

2023年5月28日 13:30-15:30

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（慶應義塾大学） hironok@sfc.keio.ac.jp

持橋大地（統計数理研究所） daichi@ism.ac.jp

概要

短歌の評価傾向の分析として、複数の短歌作品に対する複数の評価者の評価の定量的な分析手法を提案

- 短歌評価データ

大学の短歌会に所属する40名の評価者が
100首の短歌に対して
7段階の数値で「良い-悪い」「好き-嫌い」を評価

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}

- 多次元の項目反応理論を用いて短歌の潜在的な優劣、各評価者の評価傾向を分析する手法を提案

- 短歌の評価要因の分解
- 短歌の作者ごとの傾向分析
- 評価者や評価者のグループの評価傾向の相違の比較 が可能に

背景

背景 | 短歌

- 短歌とは？
 - 5-7-5-7-7 の韻律を基本とする歌体の1つ
- 短歌の評価
 - 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
 - 歌会：複数の参加者が短歌を持ち寄り相互に評価
 - 歌壇や短歌賞：複数の選者が複数の投稿作品を評価



歌会（「塔」短歌会HPより）



朝日歌壇（朝日新聞より）

背景 | 短歌の評価の俯瞰的な分析

- × 作品と評価者を一対一として考える
- 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
 - 過去の短歌に関する研究
 - 特定の作者年代の作品を対象に文学的に掘り下げる研究が大半
 - 問題点
 - 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
 - 歌会の記録の大半は電子データとして残らない・公表されない
 - 短歌賞等の選考記録は公開されない、選外の作品は掲載されない
 - 多数多の短歌作品-評価者データの分析手法が確立されていない

本研究で扱う短歌評価データ

短歌の段階反応評価データ

- 複数の短歌に対して複数の評価者が段階カテゴリーの離散数値評価を行ったデータ
- カテゴリ数 $C = 5$ の場合
 - N 件の短歌作品: $\{s_1, \dots, s_N\}$
 - J 人の評価者: $\{u_1, \dots, u_J\}$
 - 短歌 s_i への評価者 u_j の C 段階カテゴリー $c \in \{1, \dots, C\}$ の評価値: x_{ij}
 - 評価データ全体: $X = \{x_{ij}\} (i = 1 \dots N, j = 1 \dots J)$

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

項目反応理論 (IRT)

項目反応理論 (IRT)

Item Response Theory; IRT [3]

- テストを作成・運用・評価するための数理モデルを用いた現代テスト理論の1つ
 - 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
-
- IRTのメリット[6]
 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

項目反応理論 (IRT)

- 二値反応データ（正誤）を扱う場合
 - 基本的な項目反応モデル
- 反応が C 段階カテゴリー $c \in \{1 \dots C\}$ の段階データ
 - 多値型項目反応モデル
 - 段階反応モデル[5]

項目反応モデル（二値反応データ）

- 潜在的な特性(latent trait) θ を導入
 - 異なる能力・背景を持つ受験者集団の結果比較を可能に
 - 受験者 s_i の能力とテスト項目 u_j の性質(易しい/難しい)を表現
- 能力 θ の受験者(被験者)が問題 j に 正解する確率

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

- θ : 被験者母数; a, b_j : 項目母数
 - 能力 θ が問題の難易度 b_j より大きいと正解する確率が上昇
- a : 確率の上がり方の急さを表現
- b_j : 項目 j の性質(易しい/難しい)
- $\Psi(x)$: 標準正規分布の累積確率密度関数/ロジスティック関数
 - 本研究ではロジスティック関数を採用

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

段階反応モデル（多値反応データ）

- 評価値が (評価値 $\geq c$) となる確率の θ に関する関数

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

- 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

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

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

項目反応曲線

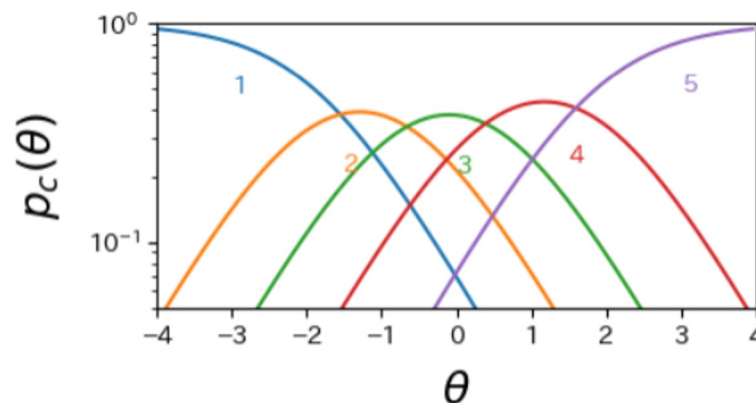
- Item Characteristic Curve; ICC
- 縦軸に $p_c(\theta)$, 横軸に θ を配して描画した曲線
 - 評価値が (評価値 $\geq c$) となる確率の θ に関する関数

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

- 評価者の傾向を分析する際に用いる



(a) 1次元の θ の項目反応曲線.



(b) 縦軸の確率を対数で表示したもの.

提案手法:

多次元項目反応理論を用いた
短歌の段階評価データの分析

多次元項目反応理論による短歌評価の分析

- 通常のIRT: 各受験者が潜在的な能力 θ を保有

- テスト項目(問題)に対する受験者の反応データ

↓推定

- 各受験者の持つ潜在的な能力 θ
- テスト項目 j の特性 a, b_j

受験者 \ テスト項目	u_1	u_2	u_3	...	u_j
s_1	3	4	5	...	x_{1j}
s_2	1	3	4	...	x_{2j}
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
s_i	x_{i1}	x_{i2}	x_{i3}	...	x_{ij}

- 提案手法: 各短歌が潜在的な優劣(良し悪し) θ を保有

- 評価者の短歌へ対する反応データ

↓推定

- 各短歌の持つ潜在的な優劣 θ
- 評価者の特性 a, b_j

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

短歌, 評価者パラメータ

- 優劣 θ 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

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

- 反応の閾値 b_{jc} は評価者 j ごと/反応カテゴリ c ごとに異なる
- 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つとする

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

- a_j は各評価者の判断のぶれの大きさに相当
- b_{jc} は評価者 j が評価 c を与える閾値に相当

θの次元数

- 通常のIRT分析ではθの次元数は1, 2次元に設定
- θの次元数の多次元化
 - θで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可能に
 - ✗ 推定に要する計算量が増大
 - ✗ 結果の解釈が困難に
- 本研究におけるθの次元数
 - 1次元のθおよび2次元の $\theta = (\theta_1, \theta_2)$ (※3次元以上も可能;検討中)

(評価値 $\geq c$) となる確率の θ に関する関数

- θ が1次元の場合

$$p_c^*(\theta) = \Psi(a(\theta - b_c)) \quad (7)$$

- θ が2次元の場合 $\theta = (\theta_1, \theta_2)$

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

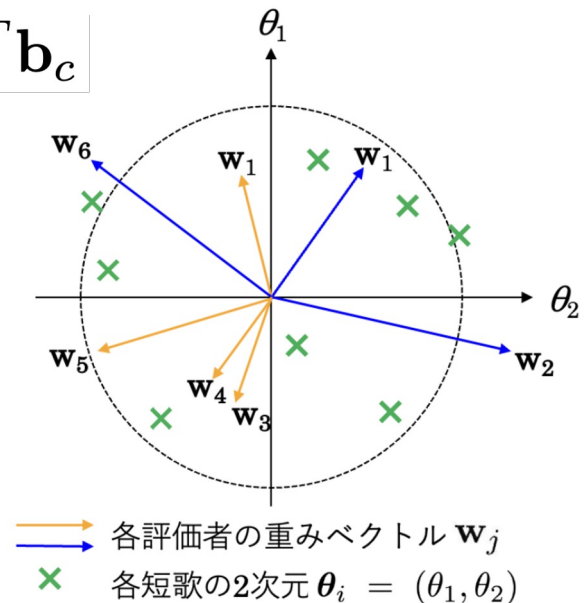
$\boldsymbol{\theta}$ の線形モデル

$$\eta_c = \mathbf{a}^T \mathbf{b}_c$$

$$\downarrow$$

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

- 学習された \mathbf{w} と $\boldsymbol{\theta}$ の内積 $\mathbf{w}^T \boldsymbol{\theta}$ が大きいほど評価値は高くなる
- 重みベクトル \mathbf{w}_j は評価者 u_j ごとに推定

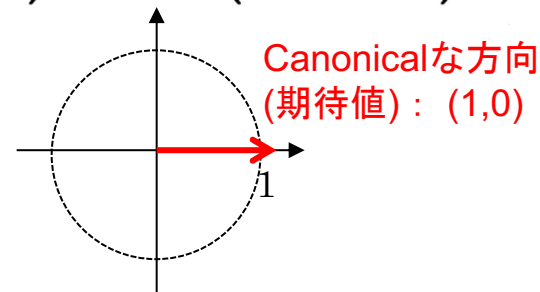


各評価者の評価方向, グループ化

- w_j は各評価者の評価方向 ϕ_j で表すことができる

$$\mathbf{w}_j = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \sqrt{w_1^2 + w_2^2} \begin{pmatrix} \frac{w_1}{\sqrt{w_1^2 + w_2^2}} \\ \frac{w_2}{\sqrt{w_1^2 + w_2^2}} \end{pmatrix} = \mathbf{a}_j \begin{pmatrix} \cos \phi_j \\ \sin \phi_j \end{pmatrix}$$

- 評価方向ベクトルの分布に関する制約:
 - 長さ是对数正規分布
 - 角度はvon Mises-Fisher分布



- 評価者の所属や性別など属性情報が明らかである場合, グループごとの ϕ_j の von Mises-Fisher分布(vMF分布)[4]を用いて評価方向の傾向を数値的に比較可能

- vMF (方向データに対する確率分布) の確率密度関数

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

- μ : 平均方向
- κ : 集中度パラメータ
- $I_0(\kappa)$: 第1種変形ベッセル関数

パラメータの推定

- 短歌評価データ $X = \{x_{ij}\} (i = 1 \dots N, j = 1 \dots J)$ から θ_i, a_j, b_{jc} を推定

- 各短歌の θ_i および各評価者の a_j, b_{jc} をまとめて

$$\Theta = \{\theta_i\}_{i=1}^N, \mathbf{a} = \{a_j\}_{j=1}^J, \mathbf{b} = \{b_{jc}\} \quad (i = j \dots J, c = 2 \dots C) \text{ とする}$$

- X, Θ の同時確率

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

- 対数尤度 $L = \log p(X, \Theta | \mathbf{a}, \mathbf{b})$

$$= \sum_{i=1}^N \left[\log p(\theta_i) + \sum_{j=1}^J \sum_{c=1}^C \mathbb{I}(x_{ij} = c) \log p_c(\theta_i) \right]$$

$$= \sum_{i=1}^N \left[-\frac{1}{2} \theta_i^2 + \sum_{j=1}^J \sum_{c=1}^C \mathbb{I}(x_{ij} = c) \right.$$

$$\left. \log \left\{ \Psi(a_j(\theta_i - b_{jc})) - \Psi(a_j(\theta_i - b_{jc+1})) \right\} \right] \quad (13)$$

データ X をもとに式(13)を
最大化する θ_i, a_j, b_{jc} を求める

- \mathbf{a} の事前分布として対数正規分布+vMF 分布を設定

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

パラメータの推定

通常 of 推定手法

- θ が正規分布 $N(0, 1)$ に従うことから θ を周辺化
- 式(15) を近似的に計算
- 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 θ を求める[1]

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

↓

- 点推定した不確定な a, b から θ を復元するよりも、直接 θ, a, b を同時に推定した方が良い推定値が求められる

本研究

- 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

提案手法を用いた分析

提案手法を用いた分析

• 分析1: 短歌の評価

大学短歌会所属者を中心に被験者を募り独自に収集

- 評価者数: 40人
- 評価対象の短歌: 100首
- 評価段階: 7 (良い7-悪い1/好き7-嫌い: 1)
- 評価数 4,000件 $\times 2 = 8,000$

<より規模が大きく疎なデータ, 欠損値も含む>

• 分析2: 映画の評価

ミネソタ大学の GroupLens が管理・公開する映画レビューのデータセット MovieLens 100K Dataset(1998/4) [2]

- 評価者数: 943人
- 評価対象の映画: 1,682件
- 評価段階: 5 (良い:5 – 悪い: 1)
- 評価数: 100,000件

分析1: 短歌評価データ

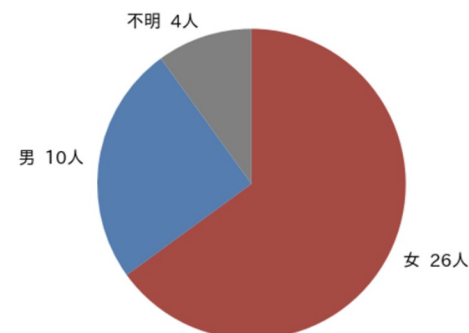
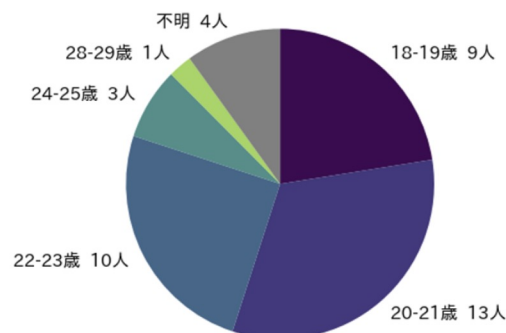
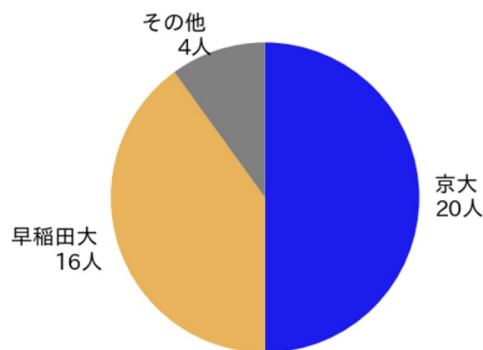
- 評価データ（計4,000件 x 2） *評価時にコメントも収集
 - 「良い（7） - 悪い（1）」 → 短歌good-bad評価
 - 「好き（7） - 嫌い（1）」 → 短歌like-dislike評価
- 評価対象の短歌：100件

世間的に評価されている作品 (1)49件、(2) 31件

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 行ってきますと自分に伝えたいいと自分を迎える单身赴任	(3) 「塔」若葉集	-
	(3) 「塔」若葉集	-

- 評価者: 40名

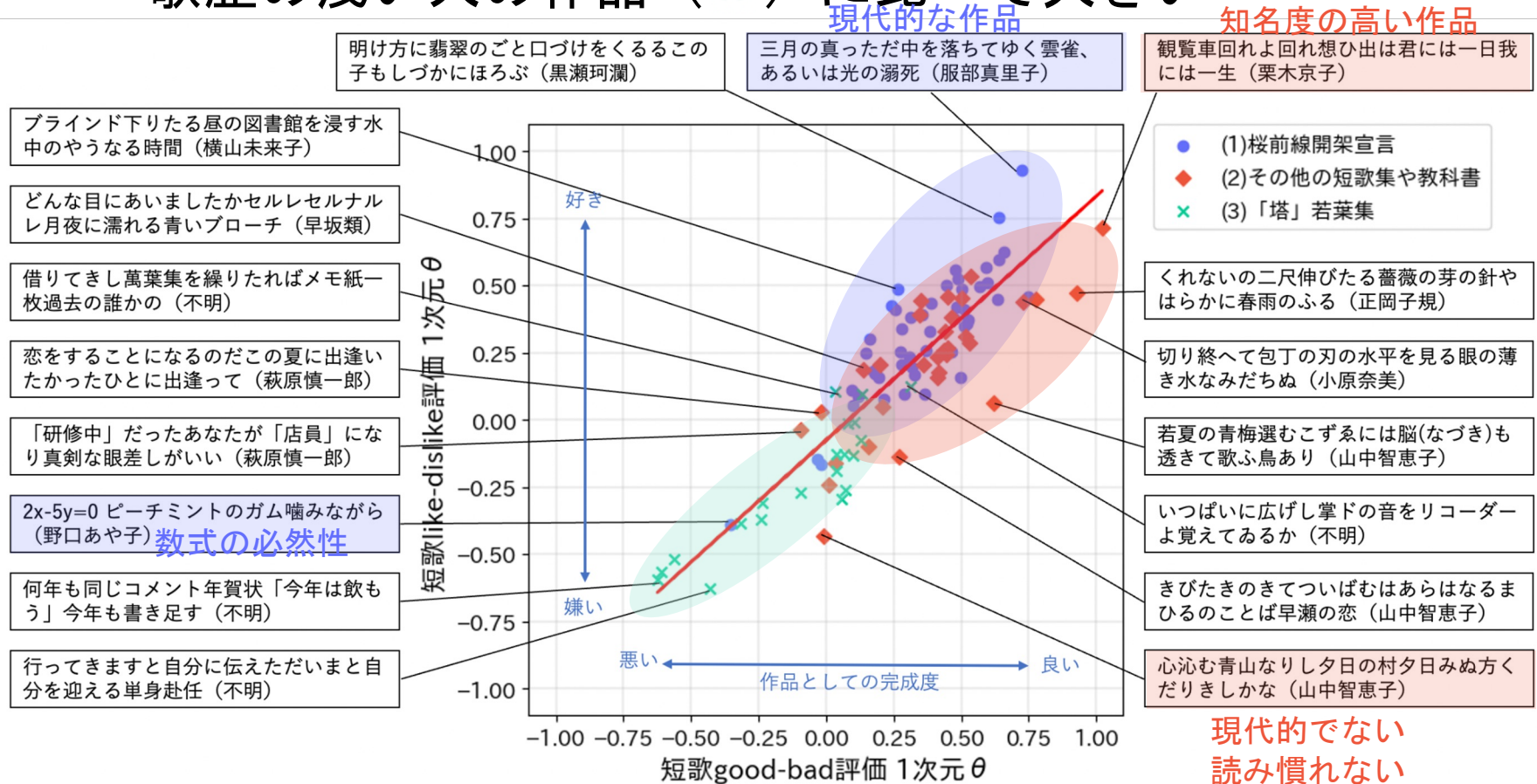
歌歴の浅い作者の作った
世間的には評価されていない作品 (3) 20件



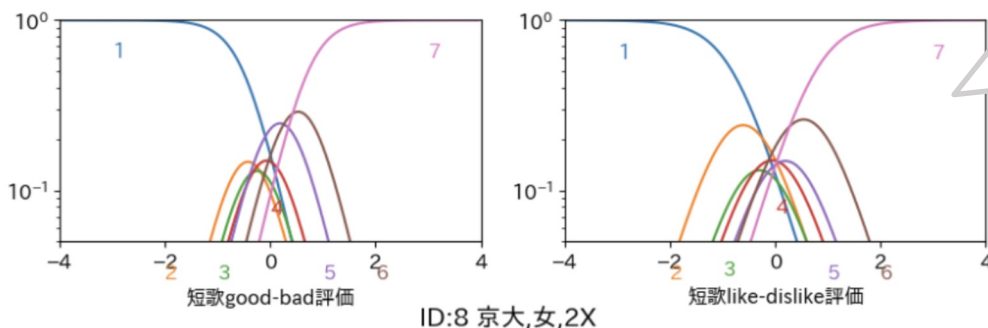
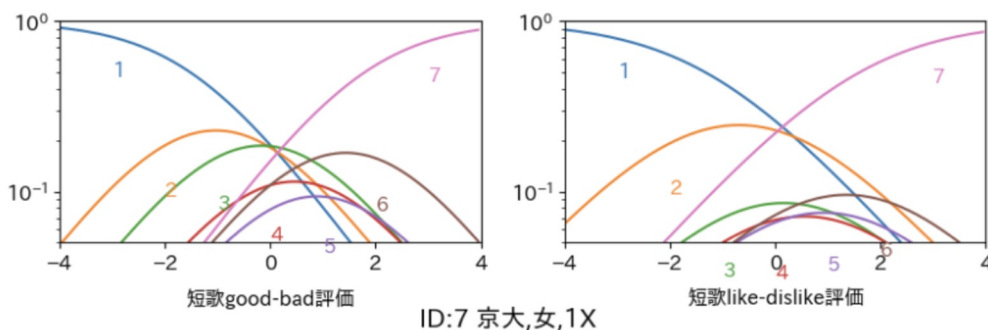
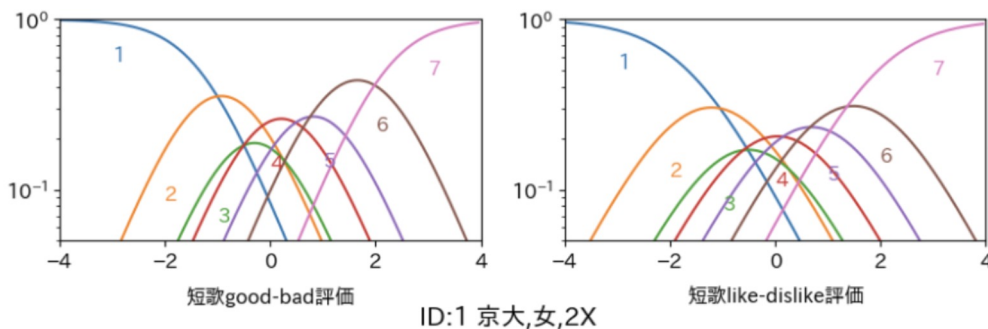
θが1次元の場合の分析結果

短歌good-bad評価, like-dislike評価の関係

- good-badの θ とlike-dislikeの θ には相関関係 ($R = 0.88$)
- 著名な歌人の作品 (●や◆) の θ 値は歌歴の浅い人の作品 (×) に比べて大きい



評価者の項目反応曲線



- θ の値が小さいほど
最低評価値 $p_1(\theta)$ の確率が高い
- θ の値が大きいくほど
最高評価値 $p_7(\theta)$ の確率が高い



good-bad, like-dislike評価
ともに安定した評価

θ 値の大小と $p_c(\theta)$ に上述の
関係性が見られない

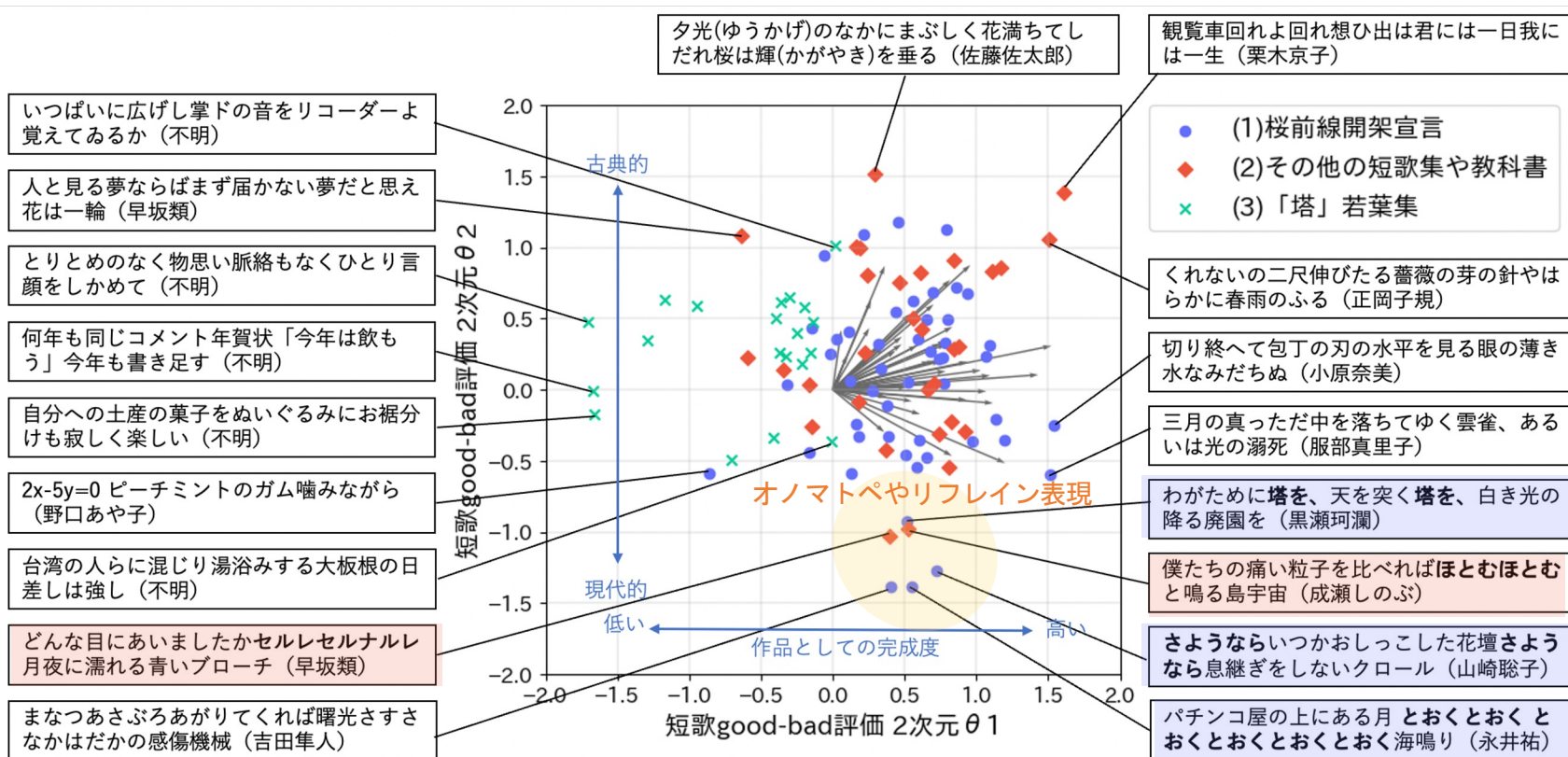


評価データにおける全体的な
評価傾向から外れた評価を
行っている

θが2次元の場合の分析結果

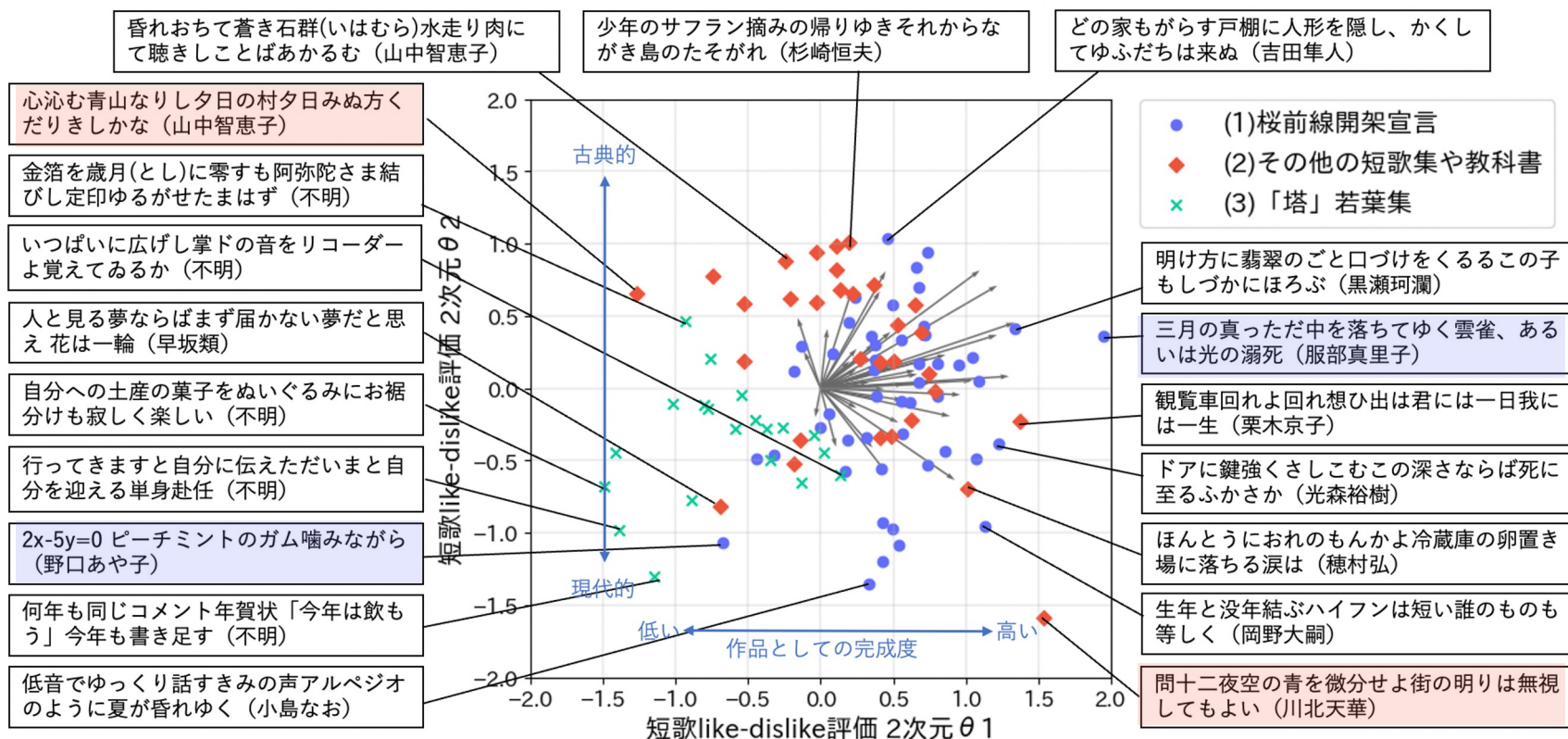
短歌good-bad評価の θ_1 , θ_2 の値

2次元 θ	小さい	大きい
θ_1	作品の質が低い	作品の質が高い
θ_2	現代的, オノマトペ・リフレイン	古典的



短歌like-dislike評価の θ_1 , θ_2 の値

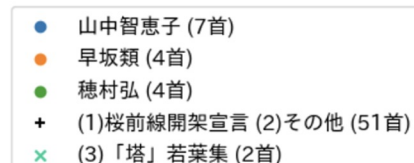
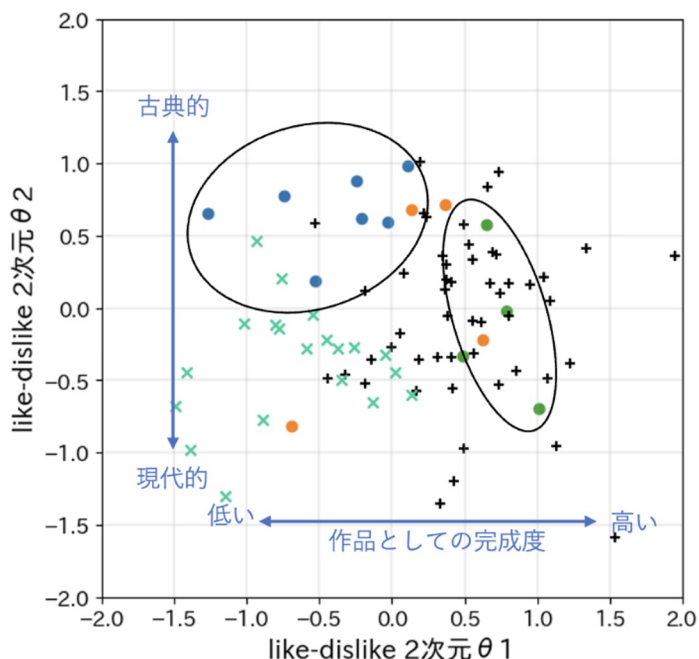
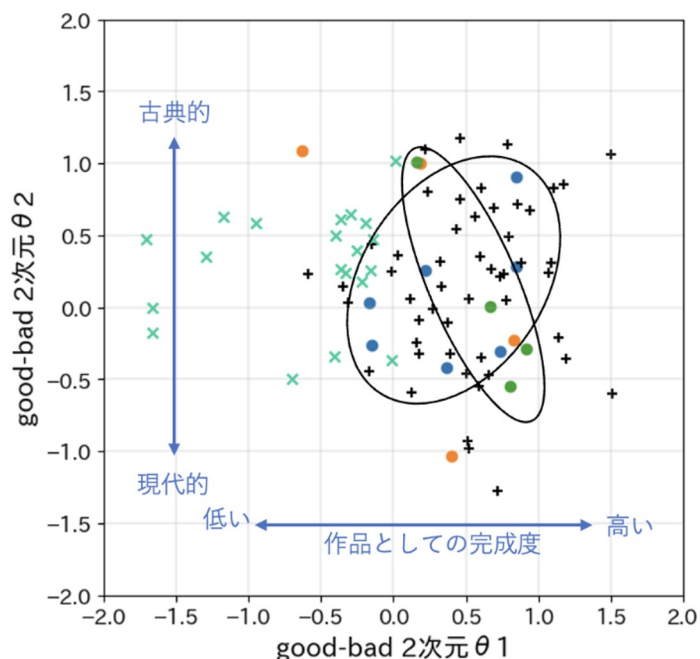
2次元 θ	小さい	大きい
θ_1	作品の質が低い	作品の質が高い
θ_2	現代的	古典的



作者に関する分析

- 評価対象100首のうち同一作者の作品
 - 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
- 山中の作品
 - Good-bad評価: θ_1, θ_2 ともに大きい
 - Like-dislike評価: θ_1 は小さく θ_2 は大きい

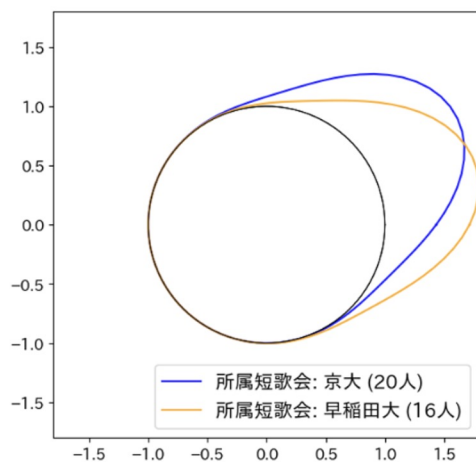
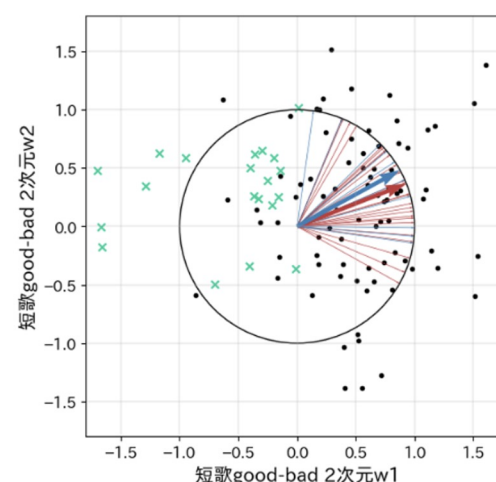
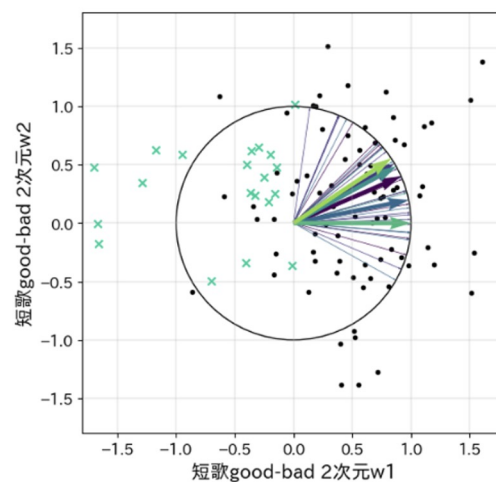
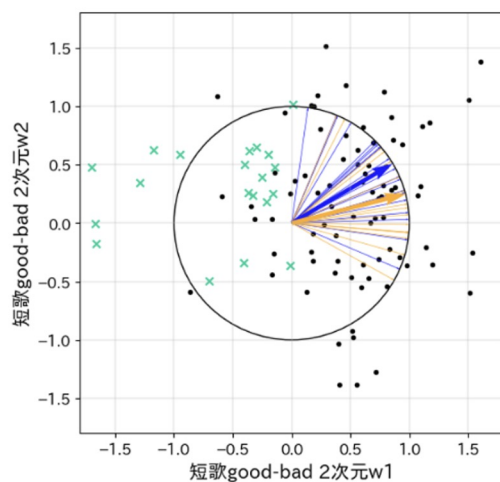
評価者のコメント：
古典的な作品に対し
「わからない」「現代短歌のほうが
読み慣れているので評価しづらい」



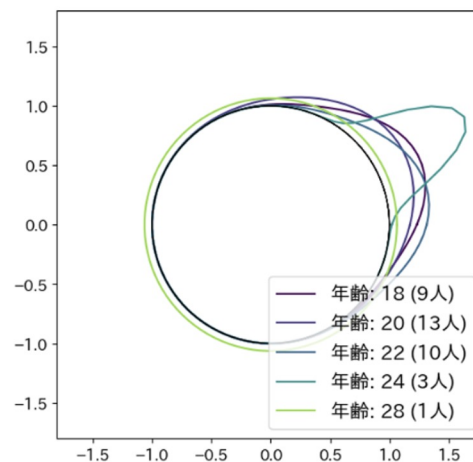
作者別の傾向の違いを分析するためには
より多くの同一作者の作品に対する評価データが必要

評価者グループの傾向

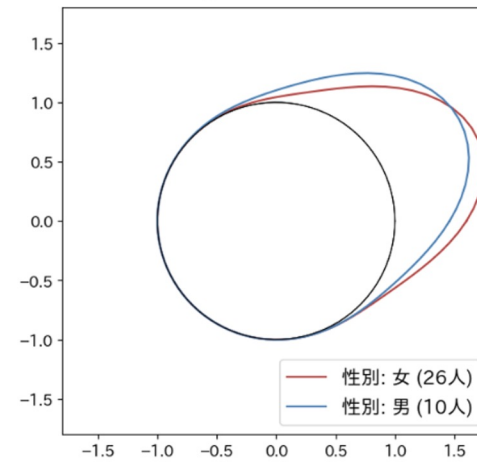
- 短歌good-bad評価の場合
 - 所属短歌会、年齢、性別のグループ間で評価傾向に差異



(a) 所属短歌会



(b) 年齢



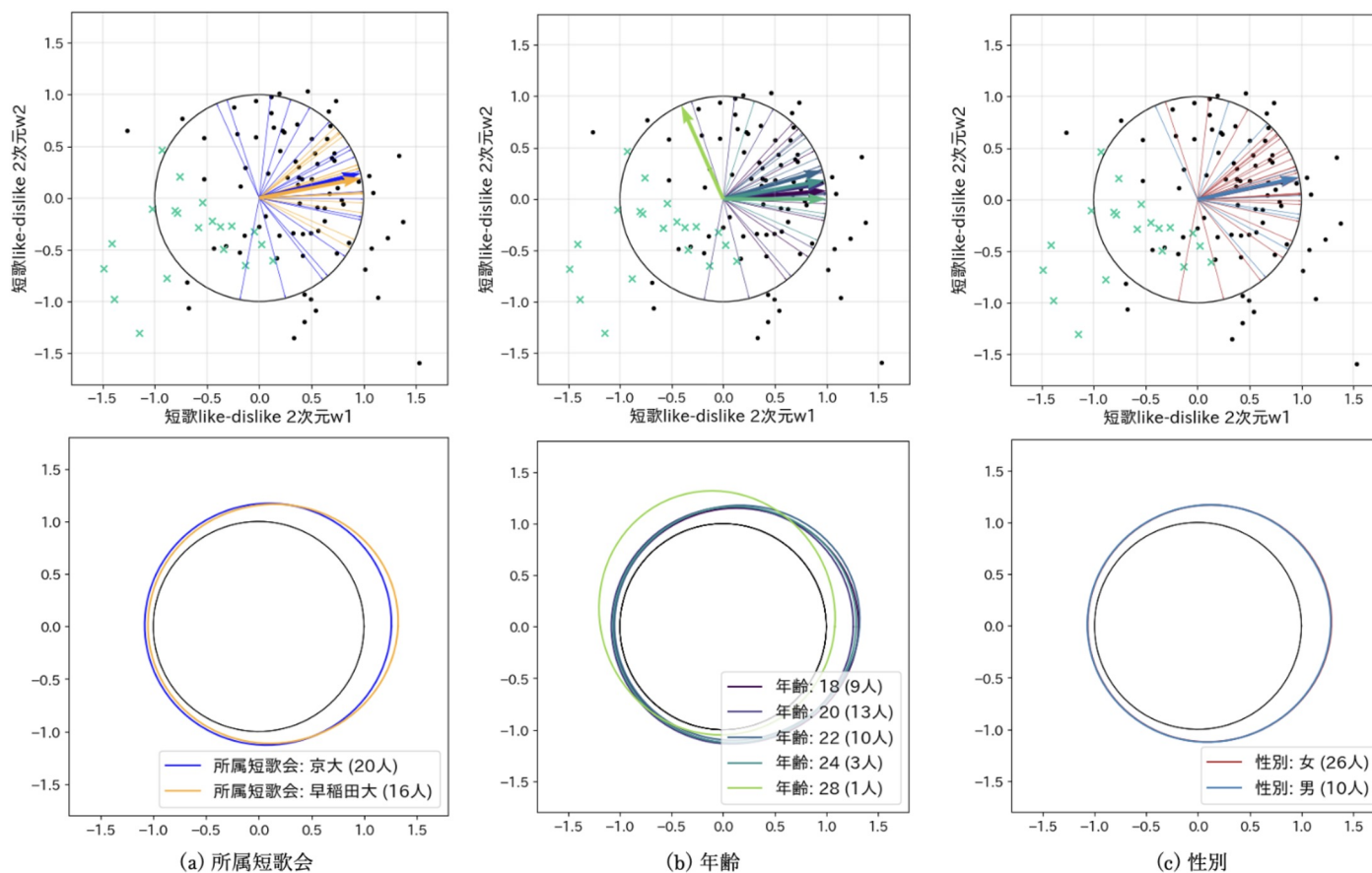
(c) 性別

評価者グループの傾向

• Like-dislike評価の場合

- Good-bad評価に比べ所属短歌会、年齢、性別グループ間ともに大きな差異は見られない

→ より個人的な好みが見られる



まとめ

まとめ

- 短歌作品の段階評価データに対して項目反応理論に基づいた分析手法を提案
- 短歌の潜在的優劣を表す θ , 評価者の傾向 a, b をパラメータとして推定
- 独自に収集した短歌データにおいて分析を実施

今後の展望・予定

- θ をより高次元（3次元以上）にした場合の各次元の要素の分析
- θ の値との短歌作品の内容（使用されている単語）との関係性に関する分析
- 短歌以外の短型文学（俳句や川柳など）で同様の分析

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

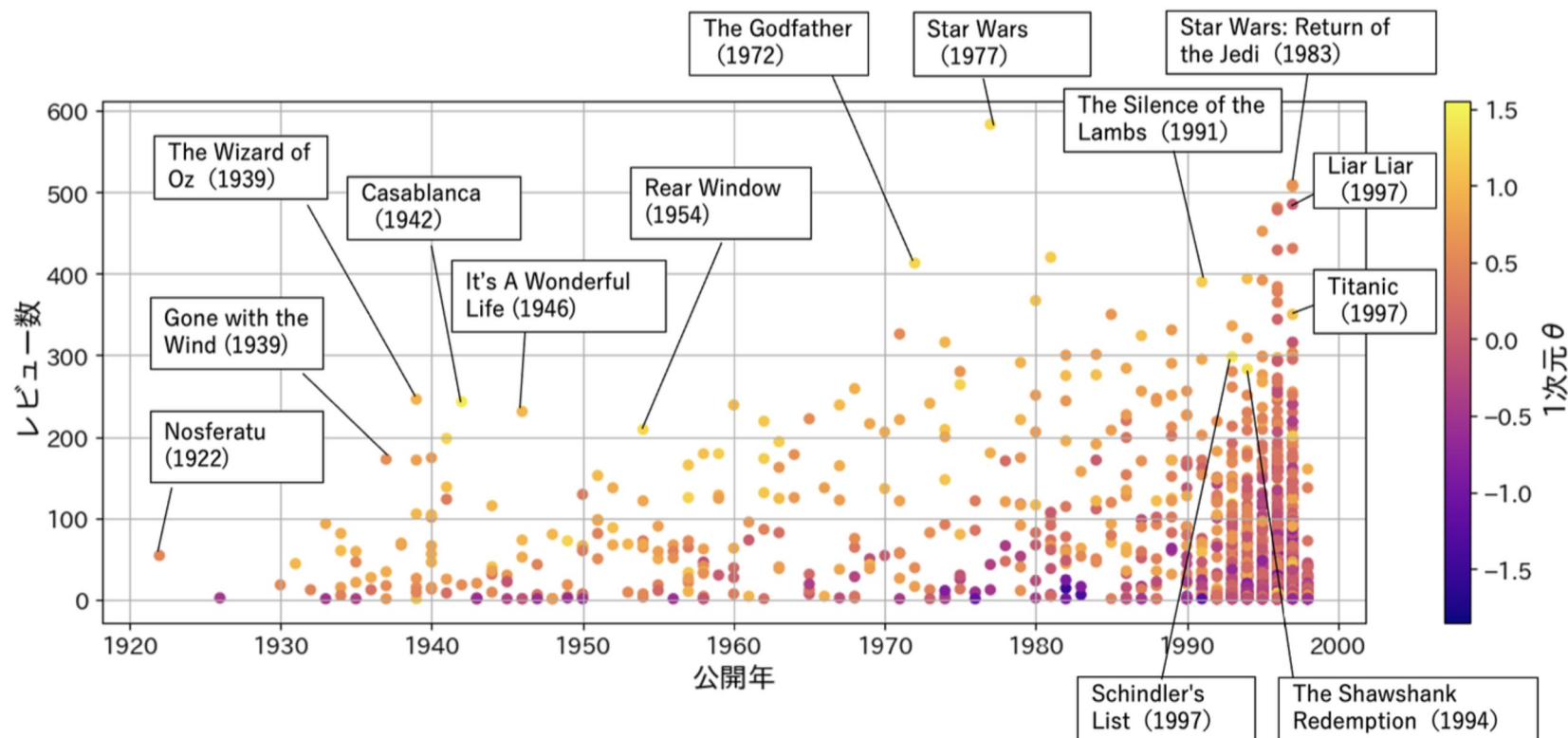
- [1] Bock, R. D. and Aitkin, M.: Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm, *Psychometrika*, Vol. 46, No. 4, pp. 443–459 (1981).
- [2] Harper, F. M. and Konstan, J. A.: The movielens datasets: History and context, *Acm transactions on interactive intelligent 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 TR
- [7] 野口あや子: 歌集『くびすじの欠片』, 短歌研究社 (2009).
- [8] 黒瀬珂瀾: 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). *ANSATIONS on Inf. Syst.*, pp. 377–388 (2008).

分析2: 映画の評価

- 評価データ 100,000件
 - 各映画に対して付与された評価数 1~583件 (平均59.5件)
 - 各評価者の評価数 20~737件 (平均106.0件)
- 評価対象の映画 : 1,682 件
 - 1922年~1998年公開の作品
 - 22種類のジャンルが0個以上付与される (最大6, 平均1.72)
- 評価者: 943 名
 - 属性情報: 性別, 年齢, 職業, Zipコード
 - 職業はartist, doctorなど21種類(other, noneを含む)

θ が1次元の場合

- レビュー数が多い作品の方が θ の値が大きい
- 公開年が古い作品の方が θ の値が大きい



θ が1次元の場合

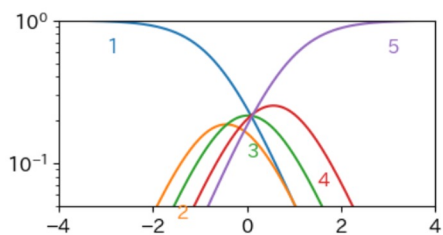
- θ の値が大きい作品
 - 古典的な名作, 戦争や生命, 子供向けのコメディ映画(一部)など
- θ の値が小さい作品
 - ホラー映画, シリーズものなど

作品名 (公開年)	レビュー数	θ
<i>Pather Panchali</i> (1996) 大地のうた	8	1.552
<i>Casablanca</i> (1942) カサブランカ	243	1.436
<i>Wrong Trousers, The</i> (1993)	118	1.419
<i>Schindler's List</i> (1993) シンドラーのリスト	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

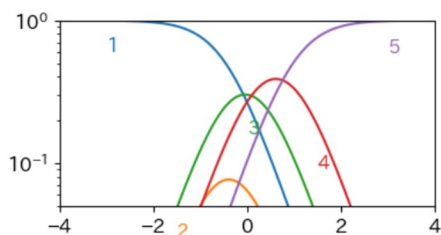
ジャンル	作品数	θ の平均
FilmNoir	24	0.583
War	71	0.464
Documentary	50	0.388
Musical	56	0.279
Mystery	61	0.277
Drama	725	0.272
Romance	247	0.225
Animation	42	0.222
Crime	109	0.211
Thriller	251	0.103
SciFi	101	0.093
Adventure	135	0.015
Comedy	505	0.006
Action	251	-0.054
Fantasy	22	-0.178
Children	122	-0.181
Horror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

θ が1次元の場合 | 評価者の項目反応曲線

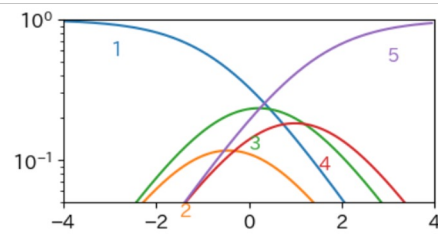
- 職業や性別特有の項目反応曲線の特徴は見られず



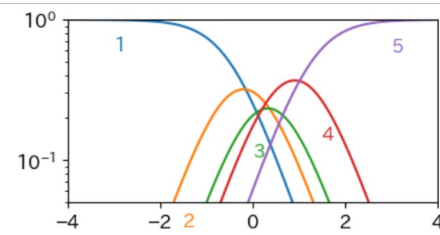
ID:102 男,38,programmer,30220(216)



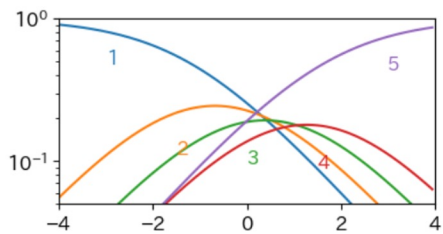
ID:171 女,48,educator,78750(28)



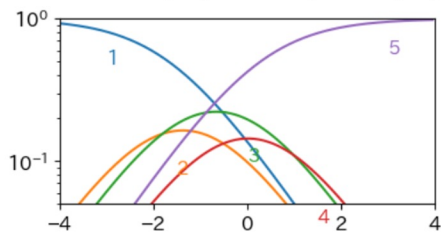
ID:1 男,24,technician,85711(272)



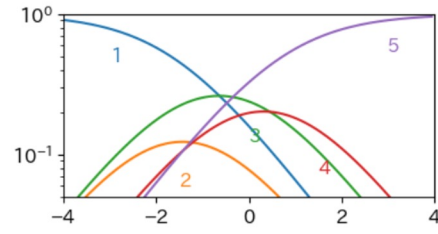
ID:9 男,29,student,01002(22)



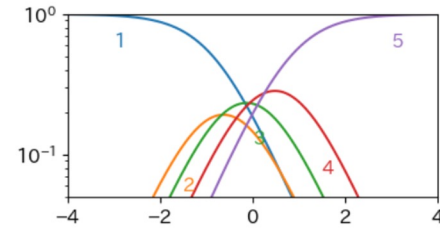
ID:155 女,32,other,11217(22)



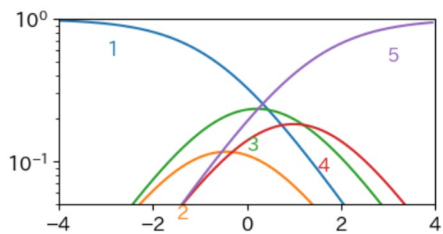
ID:27 女,40,librarian,30030(25)



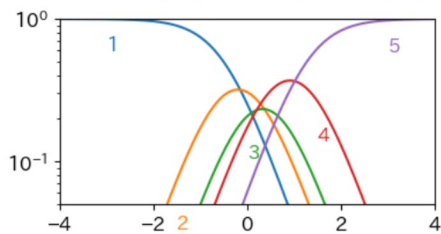
ID:109 男,29,other,55423(234)



ID:38 女,28,other,54467(121)

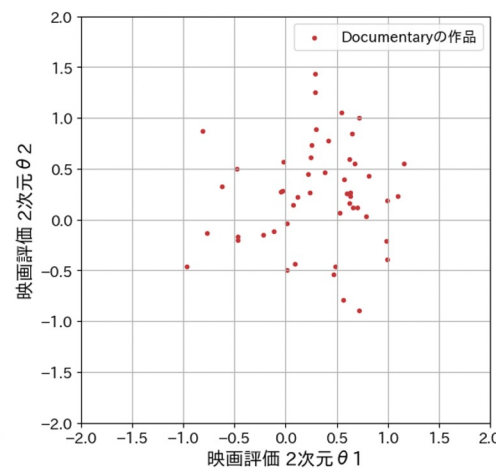
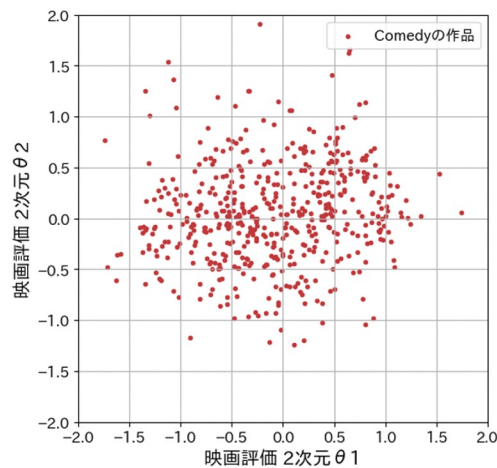
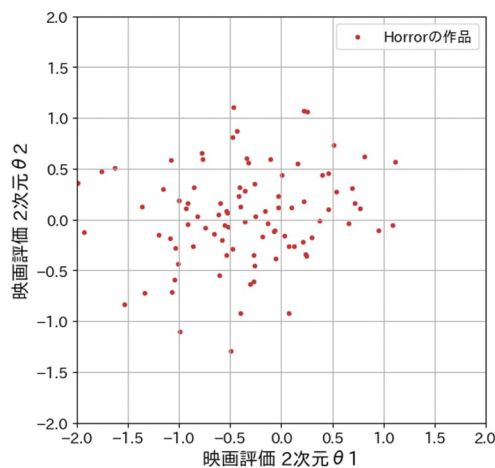


ID:1 男,24,technician,85711(272)



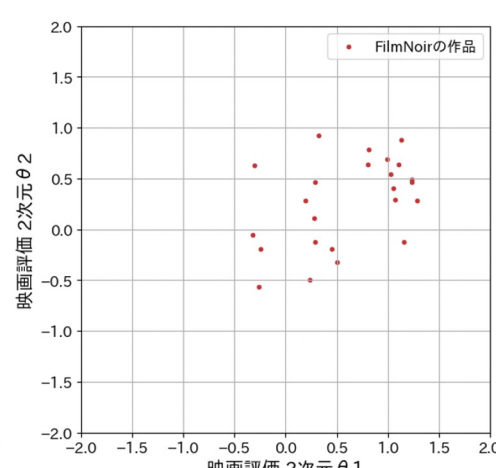
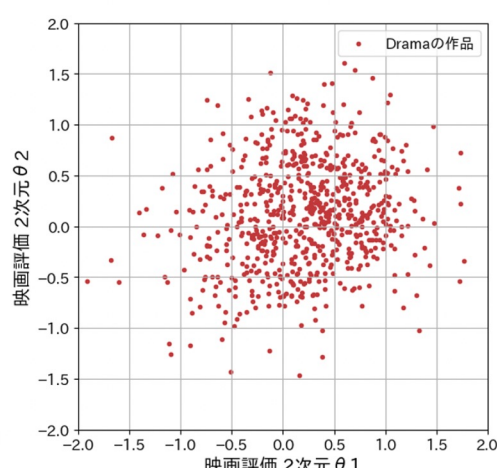
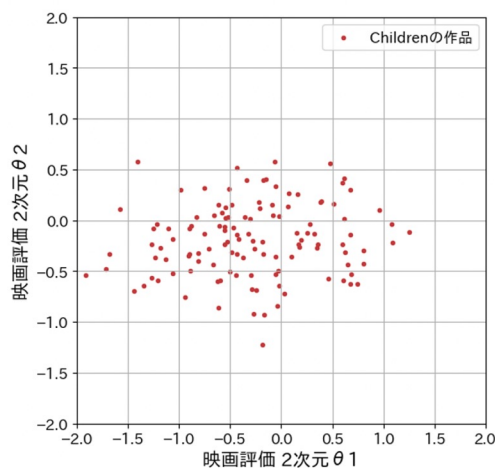
ID:9 男,29,student,01002(22)

映画評価の2次元 θ のジャンル別の表示



新しい
前衛的

古い
大衆的



新しい
前衛的

古い
大衆的

映画評価の職業別の評価方向

- Scientist, Writer, Artist など→ θ_2 の高い作品 (前衛的)を評価
- Doctor, Executive, Homemaker など→ θ_2 の低い作品 (大衆的)を評価

