

# 多次元項目反応理論による短歌の評価傾向の分析

川島 寛乃<sup>1,a)</sup> 持橋 大地<sup>2,b)</sup>

**概要：**短歌は日本の伝統的な詩形の一つであり、これまで文学としての研究は行われてきているが、その評価について定量的な研究は行われていない。そこで本研究では、複数の短歌について複数の評価者が「良い-悪い」および「好き-嫌い」の軸で  $K$  段階の評価を付与した短歌の評価データに対して、多次元の項目反応理論に基づき、短歌の潜在座標および各評価者の評価傾向を分析する手法を提案する。多次元項目反応理論における潜在変数の次元数を変化させることで、短歌の評価に影響する要因の分解や、作者ごとの短歌の傾向の分析、評価者のグループ間の相違を統計的に比較することが可能になった。

**キーワード：**短歌、評価者モデル、段階評価、項目反応理論、潜在変数モデル

## A Statistical Analysis of Ratings for Japanese Short Songs using Multi-dimensional Item Response Theory

### 1. はじめに

短歌は5-7-5-7-7の韻律を基本とする日本の最も古典的な歌体の一つであり、国語教育で扱われるだけでなく、新聞や雑誌の歌壇、宮中行事の歌会始など、作品とその評価を日常的に目にする機会が多い文学の一つである。短歌では各作品が短いため、評価の際に1つの作品のみ、または1人の評価者のみが評価を行うのではなく、複数の作品に対して複数の評価者が評価を行うことが一般的である。例えば、参加者がそれぞれ自作の短歌を持ち寄って評し合う歌会では、自分が複数首の短歌を詠むだけでなく、参加者の相互評価によって複数首の歌を選びあう。このとき、基本的に短歌作品と評価者は、多数対多数の構造において評価が行われる。また多くの短歌賞では、複数の短歌を収めた歌集<sup>\*1</sup>または複数の短歌からなる作品群<sup>\*2</sup>に対して複数人の選者が評価を行う選考形式がとられている。新聞の歌壇欄においても、毎週応募される多数の短歌作品に対してそれぞれ数名の選者が複数の句を選び、評する形式が一般的である<sup>\*3</sup>。このように、短歌の評価を俯瞰的に分析する上では作品と評価者を一対一だけで考えるのではなく、複数の作品に対する複数の評価者の評価を総合的に考慮する必要がありと考えられる。

しかし、これまでの短歌に関する研究は、特定の作者や年

代の作品の傾向について、文学的に掘り下げる研究がその大半であった。その原因としてまず第一に、十分な量の短歌の評価データが研究資源として存在しないことが挙げられる。歌会ではその都度複数の評価がされるものの、これらはデータ化されておらず、各賞や歌壇での選は一部を除いてその詳細な過程が非公開である。また各選考で共通の評価指標が使われているわけではなく、評価自体にも作品の良し悪しの他に、各評価者の好みや評価の厳正さといった傾向が多分に含まれている。このため多対多の作品-評価データは十分な量を得るのが難しく、さらにその評価に関する分析手法も確立されていないのが現状である。

そこで、本研究ではまず多数多の短歌-評価者データとして、特定の大学の短歌会所属者を中心とする45名の評価者がそれぞれ100首の短歌に対して、7段階の数値で「良い-悪い」「好き-嫌い」の評価をした短歌評価データを収集した。本研究ではこの短歌評価の段階カテゴリーデータに対して、項目反応理論 (Item Response Theory; IRT) [3] の段階反応モデル [5] を用いた短歌作品および評価者の分析方法を提案する。項目反応理論は本来はテストの問題 (項目) とその受験者の回答 (反応) データに基づいてテストを分析・運用するためのテスト理論の1つであり、受験者のテスト項目への反応を受験者の特徴 (能力) を表す潜在変数と、項目の特徴を表す困難度や識別力といったパラメータで定義される確率モデルで表現する。特に本研究では、 $\theta$  を多次元化することで単純な優劣でない軸を抽出することを目的とする。本研究ではこの考え方に沿って、短歌評価データの各評価者を項目、各短歌への評価を反応とみなす。つまり、短歌に対する評価者の反応を、各短歌作品の特徴を表す潜在変数  $\theta$  と、評価者の特性を表すパラメータ  $a, b$  で定義される確率モデルで表現する。

<sup>1</sup> 慶應義塾大学大学院 政策・メディア研究科  
Graduate School of Media and Governance, Keio University

<sup>2</sup> 統計数理研究所  
The Institute of Statistical Mathematics

a) hironok@sfc.keio.ac.jp

b) daichi@ism.ac.jp

\*1 遼空賞、詩歌文学館賞、齋藤茂吉短歌文学賞など

\*2 角川短歌賞、短歌研究賞など

\*3 朝日歌壇、読売歌壇

実験では、上述の短歌評価データ（「良い-悪い」「好き-嫌い」）について提案手法を適用し、短歌作品の潜在的な優劣および各評価者の傾向について、短歌作品や評価者の属性情報も用いながら分析を行った。またより規模が大きく、疎な段階評価データとして、1,682 件の映画作品に対して 943 人の評価者が 5 段階の評価を実施している映画評価の公開データセット MovieLens 100K[2] を用いて同様の実験を行い、作品および評価ユーザに関して統計的な傾向を取り出すことができた。

## 2. 項目反応理論 (IRT)

項目反応理論 (Item Response Theory; IRT) [3] は、テストを作成・運用・評価するための数理モデルを用いたテスト理論の 1 つであり、テスト項目に対する受験者の回答 (反応) を、受験者の能力を表す潜在変数と項目の特性を表すパラメータで定義される確率モデルで表現する。IRT は (1) 異なる項目で構成されたテストを受験した場合も同一尺度上で能力を測定可能、(2) 個々の項目やテスト全体の能力測定精度を分析可能、(3) 欠測データの扱いが容易、などの利点があり [6]、国際コミュニケーション英語能力テスト (TOEIC)\*4 や IT パスポート試験\*5 など現代の様々な大規模試験において広く実用化されている。

項目反応理論では、基礎的な項目反応モデルではテスト項目に対する受験者の反応が正誤の二値の反応データを扱う。反応が  $C$  段階カテゴリー  $c \in \{1 \dots C\}$  の段階データである場合には、多値型項目反応モデルが用いられる。

### 2.1 項目反応モデル (2 値反応データ)

項目反応理論では、異なる能力・背景を持つ受験者集団におけるテスト結果の比較を可能にするために、受験者の集団に依存しない潜在的な特性  $\theta$  を導入する。 $\theta$  は直接観測できない個人の潜在特性 (latent trait) であり、受験者  $s_j$  の能力とテスト項目  $u_j$  の性質 (その項目が易しいのか、難しいのか) を表現するために、能力  $\theta$  の個人が問題  $j$  に正解する確率を

$$p_j(\theta) = \Psi(a(\theta - b_j)) \quad (1)$$

で与える。これは項目反応関数と呼ばれる。 $\theta$  は被験者母数、 $a, b_j$  は項目母数と呼ばれ、式 (1) は能力  $\theta$  が問題の難易度  $b_j$  より大きいと正解する確率が上がることを意味している。この際、 $a$  は確率の上がり方の急さを表現している。 $\Psi(x)$  には標準正規分布の累積確率密度関数

$$\Psi(x) = \int_{-\infty}^x N(t | 0, 1) dt \quad (2)$$

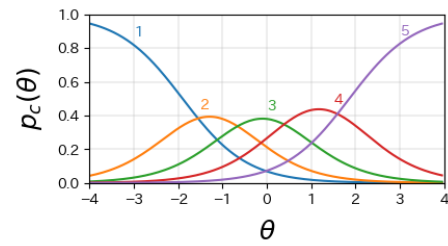
またはロジスティック関数

$$\Psi(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

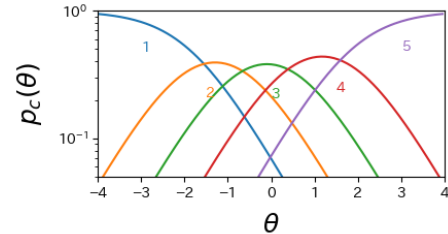
が用いられ、本研究では後者を採用した。

\*4 <https://www.iibc-global.org/toeic.html>

\*5 <https://www3.jitec.ipa.go.jp/JitesCbt/>



(a) 1次元の  $\theta$  の項目反応曲線。



(b) 縦軸の確率を対数で表示したもの。

図 1: 各段階への反応確率を表す  $\theta$  の項目反応曲線。

この関数を縦軸に  $p_j(\theta)$ 、横軸に  $\theta$  を配して描画した曲線 (図 1a) は項目特性曲線 (item characteristic curve; ICC) と呼ばれ、評価者の傾向を分析する際に用いられる。本研究では可視性向上のため、図 1a の項目反応曲線の縦軸を対数にして図 1b のようにしたものを図として用いる。

### 2.2 段階反応モデル (段階反応データ)

本研究では、短歌評価データとして段階反応データを扱うため、多値型の項目反応モデルとして知られる段階反応モデル [5] を用いる。

段階反応モデルでは、評価値が (評価値  $\geq c$ ) となる確率の  $\theta$  に関する関数を

$$p_c^*(\theta) = \Psi(a(\theta - b_c)) \quad (4)$$

と表す。このとき受験者  $s_j$  がテスト項目  $u_i$  に対してカテゴリ  $c \in \{1, \dots, C\}$  と反応する確率は

$$p(u_{j=c} | \theta) = p_c(\theta) = p_c^*(\theta) - p_{c+1}^*(\theta) \quad (5)$$

となる。ただし、

$$p_1^* = 1, p_{C+1}^* = 0 \quad (6)$$

である。

## 3. 短歌の段階評価データ

本研究では、短歌作品と評価者が多対多の関係になり、かつ各評価が段階カテゴリーの離散数値となる短歌評価データを扱う。段階カテゴリーの離散数値評価とは、例えば 5 段階の悪い-良いの反応データ「悪い (1-2-3-4-5) 良い」のような多値の段階的な評価のことをさし、評価者には最も良い悪いと思った場合には「1」、最も良いと思った場合には「5」をつけてもらうように指示する。データの一例を表 1 に示す。

このとき  $N$  件の短歌作品を  $\{s_1, \dots, s_N\}$ 、 $J$  人の評価

表 1: 短歌評価データの一例 (カテゴリ数  $C = 5$  の場合).  
短歌  $s_i$  に対する評価者  $u_j$  の評価を  $x_{ij}$  と表す.

短歌作品 \ 評価者	$u_1$	$u_2$	$u_3$	$\cdots$	$u_j$
$s_1$	3	4	5	$\cdots$	$x_{1j}$
$s_2$	1	3	4	$\cdots$	$x_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$s_i$	$x_{i1}$	$x_{i2}$	$x_{i3}$	$\cdots$	$x_{ij}$

者を  $\{u_1, \dots, u_j\}$  とし, 短歌  $s_i$  に対する評価者  $u_j$  の  $C$  段階カテゴリ  $c \in \{1, \dots, C\}$  の評価値を  $x_{ij}$  とする. 評価が存在しない場合は,  $x_{ij} = 0$  とする.  $N$  件の短歌作品に対する  $J$  人の評価者の評価データ全体を  $X = \{x_{ij}\} (i = 1 \dots N, j = 1 \dots J)$  とし, 本研究ではこれを作品-評価者数が多対多の段階評価データとして扱う.

#### 4. 多次元項目反応理論による 短歌の段階評価データの分析

本研究では項目反応理論の段階反応モデルを応用し, 前章で述べた表 1 のような, 作品-評価者が多対多である短歌評価データに対して, 作品の潜在的な特徴および評価者の傾向に関する分析を行う手法を提案する.

##### 4.1 短歌パラメータ $\theta_i$

前章で述べた通り, 通常の IRT ではテスト項目に対する受験者の反応データから, 各受験者の持つ潜在的な能力  $\theta$ , テスト項目  $j$  の特性  $a, b_j$  を推定する. 本研究では項目反応理論の基礎的な考え方を短歌の評価に適用するにあたり, 各短歌作品  $s_i$  ごとに短歌の潜在的な優劣 (能力) を表す  $\theta_i$  が存在すると考え, 表 1 のような段階評価データ  $X = \{x_{ij}\} (i = 1 \dots N, j = 1 \dots J)$  から  $\theta_i$  を推定する.

##### 4.2 評価者パラメータ $a_j, b_{jc}$

短歌の評価にあたり, 各評価者の評価は評価者の好み・嗜好に左右されるだけでなく, 「厳しい, 甘い」「ぶれやすい, ぶれにくい」などの評価の反応傾向にも影響を受けると考えられる. そこで各評価者の反応傾向を表現するために, 本研究では評価データ  $X$  より, 優劣  $\theta$  の短歌に対する評価者  $j (j = 1 \dots J)$  の各カテゴリ  $c \in \{1, \dots, C\}$  への反応曲線を

$$p_c^*(\theta) = \Psi(a_j(\theta - b_{jc})) \quad (7)$$

のように, パラメータ  $a_j, b_{jc} (c = 2 \dots C)$  で表現する. 反応の閾値  $b_{jc}$  は評価者  $j$  ごとおよび反応カテゴリ  $c$  ごとに異なるが, 曲線の傾き  $a_j$  は式 (5) の  $p_c^*(\theta) - p_{c+1}^*(\theta)$  の差が負になることを避けるために評価者  $j$  ごとに 1 つである.  $a_j$  は各評価者の判断のぶれの大きさに相当し,  $b_{jc}$  は評価者  $j$  が評価  $c$  を与える閾値に相当する.

##### 4.3 $\theta$ の次元数

$\theta$  の次元数を多次元化することによって,  $\theta$  によって表される各短歌の潜在的な「特徴」を複数の要素に分解して

表現することが可能になる. 次元数が高くなるほど複雑な要素を表現することが可能であるが, 一方で推定に要する計算量が増大し, また結果の解釈が難しくなるため, 通常の項目反応理論分析では  $\theta$  の次元数は 1 次元, または 2 次元に設定される. そこで本研究でもデータが大きくないこともあり, 分析時に,  $\theta$  の次元数を 1 次元および 2 次元  $\theta = (\theta_1, \theta_2)$  とする.

##### 4.3.1 $\theta$ が 1 次元の場合

$\theta$  が 1 次元のとき, 段階反応モデルで評価値が (評価値  $\geq c$ ) となる確率の  $\theta$  に関する関数は, 前章で述べた式 (4) となる.  $\theta$  を多次元とする場合に比べ, 1 次元の  $\theta$  は解釈がしやすく, 特に 1 つの作品に対して「良い-悪い」「好き-嫌い」など複数の評価値が得られている場合にその関係性を比較する際にも用いることができる. また, 評価者ごとに横軸に  $\theta$  の値, 縦軸に確率をとり, 図のように 5 の各評価値  $c$  への反応確率  $p_c(\theta)$  を表す項目反応曲線を描画し, 評価者の反応傾向を分析することが可能となる.

##### 4.3.2 $\theta$ が 2 次元の場合

$\theta$  が 2 次元のとき, 式 (4) は  $\eta_c = \mathbf{a}^T \mathbf{b}_c$  を用いて

$$p_c^*(\theta) = \Psi(\mathbf{a}^T(\theta - \mathbf{b}_c)) = \Psi(\mathbf{a}^T \theta + \eta_c) \quad (8)$$

となる. このとき  $p_c^*(\theta)$  は  $\theta$  の線形モデルであり, 重みベクトル  $\mathbf{w} = (w_1, w_2)$  を用いると

$$p_c^*(\theta) = \Psi(\mathbf{w}^T \theta + \eta_c) \quad (9)$$

と表現できる. 2 次元の各  $\theta_i = (\theta_1, \theta_2)$  および  $\mathbf{w}_j = (w_1, w_2)$  は図 2 のように図示し, 短歌作品や評価者の評価傾向を把握することができる.

式 (9) において, 学習された  $\mathbf{w}$  と  $\theta$  の内積  $\mathbf{w}^T \theta$  が大きいほど評価値は高くなる.  $\theta$  が 2 次元の場合の重みベクトル  $\mathbf{w}$  は,  $\theta$  が 1 次元の場合の  $a_j$  と同様に評価者  $u_j$  ごとに推定される.  $\mathbf{w}_j$  の極座標は

$$\mathbf{w}_j = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \sqrt{w_1^2 + w_2^2} \begin{pmatrix} \frac{w_1}{\sqrt{w_1^2 + w_2^2}} \\ \frac{w_2}{\sqrt{w_1^2 + w_2^2}} \end{pmatrix} = a_j \begin{pmatrix} \cos \phi_j \\ \sin \phi_j \end{pmatrix} \quad (10)$$

のようになり, 1 次元の場合と同様に識別パラメータ  $a_j$  と, 各評価者の評価方向  $\phi_j$  で表すことができる.

評価者の所属や性別などの属性情報が明らかである場合, 属性グループ  $t$  に含まれる評価者  $\{v_1, \dots, v_{m_t}\}$  の評価方向  $\{\phi_1, \dots, \phi_{m_t}\}$  より, グループごとに  $\phi_j$  の von Mises-Fisher 分布 (vMF 分布) [4] を描画して評価方向の傾向を数値的に比較することができる. vMF 分布は方向データに対する確率分布であり, 確率密度関数は

$$f(x|\mu, \kappa) = \frac{\exp(\kappa \cos(x - \mu))}{2\pi I_0(\kappa)} \quad (11)$$

となる. このとき  $\mu$  は平均方向,  $\kappa$  は集中度パラメータを表す.  $I_0(\kappa)$  は第 1 種変形ベッセル関数である.

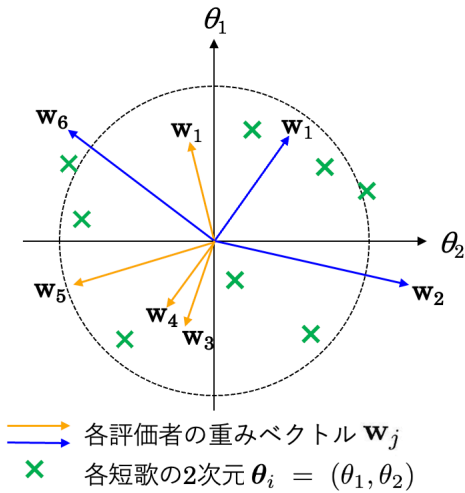


図 2: 各短歌の 2 次元  $\theta_i = (\theta_1, \theta_2)$ , 各評価者の重みベクトル  $\mathbf{w}_j$  の分布例.  $\mathbf{w}_j$  の色は評価者の属性の相違を表す.

#### 4.4 尤度関数

本研究では, 表 1 のような短歌評価データ  $X = \{x_{ij}\} (i = 1 \dots N, j = 1 \dots J)$  から,  $\theta_i, a_j, b_{jc}$  を推定する.  $x_{ij}$  は短歌  $s_i$  に対する評価者  $u_j$  のカテゴリ  $c \in \{1, \dots, C\}$  の評価値を表す. この時各短歌の  $\theta_i$  および各評価者の  $a_j, b_{jc}$  をまとめて  $\Theta = \{\theta_i\}_{i=1}^N$ ,  $\mathbf{a} = \{a_j\}_{j=1}^J$ ,  $\mathbf{b} = \{b_{jc}\} (i = j \dots J, c = 2 \dots C)$  とすると,  $X$  と  $\Theta$  の同時確率は

$$p(X, \Theta | \mathbf{a}, \mathbf{b}) = \prod_{i=1}^N p(\theta_i) \prod_{j=1}^J \prod_{c=1}^C p_c(\theta_i)^{\mathbb{I}(x_{ij}=c)} \quad (12)$$

となり, 対数尤度は

$$\begin{aligned} L &= \log p(X, \Theta | \mathbf{a}, \mathbf{b}) \\ &= \sum_{i=1}^N \left[ \log p(\theta_i) + \sum_{j=1}^J \sum_{c=1}^C \mathbb{I}(x_{ij}=c) \log p_c(\theta_i) \right] \\ &= \sum_{i=1}^N \left[ -\frac{1}{2} \theta_i^2 + \sum_{j=1}^J \sum_{c=1}^C \mathbb{I}(x_{ij}=c) \right. \\ &\quad \left. \log \left\{ \Psi(a_j(\theta_i - b_{jc})) - \Psi(a_j(\theta_i - b_{j(c+1)})) \right\} \right] \end{aligned} \quad (13)$$

となる. このとき 2 次元の  $\theta = \{\theta_1, \theta_2\}$  においては,  $\theta_1$  に全体的な優劣が反映されるようにするため,  $\mathbf{a}$  の事前分布として対数正規分布+vMF 分布

$$\log p(\mathbf{a}) = -(\log r)^2/2 + a_1/r \quad (r = \|\mathbf{a}\|_2) \quad (14)$$

を設定する. 本研究ではデータ  $X$  をもとに, 式 (13) を最大化するパラメータ  $\theta_i, a_j, b_{jc}$  を求める.

#### 4.5 パラメータの推定

項目反応理論では通常パラメータ  $\theta, a, b$  を推定する際に,  $\theta$  が正規分布  $N(0, 1)$  に従うことから,  $\theta$  を周辺化して

$$\begin{aligned} p(X|a, b) &= \int p(X, \theta|a, b) d\theta \\ &= \int p(X|\theta, a, b) p(\theta) d\theta \end{aligned} \quad (15)$$

を近似的に計算し, 式 (15) を最大化するパラメータ  $a, b$  を推定した上で潜在変数  $\theta$  を求める [1]. しかし, 点推定した不確定な  $a, b$  から  $\theta$  を復元するよりも, 直接  $\theta, a, b$  を同時に推定した方が良い推定値が求められるとの考えから, 本研究ではガウス分布のランダムウォークを用いた Metropolis-Hastings 法 (MCMC) により, 式 (15) を直接最適化する.

## 5. 分析 1: 短歌の評価

本章では, 前章で提案した項目反応理論に基づく手法を用いて短歌評価データを分析した結果について述べる.

### 5.1 短歌評価データ

本研究では短歌評価データとして, 100 件の短歌に対して 40 名の評価者が「悪い (1) – 良い (7)」, および「嫌い (1) – 好き (7)」をそれぞれ 7 段階の数値で評価した反応データを用いた. 本研究ではこれらのデータをそれぞれ短歌 good–bad 評価, 短歌 like–dislike 評価と呼び区別する. 各評価者は全ての短歌に対して必ず 2 種類の評価値を付与し, 評価数はそれぞれの評価で 4,000 件となった. また分析時に利用するために, 評価者には評価時に各短歌について思ったことや感想, 連想されたイメージなどを自由に記述してもらっている.

評価対象の短歌は多様な年代・作風・力量の作品を含めるために, 内訳は

- (1) 『桜前線開架宣言』 [11] (49 件)
- (2) その他の短歌集や教科書 (31 件)
- (3) 『「塔」若葉集』 (20 件)

の 3 つからなっている. (1) 『桜前線開架宣言』は歌人・山田航による現代の歌人 40 名の経歴および作品の解説書であり, 世間的に評価されている現代の名短歌である. (2) その他の短歌集や教科書は, (1) よりも古い時代の世間的に評価されている名短歌である. (3) の「塔」は短歌結社\*6であり, 若葉集は会誌の中で入会から 1 年と短歌暦の浅いメンバーが作品を掲載する欄である. そのため, (3) は (1), (2) よりも素人に近い人間の作った世間的には評価の定まっていない短歌である. (3) の作者は不明とする. 選出した短歌の一部を表 2 に示す.

評価者は京大短歌会および早稲田大学短歌会 (短歌会サークル) の所属者を中心に募り, 属性として (a) 所属短歌会, (b) 性別, (c) 年齢の情報を収集した. 本稿では個人情報保護のため, 評価者の年齢情報を「2X 歳」のように表記する. 各属性の人数の内訳を図 5 に示す.

### 5.2 $\theta$ が 1 次元の場合の結果

短歌 good–bad 評価, 短歌 like–dislike 評価データに対し

\*6 「塔」短歌会 <https://toutankakai.com/>

表 2: 短歌評価データに含まれる短歌, およびその属性情報の例.

作品	出典	作者
2x-5y=0 ピーチミントのガム噛みながら	(1) 桜前線開架宣言	野口あや子
三月の真ただ中を落ちてゆく雲雀、あるいは光の溺死	(1) 桜前線開架宣言	服部真里子
心沁む青山なりし夕日の村夕日みぬ方くだりきしかな	(2) その他の短歌集や教科書	山中智恵子
観覧車回れよ回れ想ひ出は君には一日我には一生	(2) その他の短歌集や教科書	栗木京子
何年も同じコメント年賀状「今年は飲もう」今年も書き足す	(3) 「塔」若葉集	-
行ってきますと自分に伝えたいまと自分を迎える単身赴任	(3) 「塔」若葉集	-

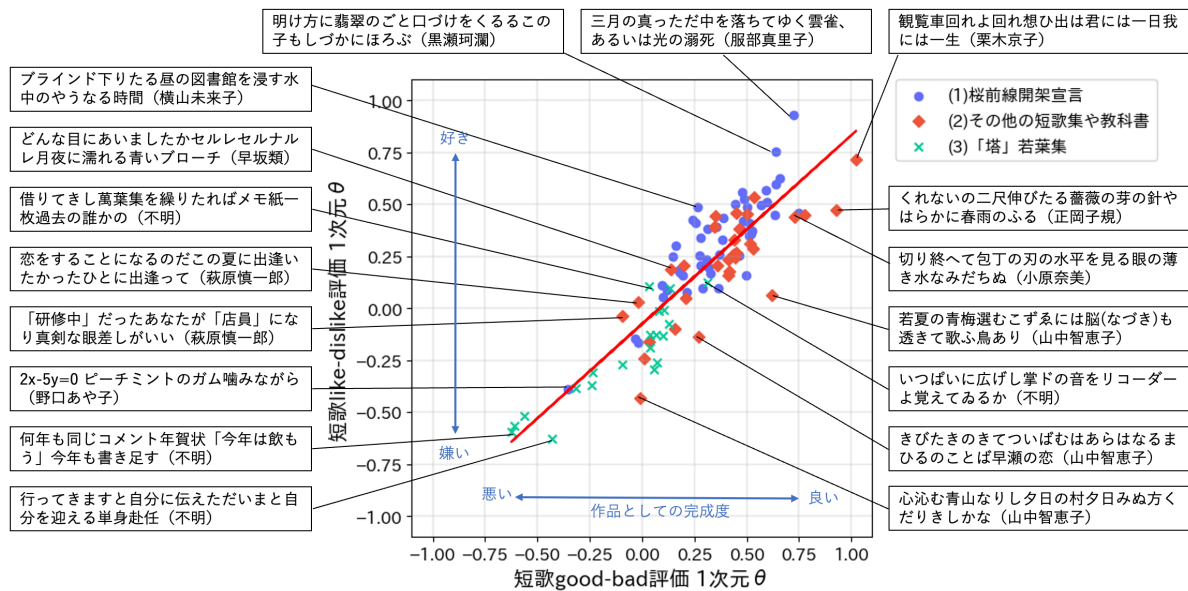


図 3:  $\theta$  が 1 次元の場合の短歌 good-bad 評価および短歌 like-dislike 評価データから推定された 100 首の短歌の  $\theta_i$  の値.

て前章で提案した多次元項目反応理論を用い, 1 次元の短歌パラメータ  $\theta$  および評価者パラメータ  $a, b$  を推定した結果について述べる. 短歌 good-bad 評価, 短歌 like-dislike 評価データにおいて, 各短歌  $s_i$  に対して推定した 1 次元の各  $\theta_i$  の値を横軸・縦軸に取り描画した結果を図 3 に示す.

### 5.2.1 good-bad 評価と like-dislike 評価の相関

結果の分析および解釈のために, 複数の点に対してその短歌内容および作者情報 (括弧の中に記載) を明記した. 図 3 のように, 短歌 good-bad 評価データから推定した  $\theta$  の値と, 短歌 like-dislike 評価データから推定した  $\theta$  の値には相関関係が見られ, 2 値間の相関係数は  $R = 0.88$  となった. また good-bad 評価, like-dislike 評価ともに, 著名な歌人の作品である (1) 『桜前線開架宣言』 (図 3, 青丸) および (2) その他の短歌集や教科書 (図 3, 赤菱形) から選出した短歌の  $\theta$  の値が大きく, 歌暦の浅い (3) 『塔』若葉集 (図 3, 緑×) の短歌は good-bad 評価, like-dislike 評価ともに  $\theta$  の値が小さい傾向が見られた.

### 5.2.2 good-bad 評価の 1 次元 $\theta$ 値について

good-bad 評価の  $\theta$  の値が最も大きくなった短歌は, 「観覧車回れよ回れ想ひ出は君には一日我には一生」 (栗木京子) [9], 次いで「くれなるの二尺伸びたる薔薇の芽の針やはらかに春雨のふる (正岡子規)」であった. 栗木京子 (1954 年-) は「読売新聞歌壇」選者, 現代歌人協会理事

長, 「塔」の選者などを務める京大短歌会出身の歌人であり, 「観覧車」の短歌は清新な青春歌を持ち味とする栗木京子の代表的な作品である. 正岡子規 (1867-1902 年) は明治を代表する歌人・俳人であり, 「くれなるの」の短歌は中学の教科書 [10] や資料集などでも引用される作品である. このように good-bad 評価の  $\theta$  値が大きくなった作品は, 短歌としての評価が確立されている, 短歌を嗜んでいる人間であれば一度は目にしたことのあるような著名な作品が多くなった.

一方で (1) 『桜前線開架宣言』, (2) その他の短歌集や教科書から選出した作品ながら good-bad 評価の  $\theta$  値が小さかった短歌として, 「 $2x - 5y = 0$  ピーチミントのガム噛みながら (野口あや子)」があった. この短歌は野口あや子 (1987 年-) の第一歌集 [7] に含まれる, 数式を含む斬新な作品であるが, 「関数」の必然性に疑問を呈するコメントが多く見受けられ, good-bad 評価値が低くなった. 同様に good-bad 評価の  $\theta$  値が小さかった短歌として, 「郊外のショッピングモールへ近づけば満州国にきた心地する」 (桜前線開架宣言, 松木秀) がある.

### 5.2.3 like-dislike 評価の 1 次元 $\theta$ 値について

like-dislike 評価の  $\theta$  の値が最も大きくなった短歌は, 「三月の真ただ中を落ちてゆく雲雀、あるいは光の溺死」 (服部真里子, 2014 年) [12], 次いで「明け方に翡翠のごと口づ

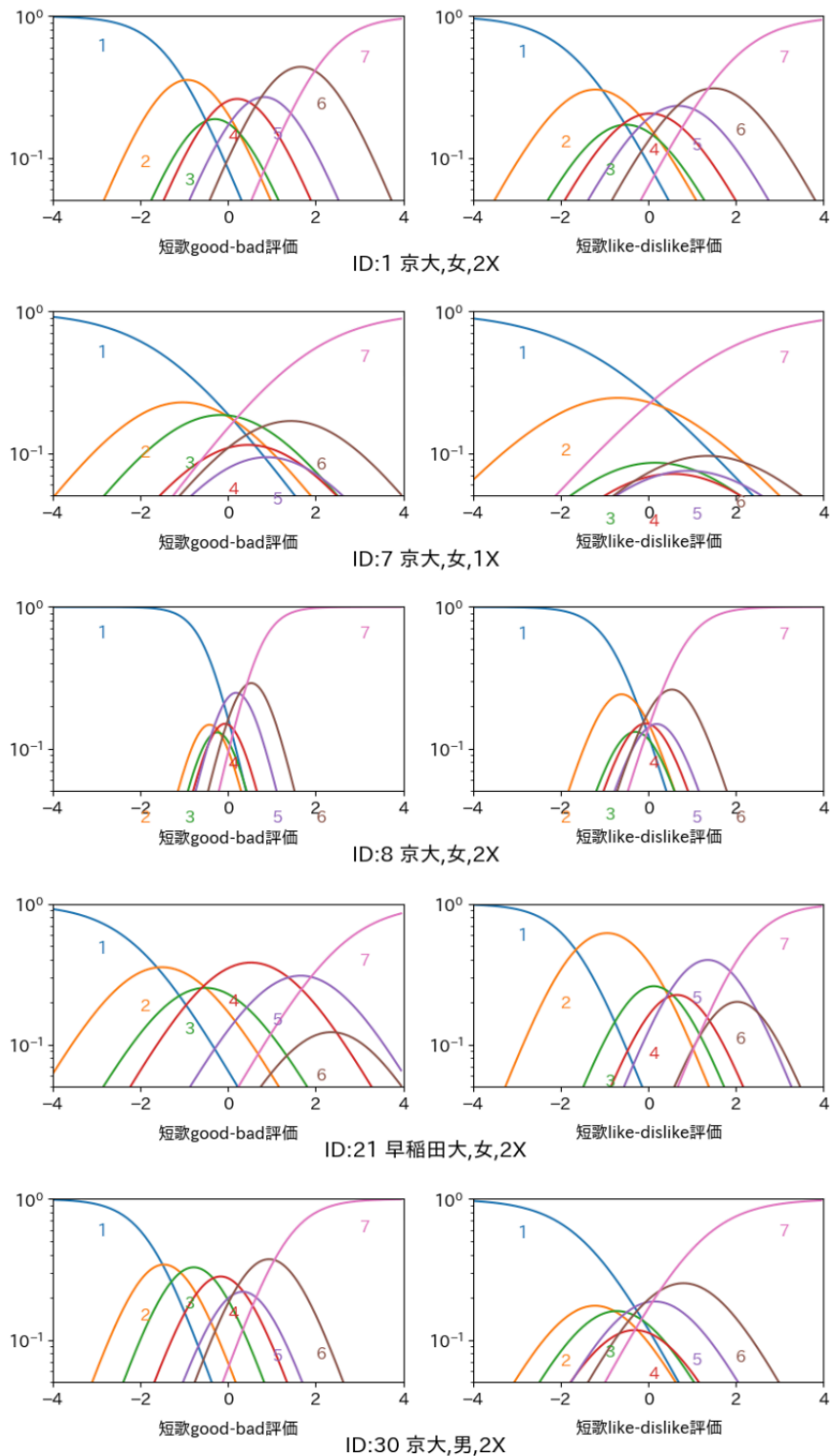


図 4: 各評価者の good-bad 評価 (左), like-dislike 評価 (右) の項目反応曲線の一例. 各グラフの横軸は  $\theta$  の値, 縦軸は評価値が  $c$  と反応する確率  $p_c(\theta)$  を表す.

けをくるるこの子もしづかにほろぶ」(黒瀬珂瀾, 2002 年) [8]であった. 服部真里子 (1987 年-) は「未来」短歌会に所属する, 早稲田大学短歌会出身の歌人である. 「三月の」の短歌は第二十一回日本歌人クラブ新人賞、第五十九回現代歌人協会賞を受賞した第一句集の巻頭を飾る代表的な作品である. 黒瀬珂瀾 (1977 年-) は「未来」短歌会に所属す

る, 京大短歌会出身の歌人である. このように like-dislike 評価の  $\theta$  値が大きくなった短歌は, good-bad 評価の  $\theta$  値が大きくなった短歌に比べるとどちらかといえば, 現代を活躍する現代的な作風の名歌ともいえる作品であった. これは, 評価者の多くが大学生であったことも大きい.

like-dislike 評価の  $\theta$  値が小さかった短歌として, 「心沁

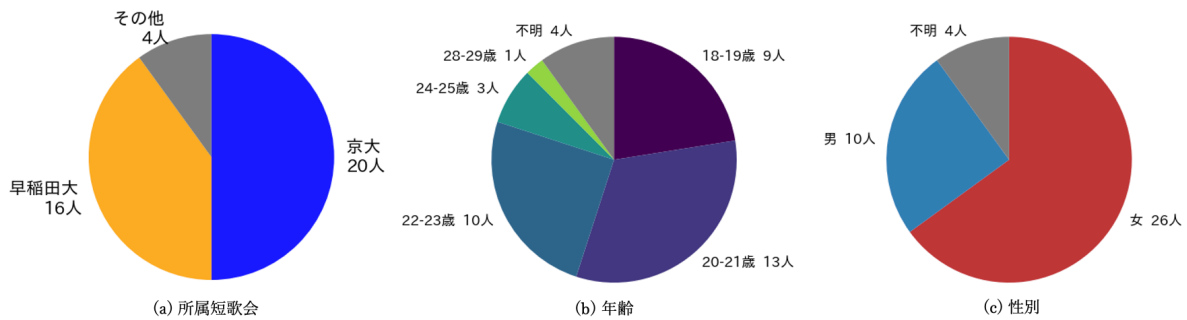


図 5: 短歌 good-bad, like-dislike 評価の評価者の属性内訳。

む青山なりし夕日の村夕日みぬ方くだりきしかな」(山中智恵子)があった。「心沁む」の短歌は、女流前衛歌人の代表的存在である山中智恵子(1925-2006年)の作品であるが、現代短歌に比べると表現が大学生にとって古風・難解であるが故に like-dislike 評価の評価値が小さくなったといえる。同じ作者の作品では「きびたきのきてつばむはあらはなるまひるのこぼ早瀬の恋」や、「若夏の青梅選むこずゑには脳(なづき)も透きて歌ふ鳥あり」などの作品も同様に、like-dislike 評価の  $\theta$  値が小さくなった。

#### 5.2.4 各評価者の項目反応曲線 $p_c(\theta)$

評価者ごとに横軸に  $\theta$  の値、縦軸に各評価値  $c$  への反応確率  $p_c(\theta)$  をとった項目反応曲線を描画すると、図 4 のようになった。図において、例えば水色の曲線「1」は、それぞれの評価者が横軸の  $\theta$  値の短歌に対して、評価値 = 1 をつける反応確率  $p_1(\theta)$  を表す。 $\theta$  の値が小さいほど最低評価値  $p_1(\theta)$  (水色) の確率が高く、 $\theta$  の値が大きいくほど最高評価値  $p_7(\theta)$  (ピンク) の確率が高くなるのが、good-bad 評価、like-dislike 評価ともに安定した評価であると言える。逆に、 $\theta$  値の大小と  $p_c(\theta)$  に上述の関係性が見られない場合は、評価データにおける全体的な評価傾向から外れた評価を行っているということが出来る。ID:1 の評価者は、good-bad 評価、like-dislike 評価ともに項目反応曲線が安定しており、全体の評価傾向と矛盾のない good-bad 評価および like-dislike 評価をしている。また、それぞれの評価値の反応確率がだいたい同じくらいの値になっていることから、7段階の評価値 1~7 をバランス良く付与した評価を行っていることがわかる。ID:7 および ID:8 評価者は、最小・最大の評価値  $p_1(\theta)$  (水色)、 $p_7(\theta)$  (ピンク) はそれぞれ矛盾なく評価しており、特に ID:8 の評価者はその傾向が強い。一方で中間の評価値  $p_2(\theta) \sim p_6(\theta)$  の曲線はそれぞれ重なりあい、明瞭な識別をしない結果となっている。ID:21 の評価者は ID:1 の評価者と同様 good-bad 評価、like-dislike 評価ともに項目反応曲線が安定しているが、それぞれの評価値の反応確率は特に like-dislike 評価においては大きな差が見受けられる。ID:30 の評価者も同様に good-bad 評価、like-dislike 評価ともに項目反応曲線が安定しているが、各評価値への反応確率が good-bad 評価では高くはっきりとしているのに対し、like-dislike 評価では低めでゆるやかな曲線となっている。これは、good-bad

評価ではより全体的な評価傾向と矛盾のない評価を行っている一方で、like-dislike ではその傾向への一致度がわずかに緩やかになっていることを表す。

図 4 の項目反応曲線はあくまでも本研究で用いた短歌評価データに対して得られたものであるため、全体的な傾向に一致していることが正しい評価を行っているということではない。特に good-bad 評価に比べ like-dislike 評価は各評価者の好みも反映されるために、good-bad 評価に比べて評価者の属性によって推定された  $\theta$  の値には大きな差が生じることが考えられる。項目反応曲線の描画では、評価者間の評価の類似性や多様性を確認し、全体的な評価傾向との相違を明示的に比較可能になることが重要である。

#### 5.3 $\theta$ が 2 次元の場合の結果

短歌 good-bad 評価における 2 次元  $\theta$  のパラメータ推定結果について、各短歌の  $\theta_i = (\theta_1, \theta_2)$ 、各評価者の  $\mathbf{w}_j = (w_1, w_2)$  の推定結果を図 6 に示す。同様に短歌 like-dislike 評価の結果を図 7 に示す。

##### 5.3.1 短歌 good-bad 評価の 2 次元 $\theta$ 値について

図 6 より、good-bad 評価において推定された 2 次元  $\theta$  値のうち、 $\theta_1$  は短歌の出典が (1) 桜前線開架宣言、(2) その他の短歌集や教科書である短歌は大きく、(3) 「塔」若葉集である短歌は小さくなる傾向にあった。最も  $\theta_1$  が大きい短歌は「観覧車」、次いで「切り終へて」、「三月の」の短歌であり、一方  $\theta_1$  が最も小さい短歌は「とりとめのなく」「何年も同じコメント」「自分への土産」の短歌であった。また審査員の評価方向を表す評価ベクトル  $\mathbf{w}_j = (w_1, w_2)$  (図 6 における矢印) はいずれも  $\theta_1$  軸に対して正の向きを指す結果となった。これより、 $\theta_1$  軸は、短歌作品の good-bad 評価に直結する表現の巧みさ・素人さ・玄人さ加減を表す評価基準として機能していると考えられる。

図 6 の縦軸  $\theta_2$  の値には、短歌の出典に関して全体的には大きな傾向の違いは見られなかった。一方で  $\theta_2$  の値が小さい短歌に着目したとき、「パチンコ屋の上にある月とおくとおくとおくとおくとおくと海鳴り(永井祐)」、「さようならいつかおしっこした花壇さようなら息継ぎをしないクロール(山崎聡子)」、「どんな目にありましたかセルレセルナルレ月夜に濡れる青いブローチ(早坂類)」、「僕たちの痛い粒子を比べればほとむほとむと鳴る島宇宙(成瀬しのぶ)」のように、オノマトペやリフレイン表現を用いて

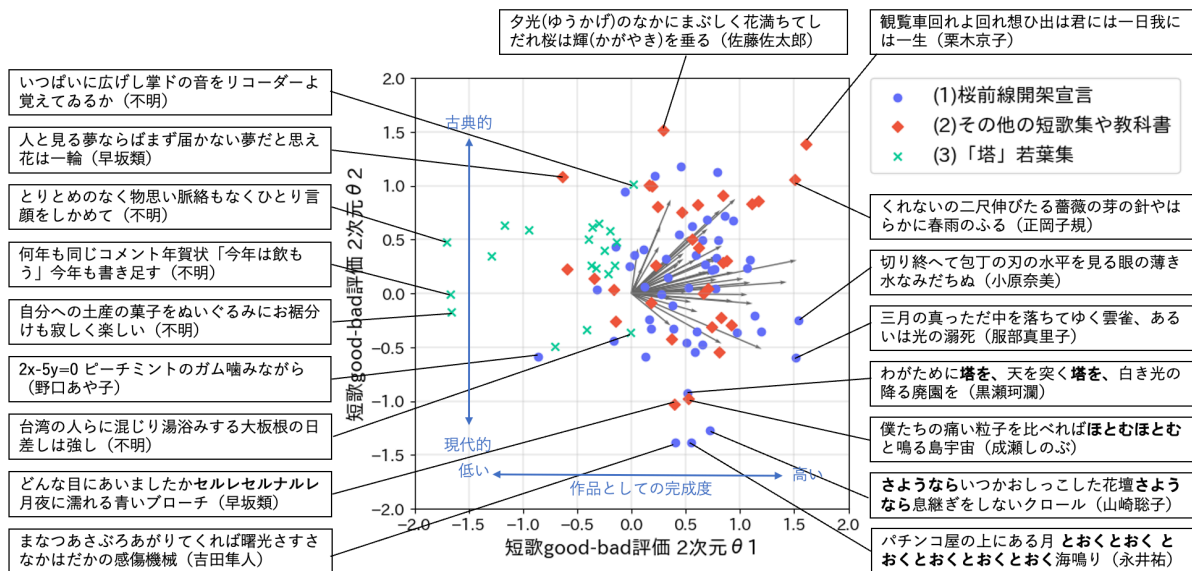


図 6: 短歌 good-bad 評価における各短歌の  $\theta_i = (\theta_1, \theta_2)$ , 各評価者の  $w_j = (w_1, w_2)$ .

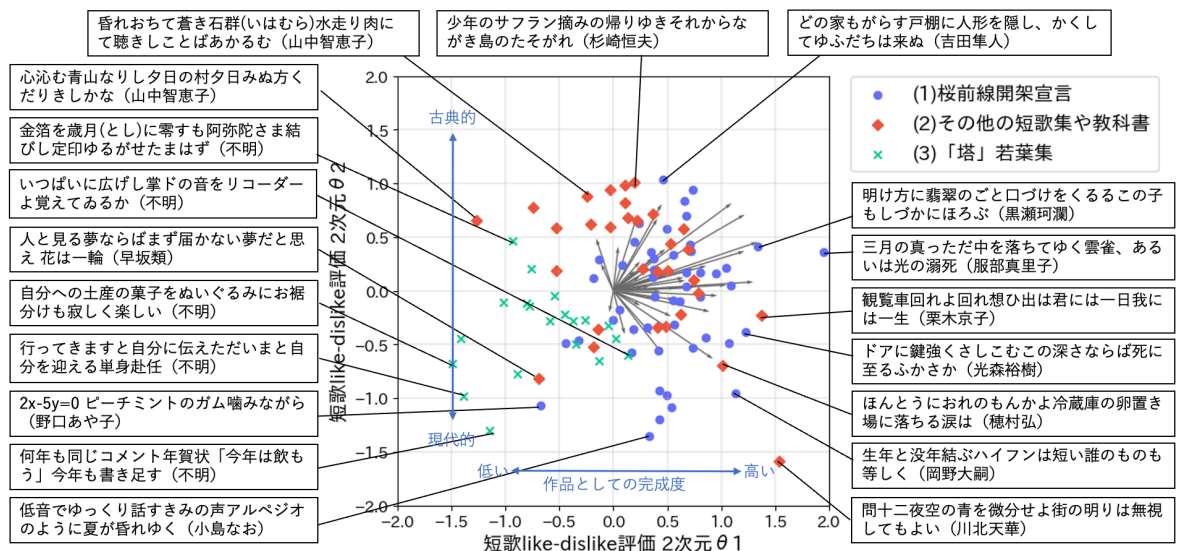


図 7: 短歌 like-dislike 評価における各短歌の  $\theta_i = (\theta_1, \theta_2)$ , 各評価者の  $w_j = (w_1, w_2)$ .

いる作品が多く見受けられた。短歌 good-bad 評価における 2次元の  $\theta_1, \theta_2$  軸の解釈を表 3 に示す。

### 5.3.2 短歌 like-dislike 評価の 2次元 $\theta$ 値について

図 7 より、like-dislike 評価において推定された 2次元  $\theta$  の  $\theta_1$  の値は、good-bad 評価の結果と同様に短歌の出典が (1) 桜前線開架宣言、(2) その他の短歌集や教科書である短歌は大きく、(3) 「塔」若葉集である短歌は小さくなる傾向にあった。これより、good-bad と同様に  $\theta_1$  軸は、短歌作品の良し悪しに関する評価基準として機能していると考えられる。最も  $\theta_1$  の値が大きい短歌は「三月の真ただ中を落ちてゆく雲雀、あるいは光の溺死 (服部真里子)」、次いで「問十二夜空の青を微分せよ街の明りは無視してもよい (川北天華)」、「観覧車回れよ回れ想ひ出は君には一日我には一生 (栗木京子)」となり、good-bad とやや異なっ

ていることは興味深い。反対に  $\theta_1$  の値が小さかった短歌のうち、(3) 「塔」若葉集以外の出典のものとして「心沁む青山なりし夕日の村夕日みぬ方くだりきしかな (山中智恵子)」、「昏れおちて蒼き石群 (いはむら) 水走り肉にて聴きしことばあかるむ (山中智恵子)」、「人と見る夢ならばまず届かない夢だと思え 花は一輪 (早坂類)」などがあつた。これらの短歌には評価コメントとして、一首内に同じ単語が複数回登場する表現の荒さや「教条的であり面白みがない」などの意見が挙げられていた。

$\theta_2$  の値については、短歌の出典が (1) 桜前線開架宣言の短歌は全体的に大きく、(3) 「塔」若葉集の短歌は全体的に小さい傾向があつた。 $\theta_2$  の値が最も大きな値となつたのは「どの家もがらす戸棚に人形を隠し、かくしてゆふだちは来ぬ (吉田隼人)」、次いで「少年のサフラン摘みの帰りゆきそれからながき島のたそがれ (杉崎恒夫)」、



表 3: 短歌 good-bad 評価の 2 次元  $\theta$  値の解釈.

2 次元 $\theta$	小さい	大きい
$\theta_1$	作品の質が低い	作品の質が高い
$\theta_2$	現代的, オノマトペ・リフレイン	古典的

表 4: 短歌 like-dislike 評価の 2 次元  $\theta$  値の解釈.

2 次元 $\theta$	小さい	大きい
$\theta_1$	作品の質が低い	作品の質が高い
$\theta_2$	現代的	古典的

きそれからながき島のたそがれ (杉崎恒夫)」であり, ほかにも旧かな遣いの古風な表現の作品が多く見受けられた. 一方  $\theta_2$  の値が最も小さくなったのは「問十二夜空の青を微分せよ街の明りは無視してもよい (川北天華)」, 次いで「低音でゆっくり話すきみの声アルペジオのように夏が昏れゆく (小島なお)」であり, ほかにも「 $2x-5y=0$  ピーチミントのガム噛みながら (野口あや子)」や「生年と没年結ぶハイフンは短い誰のものも等しく (岡野大輔)」など, 全体的に革新的で斬新な表現の作品が多い傾向にあった. これより,  $\theta_2$  軸には, 作品の古風性-現代性が反映されていると考えられる. 短歌 like-dislike 評価における 2 次元の  $\theta_1, \theta_2$  軸の解釈を表 4 に示した.

### 5.3.3 短歌の作者について

評価対象の短歌 100 首のうち, 同じ作者の作品数は山中智恵子 7 首, 穂村弘 4 首, 早坂類 4 首, その他は 26 名 2 首, 71 名 1 首であった. 含まれる作品数の多い山中, 穂村, 早坂の短歌の 2 次元  $\theta = \{\theta_1, \theta_2\}$  に色をつけ描画した結果を図 8 に示す. 短歌 good-bad 評価 (図 8a) において, 含まれる作品数が最も多かった山中智恵子の作品 (青丸) は全体的に  $\theta_1$  の値が大きい傾向があった.  $\theta_1$  軸は短歌の良し悪し評価に直結する作品の完成度を表す要素指標であると考えられ, 女流前衛歌人の代表的存在である山中の短歌作品の多くが「良い」短歌と認識・評価されていると言える. 一方短歌 like-dislike 評価 (図 8b) でも,  $\theta_1$  軸は作品の完成度を表す指標として機能していると考えられるが, 山中の作品は  $\theta_1$  値は小さく,  $\theta_2$  の値が大きい傾向にあった. これは, 現在大学短歌会に所属しており現代的な作品を目にする多くの評価者達が古典的な作品に対して「わからない」「現代短歌のほうが読み慣れているので評価しづらい」など, 必ずしも肯定的ではない評価を行っているからであると考えられる. 穂村の作品は good-bad (図 8a), like-dislike (図 8b) のいずれにおいても  $\theta_1$  値が高い傾向にあった. 早坂の作品はいずれにおいても  $\theta_1, \theta_2$  の値に明示的な傾向は見られなかった.

### 5.3.4 評価者の属性グループについて

本研究では評価者の属性として, (1) 所属短歌会, (2) 性別, (3) 年齢に基づいてそれぞれグループ化した. 各グループについて  $vMF$  分布を描画した結果を図 9, 図 10 に示す.

評価者の所属短歌会グループについては京大短歌会 (20 人), 早稲田大学短歌会 (16 人), その他 (4 人) の 3 つに

グループ化することができ (図 5a), good-bad 評価, like-dislike 評価ともに京大グループ (青) と早稲田大グループ (黄色) で評価方向の分布に違いがみられた.

評価者の年齢グループについては, 本実験で用いた短歌評価の評価者の年齢は最年少で 18 歳, 最年長で 29 歳であり, 18-19 歳, 20-21 歳, ..., 28-29 歳, 不明, の 6 つのグループに区切り (図 5b), 分析を行った. 年齢別のグループでは good-bad 評価, like-dislike 評価ともに年代ごとに評価方向の分布に違いがみられ, 特に like-dislike ではその差が大きかった.

評価者の性別グループは, 女性 (26 人), 男性 (10 人), 不明 (4 人) の 3 つにグループ化することができ (図 5c), 性別ごとのグループの評価方向には, good-bad 評価, like-dislike 評価ともに大きな違いは見られなかった.

## 6. 分析 2: 映画評価の分析

本章では, 段階反応評価データへの多次元項目反応理論分析の有効性を検証するために, より規模が大きく, 疎である段階評価データとして映画評価の公開データセット MovieLens 100K のデータ\*7 に対する同様の分析を行った.

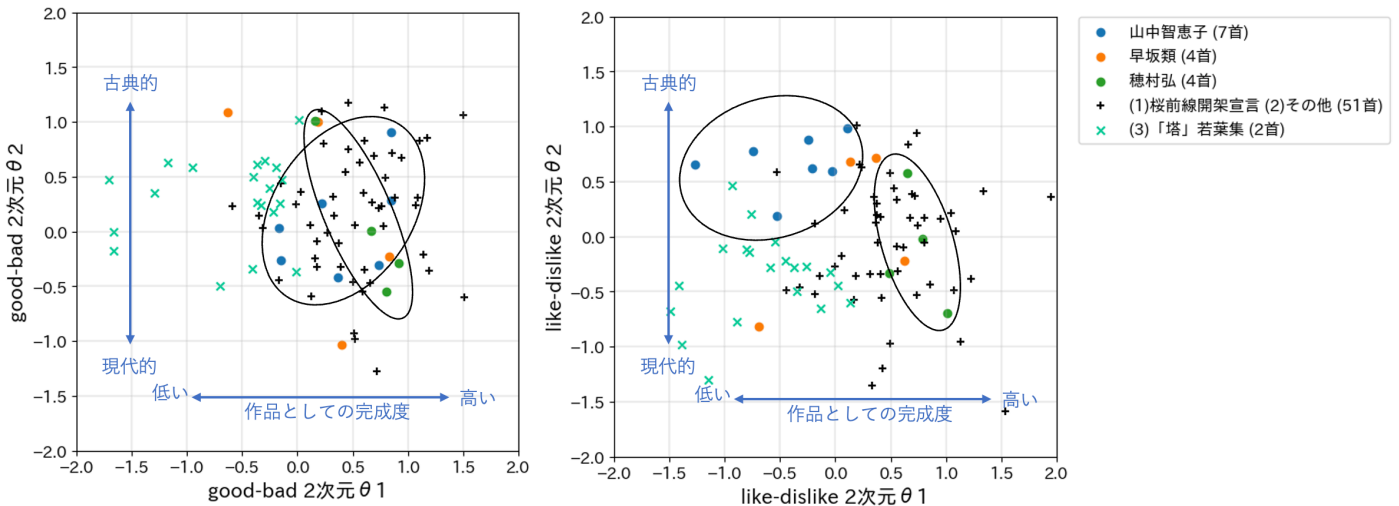
### 6.1 映画評価データ

MovieLens はミネソタ大学の研究室グループ GroupLens が管理・公開する映画レビューのデータセットである [2]. データを収集した期間に応じて複数の規模のデータセットが公開されており, 本研究では MovieLens 100K Dataset (1998 年 4 月公開) を映画評価データとして分析に用いた. 映画評価データには 1,682 件の映画に対して 943 人の評価ユーザが悪い (1)-良い (5) の 5 段階の数値で評価をした計 100,000 件の評価データが含まれる. 各映画作品に対して付与された評価数は最小で 1 件, 最大で 583 件, 平均で 59.5 件である. また各評価者が任意の映画について行った評価の数は 1 人あたり最小で 20 件, 最大で 737 件, 平均で 106.04 件である.

評価対象の映画はいずれも 1922 年から 1998 年に公開された作品であり, 属性データとして作品タイトル, 映画公開日のほかに IMDb の URL, ジャンルが含まれる. ジャンルは「Action」, 「Adventure」, 「Comedy」など 22 種類があり, 各映画に対して 0 個以上が付与されている. 作品に応じていずれのジャンルにも当てはまらない場合, もしくは複数のジャンルに当てはまる場合があり, 1 つの作品に付与されるジャンル数は最小で 0, 最大で 6, 平均で 1.72 件であった.

評価ユーザの属性情報としては, 性別, 年齢, 職業, Zip コードが含まれている. 性別は「F (女性)」, 「M (男性)」のいずれかが必ず付与され, 職業には「administrator」, 「artist」, 「doctor」など「other」, 「none」を含む 21 種類のラベルが付与されている.

\*7 <https://grouplens.org/datasets/movielens/>



(a) 短歌の good-bad 評価.

(b) 短歌の like-dislike 評価.

図 8: 短歌 good-bad, like-dislike 評価における作者別の各短歌の 2 次元  $\theta = \{\theta_1, \theta_2\}$ .

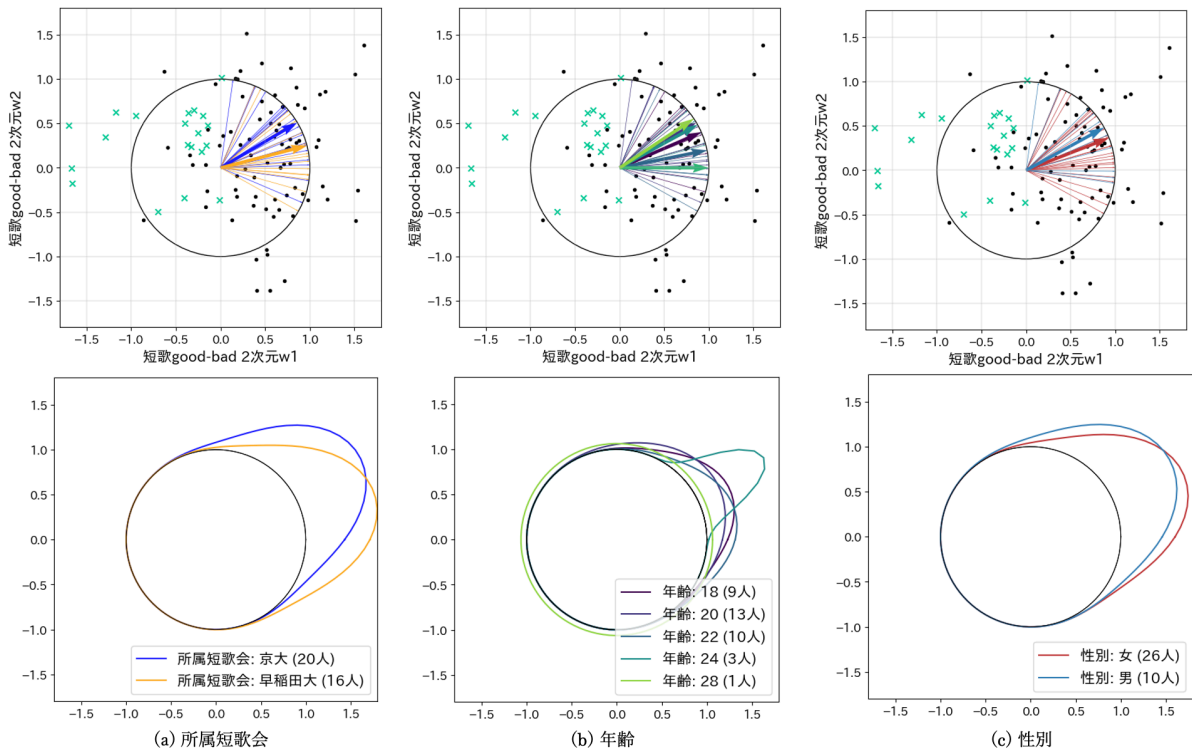


図 9: 短歌の good-bad 評価の 2 次元  $\theta$  の各グループの評価方向  $\phi$  の平均 (上段), vMF 分布 (下段).

## 6.2 $\theta$ が 1 次元の場合の結果

映画評価データに対して, 多次元項目反応理論を用い, 1 次元の映画パラメータ  $\theta$  および評価者であるユーザのパラメータ  $a, b$  を推定した結果について述べる. 各映画作品に対して縦軸に付与された評価の数 (レビュー数), 横軸に映画の公開年, 推定した 1 次元の  $\theta$  の値を色で表した結果を図 11 に示す.

### 6.2.1 映画評価の 1 次元 $\theta$ 値について

図 11 より, レビュー数の多い映画の多くは  $\theta$  の値が大

きい傾向にあった. これは興行が成功したり, 有名な名作であったり評価が高い映画ほどレビュー数が多い傾向があるためであると考えられる.  $\theta$  の値が小さかった・大きかった作品 10 件を表 5 に示す. 最も評価が高いと言える  $\theta$  値が大きかった作品には, *Pather Panchali* (大地のうた, 1955), *Casablanca* (カサブランカ, 1942), *Schindler's List* (シンドラーのリスト, 1993) など古典的な名作, 戦争や生命など壮大なテーマの作品が多くみられた. また *Santa with Muscles* (1996), *Wrong Trousers* (ペンギンに気をつけろ!, 1993) など子供向けのコメディ映画も含まれた.

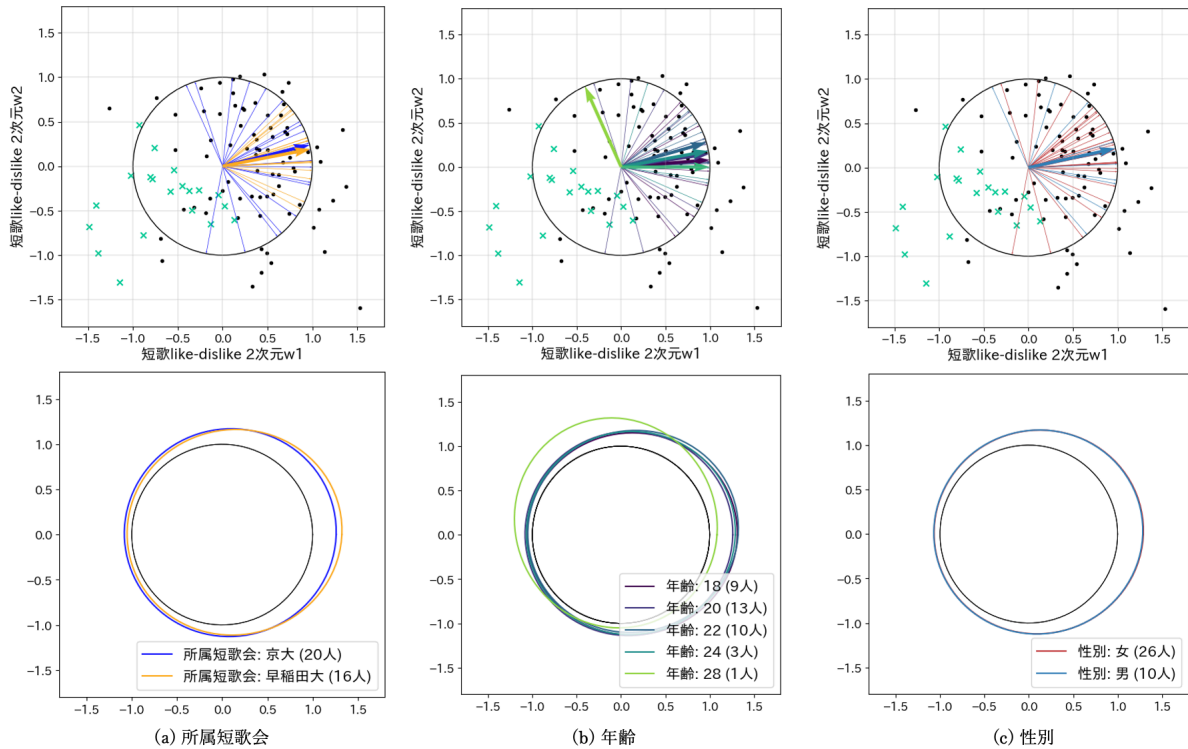


図 10: 短歌の like-dislike 評価の 2 次元  $\theta$  の各グループの評価方向  $\phi$  の平均 (上段), vMF 分布 (下段).

$\theta$  値の小さい作品はいずれも評価のレビュー数が少ない傾向にあった。

作品のジャンルについては、ジャンルごとの 1 次元  $\theta$  の平均は表 7 のようになり、Horror (ホラー), Children (子供向け), Fantasy (ファンタジー) などの  $\theta$  値が小さく, Documentary (ドキュメンタリー), War (戦争), Film Noir (フィルム・ノワール)\*8 などの  $\theta$  値が大きい傾向にあった。全体的に  $\theta$  値の高いジャンルは該当する映画数が少なく, また, Film Noir は 1940-50 年代と古い映画が中心であるため, 評価対象の映画候補が他ジャンルよりも選別されていると言える。

### 6.2.2 各ユーザの項目反応曲線 $p_c(\theta)$

評価ユーザごとの項目反応曲線の一例を図 12 に示す。各グラフ下部の括弧の中の数字はそのユーザが投稿したレビュー数を表す。図 12 の上から 1, 2 段目の 4 人は, 1-5 の評価が  $\theta$  の値と対応しており, 中でも ID:102 のユーザは 1-5 の評価値を偏らずに付与している。ID:171 のユーザも他の評価値に比べると, 中間的な評価値である 3 を付与しない傾向にある。3, 4 段目のユーザは, 最小・最大の評価値 1, 5 は矛盾なく評価しているものの, 中間の評価値 2-4 については緩やかな曲線が重なる形になっており,  $\theta$  の値と評価値が明確に対応していない結果となっている。項目反応曲線の形とユーザの属性である性別, 職業, Zip Code, レビュー数には明示的な違いは見られなかった。

\*8 1940 年代から 1950 年代後半にハリウッドでさかんに作られた犯罪映画, ハードボイルド映画 *hardboiled - detective film* のジャンル。

表 5: 映画評価の 1 次元の  $\theta$  値が大きい・小さい作品。

作品名 (公開年)	レビュー数	$\theta$
<i>Pather Panchali</i> (1996)	8	1.552
<i>Casablanca</i> (1942)	243	1.436
<i>Wrong Trousers, The</i> (1993)	118	1.419
<i>Schindler's List</i> (1993)	298	1.402
<i>Saint of Fort Washington, The</i> (1993)	2	1.395
<i>World of Apu, The</i> (1996)	6	1.381
<i>Santa with Muscles</i> (1996)	2	1.354
<i>Shawshank Redemption, The</i> (1994)	283	1.343
<i>Prefontaine</i> (1997)	3	1.338
⋮	⋮	⋮
<i>Mr. Magoo</i> (1997)	12	-1.430
<i>Amityville 1992: It's About Time</i> (1992)	5	-1.448
<i>Best of the Best 3: No Turning Back</i> (1995)	6	-1.468
<i>Being Human</i> (1993)	4	-1.469
<i>Dunston Checks In</i> (1996)	7	-1.490
<i>Bird of Prey</i> (1996)	1	-1.493
<i>Amityville II: The Possession</i> (1982)	14	-1.530
<i>Turbo: A Power Rangers Movie</i> (1997)	5	-1.554
<i>Children of the Corn: The Gathering</i> (1996)	19	-1.568
<i>Naked in New York</i> (1994)	2	-1.850

### 6.3 $\theta$ が 2 次元の場合の結果

$\theta$  が 2 次元の場合のパラメータ推定結果について, 各映画作品の  $\theta_i = \{\theta_1, \theta_2\}$ , および各評価者の  $\mathbf{w}_j = (w_1, w_2)$  の推定結果を図 13 に示す。横軸は  $\theta_1$ , 縦軸は  $\theta_2$  の値を表し, 各点および色は映画作品とその公開年, 灰色の矢印は評価者のベクトルを表す。また 2 次元  $\theta$  の推定結果について各映画とそのレビュー数を描画した結果を図 14 に示す。

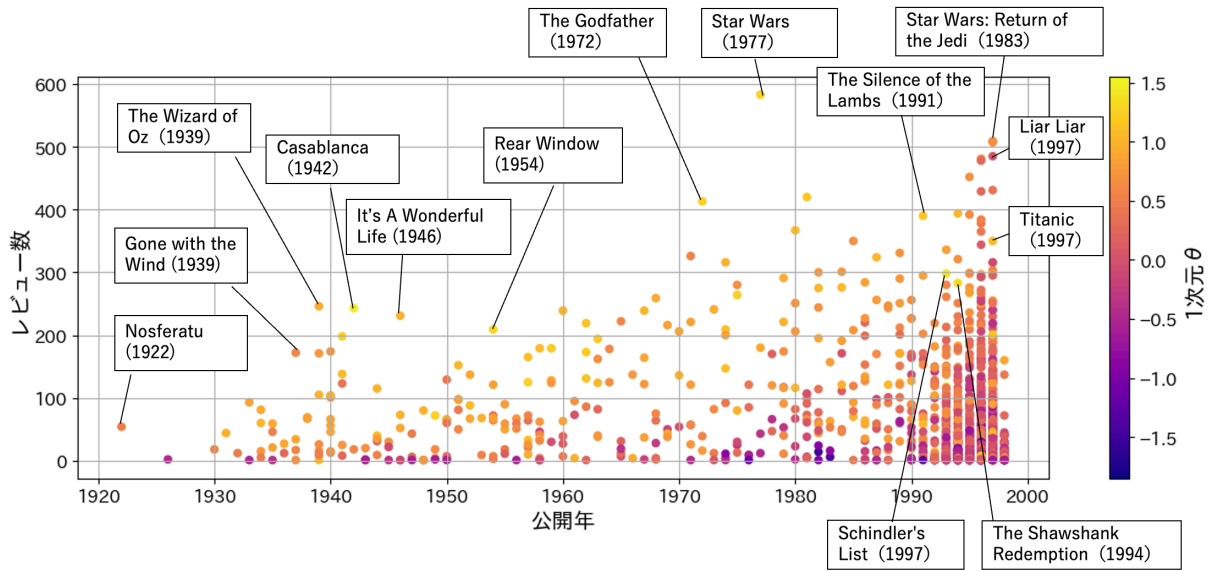


図 11: 映画評価における各作品のレビュー数 (縦軸), 映画の公開 (横軸) および 1次元  $\theta$  の値.

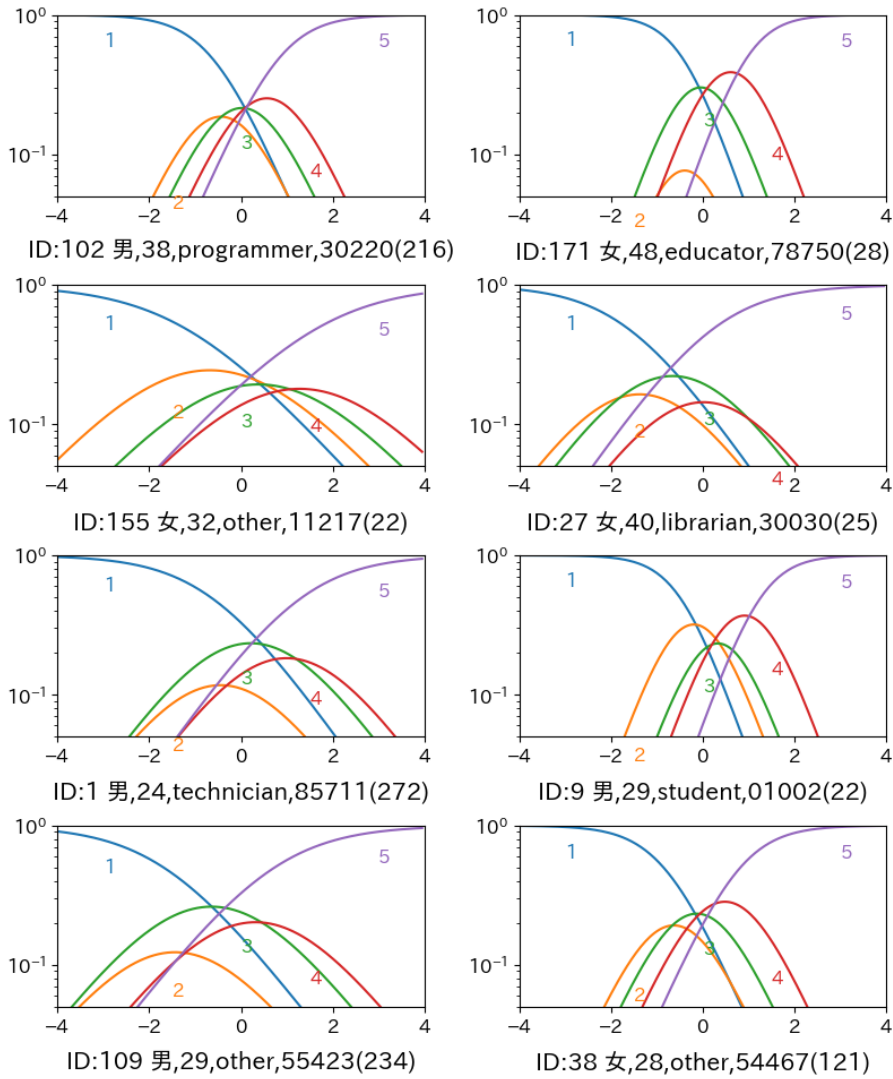


図 12: 各評価ユーザの項目反応曲線の一例. 各グラフの横軸は  $\theta$  の値, 縦軸は評価値が  $c$  と反応する確率  $p_c(\theta)$  を表す.

表 6: 映画評価データにおいて各ジャンルに含まれる作品数と 1 次元の  $\theta$  の値の平均.

ジャンル	作品数	$\theta$ の平均
FilmNoir	24	0.583
War	71	0.464
Documentary	50	0.388
Musical	56	0.279
Mystery	61	0.277
Drama	725	0.272
Romance	247	0.225
Animation	42	0.222
Crime	109	0.211
Thriller	251	0.103
SciFi	101	0.093
Adventure	135	0.015
Comedy	505	0.006
Action	251	-0.054
Fantasy	22	-0.178
Children	122	-0.181
Horror	92	-0.206

表 7: 映画評価の 2 次元  $\theta$  値の解釈.

2 次元 $\theta$	小さい	大きい
$\theta_1$	作品の質が低い	作品の質が高い
$\theta_2$	古い・大衆的	新しい・前衛的

図 13 より, 大半の評価者の  $w_j$  は  $\theta_1$  軸の値が大きくなる方向に向かっており, これより  $\theta_1$  軸が短歌分析時と同様に作品の良い, 悪い評価に直結する質の高低を表していると言える. 最も  $\theta_1$  の値が大きかった映画は戦争映画に分類される *Schindler's List* であり, 次いで子供向けのアニメーションコメディ作品 *The Wrong Trousers*, その後は *The Saint of Fort Washington* (聖者の眠る街), *Casablanca* (カサブランカ), *The Shawshank Redemption* (ショーシャンクの空に) など古典的な名作映画が続いた.  $\theta_1$  の値が最も小さかった 2 作品はホラー映画シリーズの *Amityville II: The Possession*, *Amityville: A New Generation* であり, 他にもシリーズものの作品が多い傾向があった. 公開年の古い映画は推定された  $\theta_1$  の値が大きい傾向にあり, これは評価が高くない古い映画はデータ収集時に鑑賞されることなく自然淘汰された結果であると考えられる. また図 14 より, レビュー数の多い作品は  $\theta_1$  の値が大きい傾向にあった. これは, 評価の高い映画ほど興行成績やその内容が話題となり鑑賞・レビューする人も多いためであると考えられる.

最も  $\theta_2$  の値が大きかった作品はコメディ・風刺映画の *Fear of a Black Hat* であり, 次いでミュージカル・ロマンスの *Top Hat*, 短編映画 *Grosse Fatigue* が続いた. 最も  $\theta_2$  の値が小さかった作品はコメディドラマ *The Big Bang Theory*, フランス映画の *La Collectionneuse*, ロマンス映画の *Rough Magic* などであった.

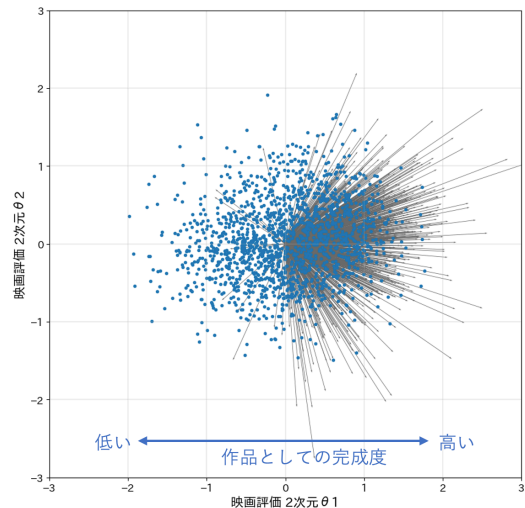


図 13: 映画評価結果に基づく各映画の  $\theta_i = (\theta_1, \theta_2)$  および各評価者の  $w_j = (w_1, w_2)$  の推定結果.

### 6.3.1 映画のジャンルについて

各映画の  $\theta_1, \theta_2$  の値についてジャンルごとに色分けした結果を図 15 に示す. ジャンルごとに明瞭な差はなかったが, 評価の著しく低い作品が含まれる Horror (左上) や Children (子供向け) (左下) は  $\theta_1$  が小さい傾向があった. 対象の作品数が多い Comedy (中央上) やドラマ Drama (中央下) は全体的に広く分布しており,  $\theta_1, \theta_2$  の値に明示的な特徴は得られなかった. 評価の高い作品が多く含まれる Documentary や FilmNoir は全体的に  $\theta_1$  の値が大きい傾向にあった.

### 6.3.2 評価者ユーザの属性グループについて

評価ユーザの属性グループのうち職業ごとのグループの傾向を分析するため, 各職業グループの評価方向  $\phi$  およびその平均を描画した結果の一部を図 16 に示す. いずれの図においても灰色の太矢印は全ての評価者の評価方向  $\phi_j$  の平均を表し, 緑色の太矢印はそれぞれの職業グループの評価者の評価方向の平均を表す. 図 16 より, Scientist, Writer, Artist などの職業グループは全体平均よりも  $\theta_2$  の高い作品を評価する傾向にあった. 一方で Doctor, Executive, Homemaker などの職業グループは全体平均よりも  $\theta_2$  の低い作品を評価する傾向にあった.

## 7. まとめ

短歌は日本の伝統的な詩形の一つであり, これまで文学としての研究は行われてきているが, その評価について定量的な研究は行われておらず, 定量的な比較・分析手法も確立されていなかった. そこで本研究では, 短歌作品への段階評価データを用いて項目反応理論に基づいた解析を行い, 短歌の評価と評価者の分析を行う手法を提案した.

項目反応理論は現代テスト理論の 1 つであり, テスト結果に基づく各問題 (項目) への受験者の回答 (反応) データを用いて, 受験者の潜在的な能力  $\theta$  およびテスト問題の傾向  $a, b$  パラメータを推定し, 問題や受験者に関する分析

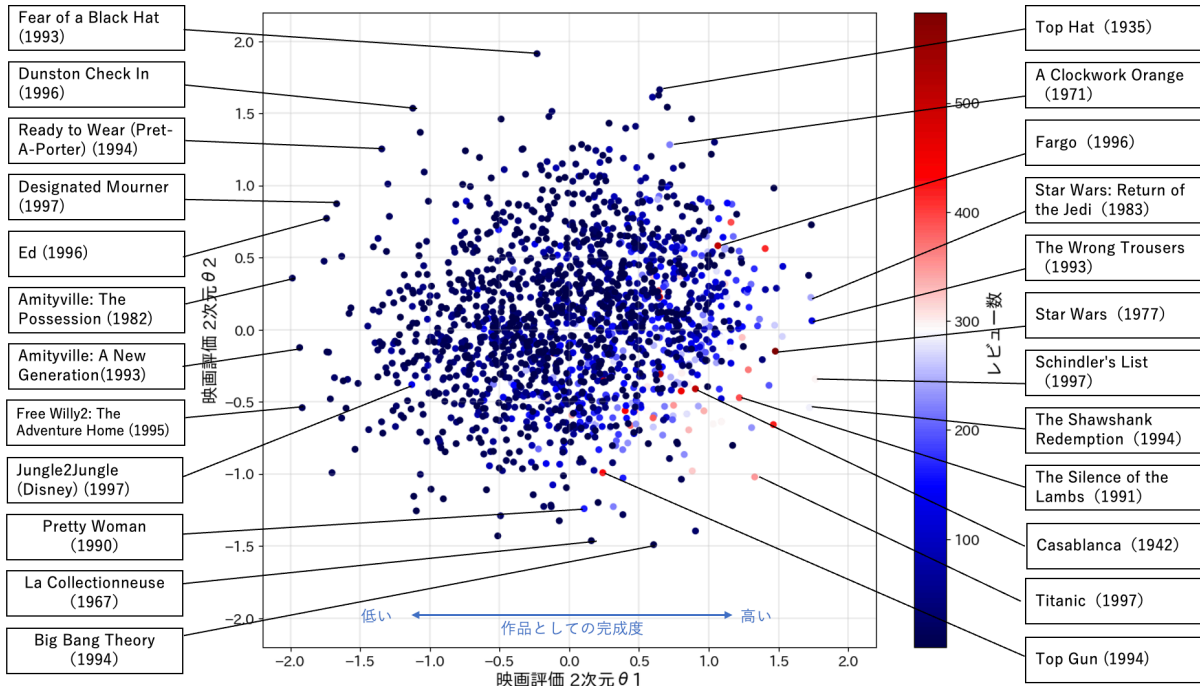


図 14: 映画評価結果に基づく各短歌の  $\theta_i = (\theta_1, \theta_2)$  および各評価者の  $\mathbf{w}_j = (w_1, w_2)$  の推定結果.

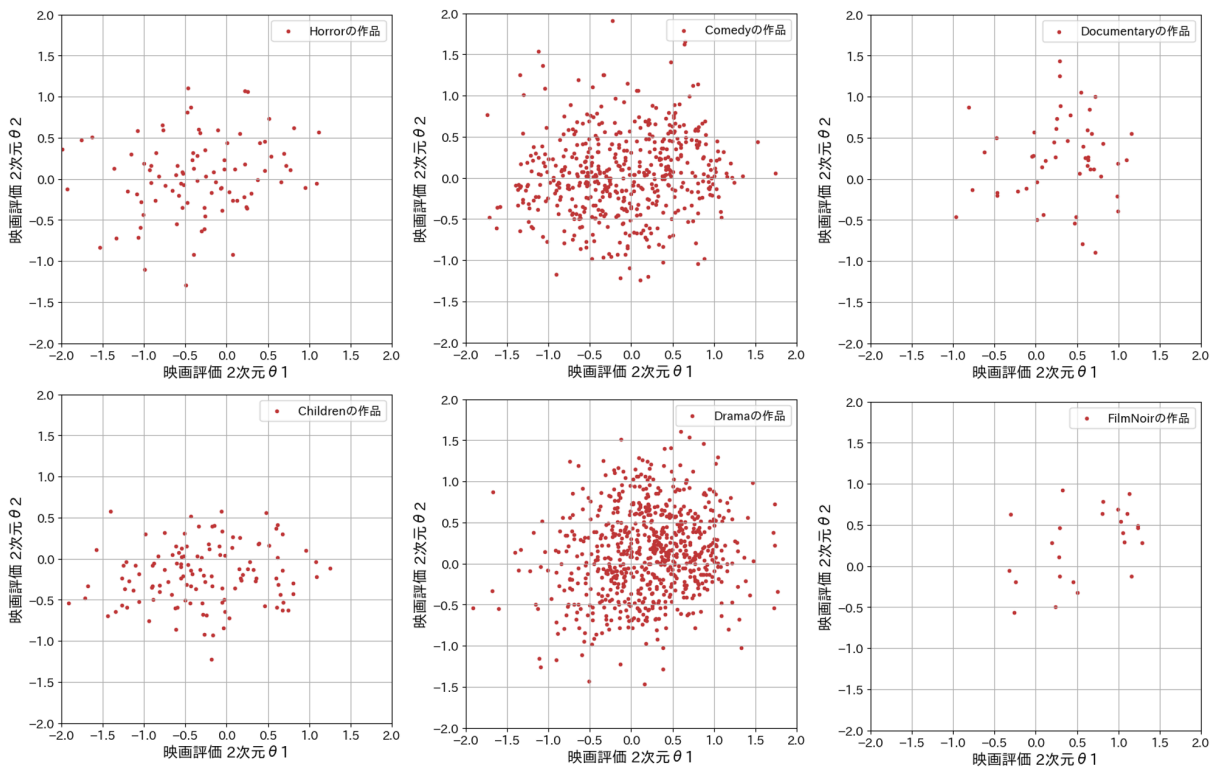


図 15: 映画評価の 2次元  $\theta$  のジャンル別の表示.

を行う。本研究ではこれを短歌の評価に応用した。本研究では複数の短歌に対する複数の評価者による7段階の「良い-悪い」および「好き-嫌い」の評価値を評価データとし、短歌の潜在的な良し悪しを表す潜在的な  $\theta$ 、評価者の傾向  $a, b$  をパラメータとして推定し、推定されたパラメータをもとに短歌情報、評価者情報と照らし合わせて、短歌作品への評価および評価者に関する分析を行った。

潜在変数  $\theta$  の次元数は1次元  $\theta$ 、2次元  $\theta = \{\theta_1, \theta_2\}$  の2通りを試し、1次元の場合には「良い-悪い」「好き-嫌い」の評価値の関係性や、評価者ごとの項目反応曲線を、2次元の場合には各次元の値の解釈と評価者の評価傾向に関する分析を行った。

また提案手法の有効性をより大規模かつ疎な段階評価データで検証するために、公開された映画評価データセッ

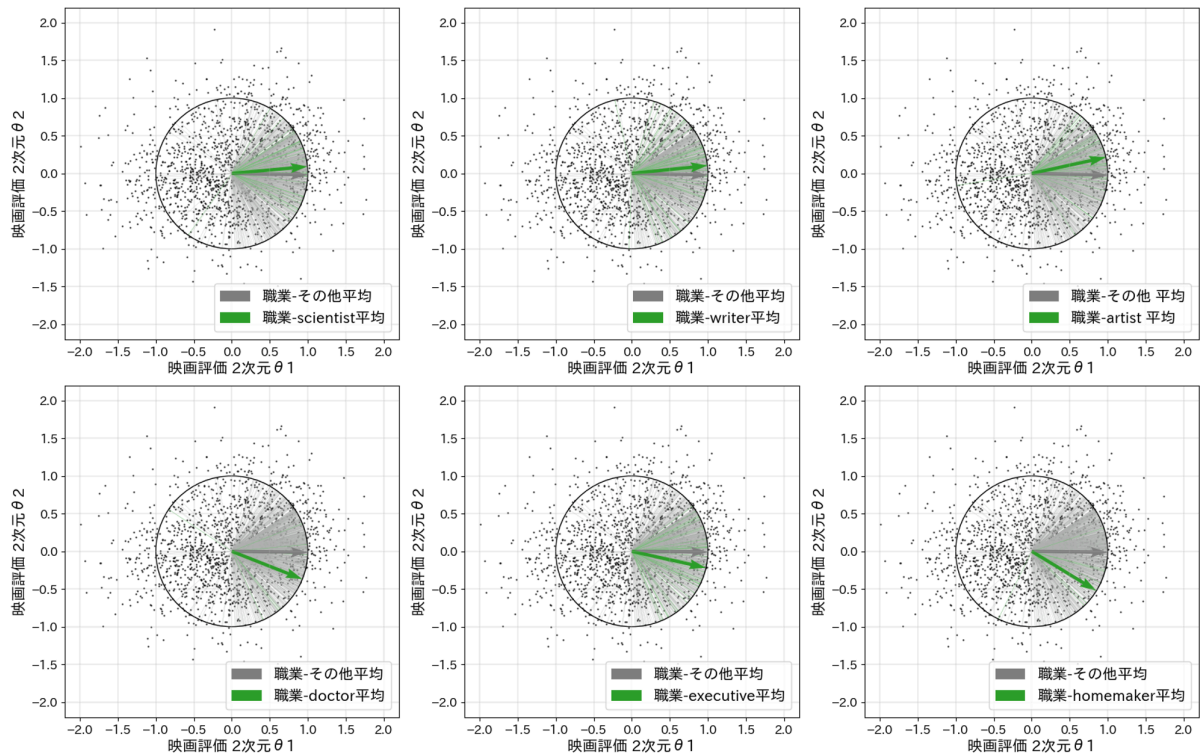


図 16: 映画評価の 2 次元  $\theta$  の評価者の評価方向  $\phi$  の職業別の平均.

ト MovieLens 100K を用いて同様の分析を行った。

その結果、1次元の $\theta$ の分析では「良い-悪い」「好き-嫌い」といった評価値の関係性を数値的に比較し、また評価者ごとの項目反応曲線を描くことで評価反応の傾向を分析することができた。2次元 $\theta$ の分析では、評価の要素を2次元に分解して解釈し、作品の作者や評価者の年齢・性別・職業などの属性のグループ別の評価傾向を分析することができた。

**謝辞** 本研究は、科研費・学術変革(B)「情動をもたらす言語芸術の脳科学」(研究課題 21H05061)に助成していただきました。

参考文献

[1] Bock, R. D. and Aitkin, M.: Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm, *Psychometrika*, Vol. 46, No. 4, pp. 443–459 (1981).  
 [2] Harper, F. M. and Konstan, J. A.: The movielens datasets: History and context, *Acm transactions on interactive intelligent systems (tiis)*, Vol. 5, No. 4, pp. 1–19 (2015).  
 [3] Lord, F. M.: *Applications of item response theory to practical testing problems*, Routledge (2012).  
 [4] Mardia, K. V., Jupp, P. E. and Mardia, K.: *Directional statistics*, Vol. 2, Wiley Online Library (2000).  
 [5] Samejima, F.: Estimation of latent ability using a response pattern of graded scores., *Psychometrika monograph supplement* (1969).  
 [6] Ueno, M., Songmuang, P., Okamoto, T. and Nagaoka, K.: Item response theory with assessors' parameters of peer assessment, *IEICE TRANSACTIONS on Inf. Syst.*, pp. 377–388 (2008).

[7] 野口あや子: 歌集『くびすじの欠片』, 短歌研究社 (2009).  
 [8] 黒瀬河瀾: 歌集『黒耀宮』, ながらみ書店 (2002).  
 [9] 栗木京子: 歌集『水惑星』, 雁書館 (1984).  
 [10] 中学教科書: 国語 2 年, 光村図書 (2020).  
 [11] 山田 航: 桜前線開架宣言, 左右社 (2015).  
 [12] 服部真里子: 歌集『行け広野へと』, 本阿弥書店 (2014).