

感性評価に基づく最適化に対するクラウドソーシングの適用

Sensory optimization using crowdsourcing

遠藤 ルッカス 良 馬場 雪乃 鹿島 久嗣
 Lucas Ryo Endo Yukino Baba Hisashi Kashima

京都大学大学院 情報学研究科 知能情報学専攻

Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

Sensory evaluation is required in many situations. The evaluation results are typically used for improving the quality of products. Because sensory evaluation needs to incorporate human evaluators, there is a demand for cost-efficient approaches. In this paper, we apply crowdsourcing and Bayesian optimization to achieve efficient sensory evaluation. Crowdsourcing provides a cost-efficient way for outsourcing evaluation tasks to a large number of people, and Bayesian optimization is a method for iteratively querying the oracle so that we can find a data point that maximize the objective function with a small number of queries. We conduct experiments with two datasets to confirm effectiveness of the proposed method. By investigating the results, we verify that the performance of our method is better than that of a random selection method.

1. はじめに

センサ技術や大規模データにより、人間の行動を分析・予測できるようになりつつある昨今、未だ人間の感覚・感性というものは推し量り辛いものである。感覚とは外界からの刺激を受けどう解釈するかということであり、感性とは様々な感覚から創発される感情やイメージという言葉で説明される [5, 6]。

これらの感覚・感性によって人間が対象物に対して下す評価を感性評価と呼ぶ。感性評価は評価者の感覚・感性に依存しているため、時に互いに矛盾する評価を下すことも起こりうる。他方感覚・感性は周囲の環境によって左右される側面もあるため、同じ文化圏や同年代などのコミュニティ内では感性評価の結果が似ることが期待される。

このような特性を持つ感性評価に基づき最適な設計を行うことが求められる事象は様々あげられる。例えばポスターのデザインでは、伝えたい意図を相手に伝えるために配色や文字の大きさなどを調整することが考えられる。このような事象に対して、専門家からの助言や通説をもとに調整を行うことや、アンケートを実施してその評価をもとに調整点を探し出すなどの手法が古典的な手法として行われてきた。

近年では、専門家の助言やアンケートの評価を手広く収集する方法としてクラウドソーシングの活用が広がっている。クラウドソーシングとは、インターネットを通じて不特定多数の人に仕事を依頼すること及びその仕組みを指す言葉である。時に匿名の相手を含む不特定多数の人に作業・業務を委託することがクラウドソーシングの大きな特徴となっている [4]。

人間の感覚や感性に基づく評価を収集する手段としてクラウドソーシングを利用することには、手軽かつ大量に評価を入力できるというメリットがあるが、その反面候補の数を増やせば増やすほどコストが増大していく。また感覚・感性には個人差があるので、真に求める評価に適合した候補を選択することは容易ではない。そのため低コストに抑えながらも真に求める候補に近い候補を効率良く発見する手法が求められている。

本研究では、感性評価の収集とそれに基づく最適化を省コストに行うことを目的としてクラウドソーシングとベイズ最適化を組み合わせた手法を提案する。クラウドソーシングでは感性評価を行い、ベイズ最適化では収集データをもとに次に探索すべき候補を予測する。この予測された候補について再び感性

評価を行い、また予測を行う。つまり評価と予測を交互に繰り返しながら逐次的に評価を収集しつつ最適な候補を探索する。

ベイズ最適化とは、ブラックボックス関数の最適化に用いられる手法である。この手法では、まず関数の入力を選択し、対象の関数から出力を得る。次にこれまでの入出力のデータから次に探索すべき入力を予測する。この選択・集計・予測のサイクルを繰り返して最適化を行っていく。

この手法の有効性を確認するため、クラウドソーシングによる感性評価タスクのデータセットを用いてシミュレーション実験を2つ行う。まず感覚に基づく単純なデータセットによる実験で、人間の評価に対してベイズ最適化が適用でき、ランダムな探索法よりも性能の良い結果をもたらすことを示す。そしてもう1つの感性評価に基づくデータセットによる実験でも良い成果をもたらすことを示す。

2. 問題設定

本研究では、パラメータを選択して評価集合を得るまでを一つのステップとして逐次的に探索を行うことを考える。 t ステップ目の入力パラメータを $\mathbf{x}^{(t)}$ とし、評価対象物 $M^{(t)}$ が生成されるとする。それを人数 w で評価して得られる評価集合を $E^{(t)} = \{e_1^{(t)}, e_2^{(t)}, \dots, e_w^{(t)}\}$ とする。このパラメータを設定すると評価集合が得られるという関係を一種のブラックボックス関数とみなして、 $E^{(t)} = \{e_i^{(t)}\}_{i=1}^w = \{f(\mathbf{x}^{(t)}) + \varepsilon_i^{(t)}\}_{i=1}^w$ という関係にあると考える。 $\varepsilon_i^{(t)}$ は人間の評価のブレや個人差を考慮したものである。そして評価結果から予測を行い次に探索するパラメータ $\mathbf{x}^{(t+1)}$ を決定する。このサイクルを繰り返し、なるべく小さい t で $\mathbf{x}^* = \operatorname{argmax} f(\mathbf{x})$ を予測するというのが本研究での問題設定である。

3. ベイズ最適化

この章ではブラックボックス関数のための最適化手法としてベイズ最適化の説明を行い、この手法で用いられているガウス過程及び獲得関数について順に説明する。

ベイズ最適化 (Bayesian Optimization, BO) とはブラックボックス関数のための最適化手法の1つである。特に対象の関数の評価に大きなコストがかかる場合に用いられる。この手

法は、データを一つずつ収集すると同時に予測を行う。対象の関数の入出力データをもとに、最適と思われる点を予測して次の入出力データを1組得る。そして1組分増えた入出力データをもとに再度予測を行う。このサイクルを繰り返して最適値を探してゆく [3]。

ベイズ最適化での予測には、ガウス過程と獲得関数を用いられる。ガウス過程では未知のデータに対する確率分布を与え、獲得関数ではその確率分布をもとにどこを選択すべきかの判断を行う。より詳細な説明は3.1節及び3.2節で行う。

3.1 ガウス過程

ガウス過程 (Gaussian Process, GP) とは確率過程の一種であり、任意の点集合 $\mathbf{x}_1, \dots, \mathbf{x}_N$ に対する関数 $f(\mathbf{x})$ の値の同時分布がガウス分布に従うとした関数 $f(\mathbf{x})$ 上の確率分布として定義される。ガウス過程では、その定義から、任意の点 \mathbf{x} に対する関数 $f(\mathbf{x})$ の確率分布はガウス分布で与えられる。つまりその分布は平均関数 $\mu(\mathbf{x})$ 及びカーネル関数 $k(\mathbf{x}, \mathbf{x}')$ の2つで記述されることになる。これを回帰として用いて解釈すると、既に観測したデータから未知のデータそれぞれに対して平均関数 $\mu(\mathbf{x})$ 及びカーネル関数 $k(\mathbf{x}, \mathbf{x}')$ の2つだけでそれらの分布を予測できるということになる。

ガウス過程で得られるのは関数上の確率分布である。ベイズ最適化では、その確率分布をもとに次に探索する点を戦略的に決定する。その決定には次節で述べる獲得関数を用いられる。

3.2 獲得関数

ベイズ最適化では、観測されたデータに対しガウス過程を用いて関数上の確率分布を与えたのち、獲得関数を用いて次に探索する点を決定する。獲得関数の主たる戦略は活用と探索のバランスをとることにある。活用とは、既知データの内最適値に近いパラメータの周辺を次の選択点として予測する戦略である。探索とは、既知データの中ではあまり探索されていない領域を次の選択点として予測する戦略である。この2つのバランスをとりより良い点を選択することと予測の精度を向上させることの両立を図ることが獲得関数の設計方針の目標となる。

ベイズ最適化では、この獲得関数にガウス過程の結果を利用して関数設計を行う。獲得関数の設計戦略における活用と探索は、ガウス過程の各点の平均 $\mu(\mathbf{x})$ と分散 $\sigma^2(\mathbf{x})$ が対応する。 N ステップ目の獲得関数を $a^{(N)}(\mathbf{x})$ とすると次の探索点の決定は $\mathbf{x}_{N+1} \leftarrow \operatorname{argmax} a^{(N)}(\mathbf{x})$ により行われる。

獲得関数の設計方法は様々存在する。本研究では代表的な関数の設計として、Expected Improvement (EI) と Mutual Information (MI) の2つをそれぞれ説明する。

Expected Improvement (EI) この獲得関数は、現在までに観測したデータの最大値 $f(\mathbf{x}_{\text{best}})$ の更新幅の期待値を関数にとる。この獲得関数を $a_{\text{EI}}(\mathbf{x})$ とすると、式 (1) として表されその中身をガウス過程によって得られた平均 $\mu(\mathbf{x})$ と分散 $\sigma^2(\mathbf{x})$ で表すと式 (2) のように表される。

$$\begin{aligned} a_{\text{EI}}(\mathbf{x}) &= \mathbb{E}(\max\{0, f(\mathbf{x}) - f(\mathbf{x}_{\text{best}})\}) \\ &= \begin{cases} (\mu(\mathbf{x}) - f(\mathbf{x}_{\text{best}}))\Phi(Z) + \sigma(\mathbf{x})\phi(Z) & (\sigma(\mathbf{x}) > 0) \\ 0 & (\sigma(\mathbf{x}) = 0) \end{cases} \end{aligned} \quad (1)$$

$$(2)$$

ここで Z は次式で与えられるものとする。

$$Z = \frac{\mu(\mathbf{x}) - f(\mathbf{x}_{\text{best}})}{\sigma(\mathbf{x})}$$

($\phi()$ 及び $\Phi()$) は標準正規分布の確率密度関数及び累積分布関数を表す)

Mutual Information (MI) [1] この獲得関数は平均 $\mu(\mathbf{x})$ と分散 $\sigma^2(\mathbf{x})$ の和である上側信頼区間 $\mu(\mathbf{x}) + \alpha\sigma^2(\mathbf{x})$ をベースにし、ステップが進むにつれ探索から活用へとシフトして戦略を持つ関数である。この獲得関数を $a_{\text{MI}}(\mathbf{x})$ として、ガウス過程によって得られた平均 $\mu(\mathbf{x})$ と分散 $\sigma^2(\mathbf{x})$ で表すと式 (3) のように表される。ただし、 $\sigma_t^2(\mathbf{x}_t)$ はステップ t で選択した点 \mathbf{x}_t におけるステップ t でのガウス過程によって得られた分散関数 $\sigma_t^2()$ の値である。

$$a_{\text{MI}}(\mathbf{x}) = \mu(\mathbf{x}) + \alpha \left(\sqrt{\sigma_N^2(\mathbf{x}) + A} - \sqrt{A} \right) \quad (3)$$

$$A = \sum_{t=1}^{N-1} \sigma_t^2(\mathbf{x}_t)$$

ここで α はハイパーパラメータである。

4. 提案手法

本研究では2節で設定される問題に対して、ベイズ最適化とクラウドソーシングを組み合わせて解決する手法を提案する。

本研究では、人間の感性評価の過程をブラックボックス関数と見なしてベイズ最適化を適用する。また、ベイズ最適化では逐次的に探索を行うことを考えるため、逐次的に感性評価を収集することが必要となる。そのプラットフォームとして本研究ではクラウドソーシングを活用して感性評価を収集を行う。

本研究で提案する感性評価に基づく最適化に対する手法は、次の手順で行われる。まず初期パラメータ $\mathbf{x}^{(1)}$ をランダムに選択して評価対象物 $M^{(1)}$ を生成する。これに対する感性評価をクラウドソーシングを用いて w 人から評価集合 $E^{(1)} = \{e_1^{(1)}, e_2^{(1)}, \dots, e_w^{(1)} | e_i^{(1)} \in \mathbb{R}\}$ を得る。次に評価集合をベイズ最適化に用いることができるように、評価の集約を行う。集約した結果をスコアと呼ぶこととし、 $S(E^{(1)}) = s^{(1)} \in \mathbb{R}$ と置く。スコア関数 $S()$ は評価の集約を行い値を返す関数である。そして、 $\mathbf{x}^{(1)}, s^{(1)}$ をもとにベイズ最適化を実行し、獲得関数 $a^{(1)}(\mathbf{x})$ を用いて次の選択点を $\mathbf{x}^{(2)} \leftarrow \operatorname{argmax} a^{(1)}(\mathbf{x})$ により決定する。そして次の探索点で同様に評価物を生成、クラウドソーシングによる評価、ベイズ最適化による予測を行う。この生成・評価・予測の一連の流れを1つのステップとし所与の回数これを繰り返す。

本研究ではこの手法を実行するにあたってあらかじめ設定しておかなければならない点が3つある。まず評価を収集する際の人数 w である。次にスコア関数 $S()$ も設定しておかなければならない。最後にベイズ最適化の獲得関数 $a()$ の設計もあらかじめ決定しておかなければならない。

5. 実験

本実験の目的は、人間の感性評価に対して提案手法を用いた最適化がどのような性能を発揮するかを検証することである。そのために2種類のデータセットを用いて、提案手法と比較手法の性能比較実験を行った。この章ではまず実験の設定を述べ、次にデータセットについての説明を行う。その後実験の結果を提示し、最後に結果の分析・議論を行う。

5.1 実験の設定

本研究では以下の流れで実験を行った。

1. 初期パラメータの選択
2. パラメータをもとに評価対象物を生成
3. 対象物を評価
4. 評価結果をもとに次のパラメータを選択
5. 規定回数に達するまで 2. に戻る

以下ではそれぞれの詳細を述べてゆく。
本研究で実験に用いた手法は以下の 3 つである。

BO-MI(提案手法) 収集した評価結果をもとに獲得関数に Mutual Information を使用したベイズ最適化を用いて次のパラメータを選択する手法。

BO-EI(提案手法) 収集した評価結果をもとに獲得関数に Expected Improvement を使用したベイズ最適化を用いて次のパラメータを選択する手法。

rand(比較手法) ランダムサンプリングである。収集した評価結果に依らずランダムに次のパラメータを選択する。

共通して、対象物の評価の集計には 1, 3, 5, 10 人の平均値及び中央値を使用した。色の組み合わせ評価タスクでは評価値が 7 つであるので平均値でのみ集計を行った。

評価指標には平均二乗誤差 (MSE : Mean Square Error) を用いた。各ステップ毎に、MSE を用いてそのステップを含むそれまでのデータで評価を行う。対象の関数を $f(\mathbf{x})$ として目的の最適値を $f(\mathbf{x}^*)$ 、 i ステップ目の入力を $\mathbf{x}^{(i)}$ とした時の k ステップ目の MSE 評価値 $\text{MSE}^{(k)}$ は式 (4) で表される。

$$\text{MSE}^{(k)} = \frac{1}{k} \sum_{i=1}^k (f(\mathbf{x}^*) - f(\mathbf{x}^{(i)}))^2 \quad (4)$$

この評価指標は値が小さいほど性能が良いということを表し、最適値に近づいていく手法はステップを更新するたびに値が減少していくことが期待される。

本研究では初期値として BO-MI, BO-EI, rand のいずれも探索するパラメータ空間内からランダムに 1 つ選択した。

5.2 データセット

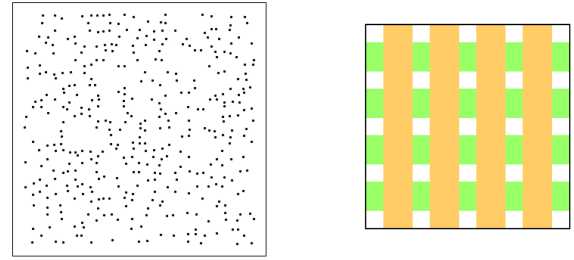
本研究では実験のために、感覚及び感性評価に関するデータセットを合わせて 2 つ用意した。以下ではそのそれぞれについて概要を述べる。

ドット数え上げタスク このデータセットは感覚に関するデータセットである。多数のドットが描画された画像 (画像例 : 図 1a) を人間に見せ、画像上にドットがいくつ表示されていると思うかを回答させた。このデータセットの作成には Horton の方法 [2] を参考にした。

画像に描画されるドットの数 y は入力パラメータ x 及びこちらで指定する関数 $f(x)$ に従い $y = f(x)$ という関係にある。関数 $f(x)$ は式 (5) で定義される Himmelblau 関数を用いて式 (6) の整数部分とした。Himmelblau 関数は最適化アルゴリズムのベンチマークにしばしば使われる多峰性関数である。実験では入力パラメータは $x \in [0, 2\pi]$ としその中から 999 個のパラメータを等間隔に選択し用いた。 $f(x^*) = 500$ である。各パラメータにつき 10 人から回答を得た。

$$h(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2 \quad (5)$$

$$f(x) = h(3 \cos(x), 3 \sin(x)) + 300 \quad (6)$$



(a) ドット数え上げタスク (b) 色の組み合わせ評価タスク

図 1: 各データセットの画像例

色の組み合わせ評価タスク このデータセットは感性評価に関するデータセットである。2 色の色が交差した画像 (画像例 : 図 1b) を人間に見せ、それらの組み合わせが良いものかどうかを 1 から 7 までの 7 段階で評価させた。

画像に用いられる 2 色の色は入力パラメータ x_1, x_2 によって決定される。色の決定には HSB 色空間を用いた。色 C_i ($i = 1, 2$) の色相、彩度、明度をそれぞれ h_i, s_i, b_i とすると本研究の実験では $(h_i, s_i, b_i) = (x_i, 0.6, 1.0)$ と設定し色を決定した。入力パラメータの範囲は $x_i \in [0, 360]$ とし、その中から 1,296 個のパラメータ組を等間隔に選択し用いた。 $f(\mathbf{x})$ については次のように設定した。まずワーカーを 2 つのグループに分割し、一方のグループの評価の平均値を $f(\mathbf{x})$ とした。実験ではもう一方のグループで探索を行う。グループの分割は全体の 3 分の 2 を $f(\mathbf{x})$ の値を求めるためのグループ、残りの 3 分の 1 を探索に用いるためのグループとした。 $f(\mathbf{x}^*)$ については集約した評価の上限とし、本実験では $f(\mathbf{x}^*) = 7$ とした。各パラメータにつき 30 人から回答を得た。

5.3 実験結果

本節では行った 2 つの実験の結果をそれぞれ順に示す。どちらの実験も、40 ステップで 1 サイクルとし合計 20 サイクルずつ行い、評価は 20 サイクルのステップごとに平均を取っている。いずれの実験でも提案手法が良い成績を修めていることを示している。

まずドット数え上げタスクの実験結果を示す。図 2a は人間の評価の集約方法に中央値を採用したもので、それぞれのグラフで集約に用いる人数が異なる。いずれの結果でも BO-MI が rand よりも良い性能を発揮しており、BO-EI も集約に用いる人数を増やすと性能が改善される傾向が見取れる。図 2b は人間の評価の集約方法に平均値を採用したもので、それぞれのグラフで集約に用いる人数が異なる。こちらの集約手法でも同様に、いずれの結果でも BO-MI が rand よりも良い性能を発揮しており、BO-EI も集約に用いる人数を増やすと性能が改善される傾向が見取れる。

次に色の組み合わせ評価タスクの実験結果を示す。図 2c に実験の各手法間で比較した結果を示す。こちらの実験では提案手法の BO-MI でのみ性能の改善が見られた。

5.4 分析

この節では、それぞれの実験の結果について分析を行う。

5.4.1 ドット数え上げタスク

まず手法間で比較すると、rand より BO-MI が良い性能であることを示し、BO-EI も概ね良い性能であることが実験結果から得られた。これはベイズ最適化の過去のデータをもとに

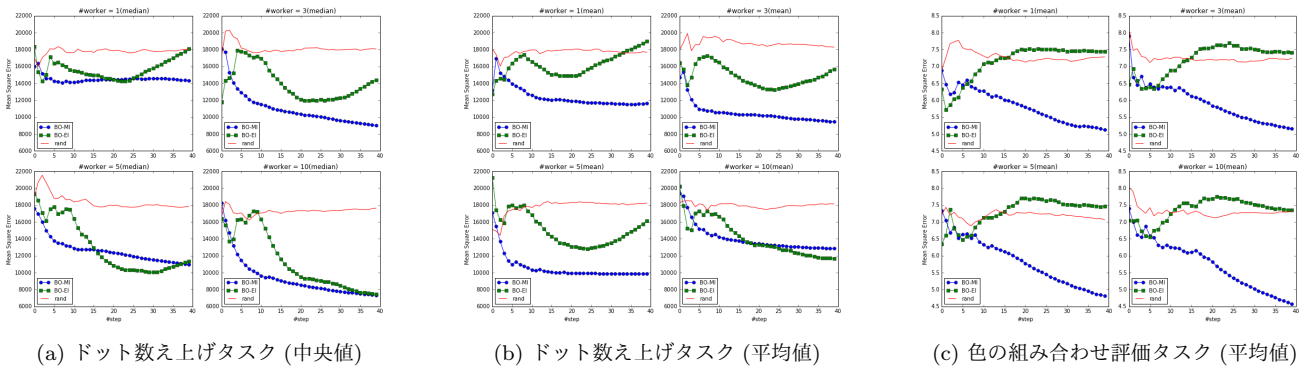


図 2: 各タスクの実験結果 (青:BO-MI, 緑:BO-EI, 赤:rand)。集約に用いた人間の数は 1(左上)、3(右上)、5(左下)、10(右下) である。いずれのグラフも横軸はステップ数、縦軸は MSE を表す。

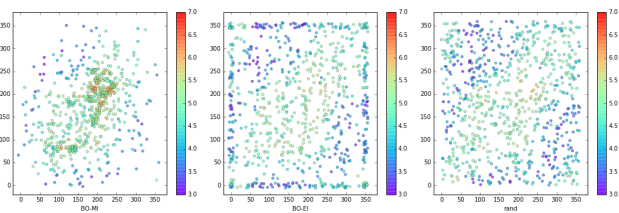


図 3: 色の組み合わせ評価タスクに対する実験において各手法が探索した点とその評価値。それぞれのグラフは BO-MI(左)、BO-EI(中)、rand(右) を表す。横軸は入力パラメータの x_1 、縦軸は入力パラメータの x_2 を表す。点の色はその点での評価値を示している。

次の探索点を決める戦略がうまく働いていると考えることができ、クラウドソーシングを用いた人間の感覚に対する評価に対してもベイズ最適化を適用できることを示している。

集約方法の違いを検討する。同じ手法、同じ集約人数間で比較すると、人数が 1、3、5 の場合では大きな差が見られないものの、人数が 10 の場合で平均値を用いた集約よりも中央値を用いた集約が良い性能を記録している。これはワーカーからの回答に他の回答とは乖離した値つまり外れ値が含まれている場合、平均値による集約はその外れ値による影響を受けているものと考えられる。また人数が 10 の場合で顕著になっている理由としては、ワーカーの人数を増やすと各ステップでの回答に外れ値が含まれやすくなるのが考えられる。

5.4.2 色の組み合わせ評価タスク

こちらの実験では、rand と比べ BO-MI が良い性能を示した一方で、BO-EI が悪い性能であることが示された。これは BO-EI の探索がこのデータセットに対してうまく働いていないということが示唆されている。図 3 に各手法が探索した点とその評価値を示す。これによると、BO-MI が中央部の評価値が比較的高いところを多く探索しているのに対し、BO-EI は評価値の比較的低い外縁部の探索が多くなっている。これは過去のデータの活用よりも探索に重きが置かれて既に探索した内部の点付近ではなく、外縁部の未探索の部分ばかり探索してしまっているものとみられる。他方 BO-MI が示すのは、過去のデータの活用と探索のバランスをうまくとることに成功し、探索が評価値の比較的高い部分に到達できているということである。

6. 結論

本研究では、感性評価に基づく最適化を省コストに行う手法として、クラウドソーシングとベイズ最適化を組み合わせた手法を提案した。この提案手法の有効性を確認するために 2 つのデータセットを用いてシミュレーション実験を行った。その結果、クラウドソーシングを用いた人間の感性評価に対して提案手法によりベイズ最適化が適用できること、及び提案手法が少ないデータ数で精度良く予測を行うことを示した。

今後の研究の展望としては以下の 2 つの方向性が考えられる。まず、集約方法を新たにすることで提案手法より良い精度をもたらす可能性があると考えられる。そして、提案手法ではベイズ最適化の獲得関数としてこれまで一般のブラックボックス関数の最適化問題に用いられてきた 2 つの関数を用いたが、これを人間の感性評価に適した形で設計することによって性能改善を図る方向性も考えられる。

参考文献

- [1] Emile Contal, Vianney Perchet, and Nicolas Vayatis. Gaussian process optimization with mutual information. In *Proceedings of the 31st International Conference on Machine Learning*, 2014.
- [2] John J. Horton. The Dot-Guessing Game : A “Fruit Fly” for Human Computation Research, 2010. Available at SSRN: <https://ssrn.com/abstract=1600372> or <http://dx.doi.org/10.2139/ssrn.1600372>.
- [3] Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning*, 2010.
- [4] 鹿島久嗣, 小山聡, 馬場雪乃. ヒューマンコンピューションとクラウドソーシング. 講談社, 2016.
- [5] 長町三生. 感性工学と新製品開発. 日本経営工学会誌, 第 41 巻, pp. 66–71, 1990.
- [6] 森典彦. デザインにおける感性工学. 日本ファジィ学会誌, 第 11 巻, pp. 52–63, 1999.