

# クラウドソーシングにおける 品質に対するワーカーの影響

金地 紗里奈<sup>†1</sup> 小坂 隆浩<sup>†1</sup>

**概要:** 現在, 人工知能技術の機械学習に用いるデータ収集などで大量のデータを集めるために, ワーカーに対して多様なタスクを容易かつ安価での委託が可能であるクラウドソーシングが利用されている。しかし, クラウドソーシングでは不特定多数のワーカーにタスクを委託することになるので, 処理結果の正解率がワーカーの能力ややる気によって大きく異なってしまい, 全てのワーカーが正しい処理結果を返してくれるとは限らない。その結果, 誤った処理結果が返ってきてしまい低品質な処理結果となってしまう。本研究では, クラウドソーシングにおける高品質な処理結果の実現を目的として, 食品認識を用いた実験を行い, ワーカーによる評価要素が品質に及ぼす影響について考察する。

**キーワード:** クラウドソーシング, Amazon Mechanical Turk, 画像認識

## 1. はじめに

クラウドソーシングとはタスクの委託を行いたい場合にオンライン上で世界中にいる不特定多数のワーカーに対して人間には可能だが, 機械による自動化が難しいとされるタスクを委託することができるシステムである。クラウドソーシングの中でもマイクロタスク型の実施形態では, 人による事前データを必要とする人工知能技術の機械学習に用いるデータ学習やデータ分類に使われている。機械学習などに用いられるための大量のデータを集めるには, 大勢のワーカーの力が必要となる。そこで, ワーカーに対して多様なタスクを容易かつ安価での委託をするためにクラウドソーシングが利用される[1]。クラウドソーシングでは図1のように, タスクを委託する人はクライアントと呼ばれ, クラウドソーシングサイトにタスクを委託し報酬の仮払いを行う。タスクを処理する人はワーカーと呼ばれ, クラウドソーシングサイトからタスクを受注し, 受けたタスクの処理結果を納品する。納品されてから一定期間の承認期間が存在し, 期間内にクライアントは納品した結果を見て満足のいく結果が納品された場合は承認, 満足のいかない結果だった場合は拒否することができる。何も行わずに承認期間が過ぎた場合は, 納品されたタスクは自動承認される。最終的に, クライアントに承認されたワーカーのみ報酬を受け取ることができる。こうしてクラウドソーシングは, オンライン上で不特定多数のワーカーに対して容易かつ安価にタスクを委託することができ, コンピュータには難しいタスクを大量に処理することができる。

しかし, 容易にタスクを不特定多数のワーカーに委託することができるので, クライアントがわかりにくいタスク

の設定を行ったり, 処理結果の正解率がワーカーの能力ややる気によって大きく異なってしまい, 低品質な処理結果となってしまうという問題が指摘されてきた。低品質な結果では機械学習の学習データに用いることができない為, クラウドソーシングにおける品質改善手法の検討がされてきた。クラウドソーシングにおける既存の品質改善手法としては, ワーカーを管理するディレクターを雇うことや, タスク設計の工夫といった安定して高品質な結果が得られる手法が用いられている。

既存の品質改善手法について述べる。ワーカー管理の改善手法として西らの研究では, 外部システムのソーシャルネットワークを用いてワーカー間でタスクを再帰的に委託可能にすることにより, 処理結果の品質向上手法を提案した。タスクに正解したワーカー及びタスクを委託したワーカーに報酬を支払うことで, より能力の高いワーカーにタスクを委託することができ, 高い品質の処理結果が得られることを示した[2]。

タスク設計の工夫による改善手法としては, 芦川らの研究ではクラウドソーシングでゲーミフィケーションを用いた処理結果の品質向上手法を提案した。好適な環境での言語, 音声資源収集のために外部システムの音声認識エンジンを用いて録音された音声の得点を表示し, 得点が高いワーカーの報酬を上げることで読み誤りの除去を行った。その結果, 正しく読み誤りの除去を行うことができ, 高品質な処理結果を得られることを示した[3]。

既存手法として挙げられる外部システムとの連携[4]やタスク設計の工夫[4-7]では, 高品質な処理結果を得ることができることが示された。しかし, 外部システムの連携を行うには外部システムの新たな開発が必要となる。また, タスク設計の工夫に関してもその工夫が他のタスクへ応用できるとは限らない。

<sup>†1</sup> 同志社大学  
Doshisha University.

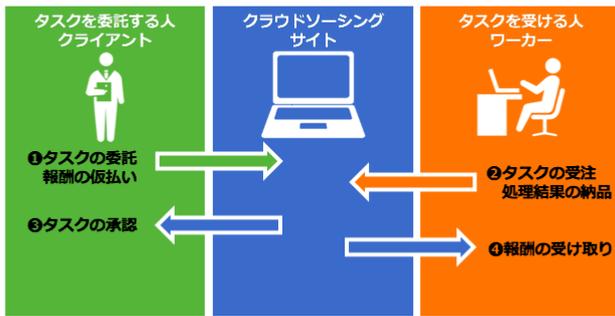


図1 クラウドソーシングの仕組み

本研究では、既存手法として挙げられたような外部システムとの連携ではやタスク設計の工夫を行わずにクラウドソーシングにおける高品質な処理結果の実現を目的として、ワーカーによる評価要素が品質に及ぼす影響を検討する。クラウドソーシングにおけるワーカーが持つ性別・年齢・タスクの承認数といった評価要素を用いて、品質に及ぼす影響を明らかにする。

## 2. クラウドソーシング

クラウドソーシングとは、匿名の不特定多数のワーカーに対してタスクを委託するという新しい仕事の形で、crowd(群衆)とsourcing(業務委託)を組み合わせた造語である。タスクの委託・受注から成果物の納品までがインターネット上で行われ、最初から最後まで非対面で仕事を終えることができる[8]。クラウドソーシングには一定のスキルを必要とするタスクから単純作業まで幅広いタスクがあり、タスクの種類に応じての実施形態を表1に示す。

表1 クラウドソーシングの実施形態

| 実施形態      | 概要               | 仕事の例    |
|-----------|------------------|---------|
| マイクロタスク型  | 短い時間で実行できる単純な仕事  | データ入力作業 |
| プロジェクト型   | 複雑で専門的な能力を要求する仕事 | ウェブ開発   |
| コンペティション型 | デザインなどの創造的な仕事    | ロゴ作成    |

本研究では、マイクロタスク型を用いて誰にでもできるような単純な作業を対象とする。マイクロタスク型のタスクでは、特別なスキルを使用しないので学生や主婦といった幅広いワーカーによる受注が可能である。マイクロタスク型では、タスクを受注することができるワーカーが多いが、タスクを処理するワーカーの能力や信頼性を計ることが難しい。

## 2.1 品質管理

品質管理は、クラウドソーシングにおける大きな課題の一つである。通常、機械学習などで用いるデータセットの作成は信頼に値するワーカーによって行われる。しかし、クラウドソーシングを用いてデータを集める場合、不特定多数のワーカーに委託することになるので委託するワーカーの能力や信頼性を把握すること難しい。他にも、タスクの委託者側の問題やタスクの形式・報酬について考える必要もあり、信頼性の高い結果を得るためには品質管理が重要である。安定した品質での処理結果を得るための品質管理手法は以下の3種類に分類される[9]。

1. クライアントに対する管理手法
2. タスクに対する管理手法
3. ワーカーに対する管理手法

クライアントに対する管理手法では、報酬が適切に支払われないことや処理結果に問題がないにもかかわらず結果を却下され、ワーカーのモチベーションが低下してしまうことが問題としてあげられる。タスクに対する管理手法では、わかりにくいタスクの設計や委託するワーカーに応じた難易度設定ができないことが問題としてあげられる。ワーカーに対する管理手法では、能力ややる気が異なるワーカーにタスクを委託するので、処理結果の品質にばらつきが起きてしまうところが問題として挙げられる。

既存研究では、ワーカーを管理するディレクターを雇うことでワーカーの信頼性の向上を図ったり[2]、ゲーミフィケーションをタスクに導入することで、ワーカーの能力の向上を図っている[3]。本研究では、処理結果自体の品質を向上しなければ高品質な結果とは呼べないので、品質に対して最も関係が深いワーカーに対する管理手法について考える。

## 3. 実験方法

本研究では、クラウドソーシングの品質管理手法の考察のために、画像を用いた認識をワーカーに依頼する実験を行う。タスクの承認数はより多いほど、高品質が期待される。実験ではワーカー自身に対する評価要素として性別・年齢・タスクの承認数の品質に及ぼす影響を食事認識の正解率により考察する。

### 3.1 実験1

実験1では、ワーカーは図2に示す画像を用いて、枠に囲まれた4品目の食品認識を行う。本研究では、食品名をワーカーが返答できれば正解として4品目の認識を行う。Amazon Mechanical Turkを用いて行い50人のワーカーに0.4ドルでタスクを委託した。表2は納品された処理結果

の一例である。4品の認識率ごとにワーカーを区別し、評価要素による正解率の変動を表す。評価要素として性別・年齢・タスクの承認数を用いて正解率との相関関係进行评估する。実験では高品質な結果を正解率100%と定義して高品質な処理結果の実現を目指す。



図2 実験に用いた食品画像

表2 処理結果の一例

|       | 1 | 2 | 3 | 4 | 承認数   | 年齢 | 性別     |
|-------|---|---|---|---|-------|----|--------|
| ワーカー1 | ○ | ○ | ○ | × | 74891 | 30 | Male   |
| ワーカー2 | ○ | ○ | ○ | ○ | 95590 | 49 | Female |
| ワーカー3 | ○ | ○ | ○ | ○ | 72394 | 46 | female |

### 3.2 実験2

実験2では、ワーカーは図3に示す画像を用いて画像に隠れた動物のイラスト4つの認識を行う。実験2では、動物名をワーカーが返答できれば正解として Amazon Mechanical Turk を用いて行い、50人のワーカーに0.4ドルでタスクを委託した。4つの認識率ごとにワーカーを区別し、評価要素による正解率の変動を表す。評価要素として性別・年齢・タスクの承認数を用いて正解率との相関関係进行评估する。実験では高品質な結果を正解率100%と定義して高品質な処理結果の実現を目指す。



図3 実験に用いた画像

### 3.3 Amazon Mechanical Turk

実験では、Amazon Mechanical Turk (Mturk) を用いる。Mturkはアメリカのアマゾン社のAmazon Web Serviceの1つであり、コンピュータプログラムを人間の知能と組み合わせ、コンピュータだけでは不可能なタスクを処理するマイクロタスク型のクラウドソーシングである。細かく分割されたタスクを少額の金額で大量のワーカーに委託することができ、人間には可能だが機械には難しいとされる自然言語理解、音声認識などの領域においての機械学習をはじめとするデータ駆使アプローチ[10]に用いられている。

## 4. 実験結果

50人のワーカーに実験方法で述べた実験を行った実験結果について述べる。

### 4.1 実験1

実験1では、食事画像を用いた食品認識のイラストの認識を行った。図4はワーカーの出身国の統計であり、全体の約9割がアメリカ出身であることがわかる。図5では各正解率による人数の変動を表している。正解率75%と正解率100%のワーカーが多く見られた。

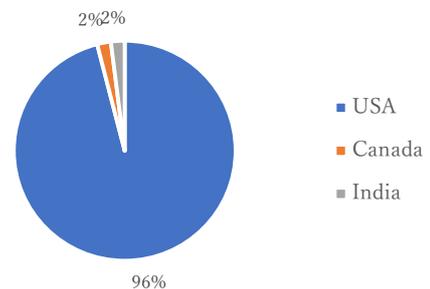


図4 ワーカーの出身国

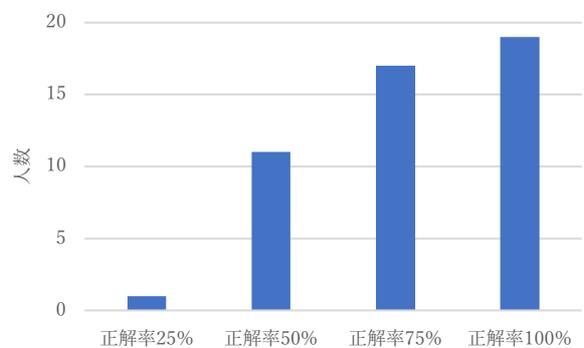


図5 正解率による人数の変動

クラウドソーシングを用いた食品認識についてワーカーのタスクの承認数による正解率の変動を箱ひげ図で表したものを図 6 に示す。正解率 100%の場合タスクの承認数が 200000 以下で初心者のワーカーが多く見られた。正解率 75%の場合タスクの承認数が 1000000 以下で初心者から熟練者までの多くのワーカーが見られた。正解率 50%の場合タスクの承認数が 400000 以下で初心者から中級者までのワーカーが見られた。結果、承認数の少ない初心者の正解率が高く承認数の多さが品質に及ぼす影響は少なかった。

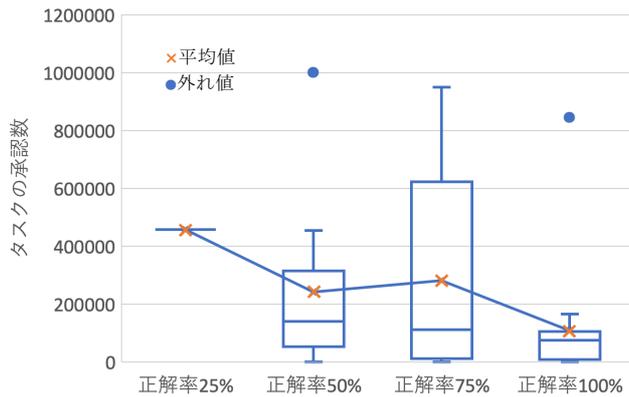


図 6 承認数による正解率の変動

クラウドソーシングを用いた食品認識についてワーカーの年齢による正解率の変動を箱ひげ図で表したものを図 7 に示す。高品質とする正解率 100%の場合、他の結果に比べ平均値や最大値が少し高かったが正解率の分散に大きな違いは見られず年齢が品質に及ぼす影響は少なかった。

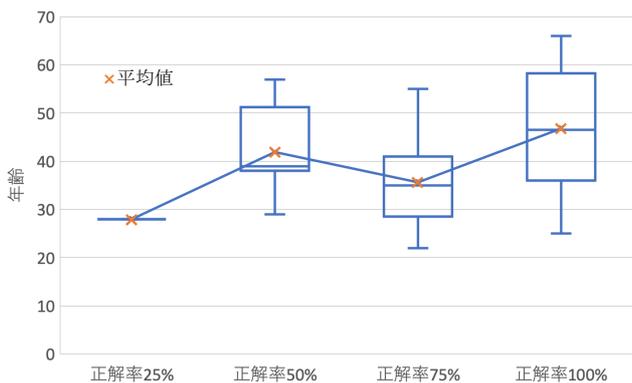


図 7 年齢による正解率の変動

クラウドソーシングを用いた食品認識についてワーカーの性別による正解率の変動を表したものを図 8 に示す。正解率が 25%の場合ワーカーは男性 1 人のみで、正解率が 100%のワーカーは女性が 15 人、男性が 3 人で圧倒的に女性のワーカーの方が多かった。女性の正解率が高かったが、女性であるからという理由だけでなく用いたタスクが食品認識だった為、性別が品質に影響した可能性が考えられる。

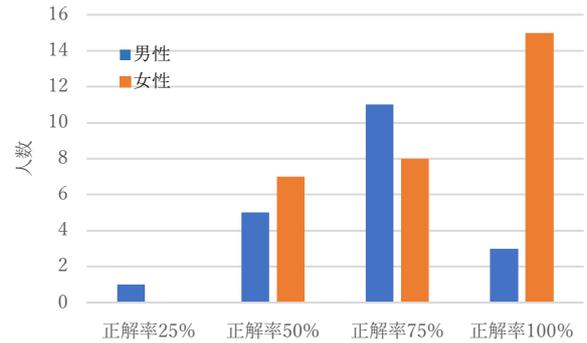


図 8 性別による正解率の変動

#### 4.2 実験 2

実験 2 では、隠し絵を用いて動物のイラストの認識を行った。図 9 はワーカーの出身国の統計であり、全体の約 7 割がアメリカ出身であることがわかる。アメリカ以外ではインド出身のワーカーも見られたがそれ以外の国はあまり見られなかった。図 10 では各正解率による人数の変動を表している。正解率 75%のワーカーが 50 名中 21 名で最も多い結果となった。

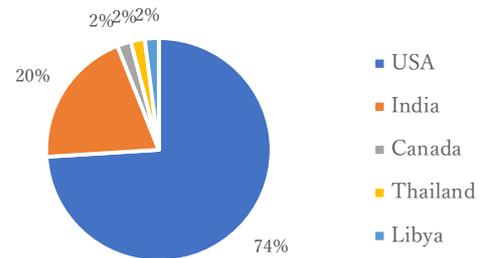


図 9 ワーカーの出身国

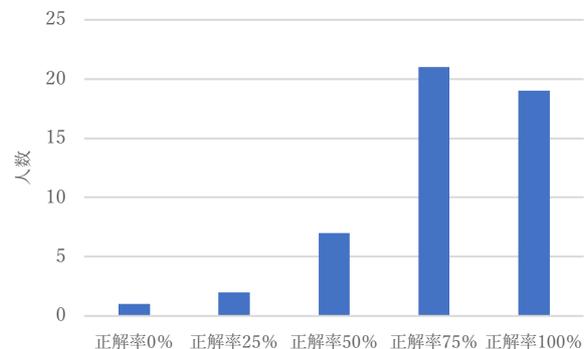


図 10 正解率による人数の変動

クラウドソーシングを用いた動物の認識についてワーカーのタスクの承認数による正解率の変動を箱ひげ図で表したものを図 11 に示す。正解率 25%から正解率が高くなるに

つれて平均値は下がっているが、分散に大きな差は見られなかった。

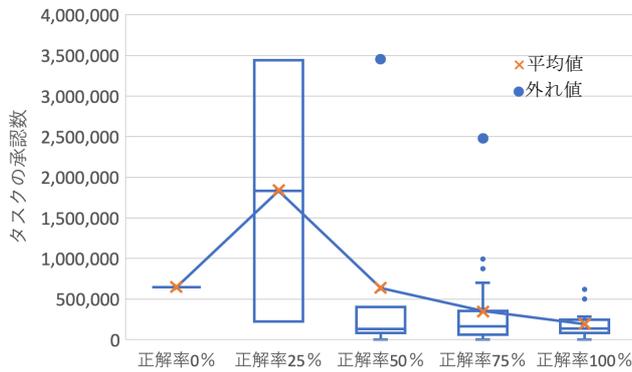


図 11 承認数による正解率の変動

クラウドソーシングを用いた動物の認識についてワーカーの年齢による正解率の変動を表したものを図 12 に示す。実験 1 と同じように正解率の分散に大きな違いは見られず年齢が品質に及ぼす影響は少なかった

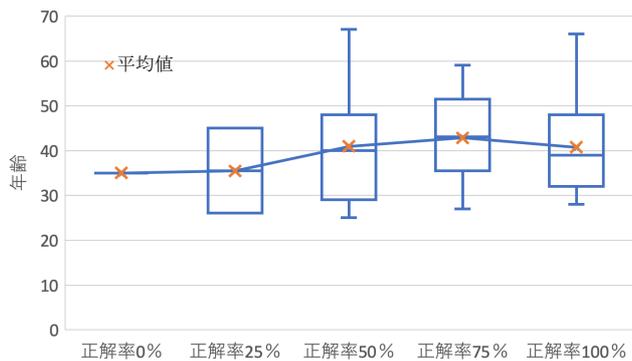


図 12 年齢による正解率の変動

クラウドソーシングを用いた動物の認識についてワーカーの性別による正解率の変動を表したものを図 13 に示す。実験 1 ほどではないが女性の処理結果に対して優位な結果が得られた。

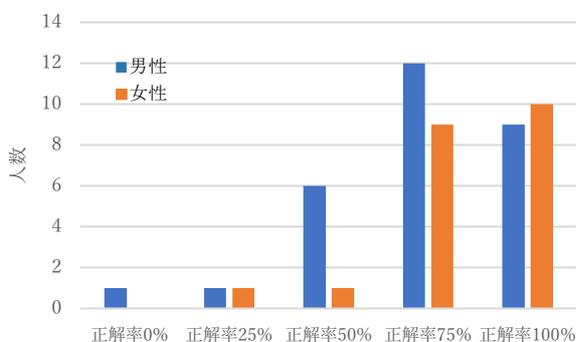


図 13 性別による正解率の変動

### 4.3 考察

タスクの承認数による正解率の変動では、正解率が 100% のワーカーに承認数が少ないワーカーが実験 1・実験 2 共に多く見られた。しかし、大きな変化が見られないことから品質に与える影響とは考えにくい。承認数の変動によって高品質な結果が返せるわけではなかった。

年齢による正解率の変動では、正解率の分散に大きな違いは見られなかったことから、ワーカーの年齢が品質に及ぼす影響は少なく有意な結果は得られなかった。

性別による正解率の変動では、実験 1 では圧倒的に女性のワーカーの方が多い結果となり、実験 2 でも女性のワーカーに優位な結果が得られた。よって、実験で用いたタスクの場合は女性に委託することで高品質な処理結果を得ることができた。

本研究ではタスクによって品質に相関関係が見える可能性があることがわかった。食品認識では、一般的に女性の方が料理に触れることが多いことから性別ではなくタスクが品質に影響を及ぼし、高品質な結果が得られたと考えられる。

### 5. まとめ

本研究では、クラウドソーシングにおける高品質な処理結果の実現を目指すため、ワーカーによる評価要素が品質に及ぼす影響を考察した。実験 1 ではタスクの承認数と性別による正解率の変動が見られ、実験 2 では性別による正解率の変動が見られた。よって、性別による正解率の変動から、タスクによっては性別が品質に及ぼす影響があることを示した。本研究では、クラウドソーシングにおける高品質な処理結果の実現を目指すため、ワーカーによる評価要素が品質に及ぼす影響を考察した。今後の課題として、他の評価要素での品質に及ぼす影響、他のタスクでの実験をする必要がある。

### 参考文献

- [1] 大澤翔吾, 柳井啓司, クラウドソーシングによる食事画像認識モデルの自動構築, DEIM2013 The 5th Forum on Data Engineering and Information Management, C4-4, 2013.
- [2] 西智樹, 小出智士, 大野宏司, 長屋隆之, ソーシャルネットワークを用いたクラウドソーシングの品質向上, 第 27 回人工知能学会, 3M3-OS-07d-4, 2013.
- [3] 中田康太, 芦川将之, PRIVATECROWDSOURCING を用いた言語、音声資源の収集 ~音声収集と品質評価~, 第 27 回人工知能学会, 3M3-OS-07d-2, 2013.
- [4] Laohaprapanon, Suriyan, Kimberly Ortleb, and Gaurav Sood. Street Sense, Learning from Google Street View. *arXiv preprint arXiv:1807.06075*, 2018.
- [5] 佐々木孝輔, 平田章, 井上智雄, マイクロタスクによる線画イラスト生成手法, 情報処理学会論文誌, Vol.57, No.1, pp. 260-269, 2016.

- [6] 高木啓伸, 井床利生, 齊藤新, 小林正朋.クラウドアクセシビリティークラウドソーシングによる障害者支援ー. 人工知能学会誌, Vol. 29, No. 1, pp. 41-46, 2014.
- [7] 巻口 誉宗, 東 正造, 下村 道夫, 金丸 直義, ヒューマンコンピューターションにおけるタスク割り当て手法の提案, JSAI2014 オーガナイズドセッション, pp. 1-4, 2014.
- [8] 堀田裕理, クラウドユーザに対する非金銭的インセンティブの効果の分析, 修士論文, 京都大学, 2015.
- [9] 芦川将之, 川村隆浩, 大須賀昭彦, マイクロタスク型クラウドソーシングプラットフォーム環境における精度向上手法の導入と評価, 人工知能学会論文誌, 29(6), pp.503-515, 2014.
- [10] 鹿島久嗣, 梶野洸, クラウドソーシングと機械学習, 人工知能学会誌, Vol. 27, No. 4, pp. 381-388, 2012.