クラウドソーシングにおけるフィードバックと追加報酬に関する研究

石崎 日香莉　　土方 嘉徳

関西学院大学商学部

contact@soc-research.org

概要　近年，ビジネスや学術研究においてクラウドソーシングが多く利用されるようになっている．そこではワーカの成果物の質を担保するための仕組みが必要とされている．特に，質の高い成果物を残すことができるワーカには継続してタスクに取り組んでもらい，そうでないワーカには早期にタスクから離脱してもらうことが理想である．クラウドソーシングでは，事例に対して正解ラベルを付与するタスクに見られるように，一つの仕事をいくつかに区切ってワーカに依頼することがある．本研究では，最初のタスクの成果物に対して評価を行いフィードバックをユーザに行うことと，その評価に基づき追加報酬を支払うことの両方が，ワーカの継続率とワーカの全体としてのパフォーマンス（正答率）に影響するかどうかを明らかにする．被験者実験より，フィードバックは，1 回目において出来の良かったワーカに対して，正答率を下げることなく，継続率を上げることが分かった．また，フィードバックと追加報酬を組み合わせることにより，出来の悪かったワーカに対して，継続率を低下させることが分かった．

キーワード　クラウドソーシング，成果物評価，フィードバック，追加報酬，継続率，正答率

1 はじめに

近年，ビジネスや学術研究においてクラウドソーシングサービスが多く利用されるようになってきている．クラウドソーシングサービスでは，委託者はインターネットを通じて作業を行うワーカの募集を行うため，ビラや求人誌等の従来手段よりも，手軽にワーカを集めることができる．また，委託を受けたワーカは，いつでも，どこでも作業が出来るので，隙間時間に依頼を済ませることが可能である．しかし，委託者はワーカを直接監視することができないため，ワーカがまじめにタスクに取り組んでいるかどうかは確認できない．そこで，クラウドソーシングサービスでは，ワーカにまじめに取り組んでもらえるように工夫することや，成果物の質を保証できるような仕組みを導入することが求められている[1,2].

特に，機械学習の学習データを集める際には，たくさんの事例に対して多くの人に正解ラベルを付与してもらう必要がある．このようなタスクにおいては，ワーカが遂行するタスクの質を高めつつ，ワーカにより多くの事例に取り組んでもらうことが望ましい．しかし，成果物の質と量はトレードオフの関係にある．すなわち，成果物の質を高めるために評価基準を高めて信頼性を担保しようとすると，収集される成果物の量は少なくなる．一方，成果物を評価することなく，かつワーカへの報酬金額を高くすると，成果物の量は多くなるが，その質は低くなる．

これをなるべく緩和し，成果物の一定の質は担保しつつ，その量を下げないようにする工夫が必要となる．一般的な方法として，タスクそのものを工夫することやタスクの実施環境を工夫することが考えられる．具体的には，タスクを営利目的ではなく，社会的に意義のあるものにすること[3]や，タスクをゲーム形式にすることで，ワーカがタスク自体を楽しみながら遂行できるようにすること[4, 5]が挙げられる．また，タスク遂行前に条件を満たさないワーカの参加を排除したり[6,7]，タスク遂行にかかった時間を計測して極端に短い時間で完了したワーカの結果を排除したりすること[8]が考えられる．さらに，ワーカが遂行したタスクの結果に対して，インセンティブにつながる効果的な報酬を設定したり[9-11]，評価基準を伝えて，ワーカがより真剣に取り組んでもらえるようにしたり[12]することが考えられる．

本研究では，機械学習用の学習データ収集のようにタスクをいくつかに分割して実施できることを前提として，すなわち同じワーカが分割されたタスクを連続して取り組む場合において，成果物の質を担保するための仕組みを考案する．具体的には，ワーカが遂行した最初のタスクの成果物を評価して，その結果をワーカにフィードバックすることを考える．ワーカが自分の成果物の評価を確認できるようにすることで，もしその評価が良かった場合は，そのワーカはそのタスクに対して，より高いモチベーションを持つようになると期待される．また，評価が良かった場合にのみ，追加で報酬を支払うことも行う．これは，ワーカは自分の頑張りにより報酬が変化することを知ると，タスク取り組みに対してより高いモチベーショ

Copyright is held by the author(s).
The article has been published without reviewing.

ンを持つようになると期待されるからである.

本研究では, このような仕組みが成果物について一定の質を担保しつつ, なおかつ多くの成果物を収集することが可能かどうかを検証する. 具体的には, 以下4つの仮説を検証する.

仮説 1:追加報酬がある方がない場合より, ワーカはタスクを継続する

仮説 2:フィードバックがある方がない場合より, ワーカはタスクを継続する

仮説 3:質の高い(成績の良かった)ワーカは, 追加報酬がある方がない場合より, 継続する

仮説 4:質の高い(成績の良かった)ワーカは, フィードバックがある方がない場合より, 継続する

論文構成は, 以下のとおりである. 2 章で本研究の実験方法を, 3 章で実験により入手したデータの分析結果とそれに対する考察を述べる. 最後に 4 章で, まとめを述べる.

2 実験方法

2.1 タスクの概要

本研究では, 実際のクラウドソーシングサービスであるクラウドワークス上で実験を行なった. 実験用に用意したタスクは, 機械学習用の学習データの収集を想定し, 図 1 で示される花の画像からその種類を特定するものとした. このタスクは, 我々が独自に PHP で実装したシステム上で実施される. タスクでは,「ハルジオン」という種の花と「ヒメジョオン」という種の花のどちらかが表示され, ワーカは表示された花がこの 2 種類のどちらかを当てる. この 2 種類の見分け方は簡単ではないため, 見分け方の説明は常に表示することにした. 1 つのタスクは, 上記のような課題が 20 個連続で出題される. 実験では, このタスクを 2 個用意し, ワーカは 1 回目のタスクに取り組んだ後, 2 回目に任意で取り組むことができる. すなわち, 1 回目で終了することもできるし, 2 回目に進んで取り組むこともできる.

本実験では, 成果物の質を担保するための仕組みとして, 1 回目のタスクの評価結果(正答率)をフィードバックすることと, さらに追加報酬を支払うことを考える. そこで, 実験では以下の 3 条件を用意した.

(1) 評価結果(正答率)をフィードバックし, なおかつ回答が優秀だった者(正答率が高かった者)には追加報酬を支払う場合

(2) 評価結果(正答率)をフィードバックするが, 追加報酬はない場合

(3) 評価結果(正答率)をフィードバックせず, 追加報酬もない場合

上記の条件(2)と(3)を比較することで, フィードバックの有無とタスクパフォーマンス(すなわち 2 回目のタスクの正答率)または継続率(タスク 1 を取り組んだのちにタスク 2 に取り組むワーカの割合)との関係を明らかにできる. また, (1)と(3)を比較することで, フィードバックと追加報酬の両方の有無とタスクパフォーマンスまたは継続率との関係を明らかにできる. 最後に, (1)と(2)を比較することで, フィードバックを行うときにさらに追加報酬があったときとなかった時で, タスクパフォーマンスまたは継続率に違いがあるかどうかを明らかにできる.

図 1 クラウドソーシングの実験システムの画面

2.2 実験条件の詳細

1 つのタスクは 20 個の問題から構成される. そのため, 正答率はそのワーカの正答数を 20 で割って計算した. 事前にパイロットテストを行い, 回答が優秀であった者(成績優秀者)と, そうでなかった者(非成績優秀者)を区別するための閾値として, 80%の正答率を採用した. すなわち, 実際の実験では 1 回目のタスクの正答率が 80%以上のワーカを成績優秀者とし, それ以外のワーカを非成績優秀者とした. 報酬は 1 回目と 2 回目のそれぞれのタスクごとに全員一律で 50 円を支払った. 追加の報酬は 25 円を支払うこととした.

クラウドソーシングでの実験においては, 明らかにやる気のないワーカや, でたらめに取り組んだワーカは, ある程度簡単な仕組みで排除することができる. 本実験では, タスク開始前に生まれた年を尋ね, タスク終了後に干支を尋ねた. また, タスク中に, 明らかに間違いと

分かる画像（大きなユリの花の画像）を途中に1問挿入した（以下，これをダミータスクと呼ぶ）．通常の問題に対する回答の選択肢は 2 択であるが，ダミータスクはこれに，[どちらでもない]を付け加えた 3 択とした．また，実験システムでは，タスク開始時刻とタスク終了時刻を計測し，所要時間を計算した．そこで，生まれた年と干支が一致しないワーカ，ダミータスクで[どちらでもない]以外を回答したワーカ，所要時間が 60 秒未満のワーカ（物理的に回答不可能な時間で回答したワーカ）は，明らかに信頼できないワーカとみなし（これらの基準をフィルタリングルールと呼ぶ），実験結果から除外した．

参加資格は年齢・性別を問わず，実験ごとに 300 人を募集した．また，参加者が被らないように，実験の間に2ヶ月ほど期間を設けた．実験条件(1)は，2019 年 12 月 6 日に「ハルジオン・ヒメジョオン画像判定タスク（1 回目）（3 分程度で終わります）」という名称でクラウドワークスに掲載した．実験条件(2)は，2019 年 10 月 21 日に「花の種類判定タスク（1 回目）（3 分程度で終わります）」という名称で掲載した．実験条件(3)は，2019 年 7 月 1 日に「植物判定タスク（1 回目）（3 分程度で終わります）」という名称で掲載した．実験条件(1)-(3)のいずれも，用いた花の画像は，提示の順番も含めて，全く同じである．実験条件(2)については，参加資格として，過去に実験条件(3)の「植物判定タスク（1 回目）（3 分程度で終わります）」に取り組んでいないことを挙げた．また，実験条件(1)については，過去に実験条件(3)の「植物判定タスク（1 回目）（3 分程度で終わります）」と実験条件(2)の「花の種類判定タスク（1 回目）（3 分程度で終わります）」に取り組んでいないことを挙げた．各実験終了後，クラウドワークスのユーザ ID を検査して，一度実験条件(2)または(3)のいずれかに参加したワーカは排除した．

3 実験結果

3.1 全体のワーカの分析

各実験条件の実験参加者（ワーカ）の数と，フィルタリングルールで信頼できないワーカとして除外されたワーカの数，分析対象として残ったワーカの数を表 1 に示す．この結果から，完全にでたらめに回答したワーカはほとんどいなかったことが分かる．ワーカの正答率に注目する．最初に，各実験条件の正答率（1 回目と 2 回目のそれぞれ）に対してコルモゴロフ・スミルノフ検定を行ったところ，正規性は見られなかった．実験条件(1)-(3)のワーカの平均正答率を表 2 に示す．実験条件(1)-(3)を合わせた 1 回目の平均正答率は 0.76 で，2 回目の平均正答率は 0.83 であった．また，実験条件(1)-(3)のいずれも，1回目より2回目の方が平均正答率が高くなった．これらのことから，実験で用いたタスクは学習効果のあるタスクと言っても良いと思われる．

表 1 実験条件別の実験参加者（ワーカ）の数

	全ワーカ	信頼できないワーカ	分析対象のワーカ
実験条件(1)-1 回目	295	8	287
実験条件(1)-2 回目	213	22	191
実験条件(2)-1 回目	268	9	259
実験条件(2)-2 回目	226	13	213
実験条件(3)-1 回目	282	11	271
実験条件(3)-2 回目	231	11	220

実験条件により2回目の正答率に違いがあるかどうかを統計的に検証した．実験条件は，(1), (2), (3)の 3 種類あるので，クラスカル・ウォリス検定を行ったところ，p 値 =0.025 < 0.05 となり，有意差が見られた．さらに，ボンフェローニ補正による多重比較を行ったところ，実験条件(1)と(3)の間において，p 値=0.033 < 0.05 となり，有意差が見られた．実験条件(1)と(2)の間と実験条件(2)と(3)の間は，p 値= 0.091, 2.17 > 0.05 となり，有意差はなかった．実験条件(1)と実験条件(3)の間で有意差が見られたことから，1 回目の正答率をフィードバックし，さらに優秀なワーカに追加報酬を払う場合は，何も行わない場合よりも正答率が高いことが分かった．ただし，実験条件(1)は 1 回目の平均正答率が，実験条件(2), (3)よりも若干高くなっているので，追加報酬の効果により 2 回目の正答率が高くなったかどうかまでは分からない．また，実験条件(2)と実験条件(3)の間では，ほとんど有意差が見られなかったことから，1 回目の正答率をフィードバックしてもしなくても，ほとんど 2 回目の正答率には影響がないことが分かる．

表 2 ワーカの平均正答率

	1 回目	2 回目
実験条件(1) FB あり・追加報酬あり	0.77	0.85
実験条件(2) FB あり・追加報酬なし	0.75	0.83
実験条件(3) FB なし・追加報酬なし	0.74	0.82

FB: フィードバック

実験条件(1)-(3)のワーカの平均継続率を表 3 に示す．実験条件により継続率に違いがあるかどうかを統計的

に検証した．実験条件の種類とタスクを継続したか否かにおいて，カイ二乗検定を行ったところ，p 値=4.84e-6<0.05 となり有意差が見られた．また，クラメールの連関係数は 0.17 であった．さらに残差分析を行ったところ，実験条件(1), (2), (3)における調整済み残差が，それぞれ 4.94，-2.73，-2.32（いずれも絶対値>1.96）となり，有意差が見られた．このことより，フィードバックと報酬のある実験条件(1)は，その他の条件よりも継続率が低いことが分かった．このことから，仮説 1 は支持されなかった．フィードバックのみのある実験条件(2)とフィードバックも報酬もない実験条件(3)の両方とも，実験条件(1)よりも継続率が高いことが分かった．このことから，仮説 2 も支持されなかった．これらについて考察すると，追加報酬を受けることができなかったユーザは，タスク遂行のモチベーションが下がり，2 回目のタスクに取り組まなかったののかもしれない．しかし，フィードバックを行うが，追加報酬がない場合である実験条件(2)は，大きく継続率が向上した．自分の回答の出来を知ることができ，2 回目も取り組んで，結果を向上させようというモチベーションが働いたのかもしれない． ただし，それでもフィードバックがない場合とは差がなかったことから，フィードバックそのものに全体の継続率を向上させる効果はなかったと言える．

表 3 ワーカの平均継続率

	継続率
実験条件(1) FB あり・追加報酬あり	0.67
実験条件(2) FB あり・追加報酬なし	0.82
実験条件(3) FB なし・追加報酬なし	0.81

FB: フィードバック

3.2 1 回目の成績ごとのワーカの分析

次に，ワーカの能力（またはやる気）により，タスクの継続率，および 2 回目の正答率に差が出るのかどうかを見る．具体的には，ワーカを 1 回目の正答率が 80%以上であった成績優秀者と，80%未満であった非成績優秀者に分けて，それぞれの平均正答率と平均継続率を算出した．その結果を表 4, 5 に示す．

まず，継続率に注目する．成績優秀者と非成績優秀者において，実験条件間で継続率が異なるかどうかを統計的に検証する．まず，成績優秀者について実験条件の種類とタスクを継続したか否かにおいて，カイ二乗検定を行ったところ，p 値=0.017<0.05 となり有意差が見られた．また，クラメールの連関係数は 0.14 であった．

表 4 成績優秀者の正答率と継続率

	1 回目 正答率	2 回目 正答率	継続率
実験条件(1) FB あり・追加報酬あり	0.87	0.85	0.71
実験条件(2) FB あり・追加報酬なし	0.86	0.86	0.85
実験条件(3) FB なし・追加報酬なし	0.86	0.86	0.80

表 5 非成績優秀者の正答率と継続率

	1 回目 正答率	2 回目 正答率	継続率
実験条件(1) FB あり・追加報酬あり	0.66	0.85	0.61
実験条件(2) FB あり・追加報酬なし	0.67	0.80	0.80
実験条件(3) FB なし・追加報酬なし	0.64	0.79	0.82

さらに残差分析を行ったところ，実験条件(1)と実験条件(2)における調整済み残差が，それぞれ 2.72，-2.06（いずれも絶対値>1.96）となり，有意差が見られた．このことから，報酬のある実験条件(1)は，報酬のない実験条件(2)よりも，継続率が低いことが分かった．この理由を推察することは困難であるが，次も続けて追加報酬をもらえる自信が持てなかったことや，追加報酬をもらうために，別のダミータスクに回答しないといけない煩雑さなどが影響した可能性がある．また，フィードバックを行わなかった実験条件(3)に対して，フィードバックを行った実験条件(2)では，継続率が高いことが分かった．このことから，フィードバックはある程度成績優秀者のモチベーションを保つことに成功したと思われる．すなわち判定能力が高いワーカやもともとやる気があったワーカは，自分の成績を知ることにより，モチベーションを保つことができたと思われる．

次に，非成績優秀者について実験条件の種類とタスクを継続したか否かにおいて，カイ二乗検定を行ったところ，p 値=5.98e-5<0.05 となり有意差が見られた．また，クラメールの連関係数は 0.22 であった．さらに残差分析を行ったところ，実験条件(1)と実験条件(3)における調整済み残差が，それぞれ 4.40，-2.37（いずれも絶対値>1.96）となり，有意差が見られた．このことから，報酬のある実験条件(1)は，他の実験条件よりも，継続率が低いことが分かった．このことから，追加報酬があると，非成績優秀者（すなわち判定能力が低いかやる気の低いワーカ）は続けてタスクに取り組むモチベーションを失ったものと思われる．また，追加報酬もフィードバックもな

い実験条件(3)が最も高い継続率となった．フィードバックのみ行う実験条件(2)においては，連関は見られなかったことから，非成績優秀者はフィードバックを見て自分の成績を知ると，わずかにモチベーションが下がるのかもしれない．

次に，正答率に注目すると，成績優秀者の 2 回目の正答率はいずれの実験条件においても，0.85 程度の安定した正答率になっている．成績優秀者の 1 回目の正答率の平均は，実験条件ごとに，0.87（実験条件(1)），0.86（実験条件(2)），0.86（実験条件(3)）であることから，1 回目も 2 回目も正答率はほとんど変化がなかったことが分かる．2 回目のタスクを行ったワーカについて，1 回目の正答率と 2 回目の正答率に違いがあるかどうかを検定した．一人のワーカが 1 回目と 2 回目と両方のタスクを行っているので，対応のあるデータとなる．そこで，ウィルコクソンの符号順位検定を行った．すると，実験条件(1), (2), (3)において，それぞれ p 値= 0.43, 1.00, 0.60（いずれも>0.05）となり有意差は見られなかった．

一方，非成績優秀者の 2 回目の正答率は，1 回目の正答率よりも高くなっている．具体的には，彼らの 1 回目の正答率の平均は，実験条件ごとに，0.66（実験条件(1)），0.67（実験条件(2)），0.64（実験条件(3)）であった．これらに比べると 2 回目の正答率は 0.13 から 0.20 も大きくなっている．ウィルコクソンの符号順位検定を行ったところ，実験条件(1), (2), (3)において，それぞれ p 値= 1.07e-5, 1.11e-14, 4.49e-15（いずれも<0.05）となり有意差が見られた．最も正答率が上がったのは，実験条件(1)（フィードバックあり・追加報酬あり）であった．驚くべきことに，2 回目の正答率は，成績優秀者とほとんど変わらない成績になった．この理由の可能性としては，追加報酬をもらうために，1 回目にまじめに取り組まなかった者も，2 回目はまじめに取り組んだことや，非成績優秀者のうち 2 回目に追加報酬をもらう自信があった者のみが取り組んだことが挙げられる．実験条件(2)（フィードバックあり・追加報酬なし）になると，正答率の増加量は大きく減った．実験条件(1)に比べて，継続率も高くなっていることを考えると，まじめに取り組まない者や判定能力の低い人も引き続き，タスクに取り組んだのだと思われる．

3.3 結果のまとめ

全ワーカの分析の結果から，実験条件の比較では，残念ながら，依然として成果物の質（すなわち正答率）と成果物の量（すなわち継続率）の間にトレードオフが存在した．1 回目のタスクの成績により，成績優秀者と非成績優秀者に分けた分析では，フィードバックは，若干ではあるが，成績優秀者の継続率向上に寄与したと思われる．しかし，非成績優秀者の継続率にほとんど影響がなかったため，全体の正答率にはそれほど大きな影響を及ぼさなかった．追加報酬は，大きく非成績優秀者のモチベーションを下げることに成功したが，同時に成績優秀者のモチベーションも下げてしまう結果になった．これらのことから，仮説 3 については完全には支持されなかったと言えるが，質の低いワーカの排除には成功したと言える．フィードバックを導入することで，質の高いワーカは継続してくれたものの，同時に質の低いワーカも継続させてしまった．すなわち仮説 4 についても完全には支持されなかったと言える．

本実験結果から，2 回目以降の成果物の質を向上させるためには，正答率のフィードバックに加えて追加報酬を与えることが必要であることが分かる．一方，成績優秀者のモチベーションも下げてしまう欠点が分かったため，モチベーションを下げないようにする工夫が必要と思われる．成績優秀者のモチベーションが下がってしまった原因を明らかにすることが今後の課題になる．

3.4 本研究の制約

本実験では，機械学習用の学習データの取得を想定し，事例に対して真か偽かを判定するような単純なタスクを設定した．これが，創造的なタスク，例えば面白い川柳を考えるタスクや，心安らぐ音楽のメロディーを作るタスクなどにも当てはまるかどうかは分からない．また，画像の判定という，比較的直感的に回答可能な課題を設定した．そのため，例えば顧客の購買データからそのユーザが早期にサービスを解約するかどうかを判定したり，商品に対するレビュー文からその著者が商品を好意的に評価しているかを判定したりするような，論理的に考えて判定しないといけないようなタスクにも適用できるかどうかは分からない．他の種類のタスクにおいて，フィードバックと報酬が継続率と正答率にどのような影響を与えるかを調べることは，今後の課題となる．

本実験におけるフィードバックの与え方は，単純に全 20 問中の正答数とその割合（正答率）を返すものであった．非常に単純で，言語的な装飾や工夫は，何も行っていないフィードバックであった．「あなたは，非常に優秀なワーカです．」や「これほど正確に判定してくれて，私どもも大変助かります．」のような，相手を褒めたり，自分たちが助かっているようなニュアンスを伝えたりするようなことはしていない．どのようにフィードバックを与えるかにより，継続率は変わってくる可能性がある．フィードバックの与え方や表示方法も，今後の課題となる．

4 まとめ

本研究は，クラウドソーシングサービスにおいて，ワーカの成果物の質を維持しつつ，その量を増加させる仕組みについて検討した．具体的には，一つの仕事をいくつかに区切ってワーカに依頼する場合を対象として，

仕事の区切りにおいて，その前の仕事に対する正答率をフィードバックすることと，その正答率に応じて追加報酬を支払うことを提案した．フィードバックと追加報酬の両方がある場合，フィードバックのみがある場合，両方ない場合の 3 つの条件で，実際のクラウドソーシングサービス上で実験を行った．

実験の結果，フィードバックと追加報酬の両方がある場合は，全体として高い正答率を達成できるが，継続率が下がり，成果物の量が少なくなってしまうことが分かった．フィードバックのみがある場合は，1 回目に高い正答率を達成したワーカについては継続率を向上させることができ，1 回目に低い正答率しか達成できなかったワーカについては少しだけ継続率を低下させることができた．しかし，全体としての正答率は，何の工夫も行わなかった場合とそれほど変わらなかった．

今後は，フィードバックにおけるメッセージを工夫することで，正答率の高いワーカのモチベーションを上げることができるかどうかを明らかにしたい．また，事例に対する真偽の判定を行うような単純なタスクだけでなく，新しい作品を生み出すような創造的タスクにおいても，提案した仕組みが機能するかどうかを確かめたいと考えている．

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テーションとクラウドソーシング，講談社，2016.

[2] Wael Soliman and Virpi Kristiina Tuunainen: Understanding Continued Use of Crowdsourcing Systems: An Interpretive Study, Journal of Theoretical and Applied Electronic Commerce Research, 10(1), pp. 1-18, 2015

[3] Aniket Kittur, et al.: An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets, Proc. of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM'11), pp. 321-328, 2011.

[4] Yoonsung Hong, Haewoon Kwak, Young-Min Baek and Sue Moon: Tower of Babel: A Crowdsourcing Game Building Sentiment Lexicons for Resource-scarce Languages, Proc. of the ACM 22nd International Conference on World Wide Web (WWW'13), pp. 549-556, 2013.

[5] Jorge Goncalves, Simo Hosio, Denzil Ferreira and Vassilis Kostakos: Game of words: tagging places through crowdsourcing on public displays, Proc. of the 2014 ACM conference on Designing Interactive Systems (DIS'14), pp. 705-714, 2014.

[6] Shigeo Matsubara and Meile Wang: Preventing Participation of Insincere Workers in Crowdsourcing by Using Pay-for-Performance Payments, IEICE Transactions on Information and Systems, E97D(9), pp.2415-2422, 2014.

[7] Mohammad Allahbakhsh, et al.:Quality Control in Crowdsourcing Systems: Issues and Directions, IEEE Internet Computing, 17(2), pp. 76-81, 2013.

[8] Dan Cosley, et al.: How Oversight Improves Member-Maintained Communities, Proc. of the 2005 Conference on Human Factors in Computing Systems (CHI'05), pp. 11-20, 2005.

[9] Duncan J. Watts and Winter Mason: Financial Incentives and the “Performance of Crowds”, Proc. of the ACM SIGKDD Workshop on Human Computation (HCOMP'09), pp.77-85, 2009.

[10] Chien-Ju Ho, et al.: Incentivizing High Quality Crowdwork, Proc. of the 24 th International Conference on World Wide Web (WWW'15), pp. 419-429, 2015.

[11] Ming Yin, Yiling Chen and Yu-An Sun: The Effects of Performance-Contingent Financial Incentives in Online Labor Markets, Proc. of the 27th AAAI Conference on Artificial Intelligence (AAAI'13), pp.1191-1197, 2013

[12] Steven Dow, Anand Kulkarni, Scott Klemmer and Björn Hartmann: Shepherding the Crowd Yields Better Work, Proc. of the ACM 2012 Conference on Computer Supported Cooperative Work, pp. 1013-1022, 2012.