

## 集合知を用いたメカニズム設計の検討

櫻井 祐子†

産業技術総合研究所†

小山 聡‡

北海道大学‡

## 1 はじめに

人々の知を利用して計算機のみでは解決できない課題解決を目指すプラットフォームとしてクラウドソーシングがある [1, 2]. インターネット上では, Amazon Mechanical Turk をはじめとした, 数多くのクラウドソーシングサービスが提供されている.

クラウドソーシングでは安価に不特定多数のワーカーを集めることが可能な一方で, 得られる作業結果の品質が問題となっている. 良く知られている問題の一つに収集データの信頼性がある. 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で得る場合, 膨大な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ーのラベルには, ワーカーの意図の有無に関わらず, ラベルに対して誤りが入る可能性がある.

このような品質管理に関する問題の解決方法の一つとして, ワーカーに対するインセンティブを適切に与えるためのメカニズム設計がある. メカニズム設計とは望ましい性質を満たすメカニズムを設計することである. これまで, クラウドソーシングを対象としたメカニズムの設計が数多く行われている. しかしながら, 提案メカニズムは, 誘因両立性や個人合理性など, メカニズムが満たすべき性質を保証するために複雑になる傾向があった. クラウドソーシングの参加者はメカニズムデザインの非専門家であることを考慮すれば, 複雑な報酬設定を含むメカニズムは受け入れられないと考える. そこで, 本研究ではクラウドソーシングに適したメカニズムを設計するために, 人々の集合知を用いたメカニズムの設計方法について検討を行う.

## 2 対象とするタスク

本研究では, 図1に示す写真をワーカーに提示し, この写真に移っているアイテムを可能な限り列挙してもらおうというアノテーションタスクを対象にメカニズムを検討する. 我々は日本のクラウドソーシングサービスの一つのランサーズ (<http://www.lancers.jp/>) を利用して



図1: 画像のタグ付け

30名のワーカーを集め, 「1,000円の予算で, 写真に載っているアイテムのタグ付けを可能な限り行いたい場合, 何人のワーカーをどのような報酬設定で雇い, どのように正解を決めるかを記述してください」というタスクを実行した. 一般に, リクエストはタスクを依頼する際, 予算制約を有していることが想定されるため, この問題設定は妥当と考える.

## 3 ワーカーが考案したメカニズムの分析

本章では, ワーカーによって考案されたメカニズムの分析結果を示す.

## 3.1 統計的分析

本節では統計的分析結果を示す. まず, ワーカーが考案した報酬設定は下記の4つに分類された.

表1: 考案報酬設定と考案ワーカー数の関係

		考案ワーカー数
固定 価格	条件なし	6
	条件あり	10
成果報酬		10
従量報酬		4

条件付き固定価格とは, アノテーション数にしきい値条件が付与されたものであり, 条件を満たさない場合は報酬を支払わないという設定になっている. そのため, 固定価格ではあるが成果報酬の一形式とみなすことができる. しきい値条件に関して, 6名のワーカーが最低10個, 残りの4名についてはそれぞれ, 1個, 5個, 20個, 30個と設定した.

†EEX 2E Mechanism design by collective intelligence

†Yuko SAKURAI ‡Satoshi OYAMA

†National Institute of Advanced Industrial Science and Technology

‡Hokkaido University

最終的に正しいアノテーションを決定する方法として、17名のワーカが多数決投票を適用した。さらに、2名のワーカは集めたアノテーションからリクエストが正しいと考えるアノテーションを選ぶという方法を提案した。残りの11名は決定方法に関する記述を行わなかった。

### 3.2 ワーカによるメカニズムの判定

本節では、クラウドソーシングのワーカらによって考案されたメカニズムの妥当性を判定させた結果を示す。20名のワーカを雇い、5段階評価で考案された30件のメカニズムの判定を行わせた。下記に、最も評価が高かったメカニズムと最も評価が低かったメカニズムを示す。

- 最高評価のメカニズム (A1) : 20名のワーカを雇用し、10個以上のアノテーションを付けてもらい、条件を満たしたワーカには50円を支払う。5人以上のワーカから得たアノテーションを正解とする。
- 最低評価のメカニズム (A2) : 4名のワーカを雇用し、30分の制限時間を設ける。各ワーカに20個以上のアノテーションを付けてもらい、条件を満たしたワーカには250円を支払う。

### 3.3 理論的分析

最高評価と最低評価の2つのメカニズムを比較すると、雇用ワーカ数が大きく異なること、制限時間の有無が顕著な差分である。この設定の差異がどのような影響をもたらすのかを理論的に分析する。

理論的分析のため、リクエストは限られた予算  $K$  の範囲内かつ制限時間  $T$  内に有効なアノテーションを多く集めたいと仮定する。考えられるアノテーション集合を  $\{1, 2, 3, \dots\}$  とする。さらに、アノテーション数は有限個に限らないため、1人のワーカが時間  $t$  までにアノテーション  $k$  を付ける確率  $p_k(t)$  は指数分布  $1 - e^{-a_k t}$  に従うものとする。  $a_k$  は各単位時間ごとに  $k$  を思いつく確率とする。また、各アノテーションは番号が大きいほど思いつきやすく、べき分布に従うものとして、  $a_k = ck^{-\alpha}, \alpha > 0$  とする。但し、  $\alpha \leq 1$  のときはタグの個数を  $1, 2, 3, \dots, M$  までと制限する。

この問題設定において、先の2つのメカニズムに対して分析を行う。ここでは、紙面の都合上、べき分布の指数が1の場合のみを示す。さらに、  $c = 1/3, \alpha = 1, M = 200, T = 30$  とする。これは、各ワーカは  $t = T (= 30)$  までに平均 34.0 個のタグを思いつくという設定となる。

従って、比較的難しい画像のタグ付けのタスクとみなすことが出来る。

表2に各メカニズムでワーカ数を変化させたときの結果を示す。A1は最高評価のメカニズム、A2は最低評価のメカニズムを示し、括弧内の人数は雇用ワーカ数を示す。作業時間は報酬を得られるときの平均作業時間を示す。

表2: 分析結果 (制限作業時間 30分)

メカニズム	作業時間	報酬を得る確率	時給	アノテーション数
A1. (10人)	6.28	1	0.016K	3.6
A1. (20人)	6.28	1	0.008K	9.0
A1. (30人)	6.28	1	0.005K	14.2
A2. (3人)	23.63	0.732	0.010K	7.1
A2. (4人)	23.63	0.732	0.007K	10.7

表2より、最高評価メカニズム (A1) では、ワーカは短時間でほぼ確実に  $K/10$  の報酬が得られることが分かる。従って、リクエストにとってはA2よりも効率的なメカニズムといえる。一方、リクエストにとって、「アノテーションを10個付与すれば「定額の報酬が得られる」というルールを設定したことで、ワーカからは10個しかアノテーションを付与しないこととなり、多数のアノテーションを得ることが難しい。最低評価メカニズム (A2) は制限時間内にアノテーション数の条件を満たすことができない確率が  $1/4$  以上あるため、ワーカにはリスクの大きいメカニズムと考えることができる。

## 4 おわりに

本論文では集合知を用いたメカニズム設計に向けて、クラウドソーシングのワーカらにメカニズムを考案してもらい、その結果について3つの方向から分析を行った。

今後は、非定型タスクなど、様々なタスクを対象として、クラウドワーカによるメカニズム設計の可能性を検証する。

## 参考文献

[1] 鹿島久嗣, 小山聡, 馬場雪乃. ヒューマンコンピュータシオンとクラウドソーシング. 講談社, 2016.

[2] Edith Law and Luis Von Ahn. *Human Computation*. Morgan & Claypool Publishers, 2011.