

回答時間を用いた一対比較法の精度向上に関する研究

坂本 佳樹^{*1} 久留島 隆史^{*1} 上田 樹美^{*1} 石井 裕剛^{*1} 下田 宏^{*1}
望月 理香^{*2} 渡辺 昌洋^{*2}

A Study on Accuracy Improvement of Pairwise Comparison by Using Response Time

Yoshiki Sakamoto^{*1}, Takashi Kurushima^{*1}, Kimi Ueda^{*1}, Hirotake Ishii^{*1}, Hiroshi Shimoda^{*1},
Rika Mochizuki^{*2} and Masahiro Watanabe^{*2}

Abstract – Pairwise comparison has been widely employed as an easy method for subjective evaluation of several candidates to develop new products or to improve service. In this paper, a method using response time has been proposed to improve the accuracy of pairwise comparison. Firstly, a pairwise comparison experiment was conducted to investigate the relationship between response time and difficulty of judgement. Then, a method to improve the accuracy of calculated scales of the objects has been proposed where the answer of the pairwise comparison was modified by using the response time and correction function. In this study, three correction functions were compared to find the difference of the accuracy improvement, and two of them showed significant improvement.

Keywords : pairwise comparison, response time, correction function

1. はじめに

今日では、新製品開発やサービス向上のために、人間の五感や感情、見やすさ、使いやすさ、好感度、快適さなどの「感じ方」を中心に製品やサービスの改善を試みる事が多くなっている。

このように人間の感性や感覚を基に製品等を評価することを感性評価と呼ぶ。感性評価は、従来食品業界や化粧品業界などを中心に広く用いられてきたが、近年では自動車業界、アパレル業界、電気機器製造業界、サービス業界など、活用される業界も多様化してきている^[1]。感性評価により、様々な業界で人間の感性のなかに隠されているニーズを見出すことが可能となった^[2]。

感性評価には様々な種類があるが、その中でも2つの比較対象を評価者に提示し、比較判断の結果で対象を評価する一対比較法は、その簡単さからしばしば利用されている。一般に一対比較法は、一度に比較をする対象が2つのみであり、また絶対的な評価ではなく相対的な評価を要求することから評価者の負担が少ないという利点がある。また、回答結果から各刺激の順位だけではなく、各刺激を数値化し間隔尺度上の尺度値で表現することが可能である。さらは一対比較法では各刺激を総当たりで比較させることから、刺激間の順位をつける順位法や各刺激について点数をつける採

点法と比べて評価結果の情報量が多く、より精度の高い要因分析が可能である。

例えば、一対比較法の精度の評価に関して、田畑らは撮像シーケンスの異なる6枚のMR画像を一対比較法により評価することで、サーストンの一対比較法による尺度値の信頼性を評価した^[3]。中江らは、温度の異なる5種類の熱刺激を一対比較法を用いて提示し、一対比較結果から脳波データの精度を検証した^[4]。一対比較法の精度の向上に関しては、鳥海らは情報を欠落させることなく比較回数を減らすための具体的な比較順序を含む逐次一対比較法のアルゴリズムを提案した^[5]。また、豊田らは、共分散構造分析を用いて一対比較法により収集したデータに対して実験計画法の観点から要因解析を行うモデルを提案した^[6]。このように、一対比較法は多様な分野で利用されている。

一対比較法には様々な手法がある。代表的なものとして、刺激間の優劣のみで判断するサーストンの一対比較法や、評点をつけるシェッフェの一対比較法がある。サーストンの一対比較法は、「左の方が好み/右の方が好み」などの二者択一で回答する2件法を採用しているため、判断が簡単である。一方で、シェッフェの一対比較法は5件法や7件法などにより多段階の評点を要求するため、判断が複雑であり、評価者への負担が大きいというデメリットがある。シェッフェの一対比較法は多段階の評点で回答されるため、サーストンの一対比較法と比べて評価者の回答の情報量が多く、より精度の高い評価結果を得られることが期待できる。しかし現実には、幼児など複雑な判断が困難な

*1: 京都大学大学院エネルギー科学研究科

*2: NTT サービスエボリューション研究所

*1: Graduate School of Energy Science, Kyoto University

*2: Organization of Service Evolution Laboratories, NTT Group

評価者が評価する場合や、スポーツの結果など勝ち、負けの二値でしかデータを収集できない場合が多くあり [7], シェツフェの対比較法が適用できない場面も多い。

以上の理由から、汎用性を考慮するとサーストンの対比較法を用いることが望ましい。しかし、サーストンの対比較法では、全回答を通して一方の選択肢のみが選択された比較対があると、尺度値計算の際に欠測値が発生し尺度値を得ることができない場合がある。したがって、一方の選択肢のみが選ばれないようにするために多くの回答を得る必要がある。また、得られた尺度値は必ず誤差を含み [3], この誤差を小さくすることが課題として挙げられる。そこで、本研究では回答にかかった時間に着目し、サーストンの対比較法の改善を試みる。

2. 対比較法に関する既往手法

2.1 サーストンの対比較法

サーストンの対比較法とは、複数の評価対象から任意の2つの対象を取り出し比較し、全ての比較結果から評価対象を評価する対比較法の中でも、「はい/いいえ」などの二者択一で回答させる方法である。ここで、評価する対象を「刺激」と呼び、比較時に評価対象から取り出した2つの刺激を「対」と呼ぶ。

n 個の刺激があり、そこから対 (j, k) を作成する $(j, k = 1, 2, \dots, n)$ 。対 (j, k) に対して、 j, k のどちらかを評価者に選択させたとき、 j が選択された確率を p_{jk} とする。各対 (j, k) について、 N 回の独立な比較判断を行ったとする。この時、尺度値 $S_D(j)$ は正規分布に従うと仮定すると、

$$p_{jk} = \frac{1}{\sqrt{2\pi}} \int_{x_{jk}}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx \quad (1)$$

$$\sum_j \{x_{jk} - S_D(k) + S_D(j)\} = 0 \quad (2)$$

したがって、尺度値 $S_D(k)$ は下記のようにして求められる。

$$S_D(k) = \frac{1}{n} \sum_{j=1}^n x_{jk} \quad (3)$$

2.2 ギューリクセンの方法

式 (1) からわかるように、 p_{jk} が 1 または 0 となったとき、 x_{jk} は $+\infty$ または $-\infty$ となり、 x_{jk} を要素とする行列 \mathbf{X} に欠測値が出てしまう。このような不完備なデータに対する計算法として、ギューリクセンの方法が挙げられる。

\mathbf{X} が不完備行列で、その中に数値の存在しない空

欄を含む場合、式 (2) は、

$$\sum_j^{n_k} \{x_{jk} - S_D(k) + S_D(j)\} = 0 \quad (4)$$

と表せる。ただし、 \mathbf{X} の k 列に存在する x_{jk} の数を n_k とし、 x_{jj} は数えない。したがってこれを行列で表すと、

$$\mathbf{Z} = \mathbf{M}\mathbf{S} \quad (5)$$

と書ける。ただし、

$$\mathbf{Z} = \begin{pmatrix} \sum_j^{n_1} x_{j1} \\ \sum_j^{n_2} x_{j2} \\ \vdots \\ \sum_j^{n_n} x_{jn} \end{pmatrix} \quad (6)$$

$$\mathbf{S} = \begin{pmatrix} S_D(1) \\ S_D(2) \\ \vdots \\ S_D(n) \end{pmatrix} \quad (7)$$

また、 \mathbf{M} は、その主対角線に n_k 、それ以外の欄には、 x_{jk} が存在するところには -1 、欠測値となっているところには 0 を入れた $n \times n$ の対称行列である。ここで、例えば $S_D(1) = 0$ とし、尺度 S_D の原点を定義すると、行列 \mathbf{S} 、行列 \mathbf{Z} から、その一番上の要素を除き、行列 \mathbf{M} から、その第 1 行、第 1 列を除いた行列を考え、それぞれ \mathbf{S}_1 、 \mathbf{Z}_1 、 \mathbf{M}_{11} と書けば、式 (5) は

$$\mathbf{S}_1 = \mathbf{M}_{11}^{-1} \mathbf{Z}_1 \quad (8)$$

と書ける。これにより、 $S_D(1)$ を原点においた $S_D(k) (k \geq 2)$ の数値が求まる。

3. 提案手法

3.1 提案手法の概要

サーストンの対比較法では、評価者は提示された2つの選択肢のうちどちらかの選択肢を選択しなければならず、評価者が迷わずに出した回答と迷ったうえで出した回答は同一の回答情報として獲得される。しかし実際は、回答には評価者が確信を持って迷わずに回答したものと確信を持たずに迷ったうえで回答したものが混在している。評価者が迷わずに出した回答は、評価者は確信を持って回答したため回答結果の信用性が高い一方で、評価者が迷ったうえで出した回答は、評価者は確信を持たず回答したため回答結果の信用性は低いと考えられる。ここで、回答時間と回答時の迷

いについて、「評価者の回答時間が短いとき、評価者は迷わずに回答したため、評価者の回答に対する確信度は高く、評価者の回答時間が長いとき、評価者は迷ったうえで回答したため、評価者の回答に対する確信度は低い」と仮定する。そのうえで、本提案手法では、回答時間から得た「迷い」の情報を回答結果に反映することで、一対比較法の精度向上を試みる。そこで、回答結果と回答時間から、回答結果を修正する関数（以下「修正関数」と呼ぶ）を考える。一対比較において、刺激対 (j, k) が与えられたときに刺激 j を選択した場合、 $p_{jk} = 1, p_{kj} = 0$ と表記する。このとき、回答結果 p と回答時間 t を入力とすると、例えば、

$$(p_{jk}, p_{kj}) = (1, 0) \xrightarrow{f} (0.6, 0.4) \quad (9)$$

のように回答結果が修正されるような修正関数 $f(p, t)$ を考える。

3.2 提案する修正関数

本研究では修正関数を複数提案し、その中から最適なモデルを生成することを目的とする。そこで、まず修正関数が満たすべき条件を以下にまとめる。

選択肢 (j, k) が与えられたとき、 j を選択した場合 $p_{jk} = 1, p_{kj} = 0$ とする。サーストンの一対比較法では、どちらかの選択肢を選ばないといけないため、以下を満たす。

$$p_{jk} + p_{kj} = 1 \quad (10)$$

$$p \in \{0, 1\} \quad (11)$$

回答結果 p と回答時間 t を入力とする修正関数 $f(p, t)$ は確率を出力する関数であるとする。したがって、

$$f(p_{jk}, t) + f(p_{kj}, t) = 1 \quad (12)$$

$$0 \leq f(p, t) \leq 1 \quad (13)$$

を満たす。また、回答結果の優劣が逆転することはないものとする。すなわち、

$$f(0, t) \leq f(1, t) \quad (14)$$

を満たす。さらに、回答時間が増加すると回答結果 p は単調に 0.5 に近づくものとする。すなわち、

$$\begin{cases} \frac{\partial f(1, t)}{\partial t} \geq 0 \\ \frac{\partial f(0, t)}{\partial t} \leq 0 \end{cases} \quad (15)$$

$$\lim_{t \rightarrow \infty} f(p, t) = 0.5 \quad (16)$$

を満たす。

式 (10) から式 (16) を満たし、一対比較法の精度向上を実現する関数の候補として、以下の「修正関数 1」

および「修正関数 2」を作成した。ただし、選択結果を p 、回答時間を t とし、 x_0, x_1 は関数のパラメータである。

3.2.1 修正関数 1

以下の修正関数 1: $f_1(p, t)$ を定義する。修正関数 1 の概形を図 1 に示す。

$$f_1(p, t) = g_1(t) \cdot (p - 0.5) + 0.5 \quad (17)$$

$$g_1(t) = \frac{1}{1 + e^{x_0 \cdot (t - x_1)}} \quad (18)$$

$$x_0 > 0 \quad (19)$$

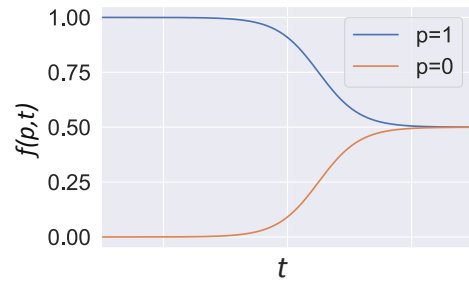


図 1: 修正関数 1 の概形
Fig. 1 Correction Function 1.

3.2.2 修正関数 2

以下の修正関数 2: $f_2(p, t)$ を定義する。修正関数 2 の概形を図 2 に示す。

$$f_2(p, t) = \frac{1}{1 + e^{-g_2(t) \cdot (p - 0.5)}} \quad (20)$$

$$g_2(t) = e^{-x_0 \cdot (t - x_1)} \quad (21)$$

$$x_0 > 0 \quad (22)$$

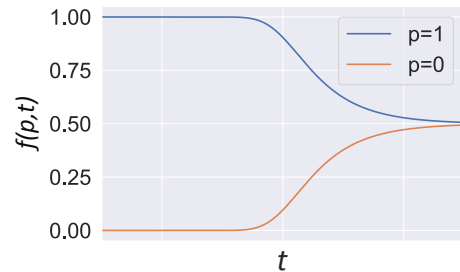


図 2: 修正関数 2 の概形
Fig. 2 Correction Function 2.

また、生理指標分野では、汎化性能に優れた手法として線形回帰法が適用されることが多い [8], [9]。そこで、式 (10) から式 (14)、式 (16) を満たす線形な関数として、以下の「修正関数 3」を作成した。

3.2.3 修正関数 3

以下の修正関数 3: $f_3(p, t)$ を定義する. 修正関数 3 の概形を図 3 に示す.

$$f_3(p, t) = g_3(t) \cdot (p - 0.5) + 0.5 \quad (23)$$

$$g_3(t) = \begin{cases} 1 & (t < -\frac{1}{2x_0} + x_1) \\ -x_0 \cdot (t - x_1) + 0.5 & (-\frac{1}{2x_0} + x_1 \leq t < \frac{1}{2x_0} + x_1) \\ 0 & (t \geq \frac{1}{2x_0} + x_1) \end{cases} \quad (24)$$

$$x_0 > 0 \quad (25)$$

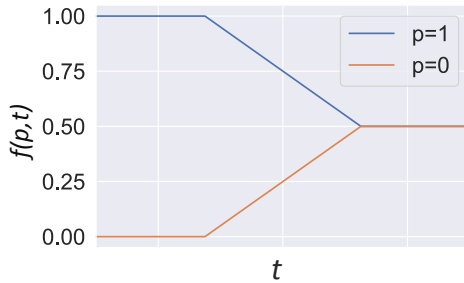


図 3: 修正関数 3 の概形
Fig. 3 Correction Function 3.

得られた回答結果行列 $\mathbf{p} = (p_{jk})$ および回答時間行列 $\mathbf{t} = (t_{jk})$ を修正関数に適用し, 修正後の結果行列 $\mathbf{f} = (f(p_{jk}, t_{jk}))$ を作成した後, 2.2 節で説明したギューリックセンの方法で尺度値を算出する.

3.3 提案手法の精度評価方法

本研究では, 一対比較法の精度を評価するため, 得られた尺度値と提示刺激を比較する必要がある. そこで, 提示する刺激は実際の物理量を数値化できるものを使用し, 得られた各刺激の間隔尺度上の尺度値と各刺激の実際の物理量がどれだけ一致しているかを比較する. この時, 算出された各刺激の尺度値と, 実際の各刺激の物理量は線形関係にあると仮定する. そこで, 尺度の精度判定に決定係数を用いる. すなわち, 実際に提示した各刺激の物理量を説明変数, 算出された尺度値を目的変数として単回帰分析をし, 単回帰式から決定係数を算出する. 一般に, i 番目 ($i = 1, 2, \dots, n$) の実測値を y_i , 単回帰式による推定値を \hat{y}_i , 全データから求められる平均値を \bar{y} とすると, 決定係数 R^2 は

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (26)$$

と表される. 決定係数 R^2 が 1 に近いほど, 尺度の精度が高いことを表す.

3.4 提案する一対比較法の精度向上方法

本研究で提案する一対比較法の精度向上方法を図 4 に示す. まず, 修正関数により得られた修正後結果行列 $\mathbf{f} = (f(p_{jk}, t_{jk}))$ から, 3.1.2 項で説明したギューリックセンの方法で尺度値を算出する. その後, 3.2.2 項で説明した修正関数の x_0, x_1 を変数として, 決定係数を最大化するよう最適化をする. なお, 同一刺激対の比較が複数回実施された場合, $f(p_{jk}, t_{jk})$ の平均の値を尺度化に用いる. 最適化には差分進化法を採用する^[10]. 決定係数が最大になるよう x_0, x_1 を最適化することで, 尺度の精度が最高となるような修正関数が作成される.

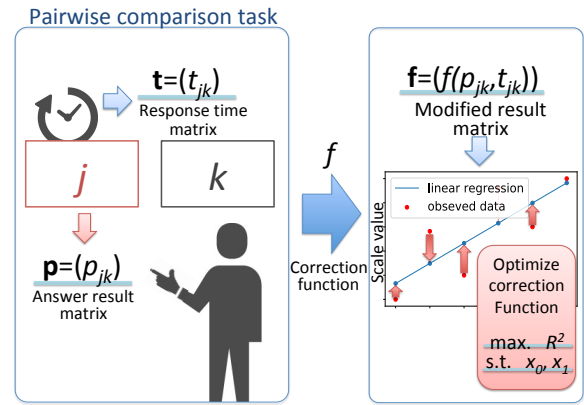


図 4: 精度向上手法の概要
Fig. 4 Accuracy Improvement Method.

4. 回答難易度と回答時間の関連を調べる実験

本実験では, 3 章で説明した手法で計算された間隔尺度上の尺度値の精度を評価する前に, 一対比較法における回答結果と回答時間を計測し, 提示された刺激の難易度とその回答時間の関連を調査することを目的とする.

4.1 実験参加者

実験には男子大学生 28 名が参加した. 参加の条件は, 裸眼または眼鏡やコンタクトにより矯正した状態で正常な視覚を有する男性とした. 参加者全 28 名の平均年齢は 21.4 歳 (SD=2.0) であった.

4.2 実験環境

実験は, 大学構内の実験室で実施した. 実験ではディスプレイに表示された一対比較問題をマウスを用いて回答してもらった. 使用したディスプレイ (I-O DATA, LCD-MF243EBR) は画面サイズ 23.6 インチ, 画面解像度 1920×1080 であった. 天井照明による机上面照度は 330lx であった.

4.3 使用タスク

本実験では一対比較問題として、2本の線を並べて表示し、より長いと感じた方を選ぶ「線の長さを問う一対比較問題」を作成した。タスクの概要を図5に示し、実際のタスク画面を図6に示す。

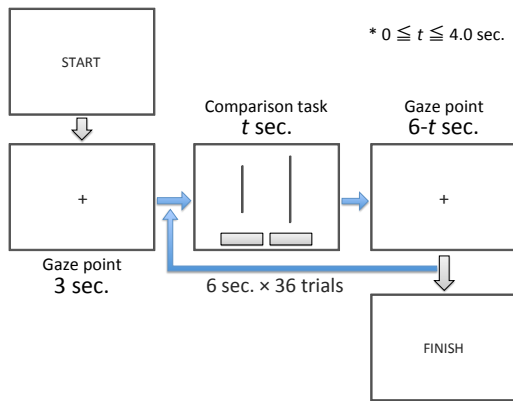


図5: タスクの概要
Fig. 5 Outline of the task.

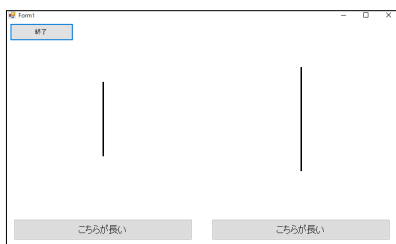


図6: 実際のタスク画面（比較課題）
Fig. 6 The actual graphical interface of the task (comparison task).

線の長さを問う一対比較問題を採用した理由は以下の2点である。まず、提示された2本の線分の実際の長さの差が小さいほどタスクの難易度が上がると考えられ、タスクの難易度すなわち迷いやすさを定量的に評価できるからである。もう1点は、実験環境の影響を受けにくいという点が挙げられる。例えば、明度の異なる2色を表示し、より濃いと感じた方を選択する「色の濃さを問う一対比較問題」は、難易度は明度差から定量的に評価できるが、刺激の感じ方はディスプレイや実験環境の明るさの影響を大きく受けてしまう。以上の理由により本実験では「線の長さを問う一対比較問題」を採用した。

本実験では、比較刺激をディスプレイ上で提示したが、比較刺激として縦ピクセル数が200px, 202px, 204px, 206px, 208px, 210pxである6種の垂直線を用

意した。実際のディスプレイ上での線の長さはそれぞれ約50.0mm, 50.5mm, 51.0mm, 51.5mm, 52.0mm, 52.5mmであった。また、実験参加者とディスプレイの距離から算出された視野角はそれぞれ約1.63°, 1.64°, 1.66°, 1.68°, 1.69°, 1.71°であった。線幅のピクセル数は全て3px (0.75mm)であった。刺激として上記の垂直線のうち2本を並べて表示し、左右の入れ替え、同一の線長の比較を含む36通りの線の対を作成した。上記の全36通りの対を比較する一連のタスクを1setとし、表示する刺激対を切り替える際には実験参加者の視点をディスプレイの中央に向けさせるために中央に十字（注視点）を表示した。比較問題はより長いと感じた方の線に対応するボタンをマウスでクリックすることで回答する。また、比較問題が提示されてから回答するまでの時間を回答時間として、回答時間を計測した。ただし、4.0秒以上無回答であった場合には、強制的に刺激対の提示を終了したのち注視点を表示し、回答結果は無回答とした。

4.4 実験手順

実験は2019年1月4日～19日の期間で実施した。まず、1分程度の間、本実験に用いるタスク画面の操作に慣れてもらうことを目的として操作を練習した。その後、4.3節で説明したタスクを12set実施した。各set間には3分程度の休憩時間を設けた。また、提示する対の順序はset毎にランダムな順番で提示した。

4.5 実験の結果

今回の実験では、全12set・計432試行のうち、36回以上制限時間内に回答せず無回答であった実験参加者、あるいは左右のうち一方を360回以上選択した実験参加者は真剣に実験に取り組んでいないとみなし、無効データとして解析対象から除いた。その結果、参加者28名のうち参加者番号5, 18の2名を無効データとして扱い、その2名を除いた26名を解析対象とした。また、無回答であった比較も無効データとし、解析の対象から外した。

各刺激対に対する回答時間の全参加者全set平均を図7に示す。縦軸は画面左に表示した線のpixel数、横軸は画面右に表示した線のpixel数を示しており、セル内の色が薄いほど回答に時間がかかったことを示している。線の長さの差が小さくなるにつれ回答時間が長くなる傾向があることが確認された。ただし、回答時間はsetごとに標準化した。これは、実験参加者ごとやsetごとの回答時間のばらつきを抑えるためである。

続いて各実験参加者の結果について述べる。ここでもsetごとの回答時間のばらつきを抑えるために、回答時間はsetごとに標準化した。結果の一例として、参加者番号14の全setの回答時間平均結果を図8に示す。線の長さの差が小さくなるにつれ回答時間

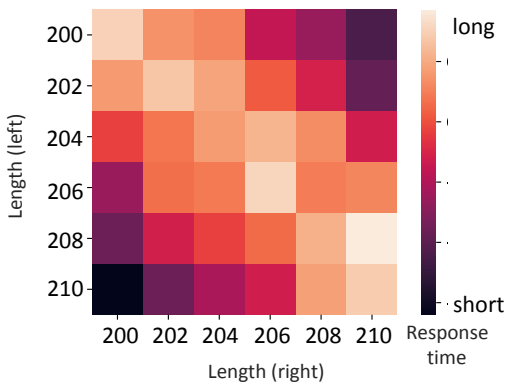


図 7: 全参加者の全 set の結果
Fig. 7 Result of all participants (average).

が長くなる傾向があることが確認された。一方で、図 9 に示す参加者番号 24 の全 set の回答時間平均結果のように、線の長さの差が小さくなるにつれ回答時間が長くなる傾向が見られない実験参加者も確認された。

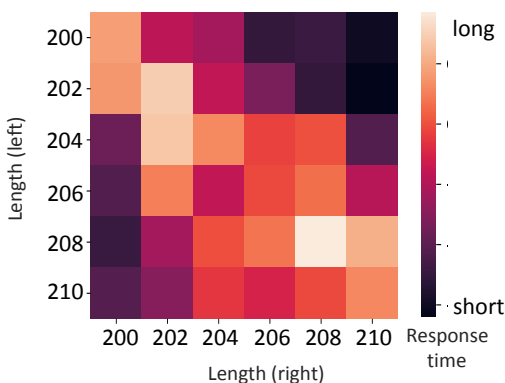


図 8: 参加者番号 10 の結果
Fig. 8 Result of participant no.10.

5. 提案手法の評価

本章では、4 章で測定した回答結果および回答時間を用いて、3.4 節で述べた提案手法の精度向上効果を評価する。

5.1 提案手法の評価方法

まず、3.4 節で説明した方法で、回答結果および回答時間データに各修正関数を適用し、実験参加者毎に決定係数を最大化するよう修正関数を最適化することにより、最適化済み修正関数を作成した。ただし、4.5 節と同様に、得られた回答時間データは set ごとに標準化した。その後、テストデータを用いて提案手法の精度向上効果を評価した。最適化済み修正関数の汎化

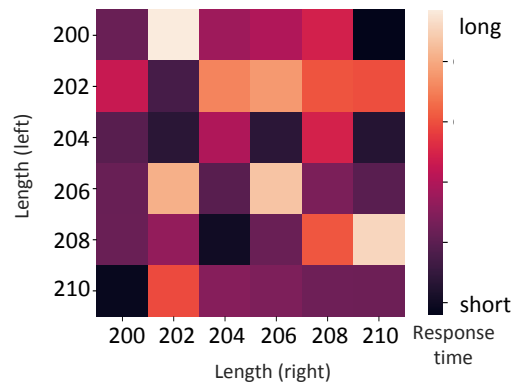


図 9: 参加者番号 24 の結果
Fig. 9 Result of all participant no.24.

性能の評価には交差検証法を用いた。交差検証法とは計測データを K 分割したうちの 1 つをテストデータ、残りを学習データとし、学習データを用いて訓練したモデルをテストデータに適用し、モデルの精度を評価する手法である。本研究では各実験参加者が実施した 12set を 4 分割した。9set の学習データから作成された最適化済み修正関数を作成し、残りの 3set をその最適化済み修正関数に適用し、決定係数を算出することで精度を評価した。その後、実験参加者あたり 4 分割したそれぞれの決定係数を算出し、平均値をその実験参加者の推定精度とした。

ただし、4.5 節と同様の理由で、実験参加者番号 5, 18 の 2 名を無効データとして扱った。また、実験参加者番号 7, 9, 14 の 3 名については、一部のテストデータの回答結果において欠測値が多く、式 (8) における M_{11} が逆行列の存在しない特異行列となった。そのため尺度値が計算できず、解析対象から除外した。その結果、残りの 23 名を有効データとして解析対象とした。

5.2 提案手法の評価結果

本節では、3.3 節で述べたように、求めた尺度値と物理的な長さとの相関の決定係数を推定精度と考えた。式 (17) から式 (19) で表される修正関数 1 に関して、修正関数 1 適用前の推定精度の全参加者平均は 0.849 であり、修正関数 1 適用後の推定精度の全参加者平均は 0.907 であった。修正関数 1 適用前の推定精度と修正関数 1 適用後の推定精度で対応のある t 検定を行った結果、有意差が認められた ($p = 0.00184$)。

式 (20) から式 (22) で表される修正関数 2 に関して、修正関数 2 適用前の推定精度の全参加者平均は 0.849 であり、修正関数 2 適用後の推定精度の全参加者平均は 0.914 であった。修正関数適用前の推定精度と修正関数適用後の推定精度で対応のある t 検定を行った結

果、有意差が認められた ($p = 0.000740$) .

式 (23), (24) で表される修正関数 3 に関して、修正関数 3 適用前の推定精度の全参加者平均は 0.849 であり、修正関数 3 適用後の推定精度の全参加者平均は 0.883 であった。修正関数適用前の推定精度と修正関数適用後の推定精度で対応のある t 検定を行った結果、有意差は認められなかった ($p = 0.0526$) .

5.3 提案手法に関する考察

5.3.1 修正関数適用前と適用後の比較

本項では修正関数適用前の推定精度と修正関数適用後の推定精度を比較し、推定精度が下がった参加者および上がった参加者の代表例を挙げ、推定精度の変化の違いを考察する。

まず、すべての修正関数で推定精度が下がった参加者番号 24 について考察する。推定精度が下がった原因として、図 9 に示すように参加者番号 24 では回答時間データにおいて線の長さの差が小さくなるにつれ回答時間が長くなる傾向が見られず、「評価者の回答時間が短いとき回答に対する確信度が高く、評価者の回答時間が長いとき回答に対する確信度が低い」という修正関数作成時の仮定が成り立たなかったことが考えられる。一方で、すべての修正関数で参加者番号 10 は推定精度が上がったが、図 8 に示すように参加者番号 10 では回答時間データにおいて線の長さの差が小さくなるにつれ回答時間が長くなる傾向が見られたため、推定精度が上がったと考えられる。

5.3.2 修正関数間の比較

続いて本項では、各修正関数の修正関数適用後の推定精度を比較し、修正関数間で推定精度の差が大きかった参加者の代表例を挙げ、各修正関数の違いを考察する。

まず、修正関数 2 適用後の推定精度が修正関数 1 適用後の推定精度よりも大きく、修正関数 1 適用後の推定精度が修正関数 3 適用後の推定精度よりも大きかった参加者番号 26 について考察する。表 1 から表 3 に各修正関数の交差番号、パラメータ x_0 , x_1 の値、修正関数適用後の算出された各尺度値、修正関数適用前の決定係数、修正関数適用後の決定係数を示す。交差番号 1 から 3 について修正関数間で修正関数適用後の決定係数に大きな違いは見られなかった。これは交差番号 1 から 3 については全ての修正関数で x_0 が 0 ないし 0 に近い値をとっており、式 (10) から式 (16) より各修正関数は回答時間データを反映せず、修正関数適用後の決定係数は修正関数適用前の決定係数と近い値をとるからである。交差番号 4 については修正関数間で修正関数適用後の決定係数に大きな違いが見られた。

修正関数 1, 修正関数 3 では、 x_0 が大きく x_1 が小

さい値となったことにより、修正関数は回答時間が長い回答に関しては回答結果によらず 0.5 となるものとなり、修正関数適用前の高精度のデータを失ったため、修正関数 1 と修正関数 3 で決定係数の値が低くなったと考えられる。一方で、修正関数 2 については $x_0 = 0$ となっていたため、回答時間データの影響を受けず決定係数が低下しなかったと考えられる。

続いて、修正関数 2 適用後の推定精度と修正関数 1 適用後の推定精度は大きくは変わらなかったものの、修正関数 1 適用後の推定精度が修正関数 3 適用後の推定精度よりも大きかった参加者番号 22 について考察する。表 4 から表 6 に各修正関数の交差番号、パラメータ x_0 , x_1 の値、修正関数適用後の算出された各尺度値、修正関数適用前の決定係数、修正関数適用後の決定係数を示す。交差番号 1, 3, 4 について修正関数間に大きな違いは見られなかったが、交差番号 2 については修正関数 3 のみ修正関数適用後の決定係数が大きく低下した。交差番号 2 では修正関数適用前の決定係数が低く、回答結果自体による誤差が大きかった。しかし、修正関数 1 や修正関数 2 のように x_1 の値が極めて小さい場合は、式 (10) から式 (16) より、多くの回答で回答時間や回答結果に関わらず修正関数は 0.5 となるため、元の低い決定係数が向上した。しかし、修正関数 3 では x_1 の値が比較的大きく、元の決定係数の低さの影響を強く受けたため、決定係数が低下したと考えられる。

6. 結論

本研究では、回答時間を用いたサーストンの一対比較法の精度向上手法を提案した。

まず、一対の線分の長さを比較する実験から回答時間を測定し、比較問題の難易度と回答時間の関係を調査した。本実験では男子大学生 28 名が参加し、2 名の無効データを除いた全 26 名を調査対象とした。その結果、実験参加者全体で平均した場合は線の長さの差が小さくなるにつれ回答時間が長くなる傾向が見られた。また、実験参加者を個別に調査したところ、その傾向が見られない実験参加者もいた。

その後、上記の実験で計測した回答結果と回答時間

表 1: 修正関数 1, 実験参加者番号 26

Table 1 Results (correction function 1, participant number 26)

iteration number	x_0	x_1	R^2 (before)	R^2 (after)
1	0.294	1.463	0.963	0.941
2	0.000	-2.076	0.862	0.943
3	0.000	-2.997	0.953	0.967
4	2.725	-1.471	0.921	0.768

表 2: 修正関数 2, 実験参加者番号 26

Table 2 Results (correction function 2, participant number 26)

iteration number	x_0	x_1	R^2 (before)	R^2 (after)
1	0.185	3.000	0.963	0.935
2	0.000	-1.524	0.862	0.942
3	0.000	2.620	0.953	0.967
4	0.000	-0.672	0.921	0.993

表 3: 修正関数 3, 実験参加者番号 26

Table 3 Results (correction function 3, participant number 26)

iteration number	x_0	x_1	R^2 (before)	R^2 (after)
1	0.073	1.000	0.963	0.939
2	0.000	0.486	0.862	0.943
3	0.000	0.572	0.953	0.967
4	4.590	-0.869	0.921	0.618

表 4: 修正関数 1, 実験参加者番号 22

Table 4 Results (correction function 1, participant number 22)

iteration number	x_0	x_1	R^2 (before)	R^2 (after)
1	3.000	0.261	0.962	0.927
2	0.393	-3.000	0.659	0.913
3	0.562	-3.000	0.450	0.960
4	3.000	-0.230	0.940	0.950

表 5: 修正関数 2, 実験参加者番号 22

Table 5 Results (correction function 2, participant number 22)

iteration number	x_0	x_1	R^2 (before)	R^2 (after)
1	0.421	-3.000	0.962	0.952
2	0.264	-3.000	0.659	0.911
3	0.448	-3.000	0.450	0.960
4	0.450	-3.000	0.940	0.973

表 6: 修正関数 3, 実験参加者番号 22

Table 6 Results (correction function 3, participant number 22)

iteration number	x_0	x_1	R^2 (before)	R^2 (after)
1	5.000	-0.296	0.962	0.953
2	5.000	0.432	0.659	0.468
3	4.985	0.309	0.450	0.934
4	4.987	-0.354	0.940	0.970

をもとに、一対比較法の精度向上を試みた。回答時間から回答結果を修正する「修正関数」を3種類作成し、その精度を評価した。その結果、2種類の修正関数では有意に精度が向上し、1種類の修正関数では有意に精度は向上しなかった。

本研究では、一対比較問題として「線の長さを問う一対比較問題」を用いた。しかし、1種類の一対比較問題についてしか調査しておらず、本研究だけでは提案手法により回答時間を用いて一対比較法の精度が向上するとは言えない。そのため、本提案手法の実用化にあたっては他に物理量が分かっている刺激を用いた一対比較問題を作成し、実験・調査する必要がある。また、本研究では3種類の修正関数を作成したが、種類を増やすなどしてより適切な修正関数を模索する必要がある。

参考文献

- [1] 長沢 伸也, 神田 太樹: 数理的感性工学の基礎, 海文堂出版, 2010
- [2] 長町 三生, 石原 茂和, 松原 行宏, 土屋 敏夫, 神田 太樹, 井上 勝雄: 商品開発と感性, 海文堂出版, 2005
- [3] 田畑 洋二, 大賀 泰文, 角田 充弘, 中前光弘, 森岡 雅幸, 宇都 文昭, 奥西 孝弘, 越智 保, 前田 要孝: 一対比較法(ケースV)における主観的尺度値の信頼性について, 日本放射線技術学会雑誌, vol.51, no.4, pp.445-449, 1995
- [4] 中江 文, 曾雌 崇弘, 次田 裕美, 岸本 千恵, 加藤 浩介: 実験的熱刺激を用いた脳波による痛みの客観的評価, PAIN RESEARCH, vol.33, no.1, pp.40-46, 2018
- [5] 鳥海 不二夫, 小林 彬, 高山 潤也, 大山 真司: 逐次一対比較法の提案, 計測自動制御学会論文集, vol.43, no.10, pp.839-846, 2007
- [6] 豊田 秀樹, 室橋 弘人, 尾崎 幸謙, 芳賀 麻誉美: 実験デザインに基づく一対比較データの解析 - 構造方程式モデリングによる表現-, 心理学研究, vol.75, no.1, pp.33-40, 2004
- [7] 宇佐美 慧: 比較判断に影響する複数の要因を考慮した一対比較データ分析法, 心理学研究, vol.79, no.6, pp.536-541, 2009
- [8] Rosalind W. Picard, Elias Vyzas, Jennifer Healey: Toward Machine Emotional Intelligence: Analysis of Affective Physiological State, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.23, no.10, pp.1175-1191, 2001
- [9] Shutaro Kunimasa, Kyoichi Seo, Hiroshi Shimoda, Hirotake Ishii: An Estimation Method of Intellectual Work Performance by Using Physiological Indices, 6th Annual International Conference on Cognitive and Behavioral Psychology, 2017
- [10] Rainer Storn, Kenneth Price: Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces, Journal of Global Optimization, vol.11, pp.341-359, 1997