2F3-05

属性付き中心順序の推定 — 手法のサーベイと比較実験

Estimating Attributed Central Orders — A Survey and an Empirical Comparison

神嶌 敏弘[*1] 賀沢 秀人[*2] 赤穂 昭太郎[*1]
Toshihiro Kamishima Hideto Kazawa Shotaro Akaho

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

[*2]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, NTT Corporation

Lists of ordered objects are widely used as representational forms. Such ordered objects include Web search results or best seller lists. In spite of their importance, the methods of processing orders have received little attention. However, research concerning object ordering is becoming more common. Some researchers have developed various methods to perform almost the same task: a learning function used for sorting objects from examples of ordered sequences. We call this task the estimation of *Attributed Central Orders*. The performance of this task is useful for sensory surveys, information retrieval, or decision making. We surveyed such methods, empirically compared their properties, and discuss their merits and demerits.

1 はじめに

本論文は順序を推定するための学習手法を実験的に比較する．

順序とは，対象を何らかの特性で並べたもので，Web の検索結果，売れ筋ランキングなど頻繁に利用されているデータの表現方法である．その重要性に比べて，順序を直接扱う研究はあまりなかったが，最近はいくつかの研究が見られるようになった．特に，属性ベクトルで表された対象の順序を学習事例として，順序付けされていない対象を整列するための規則を学習するという問題にはいくつかの手法が提案されている．ここでは，これを「属性付き順序」の推定問題と呼ぶ．この問題は，主観的な変量を扱う官能検査 [Kamishima 03b, Kamishima 03a, 神嶌 02] や，情報検索でのフィードバック [Cohen 99, Joachims 02] に利用されている．この問題の 5 種類の学習手法を実験的に比較し，その長所と短所を論じる．

2 節ではこの問題の形式的定義，3 節では手法の概要，4 節では比較実験と議論，5 節ではまとめを述べる．

2 属性付き中心順序

本節では，属性付き中心順序推定問題の形式的な定義を示す．

対象 x_j は整列されるものである．対象全集合 X^* は全ての可能な対象で構成される．各対象 x_j は属性数 k の属性値ベクトル $x_j = (x_{j1}, x_{j2}, \ldots, x_{jk})$ で表される．順序を $O{=}x_1{\succ}x_2{\succ}\cdots{\succ}x_3$ と記し，$x_1{\succ}x_2$ は x_1 が x_2 より上位であることを表す．順序 O_i を構成する全ての対象の集合を X_i と記す．集合 A の大きさを $|A|$ とすると，$|X_i|$ は順序 O_i の長さ．全対象を含む順序（O_i s.t. $X_i = X^*$）を完全順序，そうでない順序を不完全順序という．順位 $r(O_i, x_j)$ は対象 x_j の順序 O_i 中での位置を示す基数．例えば，順序 $O_i{=}x_1{\succ}x_3{\succ}x_2$ で $r(O_i, x_2)$ は 3 である．二つの順序 O_1 と O_2 について，$x_a, x_b{\in}X_1{\cap}X_2, x_a{\neq}x_b$ であるような対象の対 x_a と x_b を考える．x_a と x_b について O_1 と O_2 が一致するとは，二つの対象が同じ順序で配置されること，すなわち，$(r(O_1, x_a){-}r(O_1, x_b))(r(O_2, x_a){-}r(O_2, x_b)) \geq 0$. また，そうでないときを不一致という．$O_1$ と O_2 が一致するとは，$x_a, x_b{\in}X_1{\cap}X_2, x_a{\neq}x_b$ なる全ての対象の対について O_1 と O_2 が一致することである．

N 個の標本順序を含む集合 S に対して，中心順序は次式で定義される [Marden 95].

$$\bar{O} = \arg\min_{O'} \textstyle\sum_i^N d(O', O_i), \qquad (1)$$

ただし，$d(O_a, O_b)$ は後で述べる順序間の距離．これには統一された名前がないが，ここでは，文献 [Marden 95] にならい，中心順序と呼ぶことにする．式中の最小値を探す範囲は，対象が属性で記述されるかどうかによって決まる．「属性なし」の場合，すなわち，対象が一意な識別子によって記述される（例：果物を「リンゴ」などの名称で区別する）場合を述べる．この場合，中心順序は S 中に現れる全ての対象（$X_S \equiv \cup_{O_i \in S} X_i$）で構成され，$\bar{O}$ は X_S 中の対象の全ての順列の中から探索する．属性なし中心順序については多くの研究がある [Marden 95].

ここでは，対象が属性ベクトルで記述される属性付き中心順序を扱う．この場合，属性空間中で近傍にある対象は近くに順位付けされると仮定することで，X_S に含まれない対象の順位も推定できる．よって，対象全集合 X^* で構成される中心順序を求めること，すなわち，単に S 中の標本順序からの距離を最小化するのではなく，S の母集団から生成されるどの順序との距離をも小さくすることを目標にする．属性なし中心順序では X_S は有限なので，中心順序は X_S 中の対象の順列で表せる．しかし，属性付きの場合では，X^* が無限集合の場合にはこの順列では表せないので，別の記述方法として，整列関数 $\mathrm{sort}(X_u)$ を導入する．これは，順序なし対象集合 X_u を入力とし，X_u の要素を含み，属性付き中心順序と一致する推定順序 $\hat{O}_u$ を出力する関数である．属性付き中心順序の問題とは，標本順序集合 S から，この関数を学習することである．さらに，中規模以上の X^* に対して完全順序が得られる状況は実際には稀なので，標本順序が不完全である場合に注目する．

次に，式 (1) で用いる順序間の距離 $d(O_a, O_b)$ について述べる．ここでは，順位差の 2 乗和で定義される，一般的な Spearman の距離 d_S を用いる．これを，値域が $[-1, 1]$ となるように正規化すると，Spearman の順位相関 ρ が得られる．

$$\rho = 1 - 6d_S/(|X|^3 - |X|). \qquad (2)$$

この係数を予測精度の評価に用いる．

連絡先: 神嶌敏弘，http://www.kamishima.net/

3 手法

ここでは 5 種類の属性付き中心順序の推定手法を述べる．

Cohen の方法 (Cohen と略す)[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 $\Pr[x_a \succ x_b | x_a, x_b]$ を表す関数を標本順序から学習しておく．整列関数 $\mathrm{sort}(X_u)$ は次の手続きに相当する．まず，X_u 中の全ての対象の対について上記の確率を計算する．X_u 中の対象のある順列に対して，その順列と一致する全ての対象対 $x_a \succ x_b$ についての，確率 $\Pr[x_a \succ x_b | x_a, x_b]$ の総和を求める．この総和を最大にする順列を中心順序とする．しかし，順列の数は $|X_u|!$ と多いので，欲張り探索を行う．また，Cohen らは確率関数の推定に独自の Hedge アルゴリズムを用いたが，ここでは広く利用されている単純ベイズを用いた．

次に，SVM 系の 2 手法，賀沢らの Order SVM [賀沢 03] と Herbrich らの SVOR [Herbrich 98]，について述べる．Order SVM (OSVM) は，ある対象が，順序の中で第 j 位以上かどうかを判別するように設計された SVM を用いる．ある対象の順位は，他の対象に対して相対的に決まるので，順位を判別できても，X_u 中の対象を整列できるかどうかは自明ではない．そこで，賀沢らは，任意の対象集合の中で，対象がある順位以上になる確率関数は，その値で整列することで中心順序導くことができる効用関数と単調な関係にあることを示した．さらに，j を 1 から $|X_i|$ まで動かし，各 j 位で判別した結果が平均的に標本順序と一致するようにすることで安定的に学習を行う．Herbrich らの Support Vector Ordinal Regression (SVOR) は，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された SVM を用いる．この手法は独立に Joachims によって Ranking SVM [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，ある対象がどれだけ順序の中で上位になりやすいかを示す効用関数になり，その値で X_u 中の対象を整列することで，その中心順序が推定される．

最後に Thurstone の比較判断の法則に基づく手法を二つ述べる．この法則は，対象 i に正規分布 $N(\mu_i, \sigma)$ に従う値を割り当て，その値の順に整列することで順序が得られるとするモデルである．一つ目の手法は，Thurstone 回帰 (TR) [神嶌 02] と呼ぶ手法である．まず，Thurstone のモデルに最小 2 乗法を適用して X_S の属性なし中心順序を推定する．この中心順序中の順位を予測するための，対象の属性値の線形関数を線形回帰によって学習する．もう一つの，属性付き Thurstone モデル (ATM)[赤穂 02] は，最尤推定で線形の効用関数を求める手法である．最尤推定量は，最小 2 乗法で求めた解を初期値とした最急降下法で求める．どちらも，学習の結果，線形の効用関数が得られるので，その値の順に X_u 中の対象を整列して中心順序を推定する．

4 実験

前節の手法を，人工データと実データに適用して比較する．

4.1 人工データでの実験

4.1.1 実験手順

人工データは，属性値，中心順序，標本順序の 3 段階で生成した．最初に，2 種類のベクトル（数値とバイナリ）を生成した．数値の場合の要素数は $5(\equiv k)$ で，属性値は正規分布 $N(0,1)$ でランダムに生成した．バイナリでは $15(\equiv k)$ 個の属性からなり，全ての対象が異なる属性ベクトルで表されるようにランダムに生成した．数値とバイナリをそれぞれ *NUM* と *BIN* で記す．次に，これらの属性値を次の効用関数に与えたときの値の順に，対象を整列することで，属性付きの中心順序を生成した．

表 1: 基本条件での結果: $|X^*|$=1000, $|X_i|$=10, N=300

	LI		NL	
	NUM	BIN	NUM	BIN
Cohen	0.918 [5]	0.966 [3]	0.766 [5]	0.788 [5]
OSVM	0.933 [4]	0.933 [5]	0.931 [2]	0.924 [2]
SVOR	0.942 [3]	0.945 [4]	0.940 [1]	0.936 [1]
TR	0.993 [1]	0.985 [1]	0.790 [3]	0.810 [4]
ATM	0.992 [2]	0.985 [1]	0.789 [4]	0.811 [3]

$$\mathrm{utility}(x_j) = (1 + \textstyle\sum_{l=1}^{k} w_l x_{jl})^{\dim} \qquad (3)$$

ただし，w_{jl} は正規分布 $N(0,1)$ に従うランダムに決めた重みである．中心順序は，dim=1 の線形の場合と，dim=2 or 3 の非線型の場合で実験した．線形の場合を *LI*，非線型を *NL* と記す．重みの値による影響を避けるため，10 セットの重みを生成した．最後に，X^* から $|X_i|$ 個の対象をランダムにサンプルし，中心順序と一致するように整列することで標本順序 $O_i \in S$ を生成した．

評価尺度には，上記の真の中心順序と推定順序との間の Spearman の ρ を用いた．ρ が 1 なら，推定順序は完全に一致し，-1 なら全くの逆順である．式 (3) の効用関数の 10 セットの重みそれぞれに，10 分割の交叉確認を適用し，その平均を結果として示す．

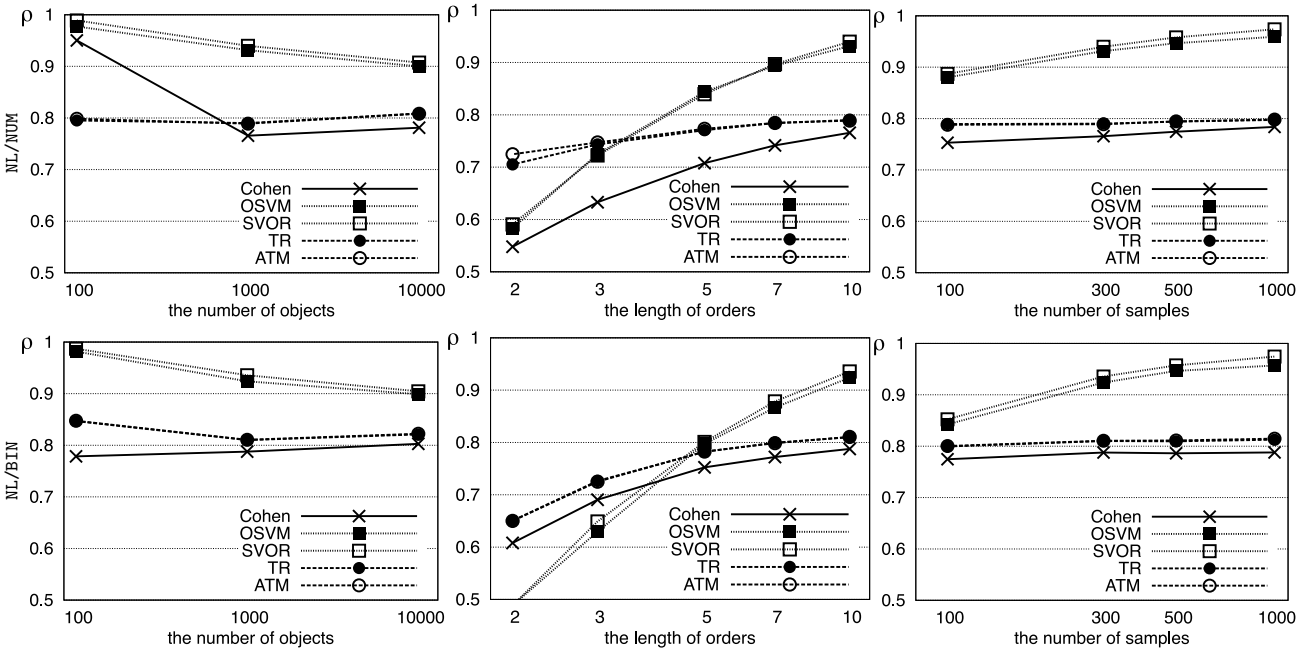
4.1.2 実験結果

基本的な実験条件を，標本順序数 N=300，標本順序長 $|X_i|$=10，全対象数 $|X^*|$=1000 とした．この条件では，テスト集合 X_u 中の対象が，標本順序のいずれにも現れない確率は約 6.6% と低い．しかし，このデータではアルゴリズムの汎化性能は検証できないと考えるのは誤りである．順序の重要な情報は，対象の相対的な位置で表されるので，少なくとも二つの対象が観測される必要がある．この基本条件では，一対の X_u 中の対象が，標本順序のいずれかに同時には含まれない確率は約 98% と高く，汎化性能を評価するのに適していると考える．以後，特に示していない条件は，この基本条件を用いる．

表 1 に，基本条件での ρ の平均を示した．第 1 列は各手法の名称，続く列は，それぞれ，LI/NUM, LI/BIN, NL/NUM および NL/BIN のデータに対する結果を示した．大括弧中の値は各手法の順位である．対応のある t 検定で Bonferroni の多重比較修正を行って検定した結果，NL/BIN データでの TR と ATM 間以外では，順位が連続している手法の間には，危険率 1%で有意な差があった．

TR と ATM は，共に Thurstone モデルを用いているので性能は同等だった．SVM 系の 2 種法も類似していた．LI/BIN データ以外では Cohen は悪かったが，最も高速だった．当然だが，線形モデルは線形データで良く，SVM 系は非線型データで優れていた．SVM 系は低バイアスモデルを用いているので線形の場合は悪いが，非線型の場合はその表現能力が有効であった．

順序ノイズに対する頑健性を調査するため，標本順序の隣接する対象をランダムに選び交換した．この交換回数によってノイズ量を調整し，交換回数が 0，15，30，72 の 4 種類のデータを用意した．予備実験では，ランダムな順序と中心順序の間の ρ がこの方法による順序と中心順序の間の ρ より小さくなる確率は，それぞれ 0%，0.1%，1% および 10% であった．図 1 には NL/NUM データで順序ノイズの量を変えた場合の



(a) $|X^*|$=$\{100, 1000, 10000\}$　(b) $|Xi|$=$\{2, 3, 5, 7, 10\}$　(c) N=$\{100, 300, 500, 1000\}$

図 3: (a) 総対象数 $|X^*|$，(b) 標本順序長 $|Xi|$ および (c) 標本順序数 N を変化させた場合

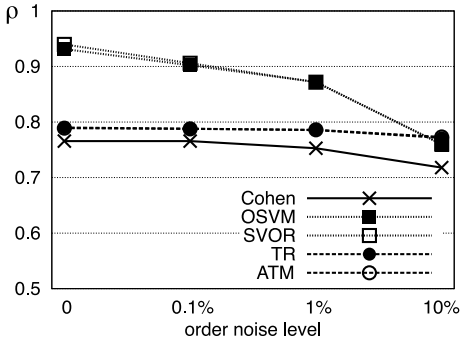


図 1: 順序ノイズの量を変化させた場合

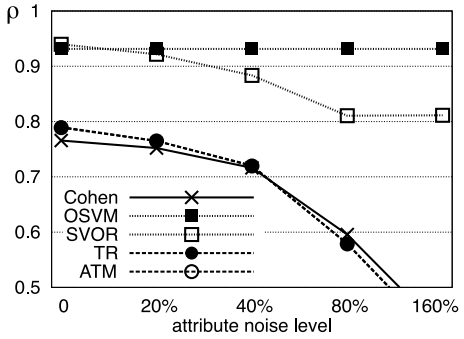


図 2: 属性ノイズの量を変化させた場合

結果を示した．TR とほぼ重なっているので ATM の結果は見ていない．

全般的に，どの方法もかなりの順序ノイズに対して頑健であった．この結果は，10% レベルのノイズではクラスタの抽出に失敗したクラスタリングの場合 [Kamishima 03b] と対照的である．個々の順序に置換があっても，中心順序は安定的に求められることが分かる．ノイズの増加に伴い，経験 ρ（ノイズを加えた標本順序と推定順序の間の ρ）は急激に悪化するが，真の ρ（ノイズのない真の順序に対する ρ）はそれほど悪くならない．例えば，10% のノイズ量では，TR の経験 ρ は 0.347 だが，真の ρ は 0.711 であった．

次に，属性値のノイズに対する頑健性を調査した．数値属性での α% のノイズとは，正規分布 $N(1, \alpha/100)$ に従うランダムな係数を乗じた場合をいう．図 2 は NL/NUM データに対する属性ノイズの変化に伴う結果である．図 1 と 2 の結果は明らかに対照的であった．SVM 系の手法は属性ノイズに頑健だが，順序ノイズにはそうではなかった．他の 3 種法はその逆であった．この現象は次のように説明できる．SVM 系の手法では，交換された順序対はサポートベクトルになりやすいので順序ノイズに影響大きく影響されるが，属性ノイズでは識別平面を超えない限り影響されない．逆に，非 SVM 系の手法では，正しい順序の対が多数派であれば影響を受けないが，少しの属性ノイズにも影響される．

図 3 は，データの他の特徴を変化させた場合の ρ の平均である．上の行が NL/NUM データで，下が NL/BIN である．第 1 列 (a) は総対象数 $|X^*|$ を変化させた場合である．$|X^*|$ を 100，1000，10000 と変えると，テスト集合中の対象対が，標本順序中に同時に現れない確率は 8.5%，98%，99.98% と増加するため，$|X^*|$ の増加に伴いより高い汎化能力が必要になる．SVM 系の手法は低バイアスモデルを用いているので，$|X^*|$ の増加にともない過適合のため性能が低下した．逆に，非 SVM 系の手法は，高バイアスな線形モデルのため，$|X^*|$ が小さくても訓練事例への適合が妨げられるので，性能が向上しない．

第 2 列 (b) は標本順序の長さの変化に伴う結果である．短くなるにつれて，SVM 系の性能は急激に悪くなる．どの手法も，推定精度は順序対の総数 $N|X_i|^2$ に依存するが，これは $|X_i|$ の減少について 2 乗で減少する．SVM 系ではパラメータ数が多いため，非 SVM 系より順序の長さの影響が大きいと考える．

第 3 列 (c) は標本順序数 N を変化させた場合である．N の減少にともない $N|X_i|^2$ も減少するので，$|X_i|$ の場合と同様に，SVM 系の手法は相対的に予測精度が大きく低下した．しかし，$N|X_i|^2$ の減少は，2 乗ではなく線形なので，$|X_i|$ の場合ほど低下は顕著ではなかった．

表 2: 寿司の嗜好調査データに対する実験結果

	100:10	1000:10	5000:10	5000:5	5000:2
Cohen	0.381 [5]	0.412 [3]	0.406 [3]	0.373 [4]	0.281 [4]
OSVM	0.393 [2]	0.424 [1]	—	0.380 [2]	0.294 [2]
SVOR	0.407 [1]	0.423 [2]	0.419 [1]	0.384 [1]	0.295 [1]
TR	0.382 [4]	0.412 [3]	0.410 [2]	0.375 [3]	0.263 [5]
ATM	0.386 [3]	0.399 [5]	0.386 [4]	0.365 [5]	0.282 [3]

表 3: 各手法の計算量

	学習	整列
Cohen	$N\lvert\bar{X}\rvert^2$	$\lvert X_u\rvert^2$
OSVM	$(N\lvert\bar{X}\rvert^2)^\gamma$	$\lvert X_u\rvert \log \lvert X_u\rvert$
SVOR	$(N\lvert\bar{X}\rvert^2)^\gamma$	$\lvert X_u\rvert \log \lvert X_u\rvert$
TR	$\max(N\lvert\bar{X}\rvert^2, \lvert X_S\rvert^2)$	$\lvert X_u\rvert \log \lvert X_u\rvert$
ATM	$\max(LN\lvert\bar{X}\rvert^2, L\lvert X_S\rvert, \lvert X_S\rvert^2)$	$\lvert X_u\rvert \log \lvert X_u\rvert$

4.2 実データに対する実験

3 節の手法を，寿司の嗜好調査データ [Kamishima 03a] に適用した．標本順序長 $|X_i|$ は 10 で，二つの名義属性と四つの数値属性で各対象は記述される．二つの名義属性のうちの一つは値域が 12 値あるので，Cohen 以外では 11 個のバイナリのダミー変数に変換した．

表 2 には標本順序と推定順序の間の ρ である経験エラーの平均を示した．100:10 の列は N=100 と $|X_i|$=10 の場合の結果である．下線の結果は最上位より有意に悪いことを示す．5000:10 では OSVM は計算時間が大きく結果を得られなかった．対象総数 $|X^*|$ が 100 と少ないので，汎化性能はこのデータではあまり検証されていない．どの手法もある程度は適切な整列関数を学習できており，人工データの場合ほど手法間の差は顕著ではない．ここでも，SVM 系の手法は N や $|X_i|$ の増加に伴い性能が向上した．ただし，バイアスが低いので結果が不安定になる場合がある．例えば，OSVM は 5000:3 の場合には他の手法より有意に悪かった．

4.3 比較結果のまとめ

ここでは，各手法の長所短所をまとめる．計算量を表 3 にまとめた．“学習” には整列関数の学習に要する計算量で，“整列” には X_u を整列する計算量を示した．総対象対数は $N|\bar{X}|^2$ ($|\bar{X}|$ は平均順序長）で近似した．表中の γ は約 2 であることが経験的に知られている．整列時間の差はそれほど大きくないが，学習時間は Cohen，Thurstone 系，SVM 系の順に遅くなる．大まかにいって，予測精度の高い手法の方が遅い．計算時間の問題から，事実上，SVM 系の手法は $N|\bar{X}|^2$ が $10^5 \sim 10^6$ 程度までしか適用できない．Thurstone 系はより大規模のデータを処理できるが，$|X_S|$ は $10^5 \sim 10^6$ が限界である．

次に，各手法の定性的な特徴を述べる．$\Pr[x_a \succ x_b | x_a, x_b]$ は [Cohen 99] のようにオンラインの分類学習手法でで更新できるので，Cohen はデータストリームの処理に適している．二つの SVM 系手法の学習の計算量は同じオーダーだが，OSVM の方が実際には遅い．また，この手法は標本順序の長さが違う場合には適用できない．しかし，属性ノイズに対しては非常に頑健である．二つの Thurstone 系手法で，ATM は TR に追加の計算が必要だが，性能は同等なので TR の方が良い．

図 1 と 2 の結果から，SVM 系と非 SVM 系では異なる型のノイズに対して頑健である．よって，順序ノイズのあるデータには非 SVM 系，属性ノイズのあるデータには SVM 系の手法がよい．図 3 から，モデルのバイアスがデータに適合するかどうかは予測精度に大きく影響することが分かる．どの手法も，バイアスの異なるモデルを用いることは可能である．Cohen は確率関数 $\Pr[x_a \succ x_b | x_a, x_b]$ を Cohen らの Hedge や一般的な SVM などで求めてもよい．非線型回帰を Thurstone 系の手法に適用することもできる．これらの変更で，低バイアスのモデルを利用できるが，計算量が増える場合がある．SVM 系の手法には高バイアスの線形カーネルなどが適用できるが，計算量を減少させることはできない．

5 まとめ

本論文では，属性付き中心順序の推定手法を比較し，その特徴を明らかにした．今後は，上述のようにバイアスを変化させた場合や，$|X^*|$ が大きな実データでの実験を行う予定である．

謝辞： 本研究は科研費萌芽研究 (14658106) の助成を受けた．

参考文献

[赤穂 02] 赤穂 昭太郎, 神嶌 敏弘：順序例からの学習のための線形モデルによるアプローチ, 人工知能学会全国大会（第 16 回）論文集, 3C1–08 (2002)

[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)

[Herbrich 98] Herbrich, R., Graepel, T., Bollmann-Sdorra, P., and Obermyer, K.: Learning Preference Relations for Information Retrieval, in *ICML-98 Workshop: Text Categorization and Machine Learning*, pp. 80–84 (1998)

[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)

[神嶌 02] 神嶌 敏弘：順序例からの学習 — 比較判断の法則の導入と嗜好調査データへの適用, ソフトウェア科学会 第 3 回データマイニング ワークショップ, pp. 61–70 (2002)

[Kamishima 03a] Kamishima, T.: 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)

[Kamishima 03b] Kamishima, T. and Fujiki, J.: Clustering Orders, in *Proc. of The 6th Int'l Conf. on Discovery Science*, pp. 194–207 (2003), [LNAI 2843]

[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)

[Marden 95] Marden, J. I.: *Analyzing and Modeling Rank Data*, Vol. 64 of *Monographs on Statistics and Applied Probability*, Chapman & Hall (1995)