

教師あり順序付け — 手法の比較実験

Supervised Ordering — An Empirical Survey

神鷹 敏弘*
Toshihiro Kamishima

賀沢 秀人†
Hideto Kazawa

赤穂 昭太郎‡
Shotaro Akaho

Abstract: Ordered lists of objects are widely used as representational forms. Such ordered objects include Web search results or best seller lists. In spite of their importance, methods of processing orders have received little attention. Only recently has research concerning object ordering become more common; in particular, some researchers have developed various methods for the task of Supervised Ordering: learning functions used to sort objects from example orders. Here, we survey such methods and propose one new method, and empirically reveal compare their characteristics.

Keywords: order, rank, survey, support vector machine, regression, boosting

1 はじめに

ここでは、順序 (order) とは、嗜好、価格といったある特性に従って整列した対象系列のことをいう。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、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) 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。この論文では、独立に提案されたこれらの問題に統一的な視点を与え、既存の統計分野の順序研究との関連について論じる。また、各手法の特性を実験により明らかにする。以前、予備実験結果を [12, 14] にて報告したが、本研究では、新手法をひとつ加え、より現実的な比較実験を行う。

以下、2 節は教師あり順序付けについて、3 節は既存手法の概要と新手法について、4 節と 5 節は人工データと実データに対する実験結果について述べ、最後の 6 節はまとめである。

*産業技術総合研究所, National Institute of Advanced Industrial Science and Technology (AIST)

e-mail: mail@kamishima.net, WWW: http://www.kamishima.net/

†日本電信電話株式会社 NTT コミュニケーション科学基礎研究所, NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation

e-mail: kazawa@cslab.kecl.ntt.co.jp

‡産業技術総合研究所, e-mail: s.akaho@aist.go.jp

2 教師あり順序付け

教師あり順序付け問題の形式的定義について述べる。まず、整列すべき対象を x_j で表す。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 $x_j = (x_{j1}, x_{j2}, \dots, x_{jk})$ で記述される。ただし、 k は属性数。順序は $O = x_{j_1} \succ x_{j_2} \succ \dots \succ x_{j_s}$ の形式で、対象 x_{j_1} がこの順序で最も前にあることを表す。ここで、添え字 j_1 は「この順序の j_1 番目の対象」ではなく「 X^* 中で番号 j_1 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 の長さになる。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、そうでないときを不完全であるという。順位 $r(O_i, x_j)$ は対象 x_j の順序 O_i 中の位置を表す基数で、例えば、 $O_i = x_1 \succ x_3 \succ x_2$ について $r(O_i, x_2) = 3$ である。二つの順序 O_1 と O_2 について、 $x_a, x_b \in X(O_1) \cap X(O_2), x_a \neq x_b$ を満たす対象の対 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で、形式的な条件は次式になる：
$$(r(O_1, x_a) - r(O_1, x_b))(r(O_2, x_a) - r(O_2, x_b)) \geq 0$$

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。

図 1 のように、教師あり順序付け (Supervised Ordering) 問題 [13] は、従属変数が順序である回帰分析問題と見なせる。入力サンプルは、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。回帰曲線に対応するのが属性付き中心順序

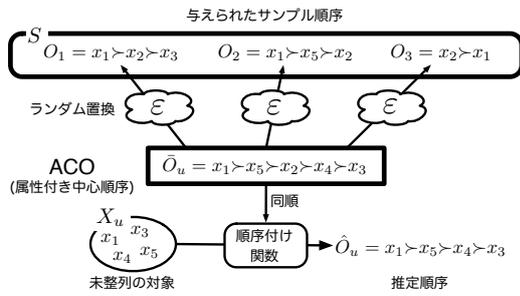


図 1: 教師あり順序付け問題

(Attributed Central Order; ACO) である。これは、 X^* の全ての対象で構成され、通常の回帰と同様に、与えられたサンプル順序だけでなく、サンプルと同じ分布からこれから発生する順序ともできるだけ同順となるように推定される。この問題は通常の回帰と以下の 2 点で異なる：第 1 に、目的変数が順序なので、ACO や順序エラーをモデル化する方法が必要になる。ACO は、順序付け関数 (ordering function) $\text{ord}(X_u)$ でモデル化する。これは、未整列の対象集合 X_u を与えると、その要素を ACO と同順に整列した推定順序 \hat{O}_u を出力する。回帰問題のエラーは、0 平均の確率変数の加算項でモデル化されるが、順序エラーはランダム置換 (random permutation) $\varepsilon(O)$ でモデル化する。これは、入力順序中の対象の順位をある分布に従ってランダムに入れ替える。これらを用いると、サンプル順序 O_i の生成モデルは $\varepsilon(\text{ord}(X_i))$ となる。第 2 に、サンプル順序は一般には不完全なので、 $X(O_i) \subset X^*$ となりうる。よって、サンプル順序集合 S に一度も現れない対象 (図 1 の x_4 など) が生じる。対応する属性空間中で近い対象は、順序中で近くに配置されるという仮定のもと、これらの対象も順序付けできる汎化能力が要求される。

ここで、順序付け関数が絶対であるとは、関数が出力する順序が互いに同順であること、またそうでない場合を相対であると定義する。これは Arrow の不可能性定理 [3] の条件 3 「The independence of irrelevant alternatives」と等価である。例えば、未整列対象集合 $\{x_1, x_2, x_3\}$ と $\{x_1, x_2, x_4\}$ を絶対順序付け関数に与えたなら、対象 x_3 や x_4 の存在とは無関係に対象 x_1 と x_2 に関して同順な順序を、この関数は各入力に対して出力する。だが、相対順序付け関数では、 x_3 や x_4 の存在が x_1 と x_2 の順序に影響する。

教師あり順位付け問題は中心順序 (central order) [17] 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。ACO は、属性で表された対象で構成され、未知の対象も含むような順序であるのに対し、この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 linear ordering 問題となり NP 困難である。その近似解法の研究は幾つかあり、採用しているモデルにより 4 種類に分類できる [5, 17]：サーストン型 (Thurstonian) では対象は内部の実数値スコアの順に整列し、一対比較 (Paired Comparison) は対象の対の順序判断に基づき、距離ベース (Distance Based) では最頻順序 (modal order) からの距離を利用し、多段階 (Multistage) 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。一対比較モデルを用いた手法は相対順序付け関数を学習し、他は絶対順序付け関数を獲得する。

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問題の順序回帰 (Ordinal Regression) [1] とも関連がある。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。例えば、上-中-下の三つの値をとりうるカテゴリ変数などである。順序カテゴリと順序とは次の二点で異なるので注意深く区別すべきである：一つは、順序は純粹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。上記の例で「上」なら全体で上位であることが分かる。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、四つの対象があったとき、上-中-下の順序カテゴリでは、どれか二つは同じカテゴリに分類され区別できないが、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。

3 手法

Cohen 法 (Cohen) [4] 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列の中で、 $\sum_{x_a \succ x_b} \Pr[x_a \succ x_b | x_a, x_b]$ 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 Rankboost (RB) [6] 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 (OSVM) [15, 16] 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 $|X_i|$ 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) [7] 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

表 1: 基本データに対する結果: $|X^*|=1000, |X_i|=10, N=300$

(a) ノイズなしの場合					
	Cohen	RB	SVOR	OSVM	ERR
li/num	0.860 [5]	0.959 [2]	0.914 [3]	0.886 [4]	0.982 [1]
li/bin	0.966 [2]	0.978 [1]	0.885 [4]	0.868 [5]	0.895 [3]
nl/num	0.682 [5]	0.763 [4]	0.911 [2]	0.878 [3]	0.935 [1]
nl/bin	0.786 [5]	0.875 [1]	0.866 [2]	0.842 [3]	0.830 [4]

(b) ノイズありの場合					
	Cohen	RB	SVOR	OSVM	ERR
nl/num	0.652 [5]	0.719 [4]	0.818 [1]	0.797 [3]	0.813 [2]
nl/bin	0.764 [5]	0.842 [1]	0.817 [2]	0.809 [3]	0.796 [4]

表 2: 実データの実験結果

		$N: X_i $	Cohen	RB	SVOR	OSVM	ERR
SUSHI		500:10(b)	0.364 [5]	0.384 [4]	0.393 [3]	0.400 [1]	0.397 [2]
		100:5(b)	0.354 [2]	0.356 [1]	0.284 [4]	0.315 [3]	0.271 [5]
		100:2(b)	0.337 [1]	0.281 [2]	0.115 [4]	0.208 [3]	0.010 [5]
		500:10(n)	0.543 [5]	0.583 [4]	0.719 [1]	0.708 [2]	0.705 [3]
		100:5(n)	0.548 [5]	0.612 [4]	0.646 [2]	0.655 [1]	0.617 [3]
		100:2(n)	0.577 [1]	0.542 [2]	0.522 [4]	0.540 [3]	0.421 [5]
NEWS		4000:7	-0.008 [5]	0.350 [3]	0.244 [4]	0.366 [2]	0.386 [1]
		1000:5	-0.009 [5]	0.340 [3]	0.362 [1]	0.353 [2]	0.312 [4]
		1000:2	-0.009 [5]	0.338 [3]	0.349 [1]	0.344 [2]	0.149 [4]

よって Ranking SVM [8] 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。

最後に期待順位回帰法 (Expected Rank Regression) (ERR) を新たに提案する。これは、[10] の手法を改良したもので、対象の期待順位を予測する回帰関数を Thurstone スコア関数として用いる方法である。観測されない完全順序 O_i^* から均一にランダムに対象が選ばれてサンプル順序 O_i が観測されると仮定する。このとき、 O_i が与えられたときの、対象 $x_j \in X_i$ の完全順序中の条件付き期待順位 $E[r(O_i^*, x_j)|O_i]$ は次式に比例する [2]:

$$\text{erank}(x_j) = r(O_i, x_j) / (|X_i| + 1) \quad (1)$$

$O_i \in S$ なる全ての順序の、 $x_j \in X(O_i)$ なる全ての対象について、 $(x_j, \text{erank}(x_j))$ なる対を生成し、これらの対を訓練事例として x_j の属性値ベクトルから $\text{erank}(x_j)$ を予測する関数を回帰分析で学習する。この関数を Thurstone スコア関数とすることで推定順序を求める。

4 人工データでの実験結果

人工データの生成方法や実験条件は文献 [14] と同様である。対象を記述する属性ベクトルが数値属性のものを num と、バイナリ属性のものを bin と記す。真の ACO を生成する関数が属性値の線形関数であるものを li と、2 次か 3 次であるものを nl と記す。推定順序と真の ACO との Spearman の順位相関 ρ によって推定精度を評価する。 ρ は対象の順位の Pearson 相関であり、

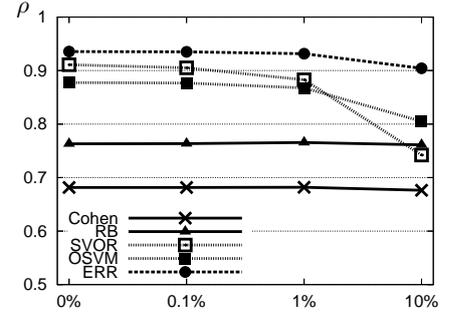


図 2: 順序ノイズに対する変化

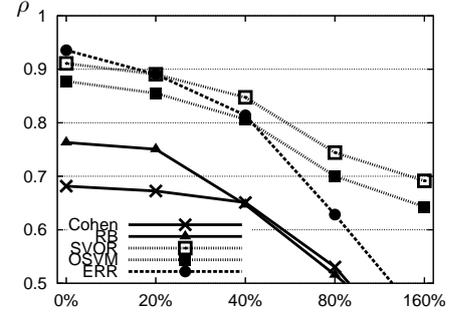


図 3: 属性ノイズに対する変化

大きいほど良い推定であることを示す。

予備実験 [12, 14] の結果に基づき、サンプル数 $N=300$ 、サンプル順序長 $|X_i|=5$ 、及び総対象数 $|X^*|=1000$ をデフォルトの基本実験条件とした。この基本条件に、後に述べるノイズを加えたデータについてアルゴリズムのパラメータを調整した。他の条件としては RB では弱順位付け器に属性の 2 多項式の項を利用し属性値は適宜 $[0, 1]$ となるように変換、SVM 系では $\sigma = 1$ のガウスカーネルを利用、ERR では回帰関数は 2 次多項式とした。よく利用される設定で、どの手法も極端に有利にならないという観点でこれらの条件を定めた。

表 1(a) はこの基本条件での ρ の平均で、行は人工データの種類を、列は手法の種類を表す。カギ括弧内の数値は各手法の順位である。nl/bin の RB と SVOR の間を除き、各手法とその次に順位付けされる手法との差は、危険率 1% の対応のある t 検定で Bonferroni の多重比較¹をしたところ統計的に有意であった。この実験結果の特徴を列挙する: Cohen が悪いのは、Hdge アルゴリズムが属性の順序情報だけを考慮するように設計されているため。ERR は bin では劣るが、num では良い。これは bin では回帰関数の重みを 0 と 1 の 2 点で推定するので不安定になるためだが、SVM 系手法は正則化項の導入でこの問題を回避している。二つの SVM 系手法は非常に類似している。RB は nl では悪いが、これは boosting の反復数 T を増やすと改善できる。しかし、遅

¹ここでは 5 群なので、危険率を $1\%/5C_2$ に調整した

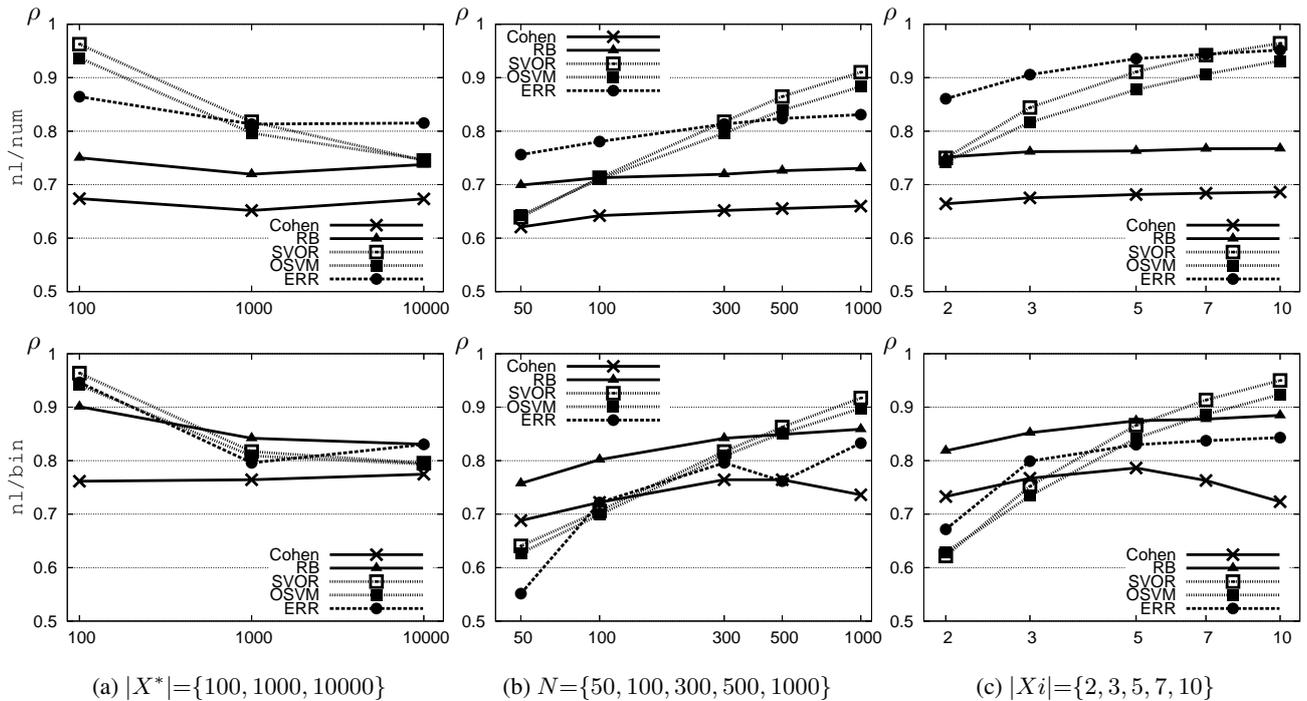


図 4: 対象数 $|X^*|$, サンプル数 N , 及びサンプル順序長 $|X_i|$ に対する変化

すぎたので $T = 100$ に統一して実験した。

表 1(a) は真の順序とサンプル順序が完全に同順なノイズのない場合であった。ここでは順序ノイズに対する頑健性を調べるため、サンプル順序中で隣接する対象をランダム置換した。この置換の回数を変えることでノイズ量を制御し、元の順序と置換後の順序の間の ρ が、元の順序とランダムな順序の間の ρ より小さくなる確率で順序ノイズ量を測った。よってノイズ量 0% はノイズなしの場合に相当する。図 2 は、 $n1/num$ データで、ノイズ量を 0% ~ 10% の範囲で変化させた結果である。ノイズの増加にともない経験 ρ (推定順序と置換後順序の間) は急激に悪化するが、真の ρ (推定順序と置換前の間) は大きくは減少しなかった。例えば、10% のノイズで $n1/num$ に ERR を適用したとき、経験 ρ は 0.409 だったが、真の ρ は 0.904 だった。

次に、属性ノイズに対する各手法の頑健性を調べた。 $N(1, \alpha/100)$ に従うランダム係数を元の属性値に乘じたノイズ量を $\alpha\%$ とした $n1/num$ での結果を図 3 に示す。予備実験 [12, 14] で確認されたように、図 2 と 3 の結果は対照的であった。SVM 系手法は属性ノイズには頑健だが、順序ノイズには弱い。一方、他の手法ではその逆である。この現象は次のように説明できるだろう：SVM 系手法が順序ノイズに敏感なのは、置換された順序対は判別境界面の逆側に現れるためサポートベクトルになりやすいが、属性値が多少変動しても判別境界面をこえなければサポートベクトルには影響しない。一方、その他

の方法では、サンプル順序の多数が正しい順序である限り正しく学習できるため、順序ノイズには頑健になる。しかし、どんなに小さな属性値の摂動も必ず影響する。

次に、実データに近いノイズがある状態での各手法の特徴を調査する。上記の実験から、手法間の相対的な推定精度はノイズの種類に影響され、どの手法も万能ではないといえる。そこで、ここでは相対的な優劣ではなく、データの変化に対する手法間の相違について調査した。結果を図 4 に示す。各行には $n1/num$ と $n1/bin$ のデータに対する結果を、(a)、(b)、及び (c) の列には、それぞれ総対象数 $|X^*|$ 、サンプル数 N 、サンプル順序長 $|X_i|$ を変化させた場合の結果を示した。ノイズ量は、ERR と SVOR とが基本条件のときにほぼ同等の性能になるようにした。具体的には、順序ノイズ量は 1% にし、属性ノイズ量は $n1/num$ では $\alpha = 40\%$ 、 $n1/bin$ では確率 1% で値を反転させた。順序ノイズ量の影響の大きさは $|X_i|$ に依存するので、図 4(c) の結果だけはノイズなしの場合を示した。表 1(b) はこのノイズ量での基本条件での実験結果である。

図 4(a) の実験では、 $|X^*|$ の増加に伴って、テスト順序中に現れる対象が訓練サンプル中に含まれていない確率は増加する。よって、より汎化性能が要求されるようになる。SVM 系手法は $|X^*|$ が小さいときには良いが、大きな $|X^*|$ では性能が低下する。 $|X^*|$ に応じてソフトマージンを適応させることが SVM 系手法が性能を発揮するために必要だろう。非 SVM 系手法の性能は $|X^*|$ に

は影響されなかった．これは SVM 系手法より学習すべき変数が少ないためだろう．

次に，図 4(b) と (c) に移る． $|X_i|$ が大きいと Cohen 法の性能は低下する．これは，この手法が採用している一対比較モデルが順序対の間の独立性を仮定しているためだろう．n1/num の場合，ERR と SVM 系手法は N や $|X_i|$ に関して振る舞いが大きく異なった．これは，サンプル集合が小さすぎる場合には，どの順序対もサポートベクトルになりやすいので，SVM 系手法は過適合しやすくなるためであろう．予測性能とモデルの複雑さとの関係を調べるため，手法のパラメータやモデルをいくらか変えて実験したが，サンプル数と学習すべき自由変数の数には単純な関係を見つけることはできなかった．

5 実データでの実験

3 節の手法を，アンケートにより獲得した実データに適用した．第 1 のデータ集合は，[9] で用いた寿司の嗜好の調査データで，SUSHI と記す．このデータ集合は $N = 500$ ， $|X_i| = 10$ ， $|X^*| = 100$ である．対象は 12 種のバイナリ属性と 4 種の数値属性で記述される．順序用のクラスタリング手法 k -o-means[11] を用いて，順序の分散が大小の 2 種類のサンプル順序を生成した．ERR 法の仮定とは異なり，ある対象がサンプル順序 O_i 中に選択される確率は均一ではなく，3.2% ~ 0.13% の幅がある．第 2 のデータ集合は，ニュース記事のタイトルを重大なものから順に並べるよう質問したデータで，NEWS と記す．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．中心順序からのサンプル順序の分散は，分散が小さい方の SUSHI データより若干大きい．新聞記事をキーワードと文書頻度で表し，LSI によって 20 個の数値属性に圧縮した．加えて，カテゴリを表す 8 個のバイナリ属性も用いた．どちらのデータ集合でも，サンプル順序や対象を取り除くことで， N や $|X_i|$ を変えたデータを生成した． N や $|X_i|$ が減少すると順序の推定はより困難になる．推定順序の良さは，人工データと違って真の順序は不明なので，サンプル順序と推定順序の間の経験 ρ を用いて評価した．

表 2 に実験結果を示す．列のラベル $N:|X_i|$ はサンプル順序の数と長さを表し，b と n の文字はそれぞれ分散の大きいもの (Broad) と小さいもの (Narrow) を表す．SUSHI データでは，人工データの場合より手法間の差は小さい．人工データの順序ノイズの実験から，分散が小さいデータで SVM は有利になると予測していたが，データの分散の違いの影響は小さかった．線形モデルを使った予備実験では，この SUSHI は線形モデルによく適

表 3: 学習と整列に必要な計算量

	Cohen	RB	SVOR	OSVM	ERR
学習	$N \bar{X} ^2k$	$N \bar{X} ^2k$	$N^2 \bar{X} ^4k$	$N^2 \bar{X} ^4k$	$N \bar{X} k^2$
整列	$ X ^2$	$ X \log X $	$ X \log X $	$ X \log X $	$ X \log X $

合していたが，このためデータの分散は影響しにくかったと考える．また，真のエラーではなく経験エラーによる評価も関係しているだろう．ERR は N や $|X_i|$ が大きい場合には良かったが，小さい場合はそうではなかった．これは 2 次多項式モデルを採用しているためで，線形モデルであれば 100:2:b の場合に対して ρ は 0.249 と向上した．逆に，RB は小さな N や $|X_i|$ で良かったが，これは boosting の反復数を $T = 100$ で止めたためである． $T = 300$ では N や $|X_i|$ が大きい場合は性能が向上したが，小さな場合には過適合のため性能が低下した．NEWS データの場合，サンプル順序の分散は小さく，この点では予測は容易だが，属性値とサンプル順序との相関は非常に弱く，どの手法も悪かった．このような弱い属性では Cohen はパラメータを調整しても良い結果は得られなかった．ここでも ERR は大きな N や $|X_i|$ に対してのみ結果が良かったが，SUSHI と同様に線形モデルを使うと性能が向上した．まとめると，Cohen 法以外の性能はほぼ同等といえるが，ソフトマージン，boosting 反復数，モデルの次数といったパラメータ設定に性能は敏感である．これらのパラメータ設定のガイドラインは今後の大きな課題である．

6 議論とまとめ

最初に学習と整列に要する計算量を表 3 にまとめる．訓練順序集合 S 中に現れる対象と対象対の数をそれぞれ $N|\bar{X}|$ と $N|\bar{X}|^2$ で近似した．ただし， $|\bar{X}|$ はサンプル順序の平均長．また，SVM の学習時間は訓練事例数に対して 2 乗と仮定した．Cohen の Hedge や RB の学習時間は，反復数 T を定数とすれば $N|\bar{X}|^2$ について線形だが， T をサンプル数に対して適応的に選ぶなら超線形になる．SVM 系手法の計算量は，本来は非 0 の属性値数に依存するので，実際には属性数 k について線形より小さい．実用的には，SVM 系の学習は遅く，Cohen と RB は中間で，ERR が最も早い．未整列の対象集合 X を整列する時間は Cohen は $O(|X|^2)$ だが，他は $O(|X|\log|X|)$ とより早い．

最後に各手法の長所と短所についてまとめる．本論文で提案した新手法 ERR は予測精度を犠牲にすることなく，実用的に最も高速だった．よって，特徴選択や学習パラメータの調整も，比較的短い時間で可能である．

SUSHI:100:2(b) のように ERR 法の性能が悪い場合でも、学習パラメータを再調整することで実際に改善できた。この方法では、対象が観測される分布が均一であると仮定しているが、この制約を満たさない SUSHI データに対する結果からこの仮定が満たされない場合に対する頑健性が示されている。しかし、 k に対して 2 乗の計算量がこの手法の欠点である。

Cohen法の最大の利点は唯一のオンライン手法であることである。実験では Cohen法の予測精度はやや悪かったが、これは Cohen法が用いている Hedge アルゴリズムが属性の順序情報だけを参照するように設計されていることの影響が大きい。単純ベイズのような数値やカテゴリ属性を扱う簡潔な手法を代わりに用いた実験では予測精度はかなり改善された。Thurston 型モデルの他の手法とは異なり、Cohen法は一対比較モデルを用いている。よって、相対順序付け関数を学習する問題がある。フィルタリングや推薦といった応用では絶対順序付け関数が望ましい。例えば、オレンジよりリンゴが好きな場合に、果物の任意の集合について嗜好順に整列したとしても、レモンなどの他の果物の存在とは関係なく、リンゴはオレンジより上位になるべきである。この性質は、まさに絶対順序付け関数の条件なので、Cohen法はこうした応用には不適合といえるだろう。

RBの特徴は多様な弱学習器を使えることである。この性質により、多様な型の属性を利用できるので、順序とカテゴリ、または、順序と数値属性とが混合した属性ベクトルで対象が表現されている場合には、この方法しか利用できない。RB 法の実験結果は若干悪いが、これは訓練サンプル数に対して適応的に反復数 T を増加させれば改善された。実験では、収束の遅さのため、初期段階の予測エラーの急速な減少が終了した時点となるような反復数で一定 $T = 100$ にした。しかし、 T を適応的に増加させると、計算量は SVM 系手法と同等になってしまうと共に、過適合も観察された。

クラス分類の SVM と同様に、属性数 k が多い場合に SVOR と OSVM は非常に有利である。実験では、SVM 系と非 SVM 系手法では異なる種類のノイズに対して頑健であった。順序中で対象が置換するノイズがあるときには非 SVM 系手法がよいが、属性値に外乱が加わる場合には SVM 系手法が良い。SVM 系手法の欠点はその学習時間の遅さにある。SVOR と OSVM の学習計算量は同じだが、実用的には OSVM の方が遅い。だが、実験では、順序ノイズについては SVOR より相対的に頑健であった。

以上、教師あり順序付け手法を実験的に比較し、その特徴を論じた。今後は、手法を改良するとともに、情報フィルタリングなどの実問題へも適用したい。

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

- [1] A. Agresti. 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).
- [2] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*. John Wiley & Sons, Inc., 1992.
- [3] K. J. Arrow. *Social Choice and Individual Values*. Yale University Press, second edition, 1963.
- [4] W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. *Journal of Artificial Intelligence Research*, Vol. 10, pp. 243–270, 1999.
- [5] D. E. Critchlow, M. A. Fligner, and J. S. Verducci. Probability models on rankings. *Journal of Mathematical Psychology*, Vol. 35, pp. 294–318, 1991.
- [6] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research*, Vol. 4, pp. 933–969, 2003.
- [7] R. Herbrich, T. Graepel, P. Bollmann-Sdorra, and K. Obermayer. Learning preference relations for information retrieval. In *ICML-98 Workshop: Text Categorization and Machine Learning*, pp. 80–84, 1998.
- [8] T. Joachims. Optimizing search engines using clickthrough data. In *Proc. of The 8th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 133–142, 2002.
- [9] T. Kamishima. Nantonac collaborative filtering: Recommendation based on order responses. In *Proc. of The 9th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 583–588, 2003.
- [10] T. Kamishima and S. Akaho. Learning from order examples. In *Proc. of The 2nd IEEE Int'l Conf. on Data Mining*, pp. 645–648, 2002.
- [11] T. Kamishima and J. Fujiki. Clustering orders. In *Proc. of The 6th Int'l Conf. on Discovery Science*, pp. 194–207, 2003. [LNAI 2843].
- [12] T. Kamishima, H. Kazawa, and S. Akaho. Estimating attributed central orders — an empirical comparison. In *Proc. of The 15th European Conference on Machine Learning*, pp. 563–565, 2004. [LNAI 3201].
- [13] T. Kamishima, H. Kazawa, and S. Akaho. Supervised ordering — an empirical survey. In *Proc. of The 5th IEEE Int'l Conf. on Data Mining*, 2005. (to appear).
- [14] 神鷹敏弘, 賀沢秀人, 赤穂昭太郎. 属性付き中心順序の推定 — 手法のサーベイと比較実験. 人工知能学会全国大会 (第 18 回) 論文集, 2F3–05, 2004.
- [15] H. Kazawa, T. Hirao, and E. Maeda. 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.
- [16] 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.
- [17] J. I. Marden. *Analyzing and Modeling Rank Data*, Vol. 64 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 1995.