人工知能学会研究会資料 SIG-KBS-A405-13 (2005)

順序中の欠損対象の補完

Filling-in Missing Objects in Orders

神嶌 敏弘 [1*]　赤穂 昭太郎 [1]
Toshihiro Kamishima[1] and Shotaro Akaho[1]

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

Abstract: Filling-in techniques are important, since missing values frequently appear in real data. Such techniques have been established for categorical or numerical values. Though lists of ordered objects are widely used as representational forms (e.g., Web search results, best-seller lists), filling-in techniques for orders have received little attention. We therefore propose a simple but effective technique to fill-in missing objects in orders. We built this technique into our collaborative filtering system.

1　はじめに

本論文では，順序中の欠損対象の補完手法 [13] を提案し，これを協調フィルタリング手法に組み込むことで，その有効性を実験的に示す．

順序とは，何らかの基準で複数の対象を並べたもので，対象の前後関係だけに意味がある．この順序に関し，まず Stevens の尺度水準 [27] の概念について述べる．これは，比率，間隔，順序，名義の四つの水準に尺度を分類したものである．比率（例:長さ）は原点がある数値尺度で，間隔（例:時間）は原点がない数値尺度，名義尺度はいわゆるカテゴリを表す．これら 3 種の水準には多くの解析・処理手法が提案されてきた．しかし，順序尺度は幅広く利用されてはいるが，その解析手法は十分ではない．例えば，Web 検索エンジンは質問への適合度に応じてページを順序付けしたリストを返す．買い物のときに売り上げランキングを参照することも多い．また，情報検索などでも，比較したときにより適合している文書を利用者に指定させるといったフィードバックなどで利用されている [6, 10]．このような重要性にも関わらず，順序の処理手法はあまり注目されてこなかった．順序中の欠損対象の補完もそのような処理の一つで，実データでは欠損対象は頻繁に生じるためこのような処理は重要である．ここでは特に，ある順序のサンプル集合について，その集合中のある順序に含まれない対象の順位を，その集合を要約量で補完する手法を提案する．これは，順序以外の尺度水準では，サンプルの平均や最頻値などの要約量で欠損値を補完することがよく行われるが，これを順序尺度において実現するものである．

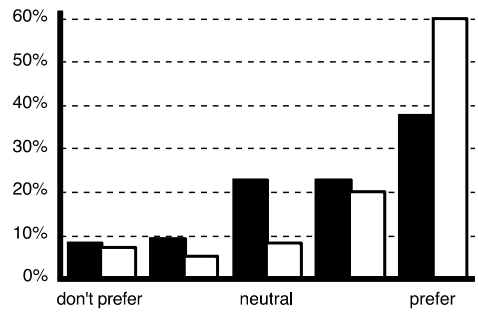

図 1: SD 法によって得られたスコアの分布

この補完手法の開発の動機は，協調フィルタリング手法の予測制度の向上であった．協調フィルタリング (Collaborative Filtering; CF)[4] とは，利用者が好むものを，他の利用者の嗜好をもとに推定するための枠組みである．この CF を実行するには利用者の嗜好の度合いを計測する必要がある．まず，この嗜好や味覚などの感覚を定量的に測る目的に順序が適していることを示す．嗜好は，Semantic Differential (SD) 法 [24] で計測されることが多い．これは，その両端が対義語で表された物差しを用いて利用者が嗜好の度合い示す方法である．例えば，両端が「好き」と「嫌い」になっている 1〜5 のスコア目盛のついた物差しなどが用いられる．こうして得られたスコアを間隔尺度として扱うために「ある物差しの目盛は等間隔」であり,「全ての物差しの全長は等しい」という二つの仮定が導入される [23]．しかし，これらの仮定を満たすために自身の感覚を利用者が定量化するのは一般には難しいだろう．このことを示すため，SD 法で得たスコアの分布を図 1 に示す．黒は寿司の嗜好調査で，白は Amazon.com の顧客によ

*連絡先：`http://www.kamishima.net/`

る評価[1]である．上記の仮定によれば，これらの分布は理想的には単峰で対称であるべきだが，実際の分布は歪みがあって単峰でもない．さらに，SD 法は心理学的な効果の影響も受けることがある．例えば，長所を積極的に認めるべきだという社会規範などの影響により実際より良く評価する寛大効果 (lenienty effect)[22] などがある．このようなスコアのシフトは Aggarwal[1] などによっても報告されている．

SD 法の代替手法の一つに順位法がある．この方法では，嗜好の度合いの順に対象を利用者に整列させた順序応答によって嗜好パターンを計測する．我々は以前に**なんとなく協調フィルタリング**[11] と呼ぶ順位法を用いた枠組みを提案し，そのの有効性を示した．しかし，利用者間の相関の評価が困難なため，順序応答の長さが短い場合には SD 法に対して優位ではなかった．このような短い順序応答は Joachims の方法 [10] により暗黙的に収集できるため，この場合での推薦の改善は重要である．これを実現するために，順序のための補完手法を開発し，それを CF に適用する．

2 節では順序の補完手法，3 節ではなんとなく CF，4 節では実験について述べ，5 節はまとめである．

1.1 関連研究

近年の順序の研究には以下のようなものがある．属性で記述された対象の順序を訓練集合とし，任意の対象集合を整列する規則を学習する教師あり順序付け問題が研究されている [2, 6, 9, 10, 12, 15, 17]．Freund は順序付けを目的とする boosting である RankBoost を考案した [8]．Mannila と Meek [19] は順序の集合から特徴的な順序構造を見つける手法を提案した．Sai ら [26] は順序変量に対する相関ルールを提案した．神嶌ら [14, 16] は順序をクラスタリングした．Lebanon と Lafferty [18] は順序変数を条件とする順序変数の条件付き確率分布を推定した．メタ検索エンジンを順序の統合問題として扱った Dwork ら [7] の研究もある．

2 順序中の欠損対象の補完

順序に関する基本的な表記をまず示す．x_j は整列される対象を表し，対象全集合 X^* は可能な対象全てで構成される．順序は $O{=}x_1{\succ}x_2{\succ}\cdots{\succ}x_3$ のように表記され，$x_1{\succ}x_2$ ははは x_1 が x_2 より順序中で前にあることを表す．また，対象 x_j の添字 j は，j 番目ではなく，X^* 中で対象 x_j を一意に定める識別子を表すことに注意されたい．X_i は順序 O_i を構成する全ての対象を含む対象集合である．よって，集合の大きさ $|X_i|$ は順序 O_i の長さに等しくなる．全ての対象を含む順序，すなわち $X_i = X^*$ なる O_i は完全順序（complete order），またそうでないとき，不完全順序（incomplete order）と呼ぶ．順位 $r(O_i\,,x_j)$ は対象 x_j の順序 O_i での位置を表す基数である．例えば，$r(O_i\,,x_2)\,,O_i{=}x_1{\succ}x_3{\succ}x_2$ は 3 である．二つの順序 O_1 と O_2 について，$x_a\,,x_b{\in}X_1{\cap}X_2\,,$ $x_a{\neq}x_b$ なる対象 x_a と x_b があるとき，x_a と x_b について O_1 と O_2 が同順（concordant）であるとはこれら二つの対象が同じ順に順序中現れることで，形式的には次の条件が満たされることである．

$$(r(O_1\,,x_a) - r(O_1\,,x_b))(r(O_2\,,x_a) - r(O_2\,,x_b)) \geq 0$$

また，そうでないとき非同順（discordant）であるという．O_1 と O_2 について，$x_a\,,x_b{\in}X_1{\cap}X_2\,,x_a{\neq}x_b$ なる全ての対象の対について同順なら，O_1 と O_2 は同順であるいう．距離 $d(O_a\,,O_b)$ は同じ対象で構成される二つの順序，すなわち，$X_a = X_b(\equiv X)$ を満たす順序間で定義される．Spearman の距離 $d_S(O_a\,,O_b)$ [20] はよく用いられる距離の一つで，順位の差の二乗和で定義される．本論文では，計算量が少なく，その性質が良く研究されているのでこの d_S を用いる．この距離を $[-1\,,1]$ の範囲に正規化したものが Spearman の順位相関 ρ で次式で定義される．

$$\rho = 1 - 6 \times d_S/(|X|^3 - |X|). \tag{1}$$

次に，欠損対象の補完について述べる．次の二つの順序の間の距離を求める場合を考える．

$$O_a = x_1{\succ}x_3{\succ}x_6 \text{ と } O_b = x_5{\succ}x_3{\succ}x_2{\succ}x_6 \tag{2}$$

距離は同じ対象で構成される順序間で定義されるが，X_a と X_b は異なっていて，どちらの順序にももう一方の順序には含まれていない欠損対象がある．すなわち，順序 O_a の欠損対象は $\tilde{X}_a = \{x|\,x{\notin}X_a{\wedge}x{\in}X_b\} = \{x_2\,,x_5\}$ で，O_b では $\{x_1\}$ である．よって，直接距離を計算することはできない．この問題の回避方法の一つは欠損対象を無視して，共通に含まれる対象 $X_a \cap X_b$ のみについて距離を計算することである．例えば，O_a と O_b について欠損対象をと除いて x_3 と x_6 だけを含む順序を作ると，どちらも $x_3{\succ}x_6$ になる．すると，Spearman の距離は $d_S = 0$ となる．しかし，欠損対象に含まれているであろう有用な情報は無視されるので，この方法では計算された距離の精度や確信度は低くなるだろう．さらに，もし $X_a \cap X_b = \emptyset$ なら，距離を計算することすらできない．そこで，順序のサンプルがある場合，より適切に距離を測るれるようにするために，欠損対象の順位をサンプルの要約統計量で補完する方法を提案する．ここで，本論文では対象は均一にランダムに欠損する，すなわち，X^* から均一に重複を許さず対象を選ぶことで X_i が生成されるものと仮定する．例

[1]KDD2003 の A. S. Weigend による招待講演 “Analyzing Customer Behavior at Amazon.com” より

えば，上位 3 位を抽出した順序などはこの仮定を満たさない．

2.1 既存の順序補完手法

心理統計の分野では，これらの欠損値は，単一の順序ではなく順序の集合を考えることで処理される．この**不完全順序集合**（Incomplete Order Set; IOS）[20] の概念[2]を示す．これは，与えられた不完全順序と同順な完全順序の集合で，形式的には，O を対象集合 X で構成される順序，$\tilde{X}$ を欠損対象の集合とすると次式で IOS は定義される：

$$\mathrm{ios}(O\,,\tilde{X})=\{O'_i|O'_i \text{は } O \text{ と同順}\,, X'_i=X\cup\tilde{X}\}$$

しかし，IOS の大きさは $|X'|!/|X|!$ のように $|X'|$ に対して指数的に増加するので，この IOS は大規模なデータには適さない．さらに，順序の集合間の距離を定義する問題もある．二つの集合間の任意の順序対の距離の平均をとる方法などが考えられるが，自身への距離 $d(\mathrm{ios}_a\,, \mathrm{ios}_a)$ が 0 にならないなどの問題があり，Hausdorf 距離などの利用が必要になる．

IOS の問題を回避するために [11] で我々はデフォルト順位の概念を提案した．これは，順序の中心部分は中立と考えられるため，これは欠損対象を順序の中央に挿入するものである．式 (2) の例の場合，欠損対象を順序の中央に挿入すると O_a と O_b 補完順序は

$$O'_a=x_1\succ x_3\succ x_2\sim x_5\succ x_6 \text{ と } O'_i=x_5\succ x_3\succ x_1\succ x_2\succ x_6$$

のようになる．ただし，$\sim$ は同順位を表す．全ての同順位対象には，これらの対象に割り当てられるべき順位の平均値である midrank 値を割り当てた．例えば，O'_a の第 3 位か 4 位になる対象 x_5 と x_2 にはこれらの順位の平均値 3.5 を両方の対象に割り当てる．しかし，このデフォルト順位をなんとなく CF で用いたところ，有効性は確認されなかった．これは，中央部分の順位が必ずしも中立を示すものではなかったからではと考えられる．例えば，ほとんどのサンプル順序中で最下位にある対象があったとする．この対象を中央部分に順位付けすると，補完後の順序ではこの対象が相対的に高く順位付けされてしまう．それゆえ，順序中で中立を表す別の方法が必要になる．

2.2 デフォルト順序

順序中での中立を表す**デフォルト順序**（default order）なる概念を提案する．数値変数や名義変数での欠損値は，平均や最頻値といったサンプルの要約統計量で補完されることが多い．同様に，欠損対象の順位をサンプル順序集合 S の中心によって補完する．中心順序 $\bar{O}_S$ [20] は次式で定義される：

$$\bar{O}_S=\arg\min_O \textstyle\sum_{O_i\in S} d(O_i\,,O).$$

ただし，$\bar{O}_S$ はサンプル順序に含まれる全ての対象 $\bar{X}_S=\cup_{O_i\in S}X_i$ で構成される．しかし，いくつかの特別な場合を除き，中心順序を厳密に求めることは計算量的に不可能である．それゆえ，順序の生成確率モデルである Thurstone の比較判断の法則 [28] に基づいた方法を用いる．このモデルでは，正規分布 $N(\mu_j\,,\sigma)$ に従う効用値で対象 x_j を整列することで，順序が生成されるとみなす．このモデルに最小 2 乗規準を適用する [21] と，μ_j を線形変換した値 μ'_j が次式で求められる：

$$\mu'_j=\tfrac{1}{|\bar{X}|}\textstyle\sum_{x\in\bar{X}_S}\Phi^{-1}\big(\Pr[x_j\succ x]\big) \tag{3}$$

ただし，$\Phi(\cdot)$ は正規累積分布関数であり，対象 x_j が x より上位になる確率 $\Pr[x_j\succ x]$ はサンプル順序中の順序対の数を数えることで求められる．ここでは，対象 x_j をこの μ'_j の値で整列して得られた順序を中心順序の近似として用いる．

デフォルト順序とは，中心順序と同順で，全ての欠損対象で構成される順序として定義される．式 (2) の例の場合で，中心順序が $x_1\succ x_5\succ x_2\succ x_3\succ x_4\succ x_6$ であるとする．O_a と O_b の欠損対象はそれぞれ $\{x_2\,,x_5\}$ と $\{x_1\}$ である．よって，O_a のデフォルト順序は $\tilde{O}_a=x_5\succ x_2$ となり，O_b では $\tilde{O}_b=x_1$ となる．このデフォルト順序によって欠損対象の順位を補完する方法を提案する．このために，観測された順序とそのデフォルト順序を併合する必要が生じる．しかし，定義によりこれらの順序は共通の対象を含まないので，既存の順序併合手法は利用できない．そこで順序統計の理論に基づく順序の新たな併合手法を提案する．

順序 O が観測され，そのデフォルト順序が $\tilde{O}$ であるとき，O と $\tilde{O}$ を併合して補完順序 O' を得る場合を考える．ただし，これら三つの順序はそれぞれ対象集合 $X, \tilde{X}$，及び X' から構成されている．定義により，$X'=\tilde{X}\cup X$ と $X\cap\tilde{X}=\emptyset$ が成立する．ここで，この併合過程を直接モデル化するかわりに，次のような順序の生成過程を考える．対象は均一に欠損すると仮定したので，X' から重複を許さずに $|O|$ 個の対象が均一分布に従ってサンプルされる．これらの対象は O' と同順となるように整列され，順序 O が観測されたとする．この場合，順序統計 [3] では，O 中の i 番目の対象 $x_{i:O}$ について，O' 中での順位の期待値が次式になることが知られている．

$$\mathrm{E}[r(O'\,,x_{i:O})]=i\times\frac{|O'|+1}{|O|+1}$$

[2] 引用文献ではこの概念は *incomplete rankings* と呼ばれているがここでは集合であることを強調するため不完全順序集合と呼ぶ

ただし，期待値は X' からサンプルされたさまざまな X についてとる．同様に，$\tilde{O}$ 中の j 番目の対象の期待順位は次式になる．

$$\mathrm{E}[r(O', x_{j:\tilde{O}})] = j\frac{|O'|+1}{|\tilde{O}|+1}$$

これらの期待順位を X と $\tilde{X}$ 中の全ての対象について求め，この期待順位の順に X' 中の対象を整列することで併合後の順序 O' を求める．式 (2) の順序 $O_a{=}x_1{\succ}x_3{\succ}x_6$ と，そのデフォルト順序 $\tilde{O}_a = x_5{\succ}x_2$ を併合する場合の例を示す．O_a の 2 番目の対象 x_3 には，期待順位 $2\times(5{+}1)/(3{+}1) = 3$ が割り当てられる．これらの期待順位を残りの対象についても同様に割り当てる．そして，これらの期待順位で整列すると補完順序

$$O'_a{=}x_1{\succ}x_5{\succ}x_3{\succ}x_2{\succ}x_6$$

が得られる．同様に O_b の補完順序は

$$O'_b{=}x_5{\succ}x_3{\succ}x_1{\succ}x_2{\succ}x_6$$

となる．

この補完手法は高速に計算可能である．まず，デフォルト順序を求めるために $|\tilde{X}|$ 個の対象を整列するために，$O(|\tilde{X}|\log|\tilde{X}|)$ の計算量が必要である．次に，$|\tilde{X}|+|X| = |X'|$ 個の対象の期待順位は $O(|X'|)$ の時間で計算できる．最後に，期待順位で全体を整列する必要がある．ここで，期待順位と O や $\tilde{O}$ 中での順位とは単調な関係にあり，かつ，X と $\tilde{X}$ 中の対象はすでに期待順位の順に整列されているので，全体の整列にかかる時間は $O(|X'|)$ で済む．よって，中心順序が事前に計算されているとした場合の計算量は

$$O(\max(|X'|, |\tilde{X}|\log|\tilde{X}|))$$

程度で済むため，大規模なデータにも適用できる．

3　対象を補完したなんとなく CF

本節では，前節の補完手法を組み込んだなんとなく CF 手法について述べる．

最初に，文献 [11] の**なんとなく協調フィルタリング**について述べる．CF の目的は，ある特定の利用者（活動利用者）の嗜好を，他の利用者について収集された嗜好データ（利用者 DB）に基づいて予測することである．まず，対象集合 X_i が提示されると，利用者 i は自身が好きなものからそうでないものへこれらの対象を順に整列した順序を返す．その順序を $O_i{=}x_1{\succ}x_2{\succ}\cdots{\succ}x_{|X_i|}$ と記す．利用者 DB$D_S = \{O_1, \ldots, O_{|D_S|}\}$ は，様々な利用者からこのような O_i を集めたものの集合とし，これらの利用者を標本利用者と呼ぶ．活動利用者の嗜好順序を O_a，この順序を構成する対象の集合を X_a と記す．O_a と D_S が与えられたとき，活動利用者が好む X_a 以外の対象を推定することがなんとなく CF の目的である．

単純相関法（SCR）はこのなんとなく CF の基本的な方法である．これは，GroupLens [25] とほぼ同じ手法で，単に利用者 i の対象 j の評価スコアを，その利用者の標本順序中の対象 j の順位 $r(O_i, x_j)$ で置き換えたものである．対象 j に対する活動利用者の嗜好は次式で推定される．

$$\hat{r}_{aj} = \frac{\sum_{i\in I_j} R_{ai}\left(r(O_i,x_j)-\bar{r}_i\right)}{\sum_{i\in I_j}|R_{ai}|} \qquad (4)$$

ただし，$\bar{r}_i$ は $X_{ai}{=}X_a{\cap}X_i$ 上での平均順位で，I_j は対象 j を順位付けした標本利用者を示す添字の集合，すなわち，$\{i|\,O_i{\in}D_S \text{ s.t. } x_j{\in}X_i\}$．$R_{ai}$ は活動利用者と標本利用者 i の，X_{ai} 中の対象についての嗜好順位の Pearson 相関係数．対象を推定嗜好の強い順に整列し，上位の対象を推薦する．ここで，もう一方の嗜好順序で欠損している対象は無視されるが，順位を繰り上げたりはしないことに注意されたい．例えば，$O_a{=}x_1{\succ}x_2{\succ}x_3$ と $O_i{=}x_3{\succ}x_1$ について，O_a 中の対象 x_2 はで O_i 中で欠損しているので，対象 x_2 は無視する．しかし，x_3 の順位 $r(O_a, x_3)$ は 2 に繰り上げず 3 のまま計算する．そのため，この相関は Spearman の ρ とは異なる．

この手法には次の問題点がある．総対象数 $|X^*|$ に対して嗜好順序の長さ $|X_i|$ が短くなると，活動利用者と標本利用者が共通の対象を順位付けしなくなる場合（$X_a{\cap}X_i{=}\emptyset$）がよく生じるようになる．この場合 R_{ai} は常に無相関 0 と扱うので，利用者の類似度は正確には計測されず不適切な推薦がなされる．このような短い，特に長さが 2 の応答順序での推薦精度の改善できれば，次の利点がある．CF を実際に運用すると，利用者が自身の嗜好を示すことを面倒だと感じ，利用者 DB の標本順序が集まらない問題が生じる．これに対し，Joachims[10] は長さが 2 の順序を暗黙的に自動で集める手法を提案しており，こうして収集した順序で高い精度の推薦ができれば非常に有用である．

SD 法での CF での評価対象が少ない場合の問題に対して，Breese ら [5] は欠損スコアを補完し，補完後の対象集合間で相関を求める方法を提案している．このアイデアをなんとなく CF に適用するため，SCR 法にて順序中で欠損した対象を補完する．利用者 DB の中心順序 $\bar{O}_S$ は事前に求めておき，相関 R_{ai} を計算する前に，応答順序中の欠損対象の順序を $\bar{O}_S$ と同順のデフォルト順序を用いて補完する．そして，補完後の順序間の Spearmanρ を R_{ai} として用いる．最後に，補完後の順序は相関 R_{ai} の計算にのみ用い，式 (4) の $\hat{r}_{aj}$ は，この R_{ai} と，補完前の順位から計算し，この $\hat{r}_{aj}$ によって推薦対象を決める．

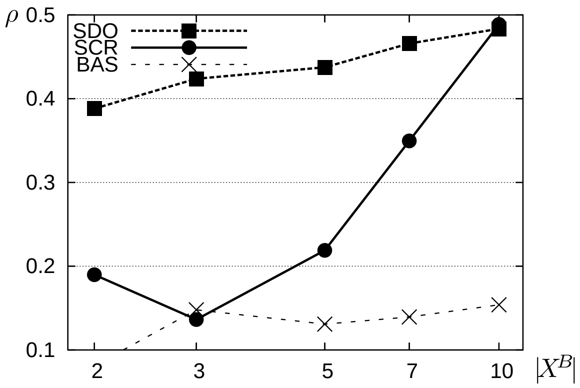

図 2: $|X^B|$ の変化にともなう推定嗜好順序と標本嗜好順序の間の ρ（大きな値ほど精度の良い推定）

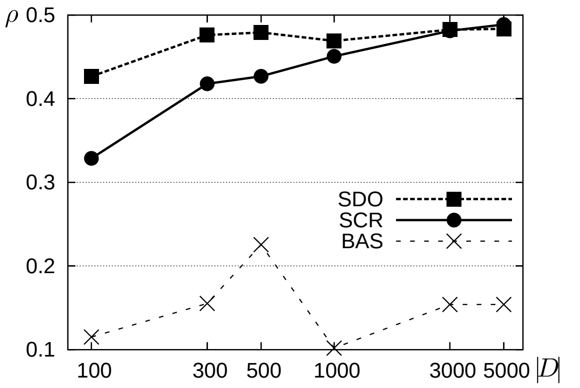

図 3: $|D|$ の変化にともなう推定嗜好順序と標本嗜好順序の間の ρ（大きな値ほど精度の良い推定）

4 実験

デフォルト順序の有効性を，寿司の嗜好データに前節の CF 手法に適用することで検証する．データの収集方法と実験手続きは文献 [11] と同じだが，データ数は 5000 件に増やした．推薦の精度は，推定したた嗜好順序と，利用者が答えた嗜好順序の間の Sparman ρ（式 (1)）を，10 分割交叉確認で評価した．よって，ρ が大きいほど精度の高い推定であるといえる．

デフォルト順序を用いたなんとなく CF 手法（SDO）を，次の二つのベースラインと比較する．一つはデフォルト順序を用いない単純相関法（SCR）．もう一つは，個人化しない推薦（BAS）で，利用者 DB の中心順序と同順となるように対象を推薦する．すなわち，中心順序を人気リストとして扱い，人気のある対象を推薦する．

図 2 に，サンプル数を $|D|$=5000 に固定し，応答順序の長さ $|X^B|$ を変えたときの結果を示す．SDO 法は明らかに元の SCR 法より優れており，その差は $|X^B| \leq 7$ で統計的に有意であった．SCR 法では応答順序が短くなるにつれ，より利用者間の類似性が不適切に計測されるので，SDO 法はより短い応答順序で SCR と比較して効果的であった．図 3 は，応答順序長を $|X^B|$=10 に固定したときに，サンプル数 $|D|$ を変化させた場合の結果である．SDO 法はサンプル数が少ない場合に SCR に対して効果的で，$|D|$=300, 500 ならその差は統計的に有意であった．サンプル数が減少すると，活動利用者が順位付けした対象を評価している標本利用者の数が減少するので，SCR 法の精度は低下する．しかし，SDO では活動利用者とどの標本利用者との間の相関も計算できるため，精度の良い推薦が可能である．

次にもう一つのベースラインである BAS 法との比較結果を検証する．もし，全ての利用者の好みが似通っていれば，人気リストである中心順序によって精度の高い推薦が可能になるだろう．しかし，この個人化していない推薦方法は明らかに SDO 法より劣っている．この結果は，デフォルト順序を採用することによる改善が，利用者の嗜好が共通しているという特徴がこのデータ集合にあるからではなく，SDO 法によって個人の嗜好をより適応的に推定できために可能となったことがわかる．

5 まとめ

本論文では，順序中での欠損対象を，順序サンプルの要約統計量である中心順序から得たデフォルト順序によって補完する方法を提案した．このデフォルト順序による補完によって，利用者間の類似性を適切に計測することが可能になったので，なんとなく CF 手法の精度を向上させることができた．

謝辞： 科研費 14658106 と 16700157 の助成を受けた

参考文献

[1] C. C. Aggarwal, J. L. Wolf, K.-L. Wu, and P. S. Yu. Horting hatches an egg: A new graph-theoretic approach to collaborative filtering. In *Proc. of The 5th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 201–212, 1999.

[2] 赤穂昭太郎, 神嶌敏弘. 順序例からの学習 —潜在変数モデルによるランク付け—. 電子情報通信学会総合大会資料集, D–4–1, 2002.

[3] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*. John Wiley & Sons, Inc., 1992.

[4] J. Ben Schafer, J. A. Konstan, and J. Riedl. E-commerce recommendation applications. *Data Min-*

ing and Knowledge Discovery, Vol. 5, pp. 115–153, 2001.

[5] J. S. Breese, D. Heckerman, and C. Kadie. Empri-cal analysis of predictive algorithms for collaborative filtering. In *Uncertainty in Artificial Intelligence 14*, pp. 43–52, 1998.

[6] W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. *Journal of Artificial Intelligence Research*, Vol. 10, pp. 243–270, 1999.

[7] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the Web. In *Proc. of The 10th Int'l World Wide Web Conf.*, pp. 613–622, 2001.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] T. Kamishima and S. Akaho. Learning from order examples. In *Proc. of The IEEE Int'l Conf. on Data Mining*, pp. 645–648, 2002.

[13] T. Kamishima and S. Akaho. Filling-in missing objects in orders. In *Proc. of The IEEE Int'l Conf. on Data Mining*, pp. 423–426, 2004.

[14] T. Kamishima and J. Fujiki. Clustering orders. In *Proc. of The 6th Int'l Conf. on Discovery Science*, pp. 194–207, 2003. [LNAI 2843].

[15] 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].

[16] 神嶌敏弘, 藤木淳. 順序のクラスタリング — 順序平均の最適性について. 電子情報通信学会技術研究報告, PRMU 2003–83, 2003.

[17] 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.

[18] G. Lebanon and J. Lafferty. Crankng: Combining rankings using conditional probability models on permutations. In *Proc. of The 19th Int'l Conf. on Machine Learning*, pp. 363–370, 2002.

[19] H. Mannila and C. Meek. Global partial orders from sequential data. In *Proc. of The 6th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 161–168, 2000.

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

[21] F. Mosteller. Remarks on the method of paired comparisons: I — the least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, Vol. 16, No. 1, pp. 3–9, 1951.

[22] 中島義明ほか（編）. 心理学辞典. 有斐閣, 1999.

[23] 中森義輝. 感性データ解析 — 感性情報処理のためのファジィ数量分析手法. 森北出版, 2000.

[24] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum. *The Measurement of Meaning*. University of Illinois Press, 1957.

[25] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. GroupLens: An open architecture for collaborative filtering of Netnews. In *Proc. of The Conf. on Computer Supported Cooperative Work*, pp. 175–186, 1994.

[26] Y. Sai, Y. Y. Yao, and N. Zhong. Data analysis and mining in ordered information tables. In *Proc. of The IEEE Int'l Conf. on Data Mining*, pp. 497–504, 2001.

[27] S. S. Stevens. Mathematics, measurement, and psychophysics. In S. S. Stevens, editor, *Handbook of Experimental Psychology*. John Wiley & Sons, 1951.

[28] L. L. Thurstone. A law of comparative judgment. *Psychological Review*, Vol. 34, pp. 273–286, 1927.