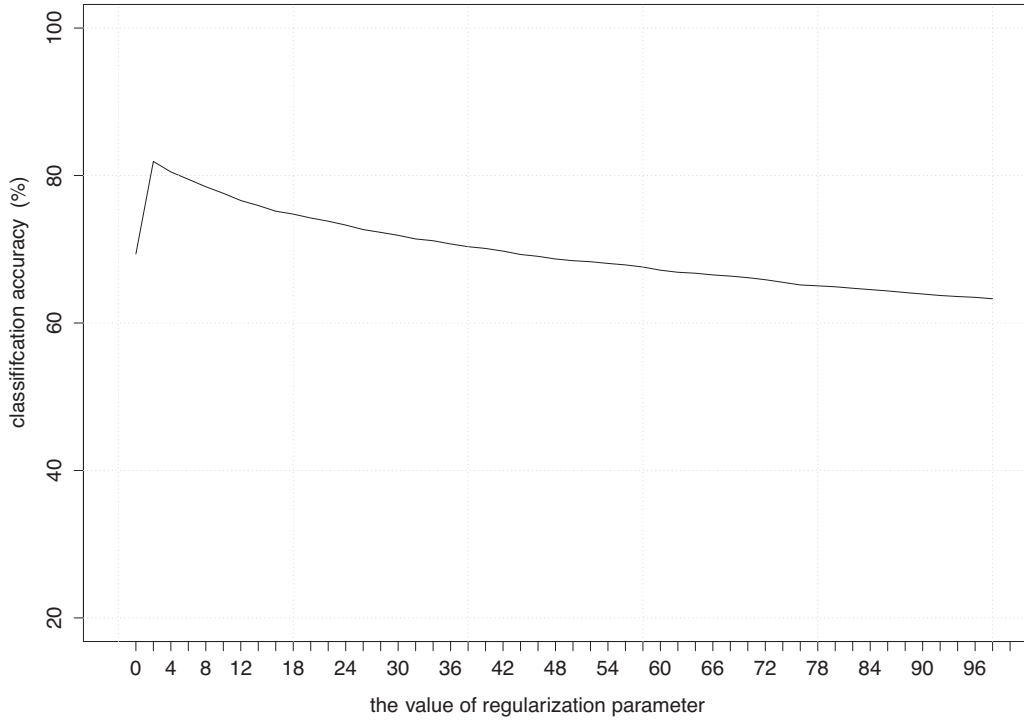


図 5.12: 正則化パラメータ η_n の変化に伴う分類精度の変化 (毎日新聞 2000 年) [56]

ただし, ρ_n を未定乗数とする. これを微分すれば,

$$\nabla_{M_n} = \alpha \frac{S_n}{|C_n|} - \frac{1-\alpha}{N-1} \sum_{n':n \neq n'} \frac{S_{n'}}{|C_{n'}|} + \eta_n I - \rho_n M_n^{-1}, \quad (5.8)$$

となる. これを $\mathbf{0}$ とおいて整理すると,

$$\rho_n M_n^{-1} = \alpha \frac{S_n}{|C_n|} - \frac{1-\alpha}{N-1} \sum_{n':n \neq n'} \frac{S_{n'}}{|C_{n'}|} + \eta_n I, \quad (5.9)$$

より,

$$\rho_n^W \det(M_n^{-1}) = \det \left(\alpha \frac{S_n}{|C_n|} - \frac{1-\alpha}{N-1} \sum_{n':n \neq n'} \frac{S_{n'}}{|C_{n'}|} + \eta_n I \right), \quad (5.10)$$

を得る。いま、式 (5.10) の行列式を考えると、 $\log \det M_n = 0$ の制約条件から、

$$\log \rho_n^W = \log \det \left(\alpha \frac{S_n}{|\mathcal{C}_n|} - \frac{1-\alpha}{N-1} \sum_{n':n \neq n'} \frac{S_{n'}}{|\mathcal{C}_{n'}|} + \eta_n I \right). \quad (5.11)$$

を得る。したがって、 ρ_n は

$$\rho_n = \left\{ \det \left(\alpha \frac{S_n}{|\mathcal{C}_n|} - \frac{1-\alpha}{N-1} \sum_{n':n \neq n'} \frac{S_{n'}}{|\mathcal{C}_{n'}|} + \eta_n I \right) \right\}^{\frac{1}{W}}, \quad (5.12)$$

となる。その結果、最適な計量行列 \hat{M}_n は

$$\hat{M}_n = \left\{ \det \left(\alpha \frac{S_n}{|\mathcal{C}_n|} - \frac{1-\alpha}{N-1} \sum_{n':n \neq n'} \frac{S_{n'}}{|\mathcal{C}_{n'}|} + \eta_n I \right)^{\frac{1}{W}} \right\} \left(\alpha \frac{S_n}{|\mathcal{C}_n|} - \frac{1-\alpha}{N-1} \sum_{n':n \neq n'} \frac{S_{n'}}{|\mathcal{C}_{n'}|} + \eta_n I \right)^{-1}, \quad (5.13)$$

となり、定理は示された。 \square

5.5.2 定理 5.2 の証明

いま、 $\text{Eigen}[\cdot]$ を最適な計量行列 \hat{M}_n の固有値を並べたものであるとする。このとき、

$$\begin{aligned} \text{Eigen} [\hat{M}_n] &= (\hat{m}_{n1}, \dots, \hat{m}_{nW})^T \\ &= \text{Eigen} \left[|A_n|^{\frac{1}{W}} A_n^{-1} \right], \end{aligned} \quad (5.14)$$

が成立する。ただし、 \hat{m}_{nk} は \hat{M}_n の k 番目の固有値であるとする。式 (5.3) より、 A_n は

$$A_n = \left(\alpha \frac{S_n}{|\mathcal{C}_n|} - \frac{1-\alpha}{N-1} \sum_{n':n \neq n'} \frac{1}{|\mathcal{C}_{n'}|} S_{n'} + \eta_n I \right),$$

であるので、 $\text{Eigen} [\hat{M}_n]$ は

$$\text{Eigen} \left[|A_n|^{\frac{1}{W}} A_n^{-1} \right] = \prod_{i=1}^W (\lambda_{ni} + \eta_n)^{\frac{1}{W}} \left(\frac{1}{\lambda_{n1} + \eta_n}, \dots, \frac{1}{\lambda_{nW} + \eta_n} \right)^T, \quad (5.15)$$

となる。ただし、 λ_{ni} は行列 A_n の固有値のうち、 i 番目に大きい固有値を示すものとする。

したがって、最適な計量行列 \hat{M}_n の k 番目の固有値 \hat{m}_{nk} は

$$\hat{m}_{nk} = \left\{ \prod_{w=1}^W (\lambda_{nw} + \eta) \right\}^{\frac{1}{W}} \frac{1}{\lambda_{nk} + \eta_n} \quad (5.16)$$

で表すことができる。上記の議論から

$$\log \hat{m}_{nk} = \frac{1}{W} \sum_{w=1}^W \log (\lambda_{nw} + \eta_n) - \log (\lambda_{nk} + \eta_n), \quad (5.17)$$

がただちに導かれる。

ここで、 \hat{m}_{nk} と $\log \hat{m}_{nk}$ の増減は一致するので $\frac{d \log \hat{m}_{nj}}{d \eta_n}$ についてその増減を検討する。いま、 $\frac{d \log \hat{m}_{nj}}{d \eta_n}$ は以下の通り表現できる。

$$\begin{aligned} \frac{d \log \hat{m}_{nj}}{d \eta_n} &= \frac{1}{W} \sum_{i=1}^W \frac{1}{\lambda_{ni} + \eta_n} - \frac{1}{\lambda_{nj} + \eta_n} \\ &= \frac{\bar{m}_n - \hat{m}_{nj}}{\left\{ \prod_{i=1}^W (\lambda_{ni} + \eta_n) \right\}^{\frac{1}{W}}}. \end{aligned} \quad (5.18)$$

式 (5.18) より、 $\bar{m}_n < \hat{m}_{nj}$ ならば、 $\frac{d \log \hat{m}_{nj}}{d \eta_n} < 0$ 、 $\bar{m}_n > \hat{m}_{nj}$ ならば、 $\frac{d \log \hat{m}_{nj}}{d \eta_n} > 0$ 。 $\bar{m}_n > \hat{m}_{nj}$ のとき、 \hat{m}_{nj} は η_n に対し単調増加であり、 $\bar{m}_n < \hat{m}_{nj}$ のとき、 \hat{m}_{nj} は η_n に対し単調減少となる。したがって、定理は証明された。 \square

5.5.3 系 5.1 の証明

一般に、計量行列 \hat{M}_n におけるすべての固有値は等しくはないため、固有値の最大値 \hat{m}_{n1} と最小値 \hat{m}_{nW} とその平均 \bar{m}_n は異なる。また、明らかに $\hat{m}_{n1} > \bar{m}_n$ かつ $\hat{m}_{nW} < \bar{m}_n$ が成立する。したがって、定理より η_n が増加すれば \hat{m}_{n1} は減少し、 \hat{m}_{nW} は増加する。その結果、それらの範囲 $\hat{m}_{n1} - \hat{m}_{nW}$ は η_n により単調減少する。同様に、 \bar{m}_n と異なる固有値は η_n の増加により \bar{m}_n に近づくため、その分散も単調に減少する。 \square

5.5.4 系 5.2 の証明

いま、 \hat{m}_{n1} (計量行列 \hat{M}_n の最大固有値) と \hat{m}_{nW} (計量行列 \hat{M}_n の最小固有値) の差について考える。これはそれぞれ、 $\hat{m}_{n1} = \left\{ \prod_{i=1}^W (\lambda_{ni} + \eta_n) \right\}^{\frac{1}{W}} \frac{1}{\lambda_{n1} + \eta_n}$ 、 $\hat{m}_{nW} = \left\{ \prod_{i=1}^W (\lambda_{ni} + \eta_n) \right\}^{\frac{1}{W}} \frac{1}{\lambda_{nW} + \eta_n}$

と書けるので、その差分を $\beta(\eta_n)$ とすれば、 $\beta(\eta_n)$ は

$$\beta(\eta_n) = \left\{ \prod_{i=1}^W (\lambda_{ni} + \eta_n) \right\}^{\frac{1}{W}} \left(\frac{1}{\lambda_{n1} + \eta_n} - \frac{1}{\lambda_{nW} + \eta_n} \right) \quad (5.19)$$

となる。式 (5.19) は、

$$\begin{aligned} \beta(\eta_n) &= \eta_n \left\{ \prod_{i=1}^W \left(1 + \frac{\lambda_{ni}}{\eta_n} \right) \right\}^{\frac{1}{W}} \left\{ \frac{\frac{1}{\eta_n}}{\frac{\lambda_{n1}}{\eta_n} + 1} - \frac{\frac{1}{\eta_n}}{\frac{\lambda_{nW}}{\eta_n} + 1} \right\} \\ &= \left\{ \prod_{i=1}^W \left(1 + \frac{\lambda_{ni}}{\eta_n} \right) \right\}^{\frac{1}{W}} \left\{ \frac{1}{\frac{\lambda_{n1}}{\eta_n} + 1} - \frac{1}{\frac{\lambda_{nW}}{\eta_n} + 1} \right\} \end{aligned} \quad (5.20)$$

と変形することができる。したがって、 $\lim_{\eta_n \rightarrow \infty} \beta(\eta_n) = 0$ が成立する。これは、固有値の最大値と最小値が等しいことと等価であるため、 $\eta_n \rightarrow \infty$ としたときには、結果としてすべての固有値は等しい。また、制約条件の $\log \det M_n = 0$ より、 $\det \hat{M}_n = 1$ が成立する。すなわち、 $\det \hat{M}_n = \prod_{j=1}^W \hat{m}_{nj}$ より、 $\forall j, \hat{m}_{nj} = 1$ となる。□

5.6 本章の結論

本章では、従来の大域的メトリックラーニングの問題点を改善するため、繰り返し処理を用いない計量行列学習法において、カテゴリ毎に異なる計量行列の存在を仮定し、これらの計量行列が繰り返し処理を用いずに導出可能となることを示した。併せて、導出したそれぞれの計量行列が持つ性質を明らかにした。

また、導出した複数個の計量行列を用いた k -NN 法を構築することにより、新規入力データの新たな分類手法について提案を行った。

機械学習におけるベンチマークデータ、新聞記事データを用いた分類実験を行うことにより、本章で提案した手法が分類精度の面で従来のメトリックラーニング手法と比較し優れていることを示した。

本手法は、学習時、分類時ともに第3章、第4章で述べた提案手法と比較し計算量を多く必要とする。その一方で、分類時の精度は他の提案手法よりも高く、十分なコンピュータ資源があり、分析実行時に計算量が多くかかったとしても高性能な分類を行いたい場合には有効な手法であると言える。

第6章 考察

6.1 本研究の考察

本論文では持橋らの手法に対し、

- 正則化項の追加による計量行列の導出
- 複数の計量行列の導出とそれらを用いた分類手法の構築

による持橋らの手法の問題点の改善と分類精度の向上を行った。

本論文で提案した手法は、それぞれ異なる特徴を持つデータや、分類時の状況に応じて使い分けることで特に有効であると考えられる。対象となるデータ、ならびに各提案手法の使い分けを以下に示す。

提案手法 1 (第 3 章) :

- 学習データの総数とその次元よりも少ない、かつできるだけ計算量をかけずに計量行列を導出したい場合。
- カテゴリ毎の学習データの特徴がそれほど大きく変わらないと想定される場合。

提案手法 2 (第 4 章) :

- 学習データベクトルの成分間に統計的関係性がないペアがあると想定される場合。
- (提案手法 1 と比較して) 多少計算量が多くなったとしても、できるだけシンプルな構造を得たい場合。
- できるだけ少ない計算量で新規入力データを分類したい場合。

提案手法 3 (第 5 章) :

- 学習データのカテゴリ毎にその統計的特徴が異なると想定される場合。

- 新規入力データの分類時にできるだけ高い分類精度を達成したい場合.
- 複数の計量行列の特徴を各々用いて分類を行いたい場合.
- 十分なコンピュータ資源が存在している場合.

既に述べている通り、提案手法1の特徴は計算量が相対的に少なく、かつ必ず一意に最適な計量行列を導出できることにある。したがって、コンピュータ資源の問題などで、学習にできるだけ時間をかけずに計量行列を得たい場合はこの手法を用いることが望ましい。特に、計量行列の導出に多く計算量がかかる高次元なデータを対象とした場合、提案手法1を使用することは有効であると考えられる。

提案手法2を用いることで、その要素の多くが0となるスパースとなる計量行列を導出することが可能となった。これは、推定するパラメータ数を削減させると共に、よりシンプルな構造で計量行列を表現することに相当する。計量行列の各要素は学習データの特徴間の関係性を表現していると解釈できることから、互いに関係がないような特徴間の重みを自動的に0とする本手法は応用面で様々な可能性があると考えられる。加えて、学習した計量行列はスパースな構造を持っていることから、密な構造を持つ計量行列と比較し、新規入力データを分類する際のマハラノビス距離の算出に必要な計算量を低減することができる。したがって、学習には時間をかけても構わないが、新規入力データの分類はできる限り短い時間で行いたい場合などにこの手法は有効であると言える。

また、提案手法3では、各カテゴリの統計的特徴を考慮した計量行列を算出できることから、学習データの統計的性質、特徴が異なることがあらかじめわかっているような場合、かつ提案手法1と比較して多少計算量、メモリ使用量が増加したとしても、高精度な予測を優先したい場合には、本手法を用いることは有効であると言える¹。加えて、複数の計量行列を用いた k -NN法を適用することができるため、局所的に学習データが分布しているような場合にも本提案手法は有効である可能性がある。

これらの手法それぞれを使い分けるための明確（定量的）な基準の決定は必要不可欠であるものの、現実存在している多種多様なデータに対して本提案手法を適用し、そのパフォーマンス

¹カテゴリ C_n における計量行列を学習するための計算コストは提案手法1と同様であるが、それを複数学習する必要があるため、提案手法3は提案手法1と比較した場合、カテゴリ数に比例した計算量が必要となる。

加えて、学習した計量行列をカテゴリ数に比例した分を保持したまま分類することから、必要とするメモリ量も提案手法1と比較して大きくなる。

を確認する必要がある。このような定量的な基準の検討に関しては今後の課題とする。

本研究で着目した持橋らの手法 [59] では、暗黙の裡にカテゴリ C_n に所属するデータが平均 μ_n 、分散共分散行列が M^{-1} の多次元正規分布に従うことを想定している。すなわち、代表元を式 (2.14) として用いることは、上記の条件下でその平均値 μ_n の最尤推定を行っていることに相当する。これは、分析に用いるデータが上記の多次元正規分布に従っている場合、もしくは十分な量の学習データが存在し、正規近似が可能となる場合には有効な手法であるが、本論文で対象としているような高次元データのように、パラメータ数に対し、十分な量の学習データを用意できない場合、もしくは明確にデータが多次元正規分布に従っていないような場合には有効ではない可能性がある。このような場合、学習データの特性を考慮し、他の（多次元）確率分布を想定する、もしくはノンパラメトリックな代表元の推定を行うことができれば、より高精度な分析が可能となる可能性があり、さらなる検討が必要である。

6.2 提案手法の実問題への適用について

本研究で提案した各手法を知識発見手法に援用することで、新たな知見を得ることができる可能性がある。具体的には、以下の二点が考えられる。

- 学習した計量行列を用いた特異な（学習）データの発見。
- 学習した計量行列の各要素に着目した学習データの各素性間の関係性の発見。

前者の「学習した計量行列を用いた特異な（学習）データの発見」に関しては、学習した計量行列に対し、学習データを再代入しその分類結果を確認することで、同一のカテゴリに所属していながらも他の同一カテゴリの学習データと異なる傾向、特徴を持ったデータを抽出できる可能性がある。このようなデータは、非常に有益な情報を有している可能性があり、その抽出を行うことで、有効であると想定される。以下にその例を示す。

例 6.1 (特異なデータの抽出 [57])

価格.comなどに代表されるユーザの口コミを投稿できるwebサイトでは、ユーザのある商品に対する評価が評価点付きで大量に蓄積されている。この評価点に着目し、例えばその高低でユーザの口コミに満足度が高い、満足度が低い、というカテゴリを付与する。これらのデータを学習データとし、計量行列を学習した後、学習した計量行列を用いて再度学習データを分類する。

これにより、誤分類されたユーザは、満足度が高い（低い）カテゴリに存在しながらも他の満足度が高い（低い）ユーザと異なる特徴を持っている可能性がある。これらのデータを個別に分析することで、分析対象者がこれまで保有していなかった新たな知見を得ることができると考えられる。

他方、後者の「学習した計量行列の各要素に着目した学習データの各素性間の関係性発見」については、計量行列の要素を分析することで、知識発見ができる可能性があることに着目したものである。既に述べている通り、計量行列の各要素は学習データの成分間の関係性を表現しているものと解釈できる。このため、学習した計量行列を分析することで、学習データ全体に亘って特徴的な関係性がある成分を抽出できる可能性がある。特に、提案手法2（スパースな計量行列の導出）を用いることで、「互いに関係性が強い」特徴量のペアを抽出するのみならず、「互いに関係がない」と想定される特徴量のペアに関しても同時に抽出できることが想定される。

例 6.2 (特徴的な単語ペアの抽出)

前述の特異なデータの発見同様、口コミサイトを対象とし、評価点の高低でカテゴリ分けされたデータについて着目する。これらについて、計量行列を学習し、その要素を分析することで目的関数で設定した条件のもとで関係性の強い単語ペア、関係性が弱い単語ペアをそれぞれ抽出することができる。加えて、提案手法3（複数の計量行列の導出）を用いることによって、満足度が高い（低い）カテゴリにのみ特徴的に共起しやすい単語のペアも抽出することが可能となる。これを分析することにより、例えば満足度の高いユーザと、満足度の低いユーザの特徴の差異が分析可能となると共に、その中でも特に注視している点を明らかにすることが可能になると考えられる。これらの情報を有効に活用することで、マーケティング分析の一助になると想定される。

第7章 結論

7.1 本研究の結論

本論文では，入力データの統計的特徴から所望の距離構造を学習することができるメトリックラーニング手法のうち，教師あり学習に基づく代表元を用いた手法 [59] をベースとし，この手法において問題となっている，以下の点について改善を行う手法を提案した．

持橋らの手法の問題点 1：

高次元なデータを対象とした場合，もしくは次元数に対し，学習データ数が著しく少ない場合，最適な計量行列の存在が保証されていない．

持橋らの手法の問題点 2：

学習データの構成要素間の一部に相関関係がないデータを対象とした場合，導出された最適な計量行列ではその構造を適切に表現できない．

持橋らの手法の問題点 3：

学習データのカテゴリ毎に統計的特徴が異なるデータを対象とした場合，その局所的構造を表現できない．

本研究で提案した各手法の概要は以下の通りである．

提案手法 1： 計量行列のトレースによる正則化法を用いた最適な計量行列の導出（第 3 章）

提案手法 2： l_1 正則化を用いたスパースな計量行列の導出（第 4 章）

提案手法 3： 複数の計量行列の導出とそれを用いた分類手法の提案（第 5 章）

上記提案手法により，本論文で明らかにされた成果をそれぞれ以下に示す．

計量行列のトレースを用いた正則化法による最適な計量行列の導出（第3章）

本論文第3章では、学習データの次元数が学習データ数よりも大きい場合、もしくは学習データ数がその次元数に対して著しく少ない場合に対する分析手法を構築した。従来の代表元を用いた計量行列の導出手法に対し、そのトレースを正則化項として追加することで、従来手法同様に解析的に最適な計量行列が導出できることを示した。また、正則化項の導入により、学習データ数がその次元数よりも少ない場合、計量行列を導出することができない従来の代表元を用いたメトリックラーニングに対し、第3章で提案した手法では、最適な計量行列が必ず一意に存在することを示した。

さらに、提案した最適化問題を解くことにより得られる計量行列に対し、正則化パラメータ η とその固有値の関係について以下の点を明らかにした。

- 正則化パラメータ η と最適な計量行列 \hat{M} の固有値の関係性
- 最適な計量行列 \hat{M} の固有値の取り得る範囲と分散
- 最適な計量行列 \hat{M} の固有値と正則化パラメータ η との関係性

また、新聞記事データを用いた分類実験により、第3章にて提案した手法が従来の手法と比較し、分類精度、計算量の面で優れていることを示した。

計量行列の l_1 ノルムを用いたスパースな計量行列の導出（第4章）

本論文第4章では、学習データの成分同士の一部に統計的関係性がないと想定されるデータに対する分析手法として、これを適切に表現する、すなわち関係性がない成分間の計量行列の値を0とするスパースな計量行列の導出方法を示した。

具体的には、代表元を用いたメトリックラーニング手法に計量行列の l_1 ノルムを正則化項として導入し、スパースな計量行列を導出する手法を示した。この手法では、非連続な関数を正則化項として追加しているため、従来の勾配法ベースのアルゴリズムでは最適な計量行列を導出することはできないが、ADMMを用いることでそれが可能になる。これにより、推定に必要なパラメータ数が減少し、シンプルな構造を持つモデルが学習できることを示した。

各種の実験結果より、第4章で提案した手法は、従来手法、ならびに第3章で提案した手法と同程度の分類精度を、スパースとなる計量行列を用いて達成できることを明らかにした。

複数の計量行列の学習法とそれを用いた距離計算手法の構築（第5章）

本論文第5章では、カテゴリ毎にその統計的特徴が異なるようなデータを対象とし、大域的に学習を行う通常のメトリックラーニング手法では各カテゴリの局所的な統計的特徴を考慮できないという問題点から、カテゴリ毎に計量行列の存在を仮定し、それらの導出方法について示した。

また、最適な計量行列の導出は第3章と同様の方法を用いることで、解析的に行うことが可能となることを示し、各計量行列の持つ特性を明らかにした。

加えて、複数の計量行列を直接用いたデータ同士の距離測定方法として、 k -NN法をベースとした分類方法を構築した。

これらの手法を用いることで、高次元、スパースなデータセット、ならびに低次元、密なデータセットの両者に対し、第5章で提案した手法が分類精度の面で有効であることを示した。

本研究の成果

本研究の成果は、代表元を用いたメトリックラーニング手法である持橋らの手法 [59] の持つ問題点を改善し、入力データの特性に合わせた距離計量の学習法を構築したことにある。持橋らの手法は、他のメトリックラーニング手法同様、比較的次元かつ密なデータを対象とした分析手法であると言え、その適用範囲を拡張したという点で本提案手法は価値があると考えられる。

前述の通り、提案した各手法は入力データの特性や分析時の状況に合わせて適切に使い分けることで分析者が望む結果を得ることができると想定される。各手法を使用すべき状況と、対象とするデータは以下の通りとなる。

提案手法 1 :

- 学習データ数とその次元数よりも少ない状況である、すなわち比較的高次元なデータを対象としている場合。
- 各カテゴリの学習データの傾向が比較的類似していると想定されるような場合。

提案手法 2 :

- 学習データの成分間に関係性がないと想定されるペアが存在するような場合。
- 学習にかかる計算量が多少増加したとしても、分類精度を維持したままシンプルな構造を学習したい場合。

表 7.1: 提案手法の持橋らの手法との相違, 各手法の特徴

	分類精度	計算量		最適性の保証	目的	目的達成のための手段
		学習時	分類時			
持橋らの手法	△	◎	○	△	<ul style="list-style-type: none"> 計量行列の学習 	<ul style="list-style-type: none"> 代表元を用いた距離計量の学習
提案手法1	○	◎	○	○	<ul style="list-style-type: none"> 持橋らの拡張(計量行列の存在保証) 	<ul style="list-style-type: none"> 計量行列のトレースを用いた正則化
提案手法2	○	△	◎	○	<ul style="list-style-type: none"> スパースな計量行列の学習 分類時の計算量低減 	<ul style="list-style-type: none"> 計量行列の l_1 ノルムを用いた正則化
提案手法3	◎	○	△	○	<ul style="list-style-type: none"> 複数の計量行列導出 分類精度向上 	<ul style="list-style-type: none"> カテゴリ間の統計的特徴の差異を考慮した最適化 それを用いた分類手法の構築

- 新規入力データの分類時の計算コストを少なく抑えたい場合.
- 短時間で多くの新規入力データを分類しなければならない応用場面に用いる場合.

提案手法 3 :

- 学習データのカテゴリ毎にその統計的特徴が大きく異なると想定される場合.
- 計量行列の学習時の計算コストが多少増加してしまったとしても, 高い分類精度を達成したい場合.

このように, 本研究で提案した手法を活用することで, 従来手法では適用が難しかった様々な対象問題や分析時の状況に対し, 重視すべき観点からの分析が可能となる. 分析対象データの多様化, 大規模化が進んでいる今日では, 提案した手法を効果的に用いることでより多くのデータを解析することが可能となる. この点からも, 本研究にて提案した手法には価値があると考えられる.

上記の議論から, 本研究で提案した各手法と従来の持橋らの手法との相違点, 各手法の特徴をまとめたものを表 7.1 に示す.

表 7.1 にある通り, 本研究で提案した各手法はそのいずれかが優れているという訳ではなく, そ

れぞれが異なる特徴を有した手法となっている。したがって、データの自動分類を提案手法を用いて行う場合、データの特性のみならず、学習にかかる計算量、分類時の計算量、分類精度のいずれを重要視するのかという応用上の評価基準の差異によって各提案手法を使い分けることが可能となる。

これらの議論より、入力データの特性、各提案手法のメリット、デメリットを考慮することで望ましい分析を実施することが可能となると考えられる。

7.2 今後の課題と展望

本研究では、代表元を用いたメトリックラーニング手法に着目し、上記の点を明らかにすることができた。しかしながら、以下の点については未解決であるため、さらなる検討が必要であると考えられる。

1. 計量行列導出時の計算量の低減
2. 代表元の計算方法
3. 実データへの適用

上記1.の計量行列導出時の計算量の低減であるが、提案手法1のように解析的に最適な計量行列を得ることができる手法でも、逆行列の計算を伴うために $O(W^3)$ の計算量が必要となる。従来の繰り返し処理を用いる手法と比較した場合には相対的に計算量は少ないと言えるものの、その導出には多くの計算時間を要してしまう。ADMMを用いた手法も同様に、毎回の更新時に固有値分解が必要となるため、 $O(W^2)$ 程度の計算コストがかかる。これらの計算量は、多種多様なデータセットへ本手法を適用する際に重要であるため、その改善が必須である。これらは、近年着目されている分散処理を用いた手法により改善できる可能性があり、その適用可能性について検討を行う必要がある。

上記2.の代表元の計算方法であるが、特に高次元なデータセットについては、各カテゴリのデータを一つの代表元で表現することで、逆に分類精度の低下を引き起こしている可能性がある。代表元に式(2.14)を用いることは、暗黙の裡に各カテゴリが多次元正規分布に従うことを想定しているが、この仮定が成り立たない場合、ならびに同一カテゴリ内のデータが局所的に複

数のまとまりをもつような場合には、異なる代表元や複数の代表元を用いた手法を検討する必要がある。

最後に、上記3.の実データへの適用であるが、本手法はベンチマークデータを用いた性能評価のみを行っており、現実中存在する多様なデータへの適用可能性、ならびにそこからの知識発見手法については未検討である。例えば、計量行列の要素は学習データの成分間の関係性を表現することが可能であるため、その要素を抽出、分析することで要素間の関係性などを把握できる可能性がある。

また、本研究では、データは必ずどこかのカテゴリに所属しているという前提のもとで分類を行ったが、実データを用いた場合にはこのような前提は成り立たない場合もある。すなわち、分類したデータがどのカテゴリの代表元からも離れたところへ位置するような外れ値である場合には、このようなデータの自動分類を諦め、人手、もしくはさらに高性能な分類器で分類を行うリジェクトルールの考え方を導入するなどの工夫が必要となる。また、データが複数のカテゴリに所属することを許容する問題におけるメトリックラーニング手法の検討についても、別途対応が必要である。これらの検討、適用についても今後の課題とする。

付録

複数の計量行列を用いたスパースな正則化法

以下では、第5章で示した複数の計量行列を用いた手法をベースとし、これに第4章と同様スパースな正則化項を追加することで複数のスパースな計量行列を導出する手法について説明を行う。

いま、式(5.1)により定義された最適化問題の正則化項を計量行列 \hat{M}_n の l_1 ノルムとした、以下のような最適化問題を考える。

$$\hat{M}_n = \underset{M_n}{\text{minimize}} \left\{ \frac{\alpha_n}{|\mathcal{C}_n|} \text{tr}(M_n S_n) - \frac{1 - \alpha_n}{N - 1} \sum_{n': n \neq n'} \frac{\text{tr}(M_n S_{n'})}{|\mathcal{C}_{n'}|} - \log \det M_n + \eta_n \|M_n\|_1 \right\}, \quad (7.1)$$

subject to $\forall n, \quad M_n \succeq 0$.

上記の最適化問題に対し、ADMMを用いるため、第4章同様以下の通り変形を行う。

$$\hat{M}_n = \underset{M_n}{\text{minimize}} \left\{ \frac{\alpha_n}{|\mathcal{C}_n|} \text{tr}(M_n S_n) - \frac{1 - \alpha_n}{N - 1} \sum_{n': n \neq n'} \frac{\text{tr}(M_n S_{n'})}{|\mathcal{C}_{n'}|} - \log \det M_n + \eta_n \|Z_n\|_1 \right\},$$

subject to $M_n = Z_n$ (7.2)

$$\forall n, \quad M_n \succeq 0$$

さらに、式(7.2)の拡張ラグランジュ関数 $L_{\rho_n}(M_n, Z_n, U_n)$ は

$$L_{\rho_n}(M_n, Z_n, U_n) = \left\{ \frac{\alpha_n}{|\mathcal{C}_n|} \text{tr}(M_n S_n) - \frac{1 - \alpha_n}{N - 1} \sum_{n': n \neq n'} \frac{\text{tr}(M_n S_{n'})}{|\mathcal{C}_{n'}|} - \log \det M_n + \eta_n \|Z_n\|_1 \right\} + \frac{\rho_n}{2} \|M_n - Z_n + U_n\|_F^2, \quad (7.3)$$

と表現できる。

したがって、各変数の更新は以下の最適化を順に繰り返すことで実行可能となる。

$$M_n^{k+1} := \arg \min_{M_n} L_{\rho_n}(M_n, Z_n^k, U_n^k), \quad (7.4)$$

$$Z_n^{k+1} := \arg \min_{Z_n} L_{\rho_n}(M_n^{k+1}, Z_n, U_n^k), \quad (7.5)$$

$$U_n^{k+1} := M_n^{k+1} - Z_n^{k+1} + U_n^k. \quad (7.6)$$

ただし、式(7.4)–(7.5)における行列 M_n, Z_n, U_n はそれぞれ、 $M_n = [M_{nij}] \in \mathcal{R}^{W \times W}$, $Z_n = [Z_{nij}] \in \mathcal{R}^{W \times W}$, $U_n = [U_{nij}] \in \mathcal{R}^{W \times W}$ とする。

上式をそれぞれ解くことにより、各変数は

$$M_n^{k+1} := Q_n \tilde{M}_n Q_n^T, \quad (7.7)$$

$$Z_n^{k+1} := S_{\eta_n/\rho} \left(M_{nij}^{k+1} + U_{nij}^k \right), \quad (7.8)$$

$$U_n^{k+1} := M_n^{k+1} - Z_n^{k+1} + U_n^k. \quad (7.9)$$

として更新することが可能となる。ただし、 Q_n は以下の式

$$\frac{\alpha_n}{|C_n|} S_n + \frac{(1 - \alpha_n)}{N - 1} \sum_{n': n \neq n'} \frac{S_{n'}}{|C_{n'}|} - \rho_n \left(Z_n^k + U_n^k \right) = Q_n \tilde{M}_n Q_n^T, \quad (7.10)$$

を満たす正規直交行列であるとし、 \tilde{M}_n は $\tilde{M}_n = Q_n M_n Q_n^T$ を満たすものとする。

謝辞

本論文は、著者が早稲田大学創造理工学研究科で行った研究成果をまとめたものである。

まず、指導教員として研究生生活の全ての過程でご指導を賜りました早稲田大学創造理工学部 経営システム工学科 後藤正幸教授に深く感謝の意を表します。著者が学部2年次にネパール研修旅行に参加して以来13年間、研究生生活のみならず、その深い懐で私生活、後輩たちへの教育指導の方法等においても十分すぎるほどのご指導、ご助言を頂きました。最後まで諦めずに博士論文をまとめることができたのも、後藤正幸先生の熱心で丁寧なご指導のお陰です。深く感謝するとともに心より篤く御礼申し上げます。

本研究をまとめるにあたって副査としてご指導を賜りました、早稲田大学 平澤茂一名誉教授、早稲田大学創造理工学部経営システム工学科 高橋真吾教授、永田靖教授、早稲田大学基幹理工学部応用数理学科 松嶋敏泰教授にはそれぞれのご専門の立場から大変有益なご意見や数多くのご助言を頂戴いたしました。深く感謝し、御礼申し上げます。特に、平澤茂一先生には著者が博士課程へ入学して以来、研究の在り方、取り組み方のみならず、教育者としてのあり方など、その広い懐で数多くのご指導を賜りました。心より御礼を申し上げます。

早稲田大学創造理工学研究科在学中の頃より助手在任期間、また、本論文の審査の過程において、研究・教務の両面で多大なご指導を賜りました、早稲田大学創造理工学部 経営システム工学科の諸先生方に深く感謝し、御礼を申し上げます。

高崎経済大学経済学部 石田崇准教授には、著者が博士後期課程入学時より、大変熱心かつ丁寧なご指導を賜りました。多くの学会発表、論文の共同執筆者であり、研究の方向性や研究活動に関する相談など、非常に多くの時間を割いていただきました。また、常日頃より、著者の研究活動を気にかけてくださると共に、公私に渡って大変お世話になりました。ここに深く感謝し御礼申し上げます。

湘南工科大学工学部 小林学教授には、ご専門とは異なる研究テーマでありながら著者の研究について多くのご指摘、ご助言を賜りました。研究の進捗や研究内容の発展、論文投稿時の検討会、

さらに博士論文をまとめるにあたっての論理構成など、本論文の多くの内容を小林先生との議論により発展させることができました。深く感謝し、御礼申し上げます。

本研究を進める上で常に温かい励ましや、日ごろのゼミで貴重なご意見、ご指摘を頂戴した、早稲田大学 酒井哲也先生、須子統太先生、堀井俊祐先生、雲居玄道氏、湘南工科大学 鈴木誠先生、電気通信大学 八木秀樹先生、東京理科大学 細谷剛先生に心より御礼申し上げます。

本論文の第4章の研究は小林学先生、八木秀樹先生、須子統太先生に開いていただいた輪講ゼミがきっかけとなり、完成させることができました。このゼミでの活発な議論、ご指導のおかげで著者の研究能力は飛躍的に向上し、博士論文執筆への大きな足掛かりとなったと感じています。

総合研究大学院大学附属図書館副館長・監査主幹 柳生修二氏には著者が学部3年次より、現在に至るまで公私に渡り大変お世話になりました。大学職員、学生という立場を超え、人生における重大な選択の際にはいつも快く相談に乗っていただき、様々なご助言を頂戴いたしました。深く感謝し、御礼申し上げます。

東京都市大学 増井忠幸名誉教授には著者の修士論文の主査として、様々なご助言を賜りました。また博士課程進学後に学会等でお会いした際にも常に研究の進捗状況について気にかけていただき、多くのご助言を頂きました。増井忠幸先生から頂いた、「常に与えられた環境で全力を出すこと」というお言葉は、研究活動でくじけそうになったときに思い出し、大きな励ましとなりました。深く感謝いたします。

聖心女子大学文学部 Brenda Bushell 教授には学部時代の指導教員として、武蔵工業大学在学中より現在に至るまで、大変お世話になりました。Brenda Bushell 先生、後藤正幸先生のご指導がなければ著者が大学院に進学し、また学位取得、大学教員への道を目指すことはありませんでした。好きな内容を研究できる環境を与えて下さり、自由な研究活動を認めていただけただからこそ、研究活動に対し、常にポジティブでいることができ、ここまで諦めることなく研究を進めることができました。また、お二人のおかげで物事に対する見方が変わり、人生観も大きく変わりました。心より深く感謝いたします。

修士課程、博士課程進学に際し、反対を押し切り、我儘ばかりだった著者をここまで育て、支えてくれた両親と弟、また、温かく見守ってくださる義母に心から敬意の気持ちを送ると共に感謝いたします。最後に、著者の研究生活を陰ながら支えてくれている妻 静に心から感謝します。

2015年9月

三川 健太

参考文献

- [1] 赤穂昭太郎, カーネル多変量解析 非線形データ解析の新しい展開, 岩波書店, 2008.
- [2] E. Alpaydin, *Introduction to Machine Learning*, The MIT Press, 2010.
- [3] 荒川貴紀, 三川健太, 後藤正幸, “未観測カテゴリーを含む文書データの自動分類手法に関する研究 (研究速報),” 信学論 (D), 情報・システム, vol.96, no.8, pp.1956–1959, 2013.
- [4] 麻生英樹, 津田宏治, 村田昇, パターン認識と学習の統計学, 岩波書店, 2003.
- [5] A. Asuncion, and D. Newman, “UCI machine learning repository,” *University of California, Irvine, School of Information and Computer Sciences*, 2007.
- [6] A. Banerjee, S. Merugu, I.S. Dhillon, and J. Ghosh, “Clustering with bregman divergences,” *Journal of Machine Learning Research*, vol.6, pp.1705–1749, 2005.
- [7] O. Banerjee, L. El Ghaoui, and A. d’Aspremont, “Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data,” *Journal of Machine Learning Research*, vol.9, pp.485–516, 2008.
- [8] D. Barber, *Bayesian Reasoning and Machine Learning*, Cambridge University Press, 2012.
- [9] A. Bellet, A. Habrard, and M. Sebban, “A survey on metric learning for feature vectors and structured data,” CoRR, vol.abs/1306.6709, 2013.
- [10] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Inc., 1995.
- [11] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag New York, Inc., 2006.
- [12] D.M. Blei, A.Y. Ng, and M.I. Jordan, “Latent dirichlet allocation,” *Journal of Machine Learning Research*, vol.3, pp.993–1022, 2003.

- [13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundation of Trends Machine Learning*, vol.3, no.1, pp.1–122, 2011.
- [14] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [15] L. Bregman, “The Relaxation Method of Finding the Common Point of Convex Sets and Its Application to the Solution of Problems in Convex Programming,” *USSR Computational Mathematics and Mathematical Physics*, vol.7, no.3, pp.200–217, 1967.
- [16] L. Breiman, “Random forests,” *Machine Learning*, vol.45, no.1, pp.5–32, 2001.
- [17] T. Choy, and N. Meinshausen, “Sparse distance metric learning,” *Computational Statistics*, vol.29, no.3-4, pp.515–528, 2014.
- [18] L. Clemmensen, T. Hastie, D. Wiiten, and E. .B., “Sparse discriminant analysis,” *Technometrics*, vol.53, no.4.
- [19] C. Cortes, and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol.20, no.3, pp.273–297, 1995.
- [20] T. Cover, and P. Hart, “Nearest neighbor pattern classification,” *IEEE Trans. Information Theory*, vol.13, no.1, pp.21–27, 1967.
- [21] T. Cover, and J.A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2006.
- [22] J.V. Davis, B. Kulis, P. Jain, S. Sra, and I.S. Dhillon, “Information-theoretic metric learning,” *Proceedings of the 24th International Conference on Machine Learning*, pp.209–216, 2007.
- [23] C. De Mol, E. De Vito, and L. Rosasco, “Elastic-net regularization in learning theory,” *Journal of Complexity*, vol.25, no.2, pp.201–230, 2009.
- [24] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman, “Indexing by latent semantic analysis,” *J. Am. Soc. Inf. Sci.*, vol.41, no.6, pp.391–407, 1990.

- [25] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, Wiley-Interscience, 2000.
- [26] J.H. Friedman, “Regularized discriminant analysis,” *Journal of the American Statistical Association*, vol.84, no.405, pp.165–175, 1989.
- [27] J. Goldberger, G.E. Hinton, S.T. Roweis, and R.R. Salakhutdinov, “Neighbourhood components analysis,” *Advances in Neural Information Processing Systems 17*, pp.513–520, 2005.
- [28] 後藤正幸, 小林学, 入門 パターン認識と機械学習, コロナ社, 2014.
- [29] D.A. Harville, *Matrix Algebra from a Statistician’s Perspective*, Springer, 1997.
- [30] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer New York Inc., 2001.
- [31] M.A. Hearst, “Untangling text data mining,” *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, pp.3–10, 1999.
- [32] 平井有三, はじめてのパターン認識, 森北出版株式会社, 2012.
- [33] T. Hofmann, “Probabilistic latent semantic indexing,” *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.50–57, 1999.
- [34] 福島雅夫, 非線形最適化の基礎, 朝倉書店, 2001.
- [35] 茨木俊秀, 最適化の数学, 共立出版株式会社, 2011.
- [36] 石井健一郎, 前田英作, 上田修功, 村瀬洋, わかりやすいパターン認識, オーム社, 1998.
- [37] 石井健一郎, 上田修功, 続・わかりやすいパターン認識, オーム社, 2014.
- [38] 石川博, 新美礼彦, 白石陽, 横山昌平, データマイニングと集合知-基礎から Web, ソーシャルメディアまで-, 共立出版, 2012.
- [39] A.K. Jain, R.P.W. Duin, and J. Mao, “Statistical pattern recognition: A review,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.22, no.1, pp.4–37, 2000.

- [40] A.K. Jain, M.N. Murty, and P.J. Flynn, “Data clustering: A review,” *ACM Computing Survey*, vol.31, no.3, pp.264–323, 1999.
- [41] P. Jain, B.J. Kulis, J.V. Davis, and I.S. Dhillon, “Metric and kernel learning using a linear transformation,” *Journal of Machine Learning Research*, vol.13, pp.519–547, 2012.
- [42] W. James, and C. Stein, “Estimation with quadratic loss,” *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp.361–379, University of California Press, 1961.
- [43] N. Jiang, W. Liu, and Y. Wu, “Order determination and sparsity-regularized metric learning adaptive visual tracking,” *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1956–1963, 2012.
- [44] R. Jin, S. Wang, and Y. Zhou, “Regularized Distance Metric Learning: Theory and Algorithm,” *Advances in Neural Information Processing Systems 22*, pp.862–870, 2009.
- [45] 寒野善博, 土谷隆, 最適化と変分法, 丸善出版, 2014.
- [46] 金明哲, 村上征勝, 永田昌明, 大津起夫, 山西健司, 言語と心理の統計, 岩波書店, 2003.
- [47] 北研二, 確率的言語モデル, 東京大学出版会, 2000.
- [48] 北研二, 津田和彦, 獅々堀正幹, 情報検索アルゴリズム, 共立出版, 2002.
- [49] 小西貞則, 多変量解析入門 –線形から非線形へ–, 岩波書店, 2010.
- [50] B. Kulis, M.A. Sustik, and I.S. Dhillon, “Low-rank kernel learning with bregman matrix divergences,” *Journal of Machine Learning Research*, vol.10, pp.341–376, 2009.
- [51] D. Lashkari, and P. Golland, “Convex clustering with exemplar-based models,” in *Advances in Neural Information Processing Systems*, pp.825–832, 2007.
- [52] E.Y. Liu, Z. Guo, X. Zhang, V. Jojic, and W. Wang, “Metric learning from relative comparisons by minimizing squared residual,” *Proceedings of 12th IEEE International Conference on Data Mining, ICDM 2012*, pp.978–983, 2012.

- [53] P.C. Mahalanobis, “On the Generalized Distance in Statistics,” *Proceedings of National Institute of Science*, vol.2, no.1, pp.49–55, 1936.
- [54] C.D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.
- [55] K. Mikawa, and M. Goto, “Regularized distance metric learning for the document classification and its application,” *日本経営工学会論文誌*, vol.66, no.2E, pp.190–203, 2015.
- [56] 三川健太, 後藤正幸, “カテゴリ毎に異なる計量行列を用いた計量距離学習に関する一考察,” *日本経営工学会論文誌*, vol.66, no.4, 2015 (掲載決定) .
- [57] 三川健太, 石田崇, 後藤正幸, “満足度を考慮したユーザレビューの分析に関する一考察,” *日本経営工学会平成 24 年度秋季研究大会予稿集*, pp.206–207, 2010.
- [58] 三川健太, 小林学, 後藤正幸, “教師あり学習に基づく l_1 正則化を用いた計量行列の学習法に関する一考察,” *日本経営工学会論文誌*, vol.66, no.3, pp.230–239, 2015.
- [59] 持橋大地, 菊井玄一郎, 北研二, “言語表現のベクトル空間モデルにおける最適な計量距離 (自然言語処理),” *信学論 (D-II), 情報・システム, II-パターン処理*, vol.88, no.4, pp.747–756, 2005.
- [60] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*, The MIT Press, 2012.
- [61] M. Nagata, “A stochastic japanese morphological analyzer using a forward-dp backward-a* n-best search algorithm,” *Proceedings of the 15th Conference on Computational Linguistics*, pp.201–207, 1994.
- [62] 永田靖, 棟近雅彦, *多変量解析法入門*, サイエンス社, 2001.
- [63] S. Parameswaran, and K.Q. Weinberger, “Large margin multi-task metric learning,” *Advances in Neural Information Processing Systems 23*, pp.1867–1875, 2010.
- [64] G.J. Qi, J. Tang, Z.J. Zha, T.S. Chua, and H.J. Zhang, “An efficient sparse metric learning in high-dimensional space via l_1 -penalized log-determinant regularization,” *Proceedings of the 26th Annual International Conference on Machine Learning*, pp.841–848, 2009.

- [65] F. Ricci, L. Rokach, B. Shapira, and P.B. Kantor, *Recommender Systems Handbook*, Springer-Verlag New York, Inc.1st edition, 2010.
- [66] S. Robertson, “Understanding inverse document frequency: On theoretical arguments for idf,” *Journal of Documentation*, vol.60, pp.503–520, 2004.
- [67] S.T. Roweis, and L.K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol.290, no.5500, pp.2323–2326, 2000.
- [68] 貞光九月, 三品拓也, 山本幹雄, “混合ディリクレ分布を用いたトピックに基づく言語モデル (音声, 聴覚),” *信学論 (D-II), 情報・システム, II-パターン処理*, vol.88, no.9, pp.1771–1779, 2005.
- [69] G. Salton, and M.J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill, Inc., 1986.
- [70] 佐藤敦, “パターン認識問題の数理,” *電子情報通信学会基礎・教会ソサイエティ Fundamentals Review*, vol.5, no.4, pp.302–311, 2012.
- [71] S. Shalev-Shwartz, and S. Ben-David, *Understanding Machine Learning From Theory to Algorithms*, Cambridge University Press, 2014.
- [72] J. Shawe-Taylor, and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, New York, NY, USA, 2004.
- [73] N. Shental, T. Hertz, D. Weinshall, and M. Pavel, “Adjustment learning and relevant component analysis,” *Proceedings of the 7th European Conference on Computer Vision-Part IV*, pp.776–792, 2002.
- [74] S. Sra, S. Nowozin, and S.J. Wright, *Optimization for Machine Learning*, The MIT Press, 2011.
- [75] 杉山将, *統計的機械学習*, オーム社, 2009.
- [76] 高村大也, 奥村学, *言語処理のための機械学習入門*, コロナ社, 2010.

- [77] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society (Series B)*, vol.58, pp.267–288, 1996.
- [78] 千野直仁, 岡田謙介, 佐部利真吾, 非対称 MDS の理論と応用, 現代数学社, 2012.
- [79] M.E. Tipping, “Sparse bayesian learning and the relevance vector machine,” *Journal of Machine Learning Research*, vol.1, pp.211–244, 2001.
- [80] 徳永健伸, 情報検索と言語処理, 東京大学出版会, 1999.
- [81] 東京大学教養学部統計学教室, 自然科学の統計学, 東京大学出版会, 1992.
- [82] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag New York, Inc., 1995.
- [83] J. Wang, H. Do, A. Woznica, and A. Kalousis, “Metric learning wit multiple kernels,” *Advances in Neural Information Processing Systems 24*, pp.1170–1178, 2011.
- [84] K.Q. Weinberger, J. Blitzer, and L. Saul, “Distance Metric Learning for Large Margin Nearest Neighbor Classification,” *Advances in Neural Information Processing Systems 18*, 2006.
- [85] K.Q. Weinberger, and L.K. Saul, “Distance metric learning for large margin nearest neighbor classification,” *Journal of Machine Learning Research*, vol.10, pp.207–244, 2009.
- [86] E.P. Xing, M.I. Jordan, S.J. Russell, and A.Y. Ng, “Distance Metric Learning with Application to Clustering with Side-Information,” *Advances in Neural Information Processing Systems 15*, pp.521–528, 2003.
- [87] E.P. Xing, M.I. Jordan, S.J. Russell, and A.Y. Ng, “Learning a Distance Metric from Relative Comparisons,” *Advances in Neural Information Processing Systems 16*, pp.41–88, 2004.
- [88] L. Yang, and R. Jin, “Distance metric learning: A comprehensive survey,” Michigan State University, 2006.

研究業績

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
論文	<p>(論文)</p> <p>[1] テキストデータに基づく顧客ロイヤルティの構造分析手法に関する一考察 日本経営工学会論文誌, vol. 58, no. 3, pp. 182–192, (2007-8) 三川健太, 高橋勉, 後藤正幸</p> <p>[2] 顧客ロイヤルティ構造図に基づく重要要因の定量化手法に関する一考察 日本経営工学会論文誌, vol. 59, no. 5, pp. 365–375, (2008-12) 三川健太, 増井忠幸, 後藤正幸</p> <p>[3] An Optimal Weighting Method in Supervised Learning of Linguistic Model for Text Classification <i>Industrial Engineering & Management Systems</i>, vol. 11 no. 1, pp. 87–93, (2012-1) K. Mikawa, T.Ishida, and M. Goto</p> <p>◎ [4] Regularized Distance Metric Learning for the Document Classification and its Application 日本経営工学会論文誌, vol. 66, no. 2E, pp. 190–203, (2015-7) K. Mikawa, and M. Goto</p> <p>◎ [5] 教師あり学習に基づく l_1 正則化を用いた計量行列の学習法に関する一考察 日本経営工学会論文誌, vol. 66, no. 3, pp. 190-203, (2015-10) 三川健太, 小林学, 後藤正幸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
<p>◎</p> <p>講演</p>	<p>[6] カテゴリ毎に異なる計量行列を用いた計量距離学習に関する一考察 日本経営工学会論文誌, vol. 66, no. 4, (採録決定) 三川健太, 後藤正幸</p> <p>(国際会議)</p> <p>[1] A Proposal of Extended Cosine Measure for Distance Metric Learning in Text Classification <i>2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC2011)</i>, pp. 1741–1746, (2011-10) K. Mikawa, T. Ishida, and M. Goto</p> <p>[2] An Optimal Weighting Method in Supervised Learning of Linguistic Model for Text Classification <i>12th Asia Pacific Industrial Engineering and Management Society (12th APIEMS)</i>, ID-141, (2011-10) K. Mikawa, T. Ishida, and M. Goto</p> <p>[3] A Proposal of Extracting Unknown Information from Customer Review for SWOT Analysis <i>2011 Asian Conference of Management Science & Applications</i>, ID-167, (2011-10) K. Mikawa, G. Kumoi, K. Suzuki, and M. Goto</p> <p>[4] An Optimal Weighting Method by Using the Category Information in Text Classification based on Metric Learning <i>13th Asia Pacific Industrial Engineering and Management Society (13th APIEMS)</i>, no. 25-1, (2012-12) K. Mikawa, T. Ishida, M. Goto, and S. Hirasawa</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[5] A Proposal of Adaptive Metric Learning to Each Category Characteristics for Text Classification <i>2013 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing</i>, pp. 544–547, (2013-3) K. Mikawa, T. Ishida, M. Goto, and S. Hirasawa</p> <p>○ [6] Regularized Distance Metric Learning and its Application to Knowledge Discovery <i>14th Asia Pacific Industrial Engineering and Management Society (14th APIEMS)</i>, (2013-12) K. Mikawa, T. Ishida, and M. Goto</p> <p>○ [7] A Proposal of l_1 Regularized Distance Metric Learning for High Dimensional Sparse Vector Space <i>2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC2014)</i>, pp. 1985–1990, (2014-10) K. Mikawa, M. Kobayashi, M. Goto, and S. Hirasawa</p> <p>(講演)</p> <p>[1] ネパールにおける環境 NGO のコミュニティマネジメントに関する一考察 日本経営工学会, 平成 17 年度秋季大会予稿集, 愛知, pp. 184–185, (2005-10) 三川健太, ブレンダ・ブッシュェル, 後藤正幸</p> <p>[2] 継続購買につながる顧客ロイヤルティの構造分析手法に関する一考察 日本経営工学会, 平成 18 年度度春季大会予稿集, 千葉, pp. 44–45, (2006-5). 三川健太, 高橋勉, 後藤正幸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[3] 大学生を対象としたネパールプロジェクトの教育効果に関する一考察 日本環境教育学会第17回大会予稿集, 北海道, 2J1, (2006-8) 三川健太, 栗島由美, ブレンダ・ブッシュェル, 後藤正幸</p> <p>[4] 顧客ロイヤルティ構造図に基づくユーザコメントの分析手法に関する一考察 日本経営工学会, 平成18年度度秋季大会予稿集, 広島, pp. 30-31, (2006-11) 三川健太, 後藤正幸</p> <p>[5] 満足度を考慮したユーザレビューの分析に関する一考察 日本経営工学会, 平成22年度度秋季大会予稿集, 福岡, pp. 206-207, (2010-11) 三川健太, 石田崇, 後藤正幸</p> <p>[6] 拡張余弦尺度を用いた距離学習に関する一考察 日本経営工学会, 平成23年度度春季大会予稿集, 愛知, pp. 56-57, (2011-5) 三川健太, 石田崇, 後藤正幸</p> <p>[7] 文書分類問題におけるカテゴリ情報を用いた適応的重み学習に関する一考察 日本経営工学会, 平成24年度度秋季大会予稿集, 大阪, pp. 206-207, (2012-10) 三川健太, 石田崇, 後藤正幸</p> <p>[8] テキスト分類問題におけるカテゴリ情報を用いた適応的距離学習に関する一考察 電子情報通信学会 技術研究報告, 情報論的学習理論と機械学習 (IBISML), pp. 83-88, (2012-11) 三川健太, 石田崇, 後藤正幸, 平澤茂一</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
○	<p>[9] 高次元かつスパースなベクトル空間における l_1 正則化に基づく計量距離学習に関する一考察 第 36 回情報理論とその応用シンポジウム (SITA2013) 予稿集, 静岡, pp. 703-706, (2013-11) 三川健太, 小林学, 後藤正幸, 平澤茂一</p> <p>[10] l_1 正則化を用いた計量距離学習による特徴選択に関する一考察 日本経営工学会, 平成 25 年度度秋季大会予稿集, 埼玉, pp. 194-195, (2013-11). 三川健太, 石田崇, 後藤正幸, 平澤茂一</p> <p>[11] カテゴリの統計的特徴を利用した適応的計量距離学習に関する一考察 日本経営工学会, 平成 26 年度度秋季大会予稿集, 埼玉, pp. 232-233, (2014-11). 三川健太, 後藤正幸</p> <p>[12] 代表元の距離構造に着目した計量距離学習に関する一考察 第 37 回情報理論とその応用シンポジウム (SITA2014) 予稿集, 富山, pp. 703-706, (2014-12) 三川健太, 小林学, 後藤正幸, 平澤茂一</p>
その他	<p>(論文)</p> <p>[1] 未観測カテゴリを含む文書データの自動分類手法に関する研究 電子情報通信学会論文誌 D, vol. J96-D, no. 8, pp. 1955-1959, (2013-8) 荒川貴紀, 三川健太, 後藤正幸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[2] エージェントベースシミュレーションによる確率潜在空間モデルを用いた推薦システムの評価 経営情報学会論文誌, vol. 22, no. 2, pp. 1-22, (2013-9) 井沢祐介, 三川健太, 後藤正幸</p> <p>[3] Multi-valued Document Classification based on coding theory China-USA Business Review, vol. 12, no. 9, pp. 911-917, (2013-9) T. Ogihara, K. Mikawa, G. Hosoya, and M. Goto</p> <p>[4] A Design of Recommendation Based on Flexible Mixture Model Considering Purchasing Interest and Post-Purchase Satisfaction 日本経営工学会論文誌, vol. 64, no. 4E, pp. 570-578, (2014-1) T. Suzuki, G. Kumoi, K. Mikawa, and M. Goto</p> <p>[5] 大規模テキストデータの分類体系化のための機械学習に基づく半自動分類法の提案 日本経営工学会論文誌, vol. 65, no. 2, pp. 51-60, (2014-7) 下村良, 三川健太, 後藤正幸</p> <p>[6] 評価と購買の両履歴データの学習による確率的潜在クラスモデルの推定精度向上に関する一考察 日本経営工学会論文誌, vol. 65, no. 4, pp. 286-293, (2015-1) 大井貴裕, 三川健太, 後藤正幸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
その他	<p>(国際会議)</p> <p>[1] Creating Learning Environments through eco-design North American Association for Environmental Education, 34th Annual Conference, (2005-10) B. Bushell, Y. Kurishima, K. Mikawa, and M. Goto</p> <p>[2] A Participatory Process toward a Model for Environmental Sustainability Salt of the earth Conference, (2006-10) K. Suzuki, R. Horiuchi, Y. Kurishima, and K. Mikawa</p> <p>[3] Multivalued document classification by maximization of posterior probability based on Relevance Vector Machine <i>12th Asia Pacific Industrial Engineering and Management Society</i> (12th APIEMS), ID-205, (2011-10) R. Odai, G. Kumoi, K. Mikawa, G. Hosoya, and M. Goto</p> <p>[4] Automated source code plagiarism detection based on coding style model probability based on Relevance Vector Machine <i>12th Asia Pacific Industrial Engineering and Management Society</i> (12th APIEMS), ID-147, (2011-10) K. Hibi, G. Kumoi, K. Mikawa, and M. Goto</p> <p>[5] A study on the recommender system based on probabilistic latent model <i>12th Asia Pacific Industrial Engineering and Management Society</i> (12th APIEMS), ID-127, (2011-10) T. Suzuki, G. Kumoi, K. Mikawa, and M. Goto</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[6] A Recommender System Considering with Item Evaluation based on Mixed Membership Stochastic Block Models <i>12th Asia Pacific Industrial Engineering and Management Society</i> (12th APIEMS), ID-118, (2011-10) Y. Izawa, H. Sakaeda, K. Mikawa, and M. Goto</p> <p>[7] The Agent-Based Simulation Analysis of Collaborative Filtering Using Mixed Membership Stochastic Block Models <i>13th Asia Pacific Industrial Engineering and Management Society</i> (13th APIEMS), no. 146-1, (2012-12) Y. Izawa, K. Mikawa, and M. Goto</p> <p>[8] A Study of Recommender System to Improve Aggregate Diversity based on Latent Class Model <i>13th Asia Pacific Industrial Engineering and Management Society</i> (13th APIEMS), no. 122-1, (2012-12) T. Suzuki, K. Mikawa, and M. Goto</p> <p>[9] A Study for Recommender System based on Mixed and Constrained Latent Dirichlet Allocation <i>13th Asia Pacific Industrial Engineering and Management Society</i> (13th APIEMS), no. 253-1, (2012-12) S. Sakamoto, Y. Izawa, K. Mikawa, and M. Goto</p> <p>[10] A Proposal of Improved Naive Bayes Method for Collaborative Filtering by Introducing Clustering <i>13th Asia Pacific Industrial Engineering and Management Society</i> (13th APIEMS), no. 172-1, (2012-12) T. Oi, T. Arakawa, K. Mikawa, and M. Goto</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[11] Multi-valued Document Classification based on Generalized Bradley-Terry Classifiers Utilizing Accuracy Information <i>13th Asia Pacific Industrial Engineering and Management Society</i> (13th APIEMS), no. 253-1, (2012-12) T. Ogihara, K. Mikawa, G. Hosoya, and M. Goto</p> <p>[12] Statistical Prediction Model of Students' Success on Job Hunting by Log Data Analysis <i>14th Asia Pacific Industrial Engineering and Management Society</i> (14th APIEMS), (2013-12) M. Hayakawa, K. Mikawa, T. Ishida, and M. Goto</p> <p>[13] A Study on Recommender System Based on Latent Class Model for High Dimensional and Sparse Data <i>14th Asia Pacific Industrial Engineering and Management Society</i> (14th APIEMS), (2013-12) S. Sakamoto, K. Mikawa, and M. Goto</p> <p>[14] Training Data Selection in Large Margin Nearest Neighbor Method for Classification Problems <i>14th Asia Pacific Industrial Engineering and Management Society</i> (14th APIEMS), (2013-12) F. Yamazaki, S. Sakamoto, K. Mikawa, and M. Goto</p> <p>[15] Multi-valued Classification of Text Data based on ECOC Approach considering Parallel Processing <i>14th Asia Pacific Industrial Engineering and Management Society</i> (14th APIEMS), (2013-12) T. Ogihara, K. Mikawa, and M. Goto</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[16] A Study of Recommender Systems Based on the Latent Class Model Estimated by Combining Both Evaluation and Purchase Histories <i>14th Asia Pacific Industrial Engineering and Management Society</i> (14th APIEMS), (2013-12) T. Oi, K. Mikawa, and M. Goto</p> <p>[17] A Prediction Method based on Weighted Ensemble of Decision Trees on Alternating Decision Forests <i>15th Asia Pacific Industrial Engineering and Management Society</i> (15th APIEMS), (2014-10) S. Misawa, N. Fujiwara, K. Mikawa, and M. Goto</p> <p>[18] A New Estimation Method of Latent Class Model with High Accuracy by Using Both Browsing and Purchase Histories <i>15th Asia Pacific Industrial Engineering and Management Society</i> (15th APIEMS), (2014-10) N. Fujiwara, K. Mikawa, and M. Goto</p> <p>[19] A Statistical Model for Recommender System to Maximize Sales Amount Focusing on Characteristics of EC Site Data <i>15th Asia Pacific Industrial Engineering and Management Society</i> (15th APIEMS), (2014-10) K. Yamagami, N. Fujiwara, K. Mikawa, and M. Goto</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[20] Distance Metric Learning with Low Computational Complexity based on Ensemble of the Low-dimensional Matrices <i>15th Asia Pacific Industrial Engineering and Management Society (15th APIEMS)</i>, (2014-10) S. Saito, F. Yamazaki, K. Mikawa, and M. Goto</p> <p>[21] Statistical Model Selection of Linear Regression for Privacy Preserving Data Mining <i>15th Asia Pacific Industrial Engineering and Management Society (15th APIEMS)</i>, (2014-10) K. Yukawa, K. Mikawa, and M. Goto</p> <p>[22] A Comparative Survey on Sustainable Tourism Development Considering Regional Characteristics in Nepal <i>The conference of the 20th IICE</i>, (2014-6) Y. Matsuzaki, S. Nagamori, K. Mikawa, M. Goto, and B. Bushell</p> <p>[23] A Modified Aspect Model for Simulation Analysis <i>IEEE International Conference on Systems, Mans and Cybanetics, (SMC2014)</i>, pp. 1306–1311, (2014-10) M. Goto, K. Minetoma, K. Mikawa, M. Kobayashi, and S. Hirasawa</p> <p>[24] An Analysis of Purchasing and Browsing Histories on an EC Site Based on a New Latent Class Model <i>The 1st East Asia Workshop on Industrial Engineering</i>, (2014-11) M. Goto, K. Mikawa, M. Kobayashi, S. Horii, T. Suko, and S. Hirasawa</p> <p>[25] Analysis of customer purchase behavior by using purchase history with discount coupon based on latent class model <i>16th Asia Pacific Industrial Engineering and Management Society (16th APIEMS)</i>, (2015-12) (発表決定) Y. Matsuzaki, K. Yamagami, K. Mikawa, and M. Goto</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[26] Multi-valued Classification of Text Data based on ECOC Approach using Ternary Orthogonal Table <i>16th Asia Pacific Industrial Engineering and Management Society</i> (16th APIEMS), (2015-12) (発表決定) L. Suzuki, K. Yamagami, K. Mikawa, and M. Goto</p> <p>[27] Adaptive Prediction Method Based on Alternating Decision Forests considering Generalization Ability <i>16th Asia Pacific Industrial Engineering and Management Society</i> (16th APIEMS), (2015-12) (発表決定) S. Misawa, K. Mikawa, and M. Goto</p> <p>[28] A Statistical Prediction Model of Students' Success Date on Job Hunting Using Internet Portal Sites Data <i>16th Asia Pacific Industrial Engineering and Management Society</i> (16th APIEMS), (2015-12) (発表決定) K. Yamagami, K. Mikawa, M. Goto, and T. Ogihara</p> <p>[29] A Finish Date Prediction of Job Hunting based on User Clustering Approach considering Time Series Variation of Entry Tendencies <i>16th Asia Pacific Industrial Engineering and Management Society</i> (16th APIEMS), (2015-12) (発表決定) S. Nagamori, K. Yamagami, K. Mikawa, M. Goto, and T. Ogihara</p> <p>[30] A Proposal for the Classification Method of Documents Data with Unobserved Categories Considering Latent Topics <i>16th Asia Pacific Industrial Engineering and Management Society</i> (16th APIEMS), (2015-12) (発表決定) Y. Yamamoto, K. Mikawa, and M. Goto</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
その他	<p>(講演)</p> <p>[1] 大学マーケティングのための市場分析に関する一考察 日本経営工学会 平成 18 年度春季大会予稿集, 千葉, pp. 36-37, (2006-5) 浜翔太郎, 三川健太, 後藤正幸</p> <p>[2] ネパール-日本の連携による小学生を対象としたプログラムの設計と評価 日本環境教育学会 第 17 回大会予稿集, 北海道, 2J5, (2006-8) 栗島由美, 三川健太, ブレンダ・ブッシュェル, 後藤正幸</p> <p>[3] 知識構造を利用した文書データの自動分析に関する一考察 日本経営工学会 平成 19 年度春季大会予稿集, 千葉, pp. 48-419, (2007-5) 渡辺智幸, 海老澤卓哉, 三川健太, 後藤正幸</p> <p>[4] 自由記述文書データ分析における知識構造の援用に関する一考察 日本経営工学会 平成 19 年度秋季大会予稿集, 北海道, pp. 74-75, (2007-11) 渡辺智幸, 海老澤卓哉, 三川健太, 後藤正幸</p> <p>[5] Web 上のユーザコメントを用いた価格プレミアム- 顧客価値構造モデルの構築 2008 年経営情報学会 秋季大会予稿集, 宮城, C3-3, (2008-11) 富田大介, 三川健太, 後藤正幸, 増井忠幸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[6] 宿泊施設を対象とした評価サイトにおけるユーザレビュー分析に関する一考察 日本経営工学会 平成 22 年度秋季大会予稿集, 福岡, pp. 192-193, (2010-10) 榮枝隼人, 三川健太, 後藤正幸</p> <p>[7] アイテム評価値の高低を考慮した混合メンバーシップブロックモデルによる推薦システム 日本経営工学会 平成 23 年度秋季大会予稿集, 愛知, pp. 36-37, (2011-5) 井沢祐介, 榮枝隼人, 三川健太, 後藤正幸</p> <p>[8] 複数要素から構成される文書データの分類手法に関する一考察 —階層潜在クラスモデルと階層 EM アルゴリズムの提案— 日本経営工学会 平成 23 年度秋季大会予稿集, 愛知, pp. 50-51, (2011-5) 荒川貴紀, 三川健太, 後藤正幸</p> <p>[9] 確率潜在空間モデルに基づく推薦システムに関する研究 日本経営工学会 平成 23 年度秋季大会予稿集, 愛知, pp. 182-183, (2011-5) 鈴木健史, 雲居玄道, 三川健太, 後藤正幸</p> <p>[10] コーディングスタイルモデルに基づく推薦システムに関する研究 日本経営工学会 平成 23 年度秋季大会予稿集, 愛知, pp. 204-205, (2011-5) 日比健太, 雲居玄道, 三川健太, 後藤正幸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[11] 二値判別器の組み合わせによる RVM 多値判別手法に関する一考察 FIT 2011 第 10 回情報科学技術フォーラム予稿集, 北海道, F-015, (2011-9) 小田井良輔, 雲居玄道, 三川健太, 後藤正幸</p> <p>[12] 商品の比較履歴とユーザーレビューに基づく推薦手法に関する一考察 FIT 2011 第 10 回情報科学技術フォーラム予稿集, 北海道, F-020, (2011-9) 榮枝隼人, 三川健太, 後藤正幸</p> <p>[13] 文書分類問題におけるカテゴリに注目した可変長 N グラム法 日本経営工学会 平成 24 年度春季大会予稿集, 東京, pp. 64–65, (2012-5) 井上大樹, 三川健太, 雲居玄道, 後藤正幸</p> <p>[14] クラスタリング手法を導入した Improved Naive Bayes 法による 推薦システム 日本経営工学会 平成 24 年度春季大会予稿集, 東京, pp. 144–145, (2012-5) 大井貴裕, 荒川貴紀, 三川健太, 後藤正幸</p> <p>[15] 混合制約付き潜在ディリクレ配分法に基づく推薦システムに関する一考察 日本経営工学会 平成 24 年度春季大会予稿集, 東京, pp. 136–137, (2012-5) 坂本俊輔, 井沢祐介, 三川健太, 後藤正幸</p> <p>[16] ユーザの評価傾向と潜在クラスを考慮した推薦手法に関する一考察 日本経営工学会 平成 24 年度秋季大会予稿集, 大阪, pp. 150–151, (2012-11) 大森悠矢, 三川健太, 後藤正幸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[17] 符号理論に基づく ECOC 法による多値パターン分類に関する一考察 日本経営工学会 平成 24 年度秋季大会予稿集, 大阪, pp. 202-203, (2012-11) 荻原大陸, 三川健太, 後藤正幸</p> <p>[18] フォークソノミーを利用した単語間の関連による Web ページ検索システム 日本経営工学会 平成 24 年度秋季大会予稿集, 大阪, pp. 270-271, (2012-11) 早川真央, 三川健太, 後藤正幸, 大和田勇人</p> <p>[19] エージェントベースシミュレーションによるモデルベース 協調フィルタリングの評価に関する一考察 第 35 回情報理論とその応用シンポジウム予稿集, (SITA2012), 大分, (2012-11) 井沢祐介, 三川健太, 後藤正幸</p> <p>[20] 確率モデルに基づく Aggregate Diversity を考慮した推薦システムに 関する一考察 第 35 回情報理論とその応用シンポジウム予稿集,)SITA2012), 大分, (2012-11) 鈴木健史, 三川健太, 後藤正幸</p> <p>[21] ECOC 法における分類器の予測精度を考慮した分類法に関する一考察 電子情報通信学会 研究技術報告 人工知能と知識処理研究会 (AI), 福岡, pp. 7-12, (2012-11) 石橋想太郎, 三川健太, 石田崇, 後藤正幸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[22] 特徴語に注目した Smith-Waterman アルゴリズムに基づく剽窃コードの自動検出手法 電子情報通信学会 研究技術報告 人工知能と知識処理研究会 (AI), 福岡, pp. 1-6, (2012-11) 日比健太, 雲居玄道, 三川健太, 後藤正幸</p> <p>[23] 三者以上でプライバシーを保護する線形回帰モデルの分散計算法について 日本経営工学会 平成 25 年度春季大会予稿集, 神奈川, pp. 172-173, (2013-5) 湯川輝一郎, 荻原大陸, 三川健太, 後藤正幸</p> <p>[24] Large Margin Nearest Neighbor の分類精度向上を目的とした学習データ選択法に関する一考察 日本経営工学会 平成 25 年度春季大会予稿集, 神奈川, pp. 178-179, (2013-5) 山崎史博, 峯苦和史, 三川健太, 後藤正幸</p> <p>[25] 就職ポータルサイトにおけるレコメンデーションモデルに関する一考察 第 36 回情報理論とその応用シンポジウム予稿集, (SITA2013), 静岡, no. 7.3.1, (2013-11) 大森悠矢, 三川健太, 石田崇, 後藤正幸, 小川晋一郎</p> <p>[26] 高次元, スパースなデータセットを対象とした潜在クラスモデルの推定アルゴリズムに関する一考察 第 36 回情報理論とその応用シンポジウム予稿集, (SITA2013), 静岡, no. 8.3.2, (2013-11) 坂本俊輔, 三川健太, 後藤正幸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[27] 層別木と混合ワイブル分布に基づく就職活動終了時期の予測モデル 第 36 回情報理論とその応用シンポジウム予稿集, (SITA2013), 静岡, no. 6.3.3, (2013-11) 早川真央, 三川健太, 石田崇, 後藤正幸, 小川晋一郎</p> <p>[28] 評価と購買の両履歴データを用いる潜在クラスモデルの 推定アルゴリズムに関する一考察 第 36 回情報理論とその応用シンポジウム予稿集, (SITA2013), 静岡, no. 8.3.1, (2013-11) 大井貴裕, 三川健太, 後藤正幸</p> <p>[29] Reed Muller 符号を用いた階層的 ECOC 法による多値文書分類 第 36 回情報理論とその応用シンポジウム予稿集, (SITA2013), 静岡, no. 5.3.1, (2013-11) 荻原大陸, 三川健太, 後藤正幸</p> <p>[30] ユーザレビューの構造を利用したマトリックス分析手法の提案 日本経営工学会 平成 25 年度秋季大会予稿集, 埼玉, pp. 190–191, (2013-11) 李昇炯, 大森悠矢, 三川健太, 後藤正幸</p> <p>[31] ランダムな次元削減とアンサンブルによるメトリックラーニングの 計算量低減法 日本経営工学会 平成 25 年度秋季大会予稿集, 埼玉, pp. 40–41, (2013-11) 齋藤洋, 山崎史博, 三川健太, 後藤正幸</p> <p>[32] 学習データの被予測性能に着目した Alternating Decision Forests の 各決定木への重み付け予測法 電子情報通信学会 研究技術報告 情報理論 (IT), 兵庫, vol. 114, no. 138, pp. 101–106, (2014-7) 三沢翔太郎, 藤原直広, 三川健太, 後藤正幸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[33] 低次元計量行列の学習とその結合による計量行列学習の計算量削減法 各決定木への重み付け予測法 電子情報通信学会 研究技術報告 情報理論 (IT), 兵庫, vol. 114, no. 138, pp. 7-12, (2014-7) 齋藤洋, 藤原直広, 三川健太, 後藤正幸</p> <p>[34] シンボルの累積出現回数を考慮したベイズ予測アルゴリズムに 関する一考察 日本経営工学会 平成 26 年度秋季大会予稿集, 広島, pp. 128-129, (2014-11) 阿内宏武, 三川健太, 雲居玄道, 後藤正幸</p> <p>[35] 就職ポータルサイトにおける潜在クラスを用いたレコメンドシステムに 関する研究 日本経営工学会 平成 26 年度秋季大会予稿集, 広島, pp. 238-239, (2014-11) 古山亮, 三川健太, 後藤正幸, 谷田部 治明</p> <p>[36] ECOC SVM におけるデータ数の偏りを考慮した多値文書分類手法に 関する一考察 日本経営工学会 平成 26 年度秋季大会予稿集, 広島, pp. 130-131, (2014-11) 安田直生, 雲居玄道, 三川健太, 後藤正幸</p> <p>[37] 潜在クラスモデルの混合比に着目したクラスタリング手法の提案 日本経営工学会 平成 26 年度秋季大会予稿集, 広島, pp. 132-133, (2014-11) 山上敢, 三川健太, 後藤正幸, 谷田部 治明</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[38] Large Margin Nearest Neighbor の分類精度向上を目的とした学習データの重み付けに関する一考察 計測自動制御学会 システム・情報部門学術講演会, SSI2014, SS27-11, 岡山, (2014-11) 山崎史博, 三川健太, 後藤正幸</p> <p>[39] 閲覧及び購買行動を同時に学習可能な潜在クラスモデルの提案 計測自動制御学会 システム・情報部門学術講演会, SSI2014, SS27-10, 岡山, (2014-11) 藤原直広, 三川健太, 後藤正幸</p> <p>[40] 低次元計量行列の学習とその結合による計量行列学習法 計測自動制御学会 システム・情報部門学術講演会, SSI2014, SS27-9, 岡山, (2014-11) 齋藤洋, 山崎史博, 三川健太, 後藤正幸</p> <p>[41] Large Margin Nearest Neighbor の分類精度向上を目的とした学習データの重み付けに関する一考察 第 37 回情報理論とその応用シンポジウム予稿集, (SITA2014), 富山, no. 3.3.1, (2014-12) 山崎史博, 三川健太, 後藤正幸</p> <p>[42] シンボルの累積出現回数を条件とするモデルクラスの ベイズ予測アルゴリズムについて 第 37 回情報理論とその応用シンポジウム予稿集, (SITA2014), 富山, no. 1.2.4, (2014-12) 阿内宏武, 三川健太, 雲居玄道, 後藤正幸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[43] 閲覧及び購買行動を同時に表現する潜在クラスモデルの提案とその学習法 第 37 回情報理論とその応用シンポジウム予稿集, (SITA2014), 富山, no. 3.3.3, (2014-12) 藤原直広, 三川健太, 後藤正幸</p> <p>[44] 類似性に基づくラベル選択法を用いたマルチトピック文書分類 第 37 回情報理論とその応用シンポジウム予稿集, (SITA2014), 富山, no. 3.3.2, (2014-12) 秋山龍太郎, 雲居玄道, 三川健太, 後藤正幸</p> <p>[45] 就職ポータルサイトにおける潜在クラスを用いた レコメンドシステムに関する研究 第 37 回情報理論とその応用シンポジウム予稿集, (SITA2014), 富山, no. 3.3.1, (2014-12) 古山亮, 三川健太, 後藤正幸, 谷田部治明</p> <p>[46] 離反顧客発見を目的とする判別分析手法に関する一考察 日本経営工学会 平成 27 年度春季大会予稿集, 東京, pp. 240-241, (2015-5) 酒井拓哉, 三川健太, 後藤正幸</p> <p>[47] 購買情報を用いた RFM 分析に基づく顧客分析手法に関する一考察 日本経営工学会 平成 27 年度春季大会予稿集, 東京, pp. 204 - 205, (2015-5) 張倩, 三川健太, 後藤正幸</p> <p>[48] 就職ポータルサイトにおける被エントリ数の予測に関する一考察 日本経営工学会 平成 27 年度春季大会予稿集, 東京, pp. 208 - 209, (2015-5) 野津琢登, 後藤正幸, 三川健太, 荻原大陸</p>

種類別	題目, 発表・発行掲載誌名, 発表・発行年月, 連名者
	<p>[49] 潜在トピックを考慮した未観測カテゴリを含む文書集合の自動分類手法の提案 日本経営工学会 平成 27 年度春季大会予稿集, 東京, pp. 206 – 207, (2015-5) 山本祐生, 後藤正幸, 三川健太, 雲居玄道</p> <p>[50] 3 元直交表に基づく ECOC 法による多値文書分類に関する一考察 日本経営工学会 平成 27 年度春季大会予稿集, 東京, pp. 210 – 211, (2015-5) 鈴木玲央奈, 山上敢, 三川健太, 後藤正幸</p> <p>[51] Information-Theoretic Metric Learning の分類精度向上を目的とした学習データペアの選択法 日本経営工学会 平成 27 年度春季大会予稿集, 東京, pp. 160 – 161, (2015-5) 馬賀嵩士, 湯川輝一郎, 三川健太, 後藤正幸</p> <p>[52] クーポン付購買履歴データを用いた顧客購買行動分析手法に関する一考察 日本経営工学会 平成 27 年度春季大会予稿集, 東京, pp. 162 – 163, (2015-5) 松寄祐樹, 山上敢, 三川健太, 後藤正幸</p> <p>[53] エントリ傾向の時間的変化を考慮したユーザクラスタリングに基づく就職活動終了時期予測 日本経営工学会 平成 27 年度春季大会予稿集, 東京, pp. 164 – 165, (2015-5) 永森誠矢, 三川健太, 後藤正幸, 荻原大陸</p>