

ヒューマンコンピューテーションとクラウドソーシング

馬場 雪乃

ヒューマンコンピューテーションは、人間と計算機の組合せにより難しい問題を解決するための方法論である。主に、計算機システムの内部に人間への問合せを組み込む形式で実装される。不特定多数の人間が参加するヒューマンコンピューテーションでは人間の信頼性が課題となり、人間から正しい回答を引き出す技術が求められる。本稿では、特に集団の力を活かして正しい回答を得る手法を紹介する。集団を組み合わせる正しい回答を得る手法や、集団の中から正しく回答できる人を見つける手法を紹介する。さらに、集団による協調問題解決を支援する手法を紹介する。

キーワード：ヒューマンコンピューテーション、クラウドソーシング、集合知

1. はじめに

ヒューマンコンピューテーションは、人間と計算機の組合せにより難しい問題を解決するための方法論である。人間には優れた認識能力・判断能力がある。計算機は、安定性やスケーラビリティの点で優れている。両者の長所を取り入れることで、人間と計算機のどちらか一方だけでは解決が難しい、複雑で大規模な問題の解決を目指すのがヒューマンコンピューテーションである。

ヒューマンコンピューテーションは、計算機システムの内部に人間への問合せを組み込む形式で実装されることが多い。たとえば、reCAPTCHA [1] は、文字認識システムに人間を組み込むことで大量の書籍の文字認識を実現している。reCAPTCHA では、まず書籍中の文字を二つの文字認識アルゴリズムに認識させる。両者の結果が一致したら、その結果を採用する。不一致のときは計算機には認識が難しいとして、人間に認識作業を依頼する。同様に VizWiz [2] は、視覚障がい者のための質問応答システムに人間を組み込んでいる。VizWiz は、音声と画像で与えられた質問に対して答えを返す（質問例：「写真中の缶のうち、コーンの缶はどれ？」、回答例：「右の缶」）。音声と画像を理解し正しい答えを生成するのは計算機には困難なので、実際には人間が回答している。

ヒューマンコンピューテーションの実装では、大量の人間の参加者が必要となる。参加者募集に用いるのが

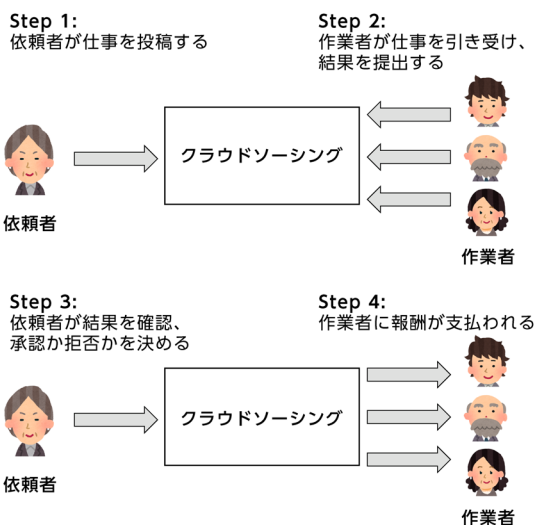


図1 クラウドソーシングの概略図

クラウドソーシングである（図1）。クラウドソーシングは、不特定多数の人に金銭報酬を支払い、単純作業を依頼するためのインターネット上のプラットフォームである。依頼者は作業内容を準備するだけで良く、作業が終われば作業結果が獲得できる。参加者の募集や報酬の支払いは、すべてプラットフォームが仲介する。特に、Amazon Mechanical Turk¹ではAPIが整備されており、計算機システムから直接、依頼の作成や結果の獲得が可能となっている。実際に VizWiz は、Amazon Mechanical Turk を利用して質問への回答を獲得している。

ばば ゆきの
筑波大学システム情報系
〒305-8573 茨城県つくば市天王台 1-1-1
baba@cs.tsukuba.ac.jp

¹ <https://www.mturk.com>

2. ヒューマンコンピューテーションの課題

不特定多数の人間に依存するヒューマンコンピューテーションでは、人間の信頼性が重要な課題となる。計算機からの問合せに対して、人間が常に誰でも正しい答えを返すとは限らない。人間はさまざまな要因でミスをする。また、人によって能力や性質が異なる。不確実で多様な人間を扱うヒューマンコンピューテーションでは、人間から正しい回答を引き出す技術が必要となる。

reCAPTCHA では、人間から正しい回答を得るために三つの工夫を採用している。一つめが、動機づけである。reCAPTCHA は、表向きはアクセス制御のしくみとして機能している。あるウェブページにアクセスを試みるユーザに対して、そのユーザが人間かロボットかを判定するために、文字認識を行わせる。ユーザには、正答できれば人間と判定されアクセスが許可されるが、間違えればロボットと判定されアクセスが許可されないと事前に伝えておく。アクセス権と引き換えにすることで、人間に真面目な作業を促す効果がある。二つめの工夫が、正解既知の問題の利用である。reCAPTCHA は、一度に二つの文字列を提示し、両方の文字列を認識するようユーザに依頼する。このうち一方については、システムが正しい認識結果を知っている。もう片方は、システムが人間に認識してもらいたいと思っている文字列であり、システムは正解を知らない。正解既知の文字列に間違えたユーザの回答は無視することで、不確実な回答の混入を防ぐ。三つめの工夫は、並列化である。reCAPTCHA では、同じ文字列の認識作業を複数人に行わせる。一定数以上の人間が同じ回答をした場合は正しい回答として採用する。

ヒューマンコンピューテーションの研究では、人間から正しい回答を得るための手法が開発されている。本稿では、reCAPTCHA の並列化のように、集団を活用する手法を紹介する。集団を組み合わせることで正しい回答を得る手法や、集団の中から正しく回答できる人を見つける手法を紹介する。さらに、集団による協調問題解決を支援する手法を紹介する。

3. 並列問合せ

3.1 回答者の信頼性を考慮

正しい答えを得るために広く用いられる手法が多数決である。すなわち、同じ問合せを複数人に行い、最も多い答えを採用する。多数決は、回答者全員の信頼性が等しいものとして扱う。しかし不特定多数の人間が参加するヒューマンコンピューテーションでは、人によって信頼性が異なる。そこで、各自の信頼性を推定した上で回答統合を行う統計的回答統合法が用いられる。古典的な手法に、DS 法がある [3]。DS 法は、元々は複数の医師が複数の患者に対して下した診断の統合のために開発された。DS 法は、各自の回答を利用して各問題の正解（たとえば、各患者の正しい診断結果）を予測する問題として定式化される。ここでは簡単のため、回答が Yes/No の二値の場合を考える。二値の DS 法では、回答者 j の信頼性を θ_j と λ_j という二つのパラメータで表現する。 θ_j は、正解が Yes のときに正しく Yes と答える確率で、 λ_j は正解が No のときに正しく No と答える確率である。回答者は、この二つのパラメータに従って回答を生成するとみなす。あたかも回答者が 2 枚のコインをもち、コインを投げて回答を決めているかのようなため、Two-coin モデルとも呼ばれている。EM アルゴリズムを利用して、各問題 i の正解 z_i と、各回答者 j の θ_j と λ_j を推定する。

DS 法を発展させ、問題の難易度も考慮して回答者の信頼性を推定する手法も提案されている [4]。この手法では、問題 i に回答者 j が正答する確率を次式で表す：

$$\Pr[y_{ij} = z_i] = \frac{1}{1 + \exp(-\theta_j \nu_i)}.$$

ここで y_{ij} は問題 i に対する回答者 j の回答、 $\theta_j \in \mathbb{R}$ は回答者の信頼性、 $\nu_i \in [0, \infty)$ は問題の簡単さである。信頼性が簡単さが 0 だと正答確率は 0.5 となり、信頼性と簡単さが大きいほど正答確率が 1 に近づく。

3.2 回答者の確信度を考慮

回答者に問題の答えだけではなく確信度を問い合わせ、確信度を利用して回答者の信頼性を推定する手法も提案されている [5]。たとえば「オーストラリアの首都はシドニーか？」という問題に回答させるだけではなく、「その回答に自信があるか？」という設問にも答えさせる。単純には、確信度が高いほど信頼性が高いと考えられるが、誤答なのに自信があると答える自信過剰の人、あるいは正答なのに自信がないと答える自信過少の人が存在する。そこで DS 法を拡張し、確信度の傾向もモデル化する。正解が Yes（あるいは No）で回答が Yes（あるいは No）のときに自信があると答える確率を確信度傾向のパラメータとして表現することで、確信度の生成モデルを構築する。DS 法の話生成モデルと組み合わせることで、回答者の信頼性と問題の正解を推定する。

Yes/No の二値で問題に回答する場合を取り上げたが、数値やランキングを答える問題 [6, 7] や一対比較 [8] など、さまざまな問題形式において回答者信頼性を考

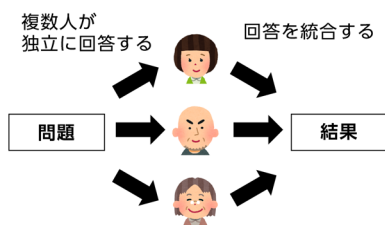


図2 並列問合せ



図3 直列問合せ

慮した統合法が開発されている。

4. 直列問合せ

4.1 他者の回答を利用

並列問合せは、複数人が同じ問題を独立に解いていた(図2)。複数人を組み合わせる別の方法に、直列問合せがある(図3)。直列問合せでは、最初にある回答者が問題(あるいは部分問題)を解き、その結果を使って別の回答者が問題を解く。これを繰り返し正しい答えを得る。たとえば、ヒューマンコンピューテーションによる文書校正システム SoyLent は、Find—Fix—Verifyの3段階で校正を実施している[9]。Findステップでは、ある回答者が文章中から問題がある箇所を検知する。それぞれについて、Fixステップでは、別の回答者が実際の校正を行う。Verifyステップでは、さらに別の回答者が校正誤りがないか確認を行う。このような逐次手続きにより多数の人が参加することで、校正精度を向上させる。

直列問合せは並列問合せと組み合わせたワークフローとして実施されることも多い。たとえば自己訂正と呼ばれる手法[10]では、最初のラウンドでは自力で問題に回答させる。次に、初回ラウンドでの自身の回答と同じ問題に対する他者の回答を提示し、必要があれば訂正させる。この二段階の手続きにより、ケアレスミスを防ぐ。また、TurKontrolは、文章やデザインなどの成果物を作成する問題において、修正と評価を逐次実行するワークフローである[11]。TurKontrolでは、ある人が成果物を生成した後に、他者に評価を依頼する。評価結果に基づいて、成果物の修正が必要かどうかを判定し、必要であれば修正を別の人物に依頼する。評価結果を状態とする部分観測マルコフ決定過程を用いることで、修正の要・不要を機械的に判定する。

4.2 他者からの評価を利用

成果物の品質を他者に評価させるワークフローをヒューマンコンピューテーションで活用する際には、評価の信頼性が問題となる。不特定多数の人間が評価をするときに、全員が真面目に評価するとは限らない。そこで、評価者の信頼性を考慮した上で評価結果から成果物の品質を推定する手法が提案されている[12]。この手法では、各評価者 r がバイアス $\eta_r \in \mathbb{R}$ と信頼性 $\kappa_r \in (0, \infty)$ のパラメータをもつと考える。さらに、成果物を作成する各作成者 a も能力 $\mu_a \in \mathbb{R}$ と信頼性 $\lambda_a \in (0, \infty)$ のパラメータをもつとする。評価の生成過程を次のようにモデル化する。各問題 t において、作成者 a が、品質 $q_{t,a}$ の成果物を作成する。品質は次式の正規分布により生成される：

$$q_{t,a} \sim \mathcal{N}(q_{t,a} \mid \mu_a, \lambda_a^{-1}).$$

これは、能力 μ_a が高い作成者ほど品質の高い成果物を作ることを表し、また、信頼性 λ_a が高い作成者は安定した品質の成果物を作ることを表している。

次に評価者が評価を行う。評価者 r は成果物を見て、まず潜在評価値 $s_{t,a,r} \in \mathbb{R}$ を決める。潜在評価値は次式の正規分布により生成される：

$$s_{t,a,r} \sim \mathcal{N}(s_{t,a,r} \mid q_{t,a} + \eta_r, \kappa_r^{-1}).$$

バイアス η_r が大きい評価者は本来の品質 $q_{t,a}$ よりも甘い評価を行い、信頼性 κ_r が小さい評価者の評価にはノイズが乗りやすいことを表している。次に評価者は、1点から K 点の範囲で点数を出力する。出力される評価 $y_{t,a,r} \in \{1, 2, \dots, K\}$ は潜在評価値に基づき段階反応モデル[13]で決まる：

$$\begin{aligned} & \Pr[y_{t,a,r} = k \mid s_{t,a,r}] \\ &= \Pr[y_{t,a,r} > k - 1 \mid s_{t,a,r}] - \Pr[y_{t,a,r} > k \mid s_{t,a,r}] \\ & \Pr[y_{t,a,r} > k \mid s_{t,a,r}] = \frac{1}{1 + \exp(-(s_{t,a,r} - b_k))} \end{aligned}$$

b_k は閾値パラメータである。観測した評価結果を用いて、成果物の品質 $q_{t,a}$ やそれ以外のパラメータを推定する。

5. 専門家発見

5.1 補助情報を利用

並列問合せと直列問合せでは、人によって信頼性は異なるとはいえ、大抵の人が回答できる問題を対象にしていた。医療に関する問題のような高度で専門性が求められる問題では、正しい答えを得るために、集団から解ける人を見つける必要がある。さまざまな手が

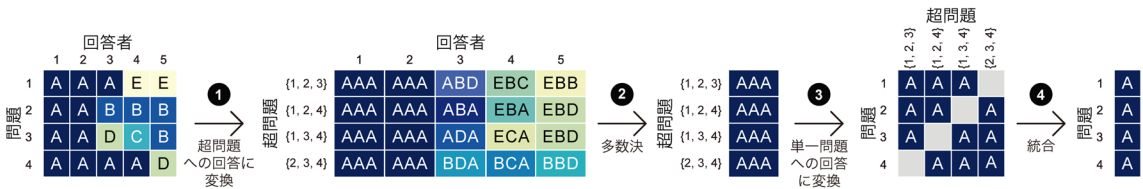


図4 Hyper-MV：超問題を利用して専門家を多数派にする

かりを利用して，そのような専門家を見つける手法が提案されている．たとえば，回答者の学歴などの属性を利用した手法がある [14]．この手法では，回答者の属性ベクトル $\mathbf{a}_j = (a_{j1}, a_{j2}, \dots), a_{jk} \in \{0, 1\}$ が与えられ，回答者の信頼性が属性ベクトルの線形和に基づいて決まる：

$$\theta_j \sim \mathcal{N}(\theta_j | \mathbf{w}^\top \mathbf{a}_j, \sigma^2).$$

DS法と組み合わせ，各自の回答から属性の重み \mathbf{w} を推定する．属性を用いることで，これまで回答していない回答者でも信頼性を推定できる．

各自が検索エンジンにおいて入力する検索クエリを専門性の手がかりにする Quizz という手法もある [15]．Quizz では，検索エンジンで提示される広告に問題へのリンクを含める．回答者がリンクをクリックすると問題が提示され，それに回答するよう指示される．Quizz は提示する問題を，回答者が入力した検索クエリに応じて決める．このしくみは，検索連動型広告と同様に実装される．検索連動型広告は，広告クリック率などに応じて広告効果の高い検索クエリを決める．Quizz では，正答確率に応じて，問題を提示する検索クエリを決める．

5.2 多数決を利用

Hyper-MV は多数決を利用して専門家を見つけ，正しい答えを得る手法である [16]．すべての問題で正解が未知の状況でも利用できる．従来の多数決は，でたために答える非専門家が多数派で正しく答える専門家が少数派の場合には，正しい答えを得ることができない．Hyper-MV は，専門家同士の回答は複数問で一致しやすいが，非専門家同士の回答は一致しづらいという事実を利用する．複数の問題の集合を超問題と呼び，Hyper-MV は超問題に対する回答の多数決をとることで，専門家が多数派になるようにする．要素数 k の超問題を k -超問題と呼ぶ，具体的な手続きを図4に示す．この例では，5名の回答者が4問に回答している．すべての問題で正解は‘A’とする．全問正解している2名が専門家である．Hyper-MV は，まず k -超問題を生成し，各自の回答を超問題に対する回答に変換す

る．この例では $k = 3$ としている．たとえば，超問題 $\{1, 2, 3\}$ への3番目の回答者の回答は‘ABD’となる．次に，超問題の上で多数決を行う．4件の超問題すべてで‘AAA’が多数決解となる．3番めのステップでは，各超問題の多数決解から，各単一問題への投票を得る．たとえば，超問題 $\{1, 2, 3\}$ の多数決解‘AAA’から，問題1, 2, 3のそれぞれに対して‘A’への投票を得る．最後に，この投票に対して多数決を実施し，各単一問題に対する回答を得る．この例では，すべての問題において正解である‘A’を回答として得ることができた．

6. 協調問題解決

6.1 集団で問題を分割

より困難で複雑な問題に対しては，人間が互いに知恵を出し合って解決にあたる必要がある．集団が効率的に問題解決できるようにするため，問題を上手く分割して複数人で解いたり，良い解をみんなで探索するためのシステマティックな手法が提案されている．Turkomatic は，問題分割を集団で行うことで大規模な問題を解決するためのワークフローである [17]．Turkomatic では，まず問題を部分問題に分割するよう人間に依頼する．分割は，十分単純な部分問題になるまで繰り返す．次に，各部分問題を人間に解かせる．最後に，部分問題の解を人間が集約し，最終的な解を出力する．

6.2 集団で解を探索

集団で解を効率的に探索するための手法として，確率的勾配法を模倣したものがある [18]．この手法は，予算案のような解が数値で表される問題で使うことができる．まず，集団の中から回答者をランダムに一人選ぶ．この回答者は，現在の値から一定範囲内で，好みの値に解を動かす．範囲を狭めながら同様の手順を繰り返すことで，多様な意見を取り入れながら解を探索することができる．

汎用的な問題では，候補となる解（アイデア）の列挙が重要となる．Innocentive²は，「フッ素パウダーを空

² <https://www.innocentive.com>

気中に拡散させずに歯磨きチューブに入れる方法は？」のような問題を解決するアイデアを世界中から収集するプラットフォームである。アイデアが採用されると賞金が支払われるため、金銭をインセンティブとすることで数多くのアイデアを収集することができる。遺伝的アルゴリズムに倣ってアイデアを生成・評価する手法も提案されている [19]。「子供向けの椅子のデザイン」のような問題において、この手法ではまず、候補となる解を人間がそれぞれ生成する。これが第一世代の個体となる。次に、投票で優良個体を選抜する。選ばれた個体からペアを作り、その交叉により新しいアイデア（第二世代の個体）を生成する。生成は人間が行う。選抜・交叉を繰り返すことで、アイデアを探索していく。

7. おわりに

本稿では、ヒューマンコンピューテーションにおいて広く研究されている、集団を活用して人間から正しい答えを引き出す手法を紹介した。ヒューマンコンピューテーションの技術は、われわれの日常や社会的問題の解決など、さまざまな場面で活用できる。より詳細な内容に興味のある読者は文献 [20] を参考にされたい。

参考文献

- [1] L. von Ahn, B. Maurer, C. McMillen, D. Abraham and M. Blum, “reCAPTCHA: Human-based character recognition via web security measures,” *Science*, **321**, no. 5895, 2008.
- [2] J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, A. Tatarowicz, B. White, S. White and T. Yeh, “Vizwiz: Nearly real-time answers to visual questions,” In *Proceedings of the 23rd annual ACM Symposium on User Interface Software and Technology (UIST)*, 2010.
- [3] A. P. Dawid and A. M. Skene, “Maximum likelihood estimation of observer error-rates using the EM algorithm,” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **28**, pp. 20–28, 1979.
- [4] J. Whitehill, P. Ruvolo, T. Wu, J. Bergsma and J. Movellan, “Whose vote should count more: Optimal integration of labels from labelers of unknown expertise,” In *Proceeding of Advances in Neural Information Processing Systems 22*, 2009.
- [5] S. Oyama, Y. Baba, Y. Sakurai and H. Kashima, “Accurate integration of crowdsourced labels using workers’ self-reported confidence scores,” In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*, 2013.
- [6] C. H. Lin, Mausam and D. S. Weld, “Crowdsourcing control: Moving beyond multiple choice,” In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2012.
- [7] T. Matsui, Y. Baba, T. Kamishima and H. Kashima, “Crowdordering,” In *Proceedings of the 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2014.
- [8] X. Chen, P. N. Bennett, K. Collins-Thompson and E. Horvitz, “Pairwise ranking aggregation in a crowdsourced setting,” In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM)*, 2013.
- [9] M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell and K. Panovich, “Soylent: A word processor with a crowd inside,” In *Proceedings of the 23th Annual ACM Symposium on User Interface Software and Technology (UIST)*, 2010.
- [10] N. Shah and D. Zhou, “No oops, you won’t do it again: Mechanisms for self-correction in crowdsourcing,” In *Proceedings of the International Conference on Machine Learning (ICML)*, 2016.
- [11] P. Dai, Mausam and D. S. Weld, “Decision-theoretic control of crowd-sourced workflows,” In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)*, 2010.
- [12] Y. Baba and H. Kashima, “Statistical quality estimation for general crowdsourcing tasks,” In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2013.
- [13] F. Samejima, “Estimation of latent ability using a response pattern of graded scores,” *Psychometrika Monograph Supplement*, **34**, pp. 1–97, 1969.
- [14] H. Li, B. Zhao and A. Fuxman, “The wisdom of minority: Discovering and targeting the right group of workers for crowdsourcing,” In *Proceedings of the 23rd International Conference on World Wide Web (WWW)*, 2014.
- [15] P. G. Ipeirotis and E. Gabrilovich, “Quiz: Targeted crowdsourcing with a billion (potential) users,” In *Proceedings of the 23rd International Conference on World Wide Web (WWW)*, 2014.
- [16] J. Li, Y. Baba and H. Kashima, “Hyper questions: Unsupervised targeting of a few experts in crowdsourcing,” In *Proceedings of the 2017 ACM Conference on Information and Knowledge Management (CIKM)*, 2017.
- [17] A. Kulkarni, M. Can and B. Hartmann, “Collaboratively crowdsourcing workflows with turkomatic,” In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW)*, 2012.
- [18] N. Garg, V. Kamble, A. Goel, D. Marn and K. Munagala, “Collaborative optimization for collective decision-making in continuous spaces,” In *Proceedings of the 26th International Conference on World Wide Web (WWW)*, 2017.
- [19] L. Yu and J. V. Nickerson, “Cooks or cobblers?: Crowd creativity through combination,” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2011.
- [20] 鹿島久嗣, 小山聡, 馬場雪乃, 『ヒューマンコンピューテーションとクラウドソーシング』, 講談社, 2016.