

# 順序の距離と確率モデル

## A Survey on Distances and Probabilistic Models for Orders

神島 敏弘\*

産業技術総合研究所

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 or bestseller lists. We survey distances and probabilistic distributions that are designed for orders or rankings.

### 1 はじめに

本稿は順序の距離と確率モデルをサーベイである。

まず **Stevens の尺度水準 (Stevens' level of measurement)** [50, 53] における順序尺度を紹介する。これは観測を数値に対応付ける尺度を次の四つに分類する。

**名義尺度 (nominal scale)** スポーツ選手の背番号など、値が一致しているかどうかだけに意味のある尺度。数値・ラベルを置換する対称群による変換に対して意味が保存される。

**順序尺度 (ordinal scale)** モースの硬度、等級、段位など、数値の大小に意味がある尺度。単調増加関数による変換に対して意味が保存される。

**間隔尺度 (interval scale)** 摂氏温度、エネルギー、日付など、原点がなく値の間隔に意味のある尺度。例えば、 $10^{\circ}\text{C}$  の2倍である  $20^{\circ}\text{C}$  は温度が2倍という意味ではない。線形変換に対してその意味が保存される。

**比率尺度 (ratio scale)** 絶対温度、長さ、重さ、密度、抵抗、音の大きさ (sone) など、比率に意味のある尺度。比例変換に対して、その意味が保存される。

その後、対数間隔尺度など他の尺度水準が提案されたが、ここでは、**順序付きカテゴリ尺度 (ordered categorical scale)** について述べておく。これは、カテゴリ尺度 (名義尺度と同義) のように有限の値のうちの一つをとるが、それらの値に順序関係があるものである。例えば、{優, 良, 可, 不可} や {松, 竹, 梅} などが該当する。順序付きカテゴリは段階が有限で、絶対的な情報を含む。例えば、{優, 良, 可, 不可} で、5個の対象があれば最低二つは大小関係が区別できない。ま

た、「優」であれば絶対的に良いことが分かる。一方、順序では段階数は無限だが、相対的な情報しか与えない。

次に記号を定義する。対象集合  $\mathcal{X}_m = \{x_1, x_2, \dots, x_m\}$  を整列したものを順序  $O = x_{i_1} \succ x_{i_2} \succ \dots \succ x_{i_m}$  と記す。全ての対象間に順序関係が定義されている全順序の場合について述べるが、後に半順序の扱いについても述べる。順位とは、指定された対象が順序中で、最上位から何番目に現れるかを示す数値である。**順位ベクトル (rank vector)** の  $i$  番目の要素は、対象  $i$  の順序中でどの順位を示す。例えば、順序  $x_3 \succ x_1 \succ x_2$  中で、 $x_1$  の順位は2であり、順位ベクトルは  $[2, 3, 1]$  となる。逆に、 $i$  番目の要素が第  $i$  にある対象の番号であるベクトルを**順序ベクトル (order vector)** と呼ぶ。例えば、 $x_3 \succ x_1 \succ x_2$  を表す順序ベクトルは  $[3, 1, 2]$  となる。

代数的には、 $\mathcal{X}_m$  中の対象が作る全ての順序に対する順位ベクトルや順序ベクトルは全体で対称群 (置換群) をなす。順位ベクトルと順序ベクトルの対称群をそれぞれ  $\mathcal{S}_m$  と  $\mathcal{T}_m$  と記す。 $\pi \in \mathcal{S}_m$  を表す順位ベクトルは  $[\pi(1), \pi(2), \dots, \pi(m)]$  であり、順序ベクトルは  $[\pi^{-1}(1), \pi^{-1}(2), \dots, \pi^{-1}(m)]$  となる。ただし、 $\pi(i)$  は対象  $x_i$  の順位である。順位ベクトルに対し、 $\eta \in \mathcal{S}_m$  を右から掛ける演算は対象の番号の付け替えに、左から掛ける演算は順位の入れ替えに相当する。例えば、 $x_2 \succ x_1 \succ x_3$  を表す順位ベクトル  $\pi = [2, 1, 3]$  に、右から  $\eta = [3, 1, 2]$  を掛けると  $\pi\eta = [1, 3, 2]$  が得られる。これは  $x_1 \rightarrow x'_3, x_2 \rightarrow x'_1, x_3 \rightarrow x'_2$  と対象の番号を付け替えて  $x'_1 \succ x'_3 \succ x'_2$  にする操作に該当する。一方で、この  $\pi$  に、左から  $\eta$  を掛けると  $\eta\pi = [3, 2, 1]$  が得られ、順序  $x_3 \succ x_2 \succ x_1$  が得られる。

以後、2節では順序間の距離について、3節では順序の確率的生成モデルについて、4節では不完全な順序の扱いについて、5節ではこれらの距離や確率モデルの応用問題について述べる。

\*連絡先: <http://www.kamishima.net/>

## 2 順序間の距離

同じ長さの二つ順序間の類似度を測る距離をいくつか紹介する。順序間の距離関数

$$d(\cdot, \cdot): \mathcal{S}_m \times \mathcal{S}_m \rightarrow \mathbb{R}$$

は、(1) 対称性、(2) 正定値性、(3) 三角不等式に加え<sup>1</sup>、ラベルの付け替えに対する不変性をもつ。形式的には、順位ベクトルで表された二つの順序  $\pi, \xi \in \mathcal{S}_m$  について、ラベルの置換  $\eta \in \mathcal{S}_m$  があるとき次式を満たす。

$$d(\pi, \xi) = d(\pi\eta, \xi\eta)$$

この性質は**ラベル不変性**や群論の用語で右不変性と呼ばれる。対称群の単位元  $e_m = [1, 2, \dots, m]$  を考えると、この右不変性により  $d(\pi, \xi) = d(\pi\xi^{-1}, e_m)$  となり、 $e_m$  から距離だけを考えればよい。以下、順序間の距離を順位ベクトル型と編集型に分けて述べたあと、順位相関係数と順序間の不等式を紹介する。

### 2.1 順位ベクトル型

長さ  $m$  の順序を表した順位ベクトルを、 $m$  次元空間中の点を表すベクトルとみなして距離を定義する。二つの順位ベクトル  $\pi, \xi \in \mathcal{S}_m$  間の Minkovski 距離は次式である。

$$\left( \sum_{i=1}^m |\pi(i) - \xi(i)|^p \right)^{1/p}$$

$p = 2$  で Euclid 距離、 $p \rightarrow \infty$  で Max 距離 (Chebyshev 距離) というのは通常のベクトル間の距離と同じである。だが、 $p = 1$  の場合は **Spearman footrule**、2 乗 Euclid 距離の場合は **Spearman 距離** と呼び、それぞれを  $d_{\text{Foot}}$  と  $d_{\text{Spear}}$  と記す。

その他、二つの順位ベクトルの間で、同じ要素の内容が一致していないもの数を Hamming 距離と呼ぶ。例えば  $[4, 2, 1, 3]$  と  $[4, 3, 2, 1]$  の間では 1 番目の 4 以外は一致していないので、Hamming 距離は 3 となる。

### 2.2 編集型

何らかの一連の操作を一方の順序に適用し、もう一方の順序に変換するのに必要な最小操作数を編集距離という。最もよく用いられるのは、隣接する対象を交換する操作を用いた **Kendall 距離** ( $d_{\text{Ken}}$  と記す) である。また、隣接してなくてもよい任意の対象対を交換する操作を用いるものを **Cayley 距離** ( $d_{\text{Cay}}$  と記

<sup>1</sup>Spearman 距離は三角不等式を満たさないが慣習的に距離として扱われる

す) と呼ぶ。Ulam 距離での操作は「本棚の本の入れ替え」に例えられる。あるの順序で本棚に並んでいる本を、目的の順序に並び替えるとする。このとき、本を一冊取り出し、目的の順序の正しい本の位置を押し開けるように他の本をずらし、そこに取り出した本を収める。  $x_2 \succ x_5 \succ x_3 \succ x_1 \succ x_4 \succ x_6$  を  $x_1 \succ x_2 \succ x_3 \succ x_4 \succ x_5 \succ x_6$  に変える例を示す。まず、 $x_1$  を取り出し  $x_2, x_5, x_3$  を右にずらして  $x_1$  を正しい先頭の位置に挿入すると  $x_1 \succ x_2 \succ x_5 \succ x_3 \succ x_4 \succ x_6$  になる。次は、 $x_5$  を取り出して、 $x_3, x_4$  を左にずらした後に、5 番目に挿入すれば目標の順序になる。この例では、2 回の操作で目標の順序になったので、Ulam 距離は 2 となる。

これらの距離について別の解釈を示しておく。対象  $x_i$  と  $x_j$  が、二つの順序中  $\pi$  と  $\xi$  中で**同順 (concordant)** であるとは、

$$(\pi(i) - \pi(j))(\xi(i) - \xi(j)) \geq 0$$

が成立すること。そうでないとき**逆順 (discordant)** であるという。  $m(m-1)/2$  個の全ての対象対のうち、逆順になっている対の数は Kendall 距離に等しい。また Ulam 距離は、二つの順序間で同じ順番に並んでいる対象の極大な系列、すなわち、最長共通部分列の長さを順序長  $m$  から引いた値でもある。

### 2.3 順位相関と不等式

まず Spearman と Kendall の順位相関係数について述べる。Spearman の順位相関係数  $\rho$  は、Spearman 距離を  $[-1, 1]$  の範囲に正規化したものである。

$$\rho = 1 - \frac{6d_{\text{Spear}}(\pi, \xi)}{m^3 - m}$$

これは、二つの順位ベクトル  $\pi$  と  $\xi$  の Pearson 相関係数に等しい。次の統計量は、二つの順序が無相関という帰無仮説の下で、自由度  $m-2$  の Student の  $t$  分布に従うことが知られている。

$$\frac{\rho\sqrt{m-2}}{\sqrt{1-\rho^2}}$$

**Kendall の順位相関係数**も同様に、Kendall 距離を  $[-1, 1]$  の範囲に正規化したものである。

$$\tau = 1 - \frac{4d_{\text{Ken}}(\pi, \xi)}{m(m-1)}$$

これも無相関の帰無仮説の下で平均が 0、分散が  $\frac{2(2m+5)}{9m(m-1)}$  の正規分布に従うことが知られている。

次に、これまでで紹介した距離や相関係数の間に成立する不等式を紹介する。順位相関  $\tau$  と  $\rho$  の間には次の **Daniels の不等式**が成立する

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

これは、 $m \rightarrow \infty$  の極限において  $-1 \leq 3\tau - 2\rho \leq 1$  となり、二つの相関係数の間に高い相関があることを示している。**Diaconis-Graham の不等式** も著名である。

$$d_{\text{Ken}} + d_{\text{Cay}} \leq d_{\text{Foot}} \leq 2d_{\text{Ken}}$$

これにより、 $d_{\text{Ken}}$  と  $d_{\text{Foot}}$  の差はたかだか2倍となり、距離としての振る舞いはほぼ等しいことが分かる。他に **Durbin-Stuart の不等式** もある。

$$d_{\text{Spear}} \geq \frac{4}{3} d_{\text{Ken}} \left( 1 + \frac{d_{\text{Ken}}}{m} \right)$$

なお、本節の事項に関する詳細は [38] や [42] の 2.5 節などを参照されたい。

### 3 順序の確率的生成モデル

ここでは順序の確率的生成モデルを概観する。なお、本節の項目については [42, 14] を参考にされたい。

$m$  個の対象を含む順序  $O$  は全部で  $m!$  個ある。この順序の分布で、最も一般的な飽和モデルは  $m! - 1$  個のパラメータをもち、これらのパラメータは  $m! - 1$  次元の単体上に分布する。また、検定の帰無仮説などとして用いられる均一分布は、全ての順序  $O$  に対し  $\Pr[O] = 1/m!$  となる。これら  $m! - 1$  個のパラメータ間に、統計分野の分割表のような制約を考えるモデルもあるが、ここでは順序の生成過程と結びついたモデルを紹介する。順序の生成モデルは次ようにの四種類に分類でき、これらを順次紹介する。

- **Thurstone 型 (Thurstonian)** : 各対象に付随した確率分布に従って発生したスコアの順に、対象を整列する。
- **一対比較型 (paired comparison)** : 対ごとに対象を比較して、その比較結果に基づいて全対象を整列する。
- **距離ベース型 (distance-based)** : モード順位付けからの距離で決まる確率分布に従って順序を生成する。
- **多段階型 (multistage)** : 先頭から末尾に向かって対象を逐次的に整列する。

#### 3.1 Thurstone 型 (Thurstonian)

最初に提案した Thurstone [52] にちなんで Thurstone 型と呼ばれるこのモデルは、順序統計量モデル (order statistics model) と呼ばれる。このモデルで

は、対象  $x_i \in \mathcal{X}_m$  を、それに付随するスコアの順に整列する。このスコアは実数で、次式で表される。

$$\text{score}(x_i) = \mu(x_i) + \epsilon(x_i)$$

ただし、 $\mu(x_i)$  は対象  $x_i$  の平均スコアであり、 $\epsilon(x_i)$  は平均 0 の確率分布に従う確率変数である。Thurstone が提案したモデルでは  $\epsilon(x_i)$  には正規分布が使われた。特に、全ての対象で分布の分散が等しく、互いの分布の相関が等しいモデルは **Thurstone の比較判断の法則タイプ V (Thurstone's law of comparative judgment, case V)** と呼ばれ、広く使われている。この場合、対象  $x_i$  が  $v_j$  より前になる確率は次式で計算される [44]。

$$\begin{aligned} \Pr[x_i \succ x_j] &= \Pr[\text{score}(x_i) > \text{score}(x_j)] \\ &= \Phi\left(\frac{\mu(x_i) - \mu(x_j)}{\sqrt{2\sigma}}\right) \end{aligned}$$

ただし、 $\Phi(\cdot)$  は正規分布の累積密度関数で、 $\sigma^2$  は分散である。他に、 $\epsilon(x_i)$  の分布として、次の累積密度関数をもつ Gumbel 分布 (二重指数分布) も利用される。

$$F(x) = 1 - \exp\left(-\exp\left(\frac{x}{\sigma_i}\right)\right)$$

スケールパラメータ  $\sigma_i$  が全ての対象  $x_i$  において等しいとき、この Gumbel 分布を用いた Thurstone 型モデルは、後に述べる Plackett-Luce モデル等価であることが知られている。

#### 3.2 一対比較型 (paired comparison)

一対比較モデルでは、対象の各対について、どちらの対象が前になるかを最初に決め、 $m(m-1)/2$  個の順序対を生成する。もしこの順序対の集合に循環がある、すなわち、対象  $x_i, x_j, x_k$  間に  $(x_i \succ x_j) \wedge (x_j \succ x_k) \wedge (x_k \succ x_i)$  のような関係があれば、全ての対象対を破棄し、順序対の生成をやり直す。もし順序対の集合が非循環であれば、すなわち、Condorcet 選択であれば  $m$  個の対象は一意に整列することができ、順序が生成される。この一対比較型モデルの飽和モデルには  $m(m-1)/2$  個のパラメータがあり、これらのパラメータは対象  $x_i$  が  $x_j$  より前にある度合いを表す。最初にモーメントを計算した寄与により、この飽和モデルは **Babington Smith モデル** [3] と呼ばれる。

飽和モデルよりパラメータの少ないモデルとしては、 $m$  個のパラメータをもつ **Bradley-Terry モデル** [4] がある。

$$\Pr[x_i \succ x_j] = \frac{f(x_i)}{f(x_i) + f(x_j)}$$

ただし、 $f(\cdot)$  は正実関数。Bradley と Terry の寄与は対の生成だけだったため、順序の生成モデルとしては Bradley-Terry/Mallows モデルと呼ぶこともある。

Mallows は他にも、パラメータ数が 2 個だけの **Mallows モデル** [41] も提案している。対象  $x_i$  と  $x_j$  の順位をそれぞれ  $\pi(i)$  と  $\pi(j)$  で表すと、Mallows モデルは次のように定義される：

$$\Pr[x_i \succ x_j] = \frac{\theta^{\pi(j)-\pi(i)} \phi}{\theta^{\pi(j)-\pi(i)} \phi + \theta^{\pi(i)-\pi(j)} \phi^{-1}}$$

ただし、 $\theta$  と  $\phi$  は正実数のパラメータ。

### 3.3 距離ベース型 (distance-base)

距離ベース型モデルで、順序  $O$  が生成される確率分布は次式：

$$\Pr[O] = \frac{1}{Z(\lambda)} \exp(-\lambda d(O_0, O))$$

ただし、 $\lambda$  は非負実数の集中度 (concentration) パラメータ、 $O_0$  はモード順位付け (modal ranking)、および  $Z(\lambda)$  は正規化定数である。確率質量が最も大きな順序は  $O_0$  であり、このモード順位付けからの距離  $d(O_0, O)$  の増加に伴って低確率で生じるようになる。距離  $d(\cdot, \cdot)$  には 2 節の任意の距離が適用可能である。特に、Kendall 距離と Spearman 距離を用いた場合は、Mallows モデルで  $\theta = 1$  と  $\phi = 1$  として消した場合と等価になるため、それぞれ Mallows $\phi$  モデルと Mallows $\theta$  モデルと呼ばれる。

### 3.4 多段階型 (multistage)

まず **Plackett-Luce モデル** [45] を紹介する。順序  $O = x_{i_1} \succ x_{i_2} \succ \dots \succ x_{i_m}$  が生成される確率を考える。このモデルでは、最上位の対象  $x_{i_1}$  が選ばれる確率は、実関数  $f(x)$  を用いて

$$\frac{f(x_{i_1})}{\sum_{x \in \mathcal{X}_m} f(x)}$$

となり、第 2 位の対象  $x_{i_2}$  が選ばれる確率は

$$\frac{f(x_{i_2})}{\sum_{x \in \mathcal{X}_m, x \neq x_{i_1}} f(x)}$$

となる。第 3 位から第  $m-1$  位までも同様に、すでに整列が済んだ対象の要素と除いた和を分母とする。順序  $O$  が生成される確率は、これらの確率の積となる。

この手続きで生成された順序は **Luce の選択公理 (Luce's choice axiom)** を満たす。これは、ある対象が最上位になる確率は、その対象がある対象の部分集合に含まれる確率と、その対象がその部分集合中で最上位になる確率の積になるという条件である。より形式的に、ある対象  $x$  が対象集合  $\mathcal{X}$  中で最上位にな

る確率を  $\Pr_{\mathcal{X}}[x]$  とし、 $\Pr_{\mathcal{X}'}[x]$ 、 $\mathcal{X}' \subset \mathcal{X}_m$  を  $\mathcal{X}_m$  中の最上位対象が部分集合  $\mathcal{X}'$  に含まれる確率とする。選択確率が Luce の選択公理を満たすとは、 $\forall \mathcal{X}' \subset \mathcal{X}_m$  について少なくとも 2 個の対象  $x_i$  と  $x_j$  が次の条件を満たすことである。

- If  $\Pr_{\{x_i, x_j\}}[x_i] > 0 \forall x_i, x_j \in \mathcal{X}$   
then  $\forall x_k \in \mathcal{X}' \subset \mathcal{X}, \Pr_{\mathcal{X}}[x_k] = \Pr_{\mathcal{X}'}[x_k] \Pr_{\mathcal{X}}[\mathcal{X}']$
- If  $\Pr_{\{x_i, x_j\}}[x_i] = 0 \exists x_i, x_j \in \mathcal{X}$   
then if  $x_j \in \mathcal{X}, x_j \neq x_i, \Pr_{\mathcal{X}}[x_j] = \Pr_{\mathcal{X} \setminus \{x_i\}}[x_j]$

すでに述べたように、このモデルは Gumbel 分布を用いた Thurstone 型モデルと等価である。

もう一つの多段階モデルとして  **$\phi$ -component モデル** [19, 20] を紹介する。順序  $O$  を、 $x_i$  が  $i$  番目である順序  $e_m = x_1 \succ x_2 \succ \dots \succ x_m$  に、Ulam 距離と同様の手順で並び替えることを考える。例えば、順序  $O = x_3 \succ x_4 \succ x_1 \succ x_2$  の  $x_1$  を先頭に動かすと  $x_3$  と  $x_4$  の 2 つがずれて  $x_1 \succ x_3 \succ x_4 \succ x_2$  となる。このときずれた個数を  $u_1 = 2$  とする。同様に、 $x_2$  と  $x_3$  も動かすとそれぞれ 2 個と 0 個の対象を移動させることになる。これらをまとめた  $(u_1, u_2, u_3) = (2, 2, 0)$  に基づいて順序  $O$  が生じる確率を次式で定める。

$$\Pr[O] = \prod_{i=1}^{m-1} \exp(\theta_i u_i - \phi(\theta_i; m - i + 1))$$

$$\phi(\theta; q) = \log \left( \frac{1 - \exp(\theta q)}{q(1 - \exp(\theta))} \right)$$

パラメータ  $\theta_i$  が  $i = 1, \dots, m-1$  で全て等しいとき、このモデルは Mallows $\phi$  モデルと等価になる。このため、 $\phi$ -component モデルは一般化 Mallows モデルとも呼ばれる。

## 4 不完全な順序

ここまでは対象集合  $\mathcal{X}_m$  中の全ての対象間に全順序が定義されている場合を扱った。すなわち、 $\mathcal{S}_m$  中のただ一つの要素として指定されていた。だが、 $\mathcal{X}_m$  の部分集合にしか順序関係が分からない場合や、一部の対象については順序関係が明確でなく同順位がある場合などがある。こうした順序を、この順序と無矛盾な全順序全体を含むような  $\mathcal{S}_m$  の部分集合をもって表現する。これを **不完全順序 (不完全順位付け; incomplete rankings)** と呼ぶ。例えば、3 個の対象のうち、 $x_3$  が最下位であることは分かっているが、 $x_1$  と  $x_2$  の間には順序関係はない場合、不完全順序は集合  $\{[1, 2, 3], [2, 1, 3]\}$  となる。

不完全順序  $\mathcal{R}$  と  $\mathcal{Q}$  の間の距離を考えよう。単純に要素間の距離の算術平均を取ってみる。

$$d(\mathcal{R}, \mathcal{Q}) = \frac{1}{|\mathcal{R}||\mathcal{Q}|} \sum_{\pi \in \mathcal{R}} \sum_{\xi \in \mathcal{Q}} d(\pi, \xi)$$

しかし、これでは  $d(\mathcal{R}, \mathcal{R}) \neq 0$  となる場合があり、半正定値性を満たさなくなる。そこで次の **Hausdorff 距離** が利用される。

$$d(\mathcal{R}, \mathcal{Q}) = \max\left\{\max_{\pi \in \mathcal{R}} d^*(\pi, \mathcal{Q}), \max_{\xi \in \mathcal{Q}} d^*(\mathcal{R}, \xi)\right\}$$

ただし、 $d^*(\pi, \mathcal{R})$  は距離の最小値  $\min_{\xi \in \mathcal{R}} d(\pi, \xi)$  である。その他、不完全順序に含まれるとき 1、そうでないとき 0 をとる長さ  $m!$  のベクトルを使う方法もある。

不完全順序のうち同順位を扱う方法として **midrank** がある。  $x_2$  と  $x_3$  が同順位であることを  $x_1 \succ x_2 \sim x_3 \succ x_4$  と記す。このとき、日常生活では  $x_2$  と  $x_3$  に共に順位 2 を与えることが多い。midrank では、同順位の対象の、全ての可能な並べ変えを考えたときの平均順位とする。この例では  $x_1 \succ x_2 \succ x_3 \succ x_4$  と  $x_1 \succ x_3 \succ x_2 \succ x_4$  が考えられる。すると、 $x_2$  は順位は 2 か 3 なので、これらの平均 2.5 を midrank とする。また、midrank を用いた  $x_1 \succ x_2 \sim x_3 \succ x_4$  の順位ベクトルは  $[1, 2.5, 2.5, 4]$  となる。

群論の枠組みを利用する方法もある。  $m$  個のうち、特定の順位にある  $k$  個対象が何であるか分かっている状況を考える。特に有用なのは、最上位  $k$  位の対象は分かっているが、  $k+1 \sim m$  位の対象は不明である場合で、**top- $k$  順位付け (top- $k$  ranking)** という。その他、最上位 3 個と最下位 3 個は分かっているといった状況も考えられるだろう。これら  $k$  個の対象が等しく、他は異なる場合を同値類とみなすと、これらは不完全順序になる。この特殊な不完全順序間の 2 章の距離を用いた Hausdorff 距離は容易に計算できる [13, 3 章]。例えば、  $\pi$  と  $\xi$  で共に上位  $k$  位以内にある対象の集合を  $A$ 、  $\pi$  でのみ  $k$  位以上のものを  $B$ 、  $\xi$  のみのものを  $C$  とし、  $h = |B|$ 、そして  $A$  中の対象対で、  $\pi$  と  $\xi$  に対し逆順な対象対数を  $d_{\text{Ken}}(A)$  とすると Kendall 距離は次式になる。

$$d_{\text{Ken}}(A) + h(m+k - \frac{h-1}{2}) - \sum_{i \in B} \pi(i) - \sum_{i \in C} \xi(i)$$

一方、Spearman 距離は次式になる。

$$\begin{aligned} & \sum_{i \in A} (\pi(i) - \xi(i))^2 + h^2(m-k-h) \\ & + \max\left\{\sum_{j=1}^h (m+1-j-\pi_B(j))^2 + \sum_{j=1}^h (k+j-\xi_C(j))^2, \right. \\ & \left. \sum_{j=1}^h (k+j-\pi_B(j))^2 + \sum_{j=1}^h (m+1-j-\xi_C(j))^2\right\} \end{aligned}$$

ただし、  $\pi_B(i)$  は  $B$  中の要素を順序  $\pi(i)$  と同順に整列した順序中の対象  $x_i$  の順位、  $\xi_C(i)$  も同様。

Kendall や Spearman の順位相関は順位統計量と呼ばれる。一方で、ある母集団からサンプリングした標本集合で、母集団中のある事例が第何位になるかを扱うのが順序統計である [15, 2]。観測されない母集団に  $m$  個の対象があり、これらの対象には全順序が与えられているとしよう。この順序から  $m'$  個の事例を均一にランダムにサンプリングし、元の全順序と同じ順序に並べる。標本の順序中で  $k'$  位の対象が母集団中の順序で  $k$  位である確率は、超幾何分布を用いて次式で表せる。

$$\Pr[k|k'; m, m'] = \binom{k-1}{k'-1} \binom{m-k}{m'-k'} / \binom{m}{m'}$$

さらに、  $k = 1, \dots, m$  について平均をとると、母集団中での順位の期待値が次式で得られる。

$$E[K|k'; m, m'] = k' \frac{m+1}{m'+1}$$

また、分散は次式になる。

$$\text{Var}[K|k'; m, m'] = \frac{k'(m'-k'+1)(m+1)(m-m')}{(m'+1)^2(m'+2)}$$

サンプル順序中で  $j' < k'$  という二つの順位が観測されたとき、元の順序中での共分散は

$$\text{Cov}[J, K|j', k'; m, m'] = \frac{j'(m'-k'+1)(m+1)(m-m')}{(m'+1)^2(m'+2)}$$

よって、サンプル順序中で  $j'$  と  $i'$  位になる対象の相関係数は

$$\rho[J, K] = \sqrt{\frac{j'(m'-k'+1)}{k'(m'-j'+1)}}$$

母集団の順序からの欠損が均一であるなどの仮定という欠点はあるが、群論の場合とは違って観測される順位が固定してはいなくてもよい利点がある。期待値と置き換えるのはやや強引ではあるが、我々の実験では少ない計算量でよい結果が得られた [31, 32, 34, 35]。

なお、これらの話題については [42] の 11 章や、[13] を参考にされたい。

## 5 順序の距離や確率モデルの応用

最後に、順序の距離や確率モデルを用いた応用を幾つか紹介する。

距離の応用として代表的なものに順序の中心の概念がある。これは  $n$  個の順序の集合  $\{O_1, O_2, \dots, O_n\}$  が与えられたとき次式で定義される。

$$\bar{O} = \min_O \sum_{i=1}^n d(O, O_i)$$

すなわち、 $n$  個の順序と平均的に矛盾が最も少ない順序のことである。この問題は統計の分野で扱われてきたが、最近では情報検索の分野でも rank aggregation [17, 51] と呼ばれて研究されている。また、社会選択の分野とも関連がある。最適な解は一般には NP 困難だが、Spearman 距離の場合にのみ効率的に計算可能である。アイテム  $x_j$  の順序  $O_i$  中での順位を  $\pi_i(j)$  と書き、アイテム  $x_j$  の平均順位を  $\bar{\pi}(j) = (1/n) \sum_i \pi_i(j)$  とする。この平均順位のが小さな順に対象を整理した順序は、Spearman 距離を用いた順序の中心であることが容易に示せる。Spearman 距離は他にも、 $n$  個の順序までの平均距離が容易に計算できるという性質も持つ。 $i$  番目の要素が対象  $x_i$  の平均順位である平均順位ベクトルを  $\bar{\pi}$  とすると、順位ベクトル  $\xi$  で表される順序からの平均距離は次式になる。

$$\frac{m(m-1)(2m+1)}{3} - 2\xi^\top \bar{\pi}$$

なお、この手順は Borda Count [16] と呼ばれる投票の集計手順と等価である。不完全順序に対する中心を求める研究に [39, 40, 18] などがある。さらに、中心が定義されたことによりクラスタリングも可能になる [33, 32, 7]。

統計の分野では順位相関や、順序の確率モデルは検定を中心の用いられてきた。だが、これらは予測のためにも用いられるようになってきており、次のような問題が研究されている。

**順序回帰 (Ordinal Regression)** 順序付きカテゴリを従属変数、特徴ベクトルを独立変数とする回帰問題である [43, 1]。統計分野の初期の研究ではロジスティック回帰などが用いられたが、最近ではパーセプトロンや SVM などいろいろな手法が適用されている [47, 12, 9, 6, 48, 49]。

**対象順位付け (Object Ranking)** 順序を従属変数、特徴ベクトルを独立変数とする回帰問題である [35, 29]。[34, 30, 10, 11, 27, 28, 46, 21, 22, 37, 36, 8] といった手法がある。順序回帰より一般的な問題なので、対象順位付け手法で順序回帰を解くことはできる。しかし、対象順位付け手法で順序回帰問題を解くのは一般には非効率的である。

**ラベル順位付け (Label Ranking)** 分類問題で多クラスの場合を考え、最も分類するのにふさわしいラベルから、そうでないラベルへ順位付けをする問題である [25, 23, 26, 5]。一つの対象に複数のラベルを割り当てることが可能なマルチラベル問題は、上位のラベルから適切な個数のラベルを選択することでも解くことができる。

なおこれらの予測問題については [24] などを参考にされたい。

以上、本稿では順序の距離と確率モデルを俯瞰し、それらの応用についてサーベイした。

## 参考文献

- [1] A. Agresti. *Categorical Data Analysis*. John Wiley & Sons, second edition, 1996.
- [2] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*. John Wiley & Sons, Inc., 1992.
- [3] B. Babington Smith. 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).
- [4] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs — i. the method of paired comparisons. *Biometrika*, Vol. 39, pp. 324–345, 1952.
- [5] K. Brinker. Active learning of label ranking functions. In *Proc. of The 21st Int'l Conf. on Machine Learning*, pp. 129–136, 2004.
- [6] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. In *Proc. of The 22nd Int'l Conf. on Machine Learning*, pp. 89–96, 2005.
- [7] L. M. Busse, P. Orbanz, and J. M. Buhmann. Cluster analysis of heterogeneous rank data. In *Proc. of The 24th Int'l Conf. on Machine Learning*, pp. 113–120, 2007.
- [8] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li. Learning to rank: From pairwise approach to listwise approach. In *Proc. of The 24th Int'l Conf. on Machine Learning*, pp. 129–136, 2007.
- [9] W. Chu and Z. Ghahramani. Gaussian processes for ordinal regression. *Journal of Machine Learning Research*, Vol. 6, pp. 1019–1041, 2005.
- [10] W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. In *Advances in Neural Information Processing Systems 10*, pp. 451–457, 1998.
- [11] W. W. Cohen, R. E. Schapire, and Y. Singer. Learning to order things. *Journal of Artificial Intelligence Research*, Vol. 10, pp. 243–270, 1999.
- [12] K. Crammer and Y. Singer. Pranking with ranking. In *Advances in Neural Information Processing Systems 14*, pp. 641–647, 2002.
- [13] D. E. Critchlow. *Metric Methods for Analyzing Partially Ranked Data*, Vol. 34 of *lecture Notes in Statistics*. Springer, 1980.
- [14] D. E. Critchlow, M. A. Fligner, and J. S. Verducci. Probability models on rankings. *Journal of Mathematical Psychology*, Vol. 35, pp. 294–318, 1991.
- [15] H. A. David and H. N. Nagaraja. *Order Statistics*. John Wiley & Sons, third edition, 2003.
- [16] J.-C. de Borda. On elections by ballot (1784). In I. McLean and A. B. Urken, editors, *Classics of Social Choice*, chapter 5, pp. 81–89. The University of Michigan Press, 1995.

- [17] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the Web. In *Proc. of The 10th Int'l Conf. on World Wide Web*, pp. 613–622, 2001.
- [18] R. Fagin, R. Kumar, M. Mahdian, D. Sivakumar, and E. Vee. Comparing and aggregating rankings with ties. In *Symposium on Principles of Database Systems (PODS)*, pp. 47–58, 2004.
- [19] M. A. Fligner and J. S. Verducci. Distance based ranking models. *Journal of The Royal Statistical Society (B)*, Vol. 48, No. 3, pp. 359–369, 1986.
- [20] M. A. Fligner and J. S. Verducci. Multistage ranking models. *Journal of The American Statistical Association*, Vol. 83, pp. 892–901, 1988.
- [21] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. In *Proc. of The 15th Int'l Conf. on Machine Learning*, pp. 170–178, 1998.
- [22] 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.
- [23] J. Fürnkranz and E. Hüllermeier. Pairwise preference learning and ranking. In *Proc. of the 14th European Conf. on Machine Learning*, pp. 145–156, 2003. [LNAI 2837].
- [24] J. Fürnkranz and E. Hüllermeier, editors. *Preference Learning*. Springer, 2010. [to appear].
- [25] J. Fürnkranz, E. Hüllermeier, W. Cheng, and K. Brinker. Label ranking by learning pairwise preferences. *Artificial Intelligence*, Vol. 172, pp. 1897–1916, 2008.
- [26] S. Har-Peled, D. Roth, and D. Zimak. Constraint classification for multiclass classification and ranking. In *Advances in Neural Information Processing Systems 15*, pp. 809–816, 2003.
- [27] 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.
- [28] T. Joachims. Optimizing search engines using click-through data. In *Proc. of The 8th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 133–142, 2002.
- [29] T. Kamishima. Invited talk: Object ranking. In *ECML/PKDD 2009 Workshop: Preference Learning*, 2009.
- [30] T. Kamishima and S. Akaho. Dimension reduction for supervised ordering. In *Proc. of The 6th IEEE Int'l Conf. on Data Mining*, pp. 330–339, 2006.
- [31] T. Kamishima and S. Akaho. Efficient clustering for orders. In *Proc. of The 2nd Int'l Workshop on Mining Complex Data*, pp. 274–278, 2006.
- [32] T. Kamishima and S. Akaho. Efficient clustering for orders. In D. A. Zighed, S. Tsumoto, Z. W. Ras, and H. Hacid, editors, *Mining Complex Data*, Vol. 165 of *Studies in Computational Intelligence*, chapter 15, pp. 261–280. Springer, 2009.
- [33] T. Kamishima and J. Fujiki. Clustering orders. In *Proc. of The 6th Int'l Conf. on Discovery Science*, pp. 194–207, 2003. [LNAI 2843].
- [34] T. Kamishima, H. Kazawa, and S. Akaho. Supervised ordering — an empirical survey. In *Proc. of The 5th IEEE Int'l Conf. on Data Mining*, pp. 673–676, 2005.
- [35] T. Kamishima, H. Kazawa, and S. Akaho. Object ranking — an empirical survey. In J. Fürnkranz and E. Hüllermeier, editors, *Preference Learning*. Springer, 2010. [to appear].
- [36] 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.
- [37] 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.
- [38] M. Kendall and J. D. Gibbons. *Rank Correlation Methods*. Oxford University Press, fifth edition, 1990.
- [39] 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.
- [40] G. Lebanon and J. Lafferty. Conditional models on the ranking poset. In *Advances in Neural Information Processing Systems 15*, pp. 431–438, 2003.
- [41] C. L. Mallows. Non-null ranking models. I. *Biometrika*, Vol. 44, pp. 114–130, 1957.
- [42] J. I. Marden. *Analyzing and Modeling Rank Data*, Vol. 64 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 1995.
- [43] P. McCullagh. Regression models for ordinal data. *Journal of The Royal Statistical Society (B)*, Vol. 42, No. 2, pp. 109–142, 1980.
- [44] 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.
- [45] R. L. Plackett. The analysis of permutations. *Journal of The Royal Statistical Society (C)*, Vol. 24, No. 2, pp. 193–202, 1975.
- [46] F. Radlinski and T. Joachims. Query chains: Learning to rank from implicit feedback. In *Proc. of The 11th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 239–248, 2005.
- [47] A. Shashua and A. Levin. Ranking with large margin principle: Two approaches. In *Advances in Neural Information Processing Systems 15*, pp. 961–968, 2003.
- [48] L. Shen and A. K. Joshi. Ranking and reranking with perceptron. *Machine Learning*, Vol. 60, pp. 73–96, 2005.
- [49] S. K. Shevade and W. Chu. Minimum enclosing spheres formulations for support vector ordinal regression. In *Proc. of The 6th IEEE Int'l Conf. on Data Mining*, pp. 1054–1058, 2006.
- [50] S. S. Stevens. On the theory of scales of measurement. *Science*, Vol. 103, No. 2684, pp. 677–680, 1946.
- [51] P.-N. Tan and R. Jin. Ordering patterns by combining opinions from multiple sources. In *Proc. of The 10th Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 695–700, 2004.
- [52] L. L. Thurstone. A law of comparative judgment. *Psychological Review*, Vol. 34, pp. 273–286, 1927.
- [53] 鷲尾隆, 元田浩. 尺度の理論. 日本ファジィ学会誌, Vol. 10, No. 3, pp. 401–413, 1998.