

# 教師あり順序付けのための次元縮約

## Dimension Reduction for Supervised Ordering

神島 敏弘<sup>1\*</sup> 赤穂 昭太郎<sup>1</sup>

Toshihiro Kamishima<sup>1</sup> and Shotaro Akaho<sup>1</sup>

<sup>1</sup> 産業技術総合研究所

<sup>1</sup> National Institute of Advanced Industrial Science and Technology (AIST)

**Abstract:** Ordered lists of objects are widely used as representational forms. Such ordered objects include Web search results and best-seller lists. Techniques for processing such ordinal data are being developed, particularly methods for the supervised ordering task: i.e., learning functions used to sort objects from sample orders. In this article, we propose dimension reduction methods specifically designed to improve prediction performance in supervised ordering tasks.

### 1 はじめに

順序 (order) とは、嗜好、価格といったある特性に従って対象を整列した系列のこととする。例えば、WWW 検索エンジンはクエリへの適合度で整列したページのリストを返す。売り上げ順に商品を並べた売れ筋商品リストなどもある。このように順序は幅広く利用されてきたが、そのマイニング技術が研究されはじめたのは最近である。その中でも、属性ベクトルで記述された対象を順序付けする関数を、訓練順序集合から学習する方法が幾つか提案されている。我々は、この学習問題を教師あり順序付け (Supervised Ordering) [Kamishima 05a, 神島 05b] と呼ぶことにした。この問題は、味覚などを調査する官能検査、利用者から順序フィードバックを得る情報検索、推薦システムなどで用いられている。

クラス分類などの学習問題と同様に、この教師あり順序付け問題でも、対象を記述している属性ベクトルの次元数が大きい場合に、予測精度が悪化する。その主な理由はいわゆる「次元の呪い」である。すなわち、次元数の増加にともないモデルのパラメータが増加するので、獲得された順序付け関数は過適合を生じ未知の対象を適切に整列することができなくなるからである。

このような高次元の属性ベクトルに由来する問題に対処する方法の一つは次元縮約である。次元縮約とは、情報の損失をできるだけ抑制しつつ、高次元空間中の点を低次の部分空間に射影する操作である。主成分分析は、この次元縮約の代表的な手法で、元データの分散についての情報ができるだけ保存されるように設計されている。この方法は他の機械学習問題には有効だったが、教師あり順序付けを目的とする場合には不適切である。主成分分析は分散という対象自身についての情

報を優先的に保存するので、学習すべき順序付けと対象との関連についての情報は失われる。そこで、本論文では、教師あり順序付けに特化した次元縮約法である順位相関次元縮約 (Rank Correlation Dimension Reduction; RCDR) 法を提案する。この RCDR は目標とする順序付けにとって有用な情報を保存するように設計する。

2 節では順序と教師あり順序付けの定式化について、3 節では順位相関次元縮約法について、4 節では実験結果について述べ、最後の 5 節ではまとめを述べる。

### 2 Supervised Ordering

教師あり順序付けについて述べる前に順序についての基本的な表記を定める。まず、整列すべき対象を  $x_j$  で表し、対象全集合  $X^*$  は全ての可能な対象を含む。各対象  $x_j$  は属性値ベクトル  $\mathbf{x}_j = [x_{j1}, x_{j2}, \dots, x_{jK}]^T$  で記述される。ただし、 $K$  は属性数。順序は

$$O = \mathbf{x}_a \succ \dots \succ \mathbf{x}_j \succ \dots \succ \mathbf{x}_b$$

の形式で、対象  $\mathbf{x}_a$  がこの順序で最も前にあるということを表す。ここで、添え字  $j$  は「この順序の  $j$  番目の対象」ではなく「 $X^*$  中で番号  $j$  で一意に指定される対象」を表すことに注意されたい。対象集合  $X(O_i)$  やその簡略形  $X_i$  は順序  $O_i$  を構成する全対象の集合を表す。よって  $|X_i|$  は順序  $O_i$  の長さとなり、これを  $L_i$  で表す。全ての対象を含む順序、すなわち、 $O_i$  s.t.  $X(O_i) = X^*$  を完全順序、そうでないときを不完全順序であるという。順位  $r(O_i, \mathbf{x}_j)$  やその簡略形  $r_{ij}$  は対象  $\mathbf{x}_j$  の順序  $O_i$  中の位置を表す基数で、例えば、 $O_i = \mathbf{x}_1 \succ \mathbf{x}_3 \succ \mathbf{x}_2$  について  $r(O_i, \mathbf{x}_2) \equiv r_{i2} = 3$  である。二つの順序  $O_1$  と  $O_2$  について、 $\mathbf{x}_a, \mathbf{x}_b \in X(O_1) \cap X(O_2)$ ,  $\mathbf{x}_a \neq \mathbf{x}_b$  を満たす

\*連絡先: <http://www.kamishima.net/>

対象の対  $x_a$  と  $x_b$  を考える． $x_a$  と  $x_b$  に関して  $O_1$  と  $O_2$  が同順 (concordant) であるとは，二つの対象がどちらの順序中でも同じ順番であること，形式的には

$$(r_{1a} - r_{1b})(r_{2a} - r_{2b}) \geq 0$$

を満たすことであり，そうでないときは逆順 (discordant) である．条件  $x_a, x_b \in X(O_1) \cap X(O_2)$ ,  $x_a \neq x_b$  を満たす全ての対象の対について同順であるとき， $O_1$  と  $O_2$  は同順であるという．

次に，二つの順序  $O_1$  と  $O_2$  が同じ対象集合で構成される，すなわち， $X(O_1) = X(O_2) \equiv X$  が満たされるときの，順序間の距離について述べる．多くの種類の順序間の距離が提案されている [Marden 95]．中でも Spearman 距離  $d_S(O_1, O_2)$  はよく用いられており，順位の差の二乗和で定義される：

$$d_S = \sum_{x_j \in X} (r_{1j} - r_{2j})^2 \quad (1)$$

これを値域が  $[-1, 1]$  となるよう正規化すると Spearman の順位相関  $\rho$  が得られる．

$$\rho = 1 - 6d_S(O_1, O_2)/(L^3 - L) \quad (2)$$

ただし  $L = |X|$ ．同順位を考えなければ，これは対象の順位の間 Pearson 相関係数と厳密に一致する．Kendall 距離  $d_K(O_1, O_2)$  もよく利用される距離である． $(x_a, x_b)$  と  $(x_b, x_a)$  のどちらか一方を含むような対象対の集合  $\{(x_a, x_b) \in X \times X\}$ ,  $a \neq b$ ,  $x_a, x_b \in X$  を考える．この集合中で， $x_a$  と  $x_b$  について  $O_1$  と  $O_2$  が逆順であるような対象対の数を Kendall 距離という．形式的には次式：

$$d_K = \frac{1}{2} \left( \frac{1}{2} L(L-1) - \sum_{\{(x_a, x_b) \in X\}} \text{sgn}((r_{1a} - r_{1b})(r_{2a} - r_{2b})) \right) \quad (3)$$

ただし， $\text{sgn}(x)$  は  $x > 0$  なら 1， $x = 0$  なら 0，それ以外では  $-1$  となる関数．これも，値域を  $[-1, 1]$  に正規化することで，Kendall の順位相関  $\tau$  が得られる：

$$\begin{aligned} \tau &= 1 - 4d_K(O_1, O_2)/(L(L-1)) \\ &= \sum_{\{(x_a, x_b) \in X\}} \text{sgn}((r_{1a} - r_{1b})(r_{2a} - r_{2b}))/\frac{1}{2}L(L-1). \quad (4) \end{aligned}$$

$\rho$  と  $\tau$  の計算量はそれぞれ  $O(L \log L)$  と  $O(L^2)$ ．これら，二つの順位相関の差の上限は次の Daniels の不等式 [Kendall 90] で抑えられ， $\tau$  と  $\rho$  の相関は高い．

$$-1 \leq \frac{3(L+2)}{L-2}\tau - \frac{2(L+1)}{L-2}\rho \leq 1$$

$d_K$  は metric だが，三角不等式の条件のため  $d_S$  は metric ではない．二つ以上の対象が同順位であるなら，これらの順位を midrank [Marden 95] とする．例えば，記

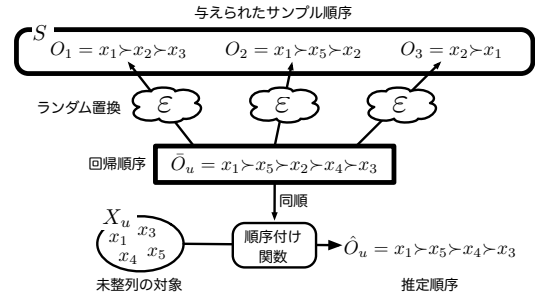


図 1: 教師あり順序付け問題

号  $\sim$  で同順位を表した順序  $x_5 \sim x_2 \sim x_3$  について，同順位の  $x_2$  と  $x_3$  は 2 か 3 番目の順位になる．このとき，これらの順位の平均である 2.5 を両方の対象に割り当てる．

図 1 のように，教師あり順序付け (Supervised Ordering) 問題は，従属変数が順序である回帰問題と見なせる．訓練サンプルは，入出力の対の集合ではなく，順序の集合  $S = \{O_1, \dots, O_N\}$  となる．ただし， $N$  はサンプル数．これらのサンプルは，どちらの対象が上位に順位付けされるかの情報を与え，サンプル中の対象は属性ベクトルにより記述されている．ここで効用値のような副次的な情報を他に与えることはしない．回帰曲線に対応するのが回帰順序 (Regression Order) である．これは， $X^*$  の全ての対象で構成され，通常回帰と同様に，与えられたサンプル順序  $S$  だけでなく，同じ情報源から将来生じる未知のサンプル順序ともできるだけ同順となるように推定される．回帰順序は順序付け関数により表される．これは，順序付けされていない対象集合が与えられると，この対象集合中の対象で構成された，回帰順序と同順となるような推定順序を出力する関数である．数値での回帰ではエラーはランダムな加算項によってモデル化するが，教師あり順序付けではランダム置換  $\varepsilon(\cdot)$  により，ランダムに順序を置換することで表す．また，教師あり順序付けはクラス分類問題と比較すると，前者の目標である順序には対称群という構造があるが，後者のクラスにはそういった構造がない点が大きな違いである．

教師あり順序付け問題は中心順序 (central order) [Marden 95] の概念と関連がある．順序集合  $S$  の中心順序  $\bar{O}$  は， $S$  中の順序までの距離の総和  $\sum_{O_i \in S} d(O_i, \bar{O})$  を最小化する順序と定義される．回帰順序は，属性ベクトルで表された対象で構成され，図 1 の  $x_4$  のような未知の対象も含む．これら未知の対象も，属性空間中で近い対象は順序中でも近くに配置されるとの仮定の下で整列される．一方，中心順序では対象は一意的な識別子で区別され，与えられた順序集合中の対象のみで構成される．

この中心順序の導出は一般には NP 困難であるが，そ

の近似手法には多くの研究があり、採用しているモデルにより次の4種類に分類できる [Critchlow 91, Marden 95]: サーストン型 (Thurstonian) [Thurstone 27] では対象を内部の実数値スコアの順に整列し、一対比較 (Paired Comparison) [Babington Smith 50] は対象の対の順序判断を基本とし、距離ベース (Distance Based) [Mallows 57] では最頻順序 (modal order) からの距離を利用し、多段階 (Multistage) [Plackett 75] では対象を先頭から順に整列する。一般に、教師あり順序付け手法はこれらの順序モデルに、属性を取り扱う方法を組み込むことで設計されている。

教師あり順序付けは、従属変数が順序カテゴリ変数である回帰である順序回帰 (Ordinal Regression) [McCullagh 80, Agresti 03] と関連がある。順序カテゴリ変数は、カテゴリ変数のように事前に定めた有限個の値の一つをとり、さらに、その値に順序関係があるものである。例えば、上-中-下の三つの値をとるカテゴリ変数などである。順序カテゴリと順序とは二つの点で異なるので注意深く区別する必要がある: 一つは、順序は純粋に相対的な情報しか含まないが、順序カテゴリでは絶対的な情報も部分的に含む。例えば「上-中-下」で「上」なら全体の上の方であることが分かるが、順序  $x_1 \succ x_2$  は  $x_1$  が相対的に  $x_2$  より上位であることしか示さない。もう一つは、順序カテゴリは事前に定めた段階でしか差をつけることができない点である。例えば、4個の対象を「上-中-下」の順序カテゴリに分類すると、どれか二つは同じカテゴリに分類されるが、順序では全ての対象間に差があるような表現が可能である。Burgessらの指摘 [Burgess 05] にあるように、教師あり順序付けは順序回帰より一般的な問題なので、教師あり順序付け手法で順序回帰問題を解ける。しかし、一般論を言えば、必要以上に一般的な問題は解くべきではない。例えば、教師あり順序付け用に設計されたSVMで順序回帰問題が解かれている [Herbrich 99] が、順序回帰に特化した手法 [Shashua 03] はより効率的である。これらのことから、教師あり順序付けと順序回帰の二つの問題は注意深く区別すべきである。

教師あり順序付け手法は幾つか開発されており、それらの長所・短所は [Kamishima 05a, 神島 05b] にまとめられている。これらの手法の概要を簡潔に述べる。Cohenらの方法 [Cohen 99] では、二つの対象  $x_a$  と  $x_b$  の属性値が与えられたとき、 $x_a$  が  $x_b$  より上位になる条件付き確率分布  $\Pr[x_a \succ x_b | x_a, x_b]$  を、Hedgeと呼ぶアルゴリズムでサンプル順序から学習しておく。そして、 $X_u$  中の対象からなる全ての順序の中で、 $\sum_{x_a \succ x_b} \Pr[x_a \succ x_b | x_a, x_b]$  を最大にするものを欲張り探索で求めて、推定順序とする。Freundらの Rank-boost [Freund 03] は、順序付けについて部分的な情報を与える弱学習器を線形結合した Thurstone スコア関数を学習する Boosting 手法である。第1のSVM系手

法である賀沢らの Order SVM [Kazawa 05, 賀沢 03] は、ある対象が、順序の中で第  $j$  位以上かどうかを、 $j$  を1から  $L_i$  まで動かして平均的に判別するように設計したSVMである。第2の手法はHerbrichらの Support Vector Ordinal Regression (SVOR) [Herbrich 98] で、二つの対象のうち、どちらが上位になるかを判別するように設計したSVMを用いる。この手法は独立にJoachimsによってRankingSVM [Joachims 02] としても提案されている。どちらの手法も、学習結果は、Thurstoneスコア関数となり、この関数を用いて  $X_u$  に対する推定順序が得られる。期待順位回帰法 (Expected Rank Regression; ERR) [Kamishima 05a, 神島 05b] は通常の最小二乗法による回帰を用いてスコア関数を求める。ある対象のスコアは、観測されない回帰順序中でのその対象の期待順位に比例する。回帰順序から対象が均等にランダムサンプリングされたとの仮定の下、訓練サンプル順序中の順位から、順序統計を用いてこの期待順位は求めることができる。

### 3 順位相関次元縮約法

前節では、教師あり順序付け問題について述べたが、この問題に特化した次元縮約法についてここでは述べる。

データマイニングや機械学習を利用して、よい結果を得るには、特徴選択、欠損値処理、そして次元縮約といった前処理手法の適用が重要になる。データを適切に前処理することで予測精度は向上し、さらに場合によってはメモリや計算コストを削減できることもある。順序を扱う学習手法でも前処理手法が幾つか提案されている。Bahamondeら [Bahamonde 04] は、教師あり順序付け問題にラッパー型の特徴選択を適用。Slottaら [Slotta 05] は順序のクラス分類のために特徴選択を行った。バイオインフォマティクス分野では、情報を含む遺伝子の選択に順位統計量が利用されている [Dettling 02, Deng 04]。順序間の類似度を測るために、欠損している対象の順位を補完する方法を神島と赤穂は提案している [Kamishima 04]。だが、我々の知る限り、教師あり順序付けのための次元縮約法はまだ提案されていない。

クラス分類や回帰のような他の学習問題と同様に、教師あり順序付け問題でも、次元縮約手法は有効だろう。特に、次元数  $K$  が大きい場合には大きく汎化能力を向上させることができるだろう。なぜなら、次元数  $K$  の増加にともなって学習すべきモデルパラメータ数も増加する。よって、有限の訓練サンプルから獲得された順序付け関数では、過適合の影響により未知の対象をうまく整列できなくなるからである。特に、複雑なモデルを用いていたたり、無関係な属性が多い場合などには、こうした問題は次元縮約によって緩和できるだろう。

表 1: 教師あり順序付け手法の計算量

Cohen	RankBoost	SVOR	Order SVM	ERR
$N\bar{L}^2K$	$N\bar{L}^2K$	$N^2\bar{L}^4K$	$N^2\bar{L}^4K$	$N\bar{L}K^2$

Note:  $\bar{L}$  はサンプル順序の平均長,  $N$  はサンプル順序数,  $K$  は属性ベクトルの次元数. SVM の計算量はサンプル数の 2 乗と見積もった. 表では Cohen 法と RankBoost の反復数は定数としているが, 実際には反復数は  $N\bar{L}^2$  に比例して増やす必要があるため, 計算量は  $N^2\bar{L}^4K$  に近づく.

しかし, 教師あり順序付け手法に入力する訓練事例に, 既存の次元縮約手法を適用しても, 次の理由により良い結果を得られない. 主成分分析は代表的な次元縮約手法である. 主成分分析では, 元の属性ベクトル空間中でのデータの分散を保存する, すなわち, データそのものの情報を保存する. しかし, 教師あり順序付けでは, 属性ベクトルと目標とする順序付けとの関係を見つけることが目的なので, 属性ベクトル中の情報を保存するだけでは不十分である. Diaconis の順序のスペクトル分析 [Diaconis 89] についても考察する. これは, 順序の分布を部分要素に分解する手法である. 例えば, 1 次要素は対象  $x_j$  が第  $l$  位になる頻度で, 2 次要素は対象  $x_j$  と  $x_k$  がそれぞれ第  $l$  位と第  $m$  位になる頻度と, 一方が第  $l$  位で, もう一方が第  $m$  位になる頻度である. しかし, ここでの目的は順序空間の分解ではなく, 属性ベクトル空間の部分空間を見つけることである.

以上の議論から, 属性ベクトルと目標順序付けとの関連の情報を保存するように設計した次元縮約手法が必要になる. これは, 属性ベクトルと目標クラスとの関係情報を保存する次元縮約である Fishser 判別分析法と, その動機において類似している.

さらに, 次元縮約に要する計算コストは, 教師あり順序付け手法のそれと比べて, 同程度に抑えるべきである. なぜなら, 次元縮約手法の計算量が教師あり順序付け手法のそれよりずっと大きければ, スケーラビリティに悪影響を与えるからである. 教師あり順序付け手法の学習に要する計算量を表 1 にまとめた. ここで, 訓練集合  $S$  中の対象対の数と対象数はそれぞれ,  $N\bar{L}^2$  と  $N\bar{L}$  で近似できるものとし, SVM の学習時間は, 訓練事例数のほぼ 2 乗とした. Cohen 法と RankBoost の計算量は, 反復数が定数との仮定の下では  $N\bar{L}^2$  である. しかし, RankBoost の論文中的実験や, 我々の実験では実験では反復数は  $N\bar{L}^2$  に応じて増やす必要があったので, 実用的には計算量は  $N^2\bar{L}^4K$  に近い.

上記のことを考慮すると, ここでの次元縮約手法には次の二つことが要求される.

1. 対象の属性と目標順序付けとの関連についての情報を保存する
2. 次元縮約に要する計算コストは, 教師あり順序付

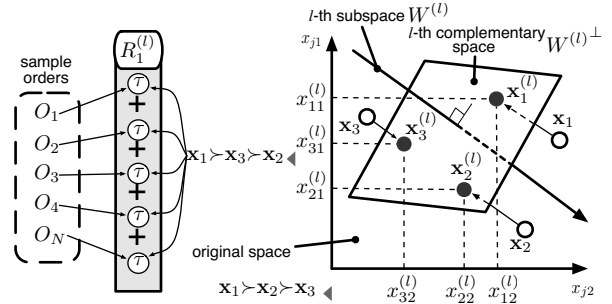


図 2: 順位相関次元縮約法の概要

### 手法のそれと同程度

これらの要求を満たす方法として, 順位相関次元縮約法 (Rank Correlation Dimension Reduction; RCDR) を提案する. この方法では,  $l$  次元の正規直交部分空間が求められている場合に, この部分空間の補空間上の属性値から決まる順序と目標順位付けとの関連を最大化するように  $l+1$  番目の基底を求める. この手続きを逐次的に反復して必要な次元数  $K'$  の部分空間を得る.

最初に, 順位相関次元縮約法の概要を述べる.  $w^{(l)} = [w_1^{(l)}, w_2^{(l)}, \dots, w_K^{(l)}]^\top$  は, 部分空間の第  $l$  番目の基底である. これらの基底は互いに正規直交である, すなわち,  $w^{(l)\top} w^{(m)} = 0, l \neq m$  かつ  $\|w^{(l)}\| = 1$ .  $w^{(1)}, \dots, w^{(l)}$  が張る部分空間を第  $l$  部分空間と呼び, 行列  $W^{(l)} = [w^{(1)}, \dots, w^{(l)}]$  で表す. さらに, これらの全ての基底と垂直な基底で構成される補空間である第  $l$  補空間を  $W^{(l)\perp}$  で表す. ここで, サンプル順序  $S = \{O_1, \dots, O_N\}$  と対象の属性ベクトル  $\{x | x_j \in X_S\}$ ,  $X_S = \cup_{O_i \in S} X_i$  が与えられ, 第  $l$  部分空間が求められているとする. この状態を図 2 に示す. 元空間中の対象  $x_j$  (図中の  $\bullet$ ) を第  $l$  部分空間の補空間に  $W^{(l)\perp}$  射影する. 対象  $x_j$  を  $W^{(l)\perp}$  に射影した属性ベクトル (図中の  $\circ$ ) を  $x_j^{(l)}$  で表す. ただし,  $x_j^{(0)} \equiv x_j$ . この射影により,  $W^{(l)}$  に含まれる目標順位付けについての情報が取り除かれる. 順序付けに関する残りの情報を取り出すため, 射影後の属性値を用いて, 補空間中から  $w^{(l+1)}$  を求める. 各  $k = 1, \dots, K$  について,  $W^{(l)\perp}$  に射影した対象の元空間での  $k$  番目の属性値によって対象を整理した順序を求める (図中の  $x_1 \succ x_3 \succ x_2$  や  $x_1 \succ x_2 \succ x_3$ ). これらの順序と, 各サンプル順序  $O_1, \dots, O_N$  との順位相関を求め, その総和を  $R_k^{(l)}$  とする (厳密な定義は後に与える). この  $R_k^{(l)}$  は, 第  $l$  補空間上に射影した対象の元空間での  $k$  番目の属性値と目標順位付けの一致度を示している. 基底  $w^{(l+1)}$  は, 元空間の各次元  $k = 1, \dots, K$  について, この  $R_k^{(l)}$  と  $w_k^{(l+1)}$  とができるだけ比例するように選ばれる.

上記の手法を形式的に述べる. サンプル順序  $S$ , 基底

$\mathbf{w}_1^{(1)}, \dots, \mathbf{w}_1^{(l)}$ , および第  $l$  補空間  $W^{(l)\perp}$  に射影された対象集合  $\{\mathbf{x}_j^{(l)} | \mathbf{x}_j \in X_S\}$  が与えられている．これらから，第  $l+1$  位の基底  $\mathbf{w}^{(l+1)}$  を以下のように求める．まず，対象の属性値と目標順序付けとの関連性を評価するため， $W^{(l)\perp}$  に射影した対象の元空間上での属性値とサンプル順序との関連性  $R_1^{(l)}, \dots, R_K^{(l)}$  を考える．まず，あるサンプル順序  $O_i$  と  $k$  番目の属性に注目する．ここで，教師あり順序付けでは順序関係を予測するので，属性値そのものではなく，それらの値の大小関係が重要になる．よって，順序  $O_i$  中に含まれる全ての対象  $\mathbf{x}_j \in X_i$  を， $W^{(l)\perp}$  上に射影した対象の元空間の  $k$  番目の属性値  $x_{jk}^{(l)}$  の降順に整列した順序  $O(X_i, x_{jk}^{(l)})$  を考える．この  $O(X_i, x_{jk}^{(l)})$  とサンプル順序  $O_i$  との一致度は，どちらも同じ対象で構成されているため順位相関  $\tau$  を用いて測ることができる．この順位相関を  $S$  中の全ての順序について求め，その総和をとることで，第  $k$  属性と目標順序付けがどれだけ一致しているかを測る．

$$R_k^{(l)} = \sum_{O_i \in S} \tau(O_i, O(X_i, x_{jk}^{(l)})) \quad (5)$$

次に，次元縮約の第一の要求を満たすため，この関連性ができるだけ保存するように第  $l+1$  番目の基底を定める．ベクトル  $\mathbf{R}^{(l)} = [R_1^{(l)}, \dots, R_K^{(l)}]^\top$  を考えると，このベクトルは，関連度  $R_k^{(l)}$  を要素としているので，属性空間中で，最も目標順序付けへの関連性の情報を保存する方向に向いていると考えられる．そこで，補空間  $W^{(l)\perp}$  上で， $\mathbf{w}^{(l+1)}$  と  $\mathbf{R}^{(l)}$  とのなすコサインを最大化するように  $\mathbf{w}^{(l+1)}$  を選ぶ．さらに， $\mathbf{R}^{(l)}$  は定ベクトルで， $\mathbf{w}^{(l+1)} = 1$  の制約があるため，コサインの最大化は  $\mathbf{R}^{(l)}$  と  $\mathbf{w}^{(l+1)}$  の内積の最大化に等しい．この最適化問題を定式化したのが次式である：

$$\mathbf{w}^{(l+1)} = \arg \max_{\mathbf{w}} \mathbf{w}^\top \mathbf{R}^{(l)}, \quad (6)$$

subject to:  $\|\mathbf{w}^{(l+1)}\| = 1, \mathbf{w}^{(l+1)\top} \mathbf{w}^{(m)} = 0, m=1, \dots, l$

ここで， $l \geq 2$  では， $\mathbf{w}^{(l)}$  は全て 0 ベクトルになると考えるかもしれない．しかし，Pearson 相関ではなく順位相関を使っているため，非線形性がある．そのため， $l \geq 2$  でも， $\mathbf{w}^{(l)}$  が 0 ベクトルになることは一般的にはない．

次に，式 (6) を解く．Lagrange の未定乗数法により次の手続きによって  $\mathbf{w}^{(l+1)}$  が導出できることが容易に示せる：順位相関の和を要素とするベクトル  $\mathbf{R}^{(l)}$  を求め，このベクトルを第  $l$  補空間へ射影し，射影したベクトルの大きさを 1 に正規化する．基底を求めたあとは，第  $l$  補空間中の対象  $\mathbf{x}^{(l)}$  を新たな第  $l+1$  空間へ射影する．このように逐次的に次の部分空間を計算できる．このアルゴリズムを図 3 に示す．3 行目で  $\mathbf{R}^{(l)}$  を計算，4 と 5 行目で現在の補空間へ射影，そして 6 行目

Input:

$S = \{O_1, \dots, O_N\}$ : a sample order set  
 $\mathbf{x}_j \in X_S \equiv \cup_{O_i \in S} X_i$ : attribute value vectors  
 $K'$ : the dimension of sub-space

Algorithm:

```

1  $\mathbf{x}_j^{(0)} \equiv \mathbf{x}_j$ 
2 for  $l$  in  $0, \dots, (K' - 1)$ 
3   compute  $\mathbf{R}^{(l)}$  s.t.  $R_k^{(l)} = \sum_{O_i \in S} \tau(O_i, O(X_i, x_{jk}^{(l)}))$ 
4   if  $l > 0$  then
5      $W^{(l)} = [\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(l)}], \mathbf{R}^{(l)} = (I - W^{(l)}W^{(l)\top})\mathbf{R}^{(l)}$ 
6      $\mathbf{w}^{(l+1)} = \mathbf{R}^{(l)} / \|\mathbf{R}^{(l)}\|$ 
7     for  $\mathbf{x}_j$  in  $X_S$ 
8        $\mathbf{x}_j^{(l+1)} = \mathbf{x}_j^{(l)} - \mathbf{w}^{(l+1)}\mathbf{w}^{(l+1)\top}\mathbf{x}_j^{(l)}$ 
9 Output  $W^{(K')} = [\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(K')}]$ 

```

図 3: Kendall 順位相関次元縮約法

でノルムの大きさを 1 にするように正規化する．7 と 8 行目では，第  $l$  補空間中の対象を第  $l+1$  補空間へ射影する．順序の一致度を Kendall  $\tau$  で測るので，この方法を Kendall RCDR 法と呼ぶ．3, 4-5, および 7-8 行目の計算量はそれぞれ  $O(N\bar{L}^2K)$ ,  $O(KK')$ , および  $O(N\bar{L}K)$  なので，各反復の計算量は  $O(N\bar{L}^2K)$  となる．よって，全体の計算量は  $O(N\bar{L}^2KK')$  となる．先に述べたように，Cohen 法と RankBoost の計算量は実用上はほぼ  $O(N^2\bar{L}^4K)$  に近いことから，この Kendall RCDR 法は ERR 以外の教師あり順序付け手法より高速である (表 1 参照)．さらに計算量を抑制するため，Kendall  $\tau$  と Spearman  $\rho$  の間の相関が高いことを利用し，図 3 のアルゴリズムの 3 行目の  $\tau$  を  $\rho$  で置き換える．この方法を Spearman RCDR 法と呼ぶ．計算量は  $O(NKK'\bar{L} \log \bar{L})$  に減るため， $K' \log \bar{L} < K$  の条件の下では ERR 法より高速になる．以上の議論から，提案した順位相関次元縮約法は，教師あり順序付け手法とただかだ同程度という第二の要求も満たしている．Kendall RCDR 法は  $L_i = 2, O_i \in S$  の特別な状況のみ，Spearman RCDR 法より高速である．こうした状況は，Joachims らの提案する暗黙的にサンプル順序収集する方法 [Joachims 02] を用いた場合や，一対比較法による官能検査を実施する場合など見られ，これらの状況では Kendall RCDR 法が有用になる．

## 4 実験

順位相関次元縮約法の簡単な適用例を示したあと，実データへの適用結果を示す．

### 4.1 順位相関次元縮約法の簡単な適用例

次元数が 5 の属性ベクトルで表された対象を  $|X^*| = 1000$  個生成する．各対象  $\mathbf{x}_j \in X^*$  の属性値を次のよう

表 2: 順位相関次元縮約法 (RCDR) と主成分分析 (PCA) により得られた基底ベクトル

第 1 基底						
method	1	2	3	4	5	
KRCDR	0.70	0.64	0.31	-0.06	-0.06	0.146
SRCDR	0.70	0.64	0.32	-0.06	-0.06	0.173
PCA	0.02	-0.74	0.54	-0.39	0.00	0.393

第 2 基底						
method	1	2	3	4	5	
KRCDR	-0.27	-0.17	0.93	-0.13	-0.13	0.007
SRCDR	-0.30	-0.15	0.94	-0.05	-0.05	0.007
PCA	-0.06	-0.18	0.39	0.90	0.00	0.213

Note: 最初の 5 列は基底  $\mathbf{w}^{(1)}$  と  $\mathbf{w}^{(2)}$  の要素. 最後の列は RCDR では  $\|\mathbf{R}^{(l)}\|/N$  を, PCA では寄与率を示した.

に定める: 第 1~4 の属性値は正規分布  $N(0, 1)$  に従ってランダムに生成し, 第 5 の属性値は第 4 と同一にした. そして, 300 個のサンプル順序を次のように生成: 5 個の対象を  $X^*$  を均一にランダムに選択し, これらの対象を理想重みベクトル  $\mathbf{w}^* = [1, 1, 0.5, 0, 0]$  と属性値ベクトルの内積  $\mathbf{w}^* \top \mathbf{x}_j$  の降順に整列する. このデータ集合に Kendall RCDR, Spearman RCDR, および PCA を適用し, 求めた第 1 と 2 の基底を表 2 に示す. 各行はそれぞれ Kendall RCDR, Spearman RCDR, および PCA により得られた基底. 最初の 5 列は基底  $\mathbf{w}^{(1)}$  と  $\mathbf{w}^{(2)}$  の要素. 最後の列は RCDR では  $\|\mathbf{R}^{(l)}\|/N$  を, PCA では寄与率を示した.

まず第 1 基底に着目する. 二つの RCDR 法による基底は非常に類似している. よって, 実用的にはより高速な手法, すなわち,  $L_i = 2$  以外の状況では Spearman RCDR 法を利用すべきである. 理想重み  $\mathbf{w}^*$  の第 4 と 5 の要素は 0 なので, 目標順序にとって有用な情報はこれらの軸には含まれない. RCDR の場合では, 第 4 と 5 の基底の要素はほぼ 0 で, 不要なこれらの軸は無視される. PCA の場合, 目標順序についての情報は全く考慮しないので, 0 よりずっと大きい. PCA は, 第 5 要素に見られるように, 属性値空間中で相関のある軸を無視するだけである. さらに, 目標順序付けは線形関数なので, 理想的には第 1 基底だけで決定可能だが, PCA の寄与率はそれほど大きくなる.

次に第 2 基底について考える. RCDR の場合は, 第 2 基底での相関ベクトルの大きさ  $\|\mathbf{R}^{(2)}\|/N$  は, 第 1 基底のそれ  $\|\mathbf{R}^{(1)}\|/N$  よりずっと小さい. これは, この適用例の場合目的の順序付けが線形関数によって行われているので, 第 2 基底は第 1 基底よりずっと少ない情報しか残っていないことを示している. PCA の場合は, この基底には有用な情報がまだ残っていることが, その寄与率から分かる.  $l$  の増加にともなって  $\|\mathbf{R}^{(l)}\|/N$  が減少することや  $\|\mathbf{R}^{(l)}\|/N$  の大きな基底の方が常に目標順序付けの予測にとってより有用であるということは

保証されてはいない. しかし, 経験的には,  $\|\mathbf{R}^{(l)}\|/N$  が非常に小さければ対応する基底にはあまり情報がなかった. この  $\|\mathbf{R}^{(l)}\|/N$  の値は, 基底の重要さを測る目安となると考える.

## 4.2 実データでの実験

3 節の方法を, 文献 [Kamishima 05a] の実験で用いた調査データに適用した. 最初のデータは, 寿司の嗜好についての調査データ SUSHI である. このデータ集合では, サンプル順序数  $N = 500$ , 順序長  $L_i = 10$ , 総対象数  $|X^*| = 100$  である. 対象は 12 個の二値属性と, 4 個の数値属性により表現されている. 第 2 のデータ集合は, 読者が考える重要性の順に新聞記事の題名を整列した NEWS データである. これらのニュース記事は「CD 毎日新聞 2003 年版」より得た. このデータ集合は  $N = 4000$ ,  $|X_i| = 7$ ,  $|X^*| = 11872$  である. 記事の題名中のキーワードの出現を 0/1 表した属性ベクトルを利用した. 18381 個のキーワードのうち, 30 回以上現れる 595 個を選んだ. さらに, 記事のカテゴリを表す 8 個の二値属性を加えて, 全部で 603 個の属性を利用した.

提案した次元縮約法の有用性を検証するため, 線形モデルを用いた期待順位回帰法 (ERR)[Kamishima 05a] をこれらのデータに適用した. サンプル順序集合をテスト用と訓練用に分け, 訓練順序集合と, 元の属性ベクトルや縮約した属性ベクトルから, 順序付け関数を獲得した. 予測性能は, テスト集合中の順序  $O_t$  と, それに対応する推定順序  $\hat{O}_t$  との間の Spearman  $\rho$  の平均で測った.  $\rho$  が大きいほど, 予測精度は高いといえる. 交差確認の分割数は SUSHI では 10, NEWS では 5 とした.

図 4 と 5 には, それぞれデータ集合 SUSHI と NEWS に対する, 縮約後の次元数  $K'$  の変化に伴う平均  $\rho$  の変化を示した. どちらのデータ集合についても, サンプル順序や, サンプル順序中の対象を間引くことでサンプル数  $N$  やサンプル順序長  $L_i$  を変化させ, それぞれ別の図に示している. 図 4 と 5 中で,  $N$  や  $L_i$  は (a) から (c) に順に増加しているので, (a) の場合に最も順序付け関数の推定が困難になる. ORIG は次元縮約を全く行わずに, 元の属性ベクトルを用いて教師あり順序付けを適用した結果であり, KRCDR, SRCDR, および PCA は, それぞれ Kendall RCDR, Spearman RCDR, および主成分分析で次元縮約をした後で, 教師あり順序付けを適用した結果である.

以上の実験結果から以下のことが分かる. 第一に, 二つの順位相関次元縮約法の予測精度は類似しているので, 場合に応じて高速な方を用いればよい. 第二に, 順位相関次元縮約は明らかに主成分分析よりよい. その

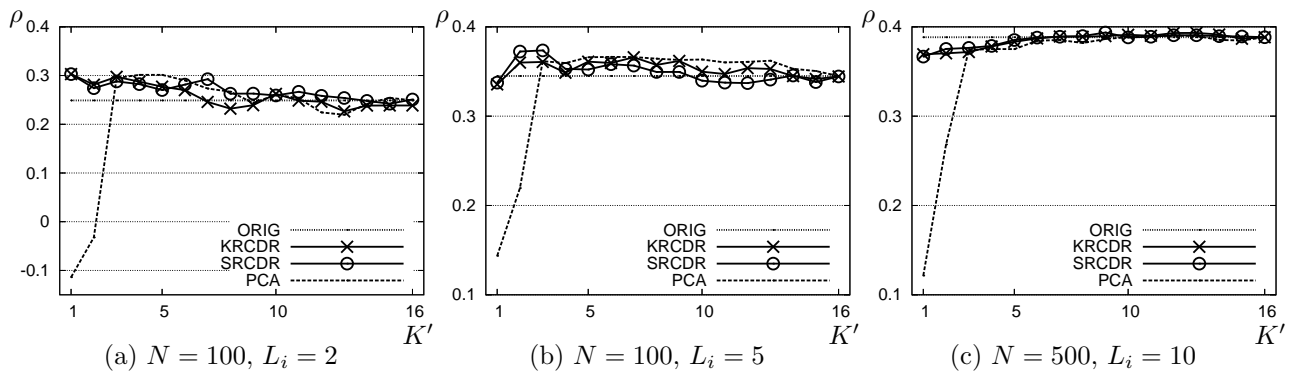


図 4: SUSHI データでの、次元数  $K'$  にもなう予測精度  $\rho$  の変化

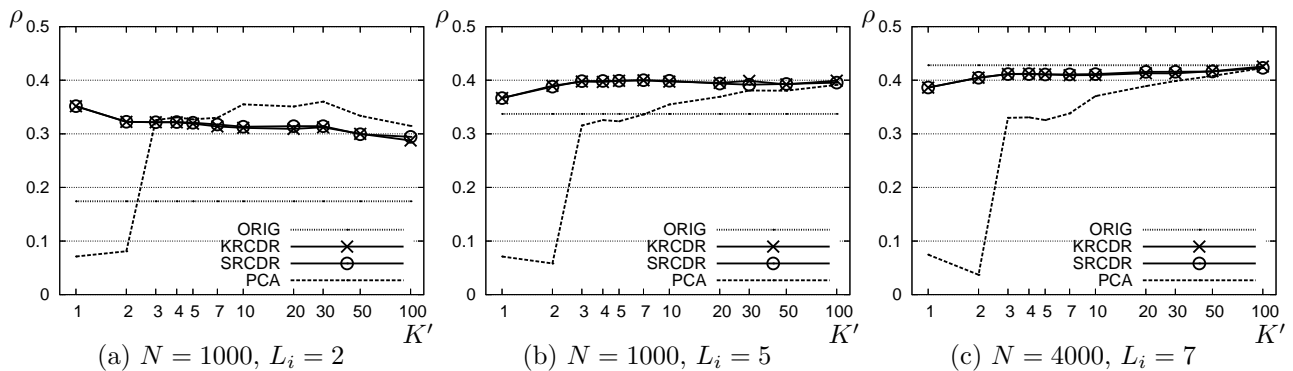


図 5: NEWS データでの、次元数  $K'$  にもなう予測精度  $\rho$  の変化

Note: サンプル順序と推定順序の一致度を Spearman  $\rho$  により評価。これらの図は、部分空間の次元数  $K'$  の変化に伴う  $\rho$  の変化を示した。  $N$  はサンプル集合の大きさで、  $L_i$  はサンプル順序の長さ。 ORIG は次元縮約を全く行わずに教師あり順序付けを適用した結果。 KRCDR, SRCDR, および PCA は、それぞれ Kendall RCDR, Spearman RCDR, および主成分分析で次元縮約をした後で、教師あり順序付けを適用した結果。

差は、次元数  $K'$  が小さい場合に特に顕著であった。これは、順位相関次元縮約法により、目標順序付けを推定するために有用な情報がうまく保存されていることを示している。よって、教師あり順序付け問題を解くにあたっては、順位相関次元縮約法は主成分分析より優れているといえる。第三に、提案した次元縮約法は予測精度の向上に寄与する。図 4 と 5 の順位相関次元縮約による結果 SRCDR/KRCDR を、元の ORIG と比較する。すると、縮約した次元は元のベクトルより良い推定ができることがあり、さらに  $K'$  の大きさが十分であれば、次元縮約をしない結果を下回ることはない。さらに、やや異なる属性集合による推定ではあるが、提案手法を用いた推定結果は、文献 [Kamishima 05a] の結果を、  $N = 100, L_i = 2$  の SUSHI データの場合を除いて上回っている。これは有用な情報を保存したまま、順序付けに用いるモデルが単純化できたためであると考えられる。単純なモデルは、サンプル数が限られている場合により高い汎化能力を発揮する。このことは、  $N$  や  $L_i$  が小さい場合に改善が顕著であることから確かめられる。よって、次元縮約により計算コストがたとえ多少増加する場合でも、RCDR には予測精度を改善でき

るという利点がある。

## 5 まとめ

本論文では、教師あり順序付け問題に特化した次元縮約法を提案した。この手法は、対象の属性ベクトルと目標順序付けの関連についての情報を保存するように設計されている。この目的で、Kendall RCDR 法と Spearman RCDR 法を開発した。これらの手法を実データに適用し、その実験結果から以下の結論を得た。第一に、教師あり順序付け問題を解く場合には RCDR 法は PCA より優れていた。第二に、特に訓練サンプルが不十分な場合には、RCDR 法によって予測性能を改善できた。最後に、提案した二つの次元縮約法の予測精度は同等だった。よって、状況に応じて高速な方を利用すればよい。すなわち、サンプル順序長が  $L_i = 2$  であるときは Kendall RCDR 法を、それ以外では Spearman RCDR 法を用いるべきである。

直観的には、RCDR の第  $l$  回目の反復では、目標順序付けに最も関連がある基底を見つける。その後、属性ベクトルを新しい補空間に射影することで、この新たな

な基底に関連している要素を、属性ベクトルから差し引いている。このとき、目標順序付け中のすでに説明された要素を、順序から差し引くことも有効であると予測される。Diaconis の順序のスペクトル解析 [Diaconis 89] のような考えを導入して、こうした改良を今後は試みたい。

## 謝辞

本研究は科研費 14658106 と 16700157 の助成を受けた。新聞記事の利用を許諾いただいた毎日新聞社に感謝する。

## 参考文献

- [Agresti 03] Agresti, A.: カテゴリカルデータ解析入門, サイエンス社 (2003), (渡邊 裕之他訳)
- [Babington Smith 50] Babington Smith, B.: Discussion on Professor Ross's Paper, *Journal of The Royal Statistical Society (B)*, Vol. 12, pp. 53–56 (1950), (A. S. C. Ross, "Philological Probability Problems", pp. 19–41)
- [Bahamonde 04] Bahamonde, A., Bayón, G. F., Quevedo, J. D. J. R., Luaces, O., del Coz, J. J., Alonso, J., and Goyache, F.: Feature Subset Selection for Learning Preferences: A Case Study, in *Proc. of The 21st Int'l Conf. on Machine Learning*, pp. 49–56 (2004)
- [Burgess 05] Burgess, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., and Hullender, G.: Learning to Rank using Gradient Descent, in *Proc. of The 22nd Int'l Conf. on Machine Learning*, pp. 89–96 (2005)
- [Cohen 99] Cohen, W. W., Schapire, R. E., and Singer, Y.: Learning to Order Things, *Journal of Artificial Intelligence Research*, Vol. 10, pp. 243–270 (1999)
- [Critchlow 91] Critchlow, D. E., Fliener, M. A., and Verducci, J. S.: Probability Models on Rankings, *Journal of Mathematical Psychology*, Vol. 35, pp. 294–318 (1991)
- [Deng 04] Deng, L., Pei, J., Ma, J., and Lee, D.-L.: A Rank Sum Test Method for Informative Gene Discovery, in *Proc. of The 10th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 410–419 (2004)
- [Dettling 02] Dettling, M. and Bühlmann, P.: Supervised Clustering of Genes, *Genome Biology*, Vol. 3, No. 12, pp. research0069.1–0069.15 (2002)
- [Diaconis 89] Diaconis, P.: A Generalization of Spectral Analysis with Application to Ranked Data, *The Annals of Statistics*, Vol. 17, No. 3, pp. 949–979 (1989)
- [Freund 03] Freund, Y., Iyer, R., Schapire, R. E., and Singer, Y.: An Efficient Boosting Algorithm for Combining Preferences, *Journal of Machine Learning Research*, Vol. 4, pp. 933–969 (2003)
- [Herbrich 98] Herbrich, R., Graepel, T., Bollmann-Sdorra, P., and Obermayer, K.: Learning Preference Relations for Information Retrieval, in *ICML-98 Workshop: Text Categorization and Machine Learning*, pp. 80–84 (1998)
- [Herbrich 99] Herbrich, R., Graepel, T., and Obermayer, K.: Support Vector Learning for Ordinal Regression, in *Proc. of the 9th Int'l Conf. on Artificial Neural Networks*, pp. 97–102 (1999)
- [Joachims 02] Joachims, T.: Optimizing Search Engines Using Clickthrough Data, in *Proc. of The 8th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 133–142 (2002)
- [Kamishima 04] Kamishima, T. and Akaho, S.: Filling-in Missing Objects in Orders, in *Proc. of The 4th IEEE Int'l Conf. on Data Mining*, pp. 423–426 (2004)
- [Kamishima 05a] Kamishima, T., Kazawa, H., and Akaho, S.: Supervised Ordering — An Empirical Survey, in *Proc. of The 5th IEEE Int'l Conf. on Data Mining*, pp. 673–676 (2005)
- [神鷹 05b] 神鷹 敏弘, 賀沢 秀人, 赤穂 昭太郎: 教師あり順序付け — 手法の比較実験, 2005 年情報論的学習理論ワークショップ, pp. 213–218 (2005)
- [賀沢 03] 賀沢 秀人, 平尾 努, 前田 英作: Order SVM: 一般化順序統計量に基づく順位付け関数の推定, 電子情報通信学会論文誌 D-II, Vol. J86-D-II, No. 7, pp. 926–933 (2003)
- [Kazawa 05] Kazawa, H., Hirao, T., and Maeda, E.: Order SVM: a kernel method for order learning based on generalized order statistics, *Systems and Computers in Japan*, Vol. 36, No. 1, pp. 35–43 (2005)
- [Kendall 90] Kendall, M. and Gibbons, J. D.: *Rank Correlation Methods*, Oxford University Press, fifth edition (1990)
- [Mallows 57] Mallows, C. L.: Non-Null Ranking Models. I, *Biometrika*, Vol. 44, pp. 114–130 (1957)
- [Marden 95] Marden, J. I.: *Analyzing and Modeling Rank Data*, Vol. 64 of *Monographs on Statistics and Applied Probability*, Chapman & Hall (1995)
- [McCullagh 80] McCullagh, P.: Regression Models for Ordinal Data, *Journal of The Royal Statistical Society (B)*, Vol. 42, No. 2, pp. 109–142 (1980)
- [Plackett 75] Plackett, R. L.: The Analysis of Permutations, *Journal of The Royal Statistical Society (C)*, Vol. 24, No. 2, pp. 193–202 (1975)
- [Shashua 03] Shashua, A. and Levin, A.: Ranking with Large Margin Principle: Two Approaches, in *Advances in Neural Information Processing Systems 15*, pp. 961–968 (2003)
- [Slota 05] Slota, D. J., Vergara, J. P., Ramakrishnan, N., and Heath, L. S.: Algorithms for Feature Selection in Rank-Order Spaces, Technical Report TR-05-08, Computer Science, Virginia Tech. (2005)
- [Thurstone 27] Thurstone, L. L.: A Law of Comparative Judgment, *Psychological Review*, Vol. 34, pp. 273–286 (1927)