

クラウドソーシングでのタスク選択に関する行動モデリング

Behavior Modeling of Selecting a Task in Crowdsourcing

松原 繁夫*1

Shigeo Matsubara

伊奈 祐輔*2

Yusuke Ina

*1 京都大学 社会情報学専攻

Kyoto University

*2 京都大学 工学部情報学科

Kyoto University

Task assignments in crowdsourcing market are an important issue to achieve efficient problem solving in crowdsourcing. However, the previous researches have not sufficiently examined the issue of workers' incentive in selecting a task. This paper proposes a method for finding appropriate amounts of rewards, which results in achieving an efficient task allocation, when a requester has different tasks in crowdsourcing.

1. はじめに

近年、クラウドソーシング市場 Amazon Mechanical Turk (MTurk) の利用増などに伴い、人工知能技術の適用対象としてクラウドソーシングが注目を集めている [小山 14]。本研究では、MTurk における複数種類タスクのワーカーへの割当問題に関して議論する。

クラウドソーシングでは、複雑なタスクを単純なサブタスクに分割して個々のワーカーに依頼し、その結果を統合する手法がよくとられる。例えば、動物画像へのアノテーションタスクについて考える。すべての動物に詳しいワーカーを見つけるよりは、イヌの名前に詳しいワーカー、ペンギンの名前に詳しいワーカーを見つける方が簡単そうである。よって、まず、(1) 動物画像をイヌの画像、ペンギンの画像などに分類し、つぎに (2) イヌの画像に対してポメラニアンやマルチーズといった記述を求める。こうすることで、より早くより正確なタスク処理が可能になると期待できる。

さて、上記の 2 種類のアノテーションタスクに関して、前者は専門知識を必要とせず多くのワーカーが請負可能であるのに対し、後者は専門知識を必要とし少数のワーカーのみが請負可能である。本稿では、前者を易タスク、後者を難タスクと呼ぶ。このとき、専門知識を持ったワーカーには難タスクを割り当て、それ以外のワーカーには易タスクを割り当てれば、効率的なタスク処理が可能となる。

しかし、MTurk ではタスクリストに様々なタスクが掲示され、ワーカーはタスクリストを見て自己の基準に従ってタスクを選択する。そのため、依頼者が強制的にタスクを割り当てることはできない。よって、報酬設定によって間接的にワーカーの振る舞いを制御する、すなわち、ワーカーに自発的にタスク選択をさせる必要がある。たとえ難タスクの報酬額が易タスクより高額であっても、処理費用を鑑みて易タスクの方が得とみなされれば、専門知識を持ったワーカーも易タスクを選択してしまう。一方、難タスクの報酬額を高く設定し過ぎれば、依頼者の支払総額が増加する。これらにより、効率的なタスク処理が困難になる。

これまでクラウドソーシングでのタスク割当てに関する研究はあるが [Ho 12]、ワーカーのインセンティブの問題は十分に議論されてこなかった。また、実践という面でも、時給 5.6 ドル相当になるようにというガイドラインがあるだけで、複数種

のタスクを持つ場合のガイドラインは与えられていない。そこで本研究では、ワーカーのインセンティブに考慮して、各ワーカーに自己のスキルに合ったタスクを自発的に選択させる報酬設定法の考案を目的とする。

2. モデル

本章では、依頼者にとって適切な報酬設定問題のモデルを示す。依頼者は n_e 個の易タスク (報酬額 r_e) と n_d 個の難タスク (報酬額 r_d) を持つとする。依頼者にとって、すべてのタスクが処理されることが必要である。

個々のワーカーは易タスク処理費用 c_e と難タスク処理費用 c_d で特徴付けられる。これをタイプと呼ぶ。このとき、ワーカーの行動は個人合理性制約と誘因両立性制約によって規定される。前者はタスクを請け負っても損をしないという制約であり、後者はワーカーにタイプを申告させる場合に真実申告するという制約である。本稿では、易タスクと難タスクの 2 種類のみが存在する場合を扱うので、難タスクの処理費用が多きいワーカーには易タスクを選択させ、難タスクの処理費用が小さいワーカーには難タスクを選択させることを意味する。

個人合理性制約 $r_e - c_e \geq 0$ と誘因両立性制約 $r_e - c_e \geq r_d - c_d$ を満たすワーカーは易タスクを選択し、 $r_d - c_d \geq 0$ 、かつ、 $r_d - c_d \geq r_e - c_e$ を満たすワーカーは難タスクを選択する。単位時間当たりのタイプ (c_e, c_d) ワーカーの出現数を $m(c_e, c_d)$ で表すと、報酬設定問題は以下の最適化問題として定式化できる。なお、*deadline* は依頼者が持つタスク完了までの制限時間を表す。また、各ワーカーはいずれか一方のタスクのみ請負可能と仮定する。

$$\min_{r_e, r_d} (n_e r_e + n_d r_d)$$

, so that

$$\int_0^{r_e} \int_{r_d - r_e + c_e}^{\infty} m(c_e, c_d) dc_d dc_e \geq n_e / \text{deadline}$$

$$\int_0^{r_d} \int_{r_e - r_d + c_d}^{\infty} m(c_e, c_d) dc_e dc_d \geq n_d / \text{deadline}$$

制約条件式において、内側の積分が誘因両立性制約に対応し、外側の積分が個人合理性制約に対応する。

ワーカーが市場に順次到着する点に着目すればポアソン分布を仮定したモデル化なども考えられるが、本稿では単純に単位時間当たりの出現数が決まっていると考える。

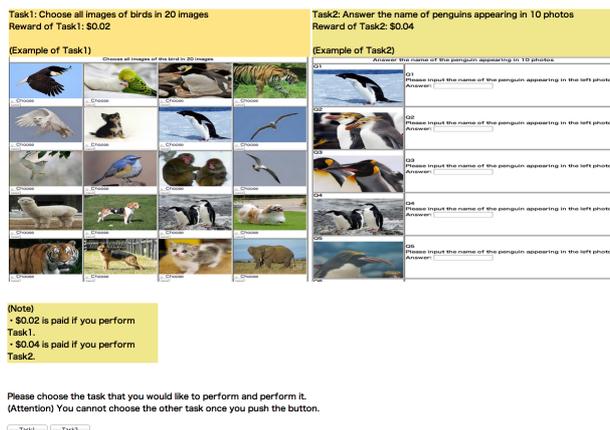


図 1: 分類・アノテーションタスク

3. 報酬設定パラメータの獲得

上述の報酬設定法の適用には、ワーカのタイプ分布の獲得が必要である。そこで、まず予備実験を行い、その結果からタイプ分布を推測し、それを用いて報酬設定を求めることにする。

3.1 予備実験の計画

予備実験に関する実験計画を述べる。ここでの目的は報酬配分法における各タイプのワーカの単位時間当たり出現数 $m(c_e, c_d)$ を推定することである。本研究では、図 1 に示すように、動物画像へのアノテーションに関する以下の易タスクと難タスクを用意し、その元で様々な報酬額設定におけるワーカのタスク選択行動を観察した。

- 易タスク：20 枚の動物画像を提示し、その中から鳥類画像を全て選択するタスクである。20 枚の画像中に、鳥類画像はランダムな場所に 10 枚表示されるようにした。作業としては、ボタンをクリックするだけである。
- 難タスク：10 枚のペンギン画像を与えて、コウテイペンギンやフンボルトペンギンといったペンギンの名前の記述を求めるタスクである。作業としては、複数の選択肢を与えて選ばせるのではなく、キーボードをタイプして答えを入力することが求められる。

さて、MTurk において易タスクと難タスクを別々のタスク (HITs) として依頼すると、ワーカが何を比較してそのタスクを選択したのかが観測できない。そのため、図 1 に示すように一つの HIT の中で、鳥類画像の分類タスクとペンギン画像に対するアノテーションタスクがあることと、各タスクの報酬額を説明し、ワーカにいずれか一方のタスクを選択させることにした。

報酬額については、易タスクと難タスク、双方最低報酬額を 2 セントとし、表 1 に示す 10 通りの報酬額の組合せを試みた。各報酬額設定において 30 回タスクを依頼し、タスク選択の様子を観察した。

3.2 予備実験の結果

MTurk 上で 2014 年 1 月に実験を行った。全体として 151 人のワーカが本タスクを実行した。その中で複数回、つまり異なる報酬額設定でのタスクを 2 回以上実行したワーカは 87 人であった。複数回タスクを実行したワーカのうち、易タスクの

表 1: 予備実験における報酬額の組合せと結果

	易タスク 報酬額	難タスク 報酬額	易タスク 選択数	難タスク 選択数	完了時間 [min.]
Case1	\$0.02	\$0.02	25(25)	5(5)	155
Case2	\$0.02	\$0.04	16(15)	14(13)	255
Case3	\$0.02	\$0.06	18(16)	12(11)	249
Case4	\$0.02	\$0.08	18(17)	12(12)	249
Case5	\$0.02	\$0.10	22(20)	8(8)	250
Case6	\$0.04	\$0.04	22(22)	8(8)	51
Case7	\$0.04	\$0.08	19(19)	11(11)	108
Case8	\$0.04	\$0.12	18(18)	12(12)	114
Case9	\$0.04	\$0.16	15(15)	15(12)	137
Case10	\$0.04	\$0.20	20(20)	10(9)	73

みを実行したワーカは 47 人、難タスクのみを実行したワーカは 7 人、両方のタスクを実行したワーカは 33 人であった。

表 1 に各場合における易タスク・難タスクの選択数と完了に要した時間を示す。各タスクの選択数欄の括弧内の数字は正解率が 60%以上のワーカの内数である。例えば、Case2 の場合、易タスクが 16 回選択され、その内 15 回は正解率が 60%以上であった。今回は正解率 60%を誠実にタスクを実行した目安としたが、正解率に関する考察は後で述べる。

実験結果で興味深い点は、易タスクと難タスクの報酬額が同額でも、少なくない数のワーカが難タスクを選択したことである。タスクの最後に任意回答としてアンケート欄を設け、なぜこちらのタスクを選択したか尋ねた。易タスクを選択したワーカの主要な回答は、易タスクの方が難タスクに比べて簡単だから、難タスクに答えられる知識を持ってないから、という十分想定されるものであった。一方、難タスクを選択したワーカの主要な回答は、難タスクの報酬額が易タスクの報酬額に比べて高いから、難タスクに答えられる知識を持っているからというものの他に、難タスクの内容に興味があるから、難タスクの内容に挑戦してみたいと思ったからという回答があった。これは、報酬という外因的動機だけでなく、内因的動機が働いていると考えられる [Rogstadius 11]。今回得られたデータは統計処理をするほどの分量はないが、挑戦してみたいといった動機が存在する点は興味深い。

易タスクと難タスクの選択比率に関しては、易タスクの報酬額が \$0.02 のときは、易タスク・難タスクで同額のときに難タスク選択数が少ないが、他の場合は大きな差はない。モデル上で考えれば、難タスクの報酬額が増えれば、難タスクの選択数が減少することはないが、例えば、易タスク報酬額が \$0.02 のとき、難タスク報酬額が \$0.08 から \$0.10 に増加すると難タスク選択数が減るなど、得られた結果は、モデルと十分に整合するものではない。この理由としては、各タイプのワーカの到着がランダム性を有し、その影響を受けたといったことが考えられる。この他に、要求品質の認識に影響を与えている可能性もある。すなわち、\$0.04 であれば高額の報酬を得ているという認識はないが、\$0.10 は高額報酬であり、それに応じた作業品質が要求されていると認識して、難タスクの処理費用を高めに見積もるといった行動である。今回得られたデータの範囲内では、どのような要因が影響しているかを十分に明らかにできないため、ランダム性を有すると解釈して、各タイプのワーカの出現数を推定することにする。

さて、MTurk においては不誠実なワーカの存在が議論されている [Ipeirotis 10]。不誠実なワーカとは報酬獲得のみを目的として、依頼者の指示に従わない作業結果を提出するワーカを指す。今回の実験においても、正解率が 0%というワーカが存在した。また、難タスクだけではなく、易タスクにおいても正解率が 10%といったワーカが存在した。どのような動物

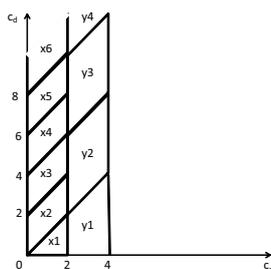


図 2: ワーカのタイプに関する領域分割

を見てきたかという生育環境による差があるかもしれないが、易タスクに関しては、ケアレスミスも考慮しても、普通に作業すれば正解率が 10% になることはないと考えられる。よって、データは不誠実なワーカを示している。

ただし、報酬額の増加に応じて不誠実なワーカ数が増加するという顕著な傾向も確認できなかった。どちらかと言えば、難タスクにおいてもかなり正解率が高かったと言える。不誠実なワーカがそれほど増えなかった理由としては以下のものが考えられる。MTurk においては、上述のようにタスクリストにタスクが掲示され、それを見てワーカがタスクを選択する。掲示の際、タスクの簡単な内容や報酬額が記述されるが、詳細な情報は個々のタスク画面に行かなければ確認できない。例えば、易タスク \$0.02 と難タスク \$0.10 の報酬組の場合、タスクリストの報酬額欄には \$0.02 が示され、クリックしてタスクの説明画面に来たところで \$0.10 の可能性が示される。よって、タスクリスト上で報酬額で並び替えを行えば、本タスクは低報酬額のタスクとして認識される。このような理由から、不誠実なワーカの増加が抑制されたのではないかと考えられる。

最後に、完了時間について見てみる。完了時間とは、各場合においてタスクを掲示してから 30 個のタスクがすべて処理されるのに要した時間である。今回得られた結果では、易タスクの報酬が \$0.02 の場合に比べて、\$0.04 の場合は半分程度の時間でタスクを完了できている。これは、単純に \$0.04 という報酬額を見て、選択するワーカが増えたものと考えられる。さて、各場合を詳細に見ると、易タスクと難タスクの報酬額が同額の場合にタスクが早く完了している。この点については、30 個のタスクがどのようなタイミングで処理されたかをより詳細に分析する必要があると考えるが、本稿ではワーカ到着のランダム性に帰せられると考えて、後の議論を行うことにする。

3.3 ワーカ分布の推定

予備実験で得られた結果を用いて単位時間当たりの各タイプのワーカの出現数 $m(c_e, c_d)$ の推定を行う。これは、図 2 に示すように領域を区切り、各領域に関して単位時間当たりのワーカ出現数を求めることになる。図の横軸は易タスクに関する処理費用 c_e を、図の縦軸は難タスクに関する処理費用 c_d を表す。

領域 x_1 から x_6 が、易タスク \$0.02 の場合に対応する。例えば、Case2 (易タスクの報酬額が \$0.02、難タスクの報酬額が \$0.04) の場合を考える。このとき、タスクリスト上で見える報酬額は \$0.02 であるため、易タスクの処理費用が \$0.02 より大きいワーカは、本タスクに反応しない。また、2. 章で述べたモデルに従ってワーカがタスク選択を行うとすると、領域 x_1, x_2 のタイプを持つワーカが難タスクを選択し、領域 x_3, x_4, x_5, x_6 のタイプを持つワーカが易タスクを選択する。ここで、Case2 では 255 分の間に難タスクを選択した誠実な

ワーカが 13 人、易タスクを選択した誠実なワーカが 15 人であるため、領域 x_1, x_2 の和、領域 x_3, x_4, x_5, x_6 の和が各々 $13/255 * 60 = 3.058823529$, $15/255 * 60 = 3.529411765$ と一致すれば良い。よって、易タスクの報酬額が r_e 、難タスクの報酬額が r_d の場合のタスク完了時間を $t(r_e, r_d)$ で表すと、以下の制約を満たす x_1, \dots, x_6 を求めればよいことになる。

$$\begin{cases} x_1 = 5/t(2, 2) \\ x_2 + x_3 + x_4 + x_5 + x_6 = 25/t(2, 2) \\ x_1 + x_2 = 13/t(2, 4) \\ x_3 + x_4 + x_5 + x_6 = 15/t(2, 4) \\ x_1 + x_2 + x_3 = 11/t(2, 6) \\ x_4 + x_5 + x_6 = 16/t(2, 6) \\ x_1 + x_2 + x_3 + x_4 = 12/t(2, 8) \\ x_5 + x_6 = 17/t(2, 8) \\ x_1 + x_2 + x_3 + x_4 + x_5 = 8/t(2, 10) \\ x_6 = 20/t(2, 10) \end{cases}$$

しかし、上記の式をすべて満たす x_1, \dots, x_6 は存在しない。よって、 x_1, \dots, x_6 から算出される出現数と実際の出現数の差の二乗和が最小化されるように x_1, \dots, x_6 を決定することにする。この問題は二次計画問題となり、CPLEX を用いることで以下の結果を得た。 $x_1 = 0.0$, $x_2 = 3.123$, $x_3 = 0.0$, $x_4 = 0.0$, $x_5 = 0.0$, $x_6 = 4.563$ 。 x_3, x_4, x_5 については 0 となっている。これは、難タスクの報酬額を \$0.06, \$0.08 と変えても、易タスクと難タスクの選択比率があまり変化しなかったためである。 x_1 については、難タスクの報酬額が \$0.02 の結果から、0 より大きい数字となるべきであるが、難タスクの報酬額が \$0.02 以外の場合の結果の類似性が高いため、差の二乗和の最小化という観点からは、 $x_1 = 0$ となった。

易タスクの報酬額が \$0.04 の場合についても同様に考えることで、 $y_1 = 2.106$, $y_2 = 0.586$, $y_3 = 0.0$, $y_4 = 0.0$, $y_5 = 0.0$, $y_6 = 10.304$ が得られる。なお、図 2 には表示されていないが、 y_5, y_6 は y_1, y_2, y_3, y_4 と同様に縦軸方向に領域を伸ばすことで得られる領域である。得られた値を用いることで、依頼者の要求に応じて適切な報酬額を算出することが可能となる。

4. 報酬設定法の評価

4.1 評価実験の計画

2. 章で提案した報酬設定法が適切に機能するかどうかを調べるための実験を行う。評価実験においても、鳥類画像を分類する易タスクとペンギンの名称記述を求める難タスクを用いた。依頼者が以下の要求を持つ場合を考える。

Case A: 易タスク 15 個、難タスク 15 個、制限時間 3 時間

Case B: 易タスク 20 個、難タスク 10 個、制限時間 5 時間

Case A は Case B に比べて、より要求が厳しいと言える。以上の場合に関して、前章で推定した各タイプのワーカ分布を用いて報酬額を計算した。ただし、MTurk では \$0.01 単位での設定となるため、丸めると、Case A では、易タスク \$0.04、難タスク \$0.08、Case B では、易タスク \$0.02、難タスク \$0.04 という報酬額が得られた。

さて、予備実験のところでも述べたように、易タスクと難タスクを個別の HIT として依頼すると、易タスクと難タスクの間でのワーカのタスク選択制御ができていないかどうかの評価が難しくなる。よって、予備実験と同様に、一つの HIT の中に易タスクと難タスクを掲示し、そこでワーカに選択させること

表 2: 評価実験における報酬額の組合せと結果

	易タスク 報酬額	難タスク 報酬額	易タスク 選択数	難タスク 選択数	完了時間 [min.]	Case A 総費用	Case A 時間 [min.]	Case B 総費用	Case B 時間 [min.]
Case11	\$0.02	\$0.02	25(23)	5(5)	119	\$0.6	357	\$0.6	238
Case12	\$0.02	\$0.04	16(15)	15(14)	87	\$0.9	93	\$0.8	116
Case13	\$0.02	\$0.06	16(14)	14(12)	143	\$1.2	179	\$1.0	204
Case14	\$0.02	\$0.08	13(13)	17(15)	155	\$1.5	179	\$1.2	238
Case15	\$0.02	\$0.10	15(14)	15(14)	170	\$1.8	182	\$1.4	243
Case16	\$0.04	\$0.04	20(19)	10(10)	116	\$1.2	174	\$1.2	122
Case17	\$0.04	\$0.08	14(14)	16(15)	130	\$1.8	139	\$1.6	186
Case18	\$0.04	\$0.12	12(12)	18(17)	147	\$2.4	184	\$2.0	245
Case19	\$0.04	\$0.16	12(12)	18(15)	135	\$3.0	169	\$2.4	225
Case20	\$0.04	\$0.20	15(14)	15(14)	131	\$3.6	140	\$2.8	187
Case21	\$0.06	\$0.06	20(19)	10(9)	120	\$1.8	200	\$1.8	133
Case22	\$0.06	\$0.12	16(15)	14(13)	128	\$2.7	148	\$2.4	171
Case23	\$0.06	\$0.18	17(16)	13(13)	123	\$3.6	142	\$3.0	154
Case24	\$0.06	\$0.24	16(14)	14(12)	131	\$4.5	164	\$3.6	187
Case25	\$0.06	\$0.30	17(15)	13(13)	134	\$5.4	155	\$4.2	179

にした。また、例えば Case A の場合、先に易タスクが 15 個選択されてしまえば、難タスクのみを選択肢として残すべきである。しかし、この制御は煩雑となるため、個々の設定で 30 回タスクを割り当て、その結果を外挿することでタスク完了時間を得ることにする。評価実験における報酬額の組合せを図 2 に示す。今回は易タスクの報酬額が \$0.06 の場合も含めている。

4.2 評価実験の結果

図 2 から、Case A の場合に最適となる、つまり制限時間内にタスクが終了し、支払額が最小となるのは、易タスク \$0.02、難タスク \$0.04 の場合であることがわかる。このとき、所要時間は 93 分で、総費用は \$0.9 である。一方、提案法で得られた易タスク \$0.04、難タスク \$0.08 という設定では、所要時間 139 分、総費用 \$1.8 である。

また、Case B の場合に最適となるのは、易タスク \$0.02、難タスク \$0.02 の場合である。このとき、所要時間は 238 分で、総費用は \$0.6 である。一方、提案法で得られた易タスク \$0.02、難タスク \$0.04 という設定では、所要時間 116 分、総費用 \$0.8 となっている。

実験結果を見ると、易タスクと難タスクの報酬が双方 \$0.02 の場合にはやや時間がかかる傾向が見られるが、それ以外の場合は報酬額を増やせば、単調に所要時間が短くなるとは言えない。そのため、提案法は大きく誤った予測をしているとも言えないが、非常に精度良く予測できたとも言えない。

例えば、今回はタスクを公開する時間などは制御しなかった。よって、タスク公開時間などをパラメータに含めることで予測精度を向上させることができるかもしれない。また、ペングインの名称記述タスクは、事前の予想以上に選択されることが多く、また、正解率が高いワーカも多かった。よって、より専門知識を持ったワーカが見つけないような課題に対しては、提案法がより有効に機能する余地があるかもしれない。これらを調べることは今後の課題である。

5. むすび

本稿では、クラウドソーシングにおいて依頼者が複数種類のタスクを持ち、タスク全体の完了に締切時刻を有する場合に、その制約を満たしつつ費用を最小にする、ワーカのインセンティブを考慮した報酬設定法を定式化し、その適用に向けた初期実験の結果を示した。提案法自体の有効性という点では更なる検討が必要であるが、報酬額が同額でも難タスクを選択するワーカがかなりの数存在するといった興味深い現象が観察された。

最後に、経済学モデルとクラウドソーシングの関係について述べておく。本稿では、経済学モデルに基づいて報酬額を算出する方法を検討した。このようなモデルを持たず、単純にいくつかの報酬額の組合せで試験をして、そこから良い設定を選択することも考えられる。ただし、タスクの種類が増えて、非専門家、準専門家、専門家にタスクを割り振るといったことを考えると、試験の回数が増えてしまう。これに対して、モデルを持っておけば、ワーカの振る舞いをよく理解でき、また、試験をするとしても、その対象を絞ることができると考えている。

謝辞

本研究は、日本学術振興会科学研究費基盤研究 (S) (24220002, 平成 24 年度 ~ 28 年度) の補助を受けた。

参考文献

- [Ho 12] Ho, C.-J. and Vaughan, J. W.: Online Task Assignment in Crowdsourcing Markets, in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI'12)* (2012)
- [Ipeirotis 10] Ipeirotis, P. G., Provost, F., and Wang, J.: Quality management on Amazon Mechanical Turk, in *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP'10)*, pp. 64-67 (2010)
- [Rogstadius 11] Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., and Vukovic, M.: An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets, in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM'11)* (2011)
- [小山 14] 小山 聡, 鹿島 久嗣, 櫻井 祐子, 松原 繁夫: 特集「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」にあたって, *人工知能学会誌*, Vol. 29, No. 1, pp. 2-3 (2014)