

クラウドソーシングでのタスク品質改善のための価格設定の検討

A Pricing mechanism for task quality-control in crowdsourcing

岡 雅晃*1 沖本 天太*2 櫻井 祐子*3 篠田 正人*4 横尾 真*1
 Masaaki Oka Tenda Okimoto Yuko Sakurai Masato Shinoda Makoto Yokoo

*1九州大学
 Kyushu University

*2新領域融合研究センター
 Transdisciplinary Research Integration Center

*3九州大学, JST さきがけ
 Kyushu University, JST PRESTO

*4奈良女子大学
 Nara Women's University

We propose a reward mechanism that utilizes workers' self-reported confidences for tasks with multiple labels in crowdsourcing. In crowdsourcing, since workers' abilities are not even, treating all labels given by different workers equally is not always a good way to infer true labels. Thus, we consider a quality control method that a requester asks workers not only for labels to the data but also for their confidence about them. In more detail, a requester proposes several reward plans to workers. Choosing a plan corresponds to selecting certain discretized confidence intervals. We ensure that choosing a plan for an interval including true confidence maximizes the worker's expected utility.

1. はじめに

ネットワーク上で提供されているサービスの一つとしてクラウドソーシングがある。日本でも、複数の企業がクラウドソーシングのサービスを提供しており、商業的に注目されている分野であり、学術的にも活発に研究が行われている [Law 11]。クラウドソーシングは、ネットワークを介して安価に不特定多数のワーカを動かせることが可能である。しかしながら、その一方で、得られる作業結果の品質が問題となっている。たとえば、画像のラベル付けなど、機械学習の訓練集合に用いるラベルをクラウドソーシングによって得る場合、大量のラベル付きデータが得られる一方で、ラベルの信頼性に関する問題が生じる。専門家等から得られるラベルとは異なり、不特定多数のワーカのラベルには、ワーカ自身の意図の有無に関わらず、ラベルに対して誤り（ノイズ）が入る可能性がある。

ノイズが含まれているラベル集合から真のラベルを推定する単純な手法の一つとして、複数のワーカに同一の画像へのラベル付けを行わせ、多数決を取る手法がある。すなわち、これはワーカ的能力（正解率）が同一とみなし、複数のワーカから提示されたラベルほど信頼性が高いと判断することになる。しかしながら、実際のクラウドソーシングではワーカに能力の差があるため、すべてのワーカのラベルに対する重みを同一として判断する手法は必ずしも有効ではない。

そこで、本論文では、作業結果の品質を保証することを目的とし、解答と共に、作業結果に対する自信（confidence）を申告させるメカニズムを提案する。自信はワーカのみが持つ確率情報であり、正直に申告させる誘因を与える必要がある。確率情報を正直に申告させるメカニズムに関する関連研究として、proper scoring rule と呼ばれるメカニズムが提案されている [Gneiting 07, Matheson 76, Savage 71]。

クラウドソーシングのワーカに直接自信を申告させる場合、proper scoring rule を適用することとなる。しかしながら、問題点としてワーカが認識する自信と正解率が必ずしも一致しな

いことである。そこで、文献 [Sakurai 13] では、2 択問題を対象に自信を直接申告させるのではなく、自信に応じて報酬プランを複数に分割し、その報酬プランリストを提示し、ワーカに報酬プランを選択させることで自信のレベルを間接的に申告させるメカニズムの提案が行われている。

本論文では、このアイデアを 2 択以上の選択肢が存在する問題に対して拡張を行う。より具体的には、3 択の問題に対して、解答方法として、(1) 選択肢から一つの解答として選択肢を選択する、(2) 選択肢から解答でない一つの選択肢を選択するという 2 つの解答方法を可能にし、それぞれの解答方法に対して自信に応じて報酬プランを選択させる。ワーカにとって、正直に解答方法と自信を申告することで期待利得が最大化されることを保証するメカニズムの提案を行う。

2. 問題設定

本章では、問題設定について説明する。本論文では、選択肢が 3 つの場合を対象とする。簡単化のため、候補補集合を $\{0, 1, 2\}$ とし、あるタスクの真の解答を $l \in \{0, 1, 2\}$ とする。

ワーカは、ネットワーク環境で各々異なる場所からタスクを実行するため、ワーカの解答は互いに独立であるとする。さらに、ワーカは合理的であり、期待利得を最大化するように行動するとする。タスクに対するワーカ i の解答を $l_i \in \{0, 1, 2\}$ とするとき、ワーカ i が正解する確率は以下のとおり定義される。

定義 1 (正解する確率) 真の解答 l に対して、ワーカ i の考える解答を l_i とするとき、ワーカ i が正解する確率は

$$a_i = Pr(l_i = l) \quad (1)$$

である。

次に、ワーカ i の自信（confidence）を次のように定義する。

定義 2 (自信) ワーカ i がタスクを実行した際に自分が正解することに対する自信を x_i とする。

連絡先: 岡 雅晃, 九州大学大学院システム情報科学府, 812-0395
 福岡県福岡市西区元岡 744 番地, oka@agent.inf.kyushu-u.ac.jp

自信と正解確率が一致するワーカにおいては、すなわち、ノイズが存在しない（認知心理学の用語でキャリブレーションされている）場合は、 $x_i = a_i$ が満たされる。自信の過大評価の場合は $x_i > a_i$ 、自信の過小評価の場合は $x_i < a_i$ である。

さらに、本論文では3択問題を対象とするため、 $1/3 \leq x_i \leq 1$ とする。つまり、何も考えずに適当に答えるランダムゲスの場合、正解確率は $1/3$ であることより、この下限値を設定することが可能である。また、ワーカ i が解答 l_i と共に申告する自信は必ずしも真の値 x_i とは限らない。したがって、申告する自信を y_i とする。

次に、報酬と期待利得に関して定義する。

定義 3 (報酬) ワーカ i が (l_i, y_i) を申告したとき、ワーカ i の解答が正しいと判断される場合の (成功) 報酬を $f(y_i)$ 、ワーカ i の解答が誤っていると判断される場合の (失敗) 報酬を $g(y_i)$ とする。

定義 4 (期待利得) ワーカ i の期待利得は、

$$F(x_i, y_i) = x_i f(y_i) + (1 - x_i) g(y_i) \quad (2)$$

と与えられる。

3. 報酬プランに基づくメカニズムの提案

本論文では、3 択問題を対象にする。3 択以上の問題では、ワーカが解答を1つだけに絞ることが難しい場合がある。そこで、解答を1つに絞れないワーカによるランダムゲスでの解答を防ぐために、解答の候補を2つ選択することを許容する。すなわち、3 択問題では解答でない一つの選択肢を選択することと同義になる。我々はワーカが一つの解答を選択する場合を1択、二つの解答を選択する場合を2択と呼ぶ。リクエスタはワーカから集めた解答によって最終的な解答を判定するが、このように、解答方法にバリエーションを有することで、リクエスタにとって、1つだけ解答させる場合よりも正解判定率を向上させることが可能となる。

さらに、ワーカの能力は同一ではないことに着目し、文献 [Sakurai 13] のアイデアを拡張し、本論文では上述の解答方法のもとで、自信に基づいてワーカの能力を区分するメカニズムの提案を行う。具体的には、ワーカに4つの価格プランを提示することで、ワーカを解答方法及び、自信がある (専門家) と自信がない (非専門家) に分類することを目標とし、次のようなプランと報酬額を設定する。

表 1: 報酬プランの例

	(正解, 不正解)
Plan 1H	(α_1, β_1)
Plan 1L	(α_2, β_2)
Plan 2H	(α_3, β_3)
Plan 2L	(α_4, β_4)

ここで、各プランについて具体的に説明する。例えば、3つの回答選択肢 0, 1, 2 のうち1つが正解である問題について、ワーカが1択または2択で解答し、各解答方法に基づいて自信によってワーカを分類する。

- 「0/1/2 しかない」というように解答を自信を持って絞れるときは Plan 1H

- 「0/1/2 だと思っけどあまり自信がない」ときは Plan 1L
- 「0 or 1/1 or 2/0 or 2 であることはほぼ間違いない」というときは Plan 2H
- 「0 or 1/1 or 2/0 or 2 だと思っけど自信がない」ときは Plan 2L

このような分類を可能にするためには、各プランでの報酬と自信の閾値をどのように設定するかが課題になる。ここでの閾値とは、各ワーカが

- 一つの解答が正しい確率がある値 c_1 より高ければ Plan 1H, c_1 より低ければ Plan 1L を選択する。
- 二つの解答のどちらかが正しい確率がある値 c_2 より高ければ Plan 2H, c_2 より低ければ Plan 2L を選択する。

という2つの値 c_1, c_2 のことを指す。

報酬に関しては、成功報酬が失敗報酬以下であるとワーカは解答を真実申告しないため、

$$\alpha_1 > \beta_1, \quad \alpha_2 > \beta_2, \\ \alpha_3 > \beta_3, \quad \alpha_4 > \beta_4$$

を仮定する。

我々の目標としては、各ワーカに解答と自信を真実申告してもらいたいというものがある。これを実現するためには、ワーカが真の解答と自信に対応するプランを選択すれば期待利得が最大化されることを保証する必要がある。

1 択のみが解答方法であれば、単純な自信に対する閾値 c_1 でプランが分割されるように報酬を設定して、ワーカに真実申告をさせることが可能である。しかしながら、本論文での設定のように2択まで解答を許容すると、我々の検討の結果、このような単純な手法では不可能であることが判明した。

定理 1 3 択問題において、1 択、2 択の自信の度合いと閾値 c_1, c_2 との大小のみによってプランが 1H, 1L, 2H, 2L の4つに分割されるように、報酬 (α_i, β_i) の組を設定してワーカに真実申告をさせることができない。

証明は紙面の都合上省略するが、直観的な証明としては、ワーカの1択、2択の自信度を総合的に判断しないと期待利得は最大化されず、そのためのプラン分割には2つの閾値だけでは不足するためである。

そこで、我々は、ワーカが真実申告を行うと期待利得が最大となるようなプラン分けとそのための報酬設定を検討する。具体的には、1 択、2 択の各プランへの (α_i, β_i) の報酬設定による期待利得は以下のように算出できるため、各プランに対応するワーカはそのプランを選択することで期待利得が最大化されるようにする。ワーカ i が最も正解の可能性が高いと思う選択肢に対する自信を $x_{i,1}$ 、次に正解である可能性が高いと選択肢に対する自信度を $x_{i,2}$ とする。すなわち、1 択で答えるときは正解率が $x_{i,1}$ 、2 択で答えるときの正解率は $x_{i,1} + x_{i,2}$ とする。なお、解答の正解率の比較から

$$x_{i,1} \geq x_{i,2} \geq 1 - (x_{i,1} + x_{i,2}) \quad (*)$$

は仮定される。このワーカが各プランを選択するときの利得の期待値は以下のように計算される。

- Plan 1H: $f_{1H}(x_{i,1}) \equiv x_{i,1}\alpha_1 + (1 - x_{i,1})\beta_1$

- Plan 1L: $f_{1L}(x_{i,1}) \equiv x_{i,1}\alpha_2 + (1 - x_{i,1})\beta_2$
- Plan 2H: $f_{2H}(x_{i,1} + x_{i,2}) \equiv (x_{i,1} + x_{i,2})\alpha_3 + (1 - x_{i,1} - x_{i,2})\beta_3$
- Plan 2L: $f_{2L}(x_{i,1} + x_{i,2}) \equiv (x_{i,1} + x_{i,2})\alpha_4 + (1 - x_{i,1} - x_{i,2})\beta_4$

このとき、ワーカの期待利得はどのプランを選択するかによって決まり、ワーカの自信によって

$$\arg \max_{1H, 1L, 2H, 2L} \{f_{1H}, f_{1L}, f_{2H}, f_{2L}\}$$

がどのプランになるか、がプランの選択条件となる。

以下、具体例を用いて説明を行う。

例 1 リクエストは、1 択への解答に対する自信 $x_{i,1}$ が閾値 $c_1 = 2/3$ 以上かどうか、2 択への解答に対する自信 $x_{i,1} + x_{i,2}$ が閾値 $c_2 = 3/4$ 以上かどうかを基準として、ワーカを分類したいとする。

このとき、報酬プランは以下のように決定できる。このとき、

表 2: 例 1 の報酬プラン

	(正解, 不正解)
Plan 1H	(14, 0)
Plan 1L	(12, 4)
Plan 2H	(10, 1)
Plan 2L	(8, 7)

各プランでの期待利得は、Plan 1H では $14x_{i,1}$ 、Plan 1L では $12x_{i,1} + 4(1 - x_{i,1}) = 8x_{i,1} + 4$ 、Plan 2H では $10(x_{i,1} + x_{i,2}) + (1 - x_{i,1} - x_{i,2}) = 9(x_{i,1} + x_{i,2}) + 1$ 、Plan 2L では $8(x_{i,1} + x_{i,2}) + 7(1 - x_{i,1} - x_{i,2}) = (x_{i,1} + x_{i,2}) + 7$ と算出される。

自信の数値が 10% 刻みの場合、自信に対するプランの選択分布は表 3 で示される。横が 1 択への自信 $x_{i,1}$ 、縦が 2 択への自信 $x_{i,1} + x_{i,2}$ を示す (空欄は (*) の条件をみたさない)。

表 3: ワーカのプラン選択分布

1 \ 2	40%	50%	60%	70%	80%	90%	100%
100%		2H	2H	2H	1H	1H	1H
90%		2H	2H	1H	1H		
80%	2H	2H	1L				
70%	2L						

この表のように、4 つのプランすべてがワーカに選ばれる可能性がある。例えば、あるワーカが選択肢 0, 1, 2 に対して、0 が 10%、1 が 70%、2 が 20% の可能性で正しいと自信を持っている場合、1 が正解とする Plan 1H を選択するのが最適となる。

こうした、2 択の解答をも考慮に入れた選択肢を導入することでリクエストはワーカの正解率を詳しく知ることができ、正解判定率を向上させることができる。

4. おわりに

本論文では、クラウドソーシングにおいて、3 択以上の複数選択肢が存在する場合の選択問題を対象とした場合、ワーカの能力を分類し、ワーカにとって能力を真実申告することが最適戦略となる報酬プランの設定方法を提案した。能力の分割方法として、自信に応じた報酬プランを選択させることでワーカに自信を間接的に申告させることが可能となっている。さらに、解答方法として、1 択だけでなく、2 択まで解答を絞れる場合の解答についても答えることができるように複数の解答手法を与えることで効率的にワーカの能力を分類することを可能にした。

今後の研究として、ワーカのプラン分割に応じた、リクエストの最終的な判定をどのように決定するかの判定方法の提案と、Amazon Mechanical Turk など、実際のクラウドソーシングサービスにおいて実験を行い、提案手法の有用性を検討することがある。

参考文献

- [Gneiting 07] Gneiting T. and Raftery, R. E.: Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association*, Vol. 102(477) pp. 359–378 (2007)
- [Law 11] Law, E. and Ahn, L. V.: *Human Computation*, Morgan & Claypool Publishers (2011)
- [Matheson 76] Matheson, J. E. and Winkler, R. L.: Scoring rules for continuous probability distributions, *Management Science*, Vol. 22(10), pp. 1087–1096 (1976)
- [Sakurai 13] Sakurai, Y., Okimoto, T., Oka, M., Shinoda, M., and Yokoo, M.: Quality-Control Mechanism utilizing Worker’s Confidence for Crowdsourced Task, *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)* (2013)
- [Savage 71] Savage, L. J.: Elicitation of personal probabilities and expectations, *Journal of the American Statistical Association*, Vol. 66(336), pp. 783–801 (1971)