

## クラウドソーシングにおける協調作業の実践と評価

## Lessons Learned from Collaborative Crowdsourcing

大谷直樹 \*1  
Naoki Otani中山愛 \*2  
Megumi Nakayama飯尾英晃 \*3  
Hideaki Iio伊藤聡比古 \*1  
Akihiko Ito林冬恵 \*1  
Donghui Lin村上陽平 \*4  
Yohei Murakami田中優子 \*5  
Yuko Tanaka\*1 京都大学情報学研究科  
Graduate School of Informatics, Kyoto University\*2 神戸大学農学部  
Faculty of Agriculture, Kobe University\*3 京都大学経営管理大学院  
Graduate School of Management, Kyoto University\*4 京都大学学際融合教育研究推進センター  
Center for the Promotion of Interdisciplinary Education and Research, Kyoto University\*5 株式会社クラウドワークス  
Crowdworks Inc.

Companies and free-lance workers who utilize crowdsourcing can enjoy a flexible working style which has very few locational or time constraints. However, crowdsourcing still has its shortcomings. One of these shortcomings is the lack of face-to-face communication between workers. Sometimes, it can be a challenge to find ways for workers to build relationships and work together properly when they have never actually physically met each other. In this paper, we explore the pros and cons of collaborative crowdsourcing by using past examples of the process being put to use. Experimental results on document simplification show that average quality increases when tasks are completed through a collaborative process, as opposed to an individual effort.

## 1. はじめに

これまで専門家に依頼されていた仕事の一部が、クラウドソーシングで行われるようになってきた。例えば、デザインやソフトウェア開発、翻訳などの業務や、研究開発のためのデータ作成などが行われている。一方で専門的なスキルを持たないワーカーを対象とした仕事も数多く発注されており、特別な機材を必要とせず参入障壁の低いライティング(文書作成)は、そうしたタスクにおける主要なもののひとつである。中でも文書の書き換えはゼロから書き始めることに比べて作業の難度が低く、多数のタスクがクラウドソーシングで発注されている。

行政機関などでは、一般市民にとって読みやすく、かつ必要な情報が正確に伝わるように行政文書の構成や表現を平易化する作業が日常的に行われている。これらは書き換え作業としてクラウドソーシングを適用できる潜在的なニーズを秘めている。近年のクラウドソーシング研究により有用性を示す学術的な知見は蓄積されているが、それを実際に運用するまでにはギャップがある。例えば、文を易しく言い換える目的にクラウドソーシングが有効に用いられることは示されている [Kajiwara 15] が、元の文の情報を過不足なく伝えることも実用上では重要である。このような運用を見据えた実践的な評価の不足は、非専門家のワーカーによるクラウドソーシングを業務に活用する際の課題となっている。

そこで我々は、協調作業によって行政文書を平易化する作業をクラウドソーシングで実践し、それを実際の運用を見据えて評価した。評価では文の読みやすさだけでなく、文章読解問題をワーカーに解かせることにより文書が伝える情報の正確さ

連絡先: 大谷直樹, 京都大学情報学研究科, 606-8501 京都市左京区吉田本町 36-1, otani.naoki.65v@st.kyoto-u.ac.jp

を調べた。協調作業の枠組みとしては、標準的なワークフローである分割-実行-統合型の設計を用いた。本稿では、協調作業の平易化作業結果を、原文や単独のワーカーによる作業結果と比較することで、その特徴や課題点を示す。

## 2. 協調作業による文の平易化ワークフロー

この節では、文の平易化に用いるワークフロー(図1)を説明する。このワークフローは標準的なフローのひとつである分割-実行-統合アプローチ [Kittur 11] に基づく。

## 2.1 原文の分割

複数のワーカーが並列的に平易化作業が行えるように原文を分割する。分割数がワーカーごとに大きく異なるのを防ぐために、予め文字数から計算した目安を提示する。また、分割された各部分に対して平易化作業を行うことを説明文に明記し、(1) 各部分に含まれる文字数がほぼ等しく、かつ (2) 意味のまとまりが保たれるように分割するよう指示を与える。複数の

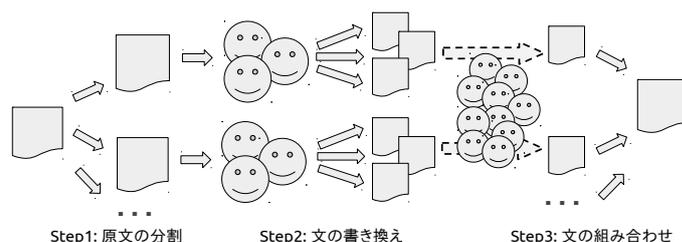


図1: 分割-実行-統合アプローチに基づく文の平易化

前後の文

「**黒**」をクリックすると前後の文が表示されます。**赤**色はあなたが修正する文、**青**色は他の人によって修正された文です。

### 提出フォーム

あなたのクラウドワークスユーザー名 (メールアドレスではありません)

修正前の文

国民権に属しない番号を付与し、それぞれの個人情報をこれに照合させることで国民全体の個人情報管理の効率化を図ろうとするものである。氏名、登録出生地、住所、性別、生年月日を中心とした国民の管理対象となる個人情報は、社会保険制度、納税、各種免許、犯罪歴、金融口座、経歴関係などがあられる。多くの情報を本制度によって管理すればそれだけ行政コストが下がりが、国民にとっても自分の情報を確認や訂正がやりやすくなる。一方、国民の基本的な権利が守られず、行政機関による過剰な監視、官公庁関係者(非正規雇用職員)の恣意や、不正に情報を入手し、第三者による情報漏洩の可能性があること、公平の原則のもと国民の資産を把握し上げられた政府債務の解消のために現金封鎖を容易にすることも懸念する意見もある。

修正後の文

### 他の人の結果

他の人が修正した結果が「あられ」下に表示されます。この文をもとに修正文を持っていないと編集できません(その場合はフォームの「修正済みの文を利用」にチェックを入れてください)。

1. この制度の目的は、国民一人一人に異なる番号を割り当て、それぞれの個人情報をこの番号に関連づけることで、個人情報管理を効率化しようとするところにある。基本的な情報は、氏名、出生地、住所、性別、生年月日である。これに加えて、社会保険制度、納税、各種免許、犯罪歴、金融口座、経歴関係などの個人情報を管理する必要がある。この制度の主なメリットは、行政のコストが削減できると国民の側が自分の情報を確認したり訂正することが容易になる点である。一方で、デメリットとしてさまざまな懸念が指摘されている。過剰な個人情報を管理されることによる監視、行政機関による不正な監視や管理者が情報を盗み出すかもしれないという危険性がある。外部に情報が流出して第三者によって犯罪などに利用されるおそれもある。また、将来の国民の基本的な権利が守られず、官公庁関係者(非正規雇用職員)の恣意や、不正に情報を入手し、第三者による情報漏洩の可能性があること、公平の原則のもと国民の資産を把握し上げられた政府債務の解消のために現金封鎖を容易にするのではないかという声もある。

2. この制度は、国民一人一人に対して異なる番号を付与し、国民それぞれの個人情報を結びつけることにより、国民の個人情報を容易に管理できるようにすることが目的である。氏名、出生地、住所、性別、生年月日など国民の基本的な個人情報を管理する。また、社会保険制度、納税、各種免許、犯罪歴、金融口座、経歴関係などの個人情報も対象となる。多くの情報を効率的に管理できるため、行政サービスの管理コストが削減される。国民の基本的な権利が守られず、官公庁関係者(非正規雇用職員)の恣意や、不正に情報を入手し、第三者による情報漏洩の可能性があること、公平の原則のもと国民の資産を把握し上げられた政府債務の解消のために現金封鎖を容易にするのではないかという声もある。

表 1: 協調作業の実践に使用した費用

対象文書	文字数	分割数	分割	平易化	統合	合計
A, B	約 3,300	9	500	3,000	600	4,100
C	約 2,300	2	500	1,800	600	2,900
D	約 2,200					

(単位: 円)

文書の分割作業においてワーカーに指示する分割数の目安は、400 字を 1 分割の基準にして計算した。その結果、マイナンバー、健康保険は 10 個、酒税法と雇用保険法は 6 個に分割された。

分割作業は各文書 10 人、平易化作業は各文書各パート 3 人、統合作業は各文書 5 人に依頼した。費用の合計を表 1 に示す。

本実験では、得られた組み合わせの中から平均ハミング距離が最小である 2 候補 (雇用保険法のみ同じ値が存在したため 3 候補) を評価に用いた。

### 3.2 比較対象

協調作業と比較するために、文書を分割せずに原文全てを単一のワーカーに平易化させた (以下、単独作業)。この作業を各文書について 3 人のワーカーに依頼した。報酬額は、協調作業の場合と単一のワーカーに依頼した場合とで各文書の費用が同額になるように設定した。具体的には、各文書において、協調作業で平易化する場合にかかる費用の金額を、3 人のワーカーに均等に配分した。例えば、文書 A なら各ワーカーに支払う金額は  $4,100/3 \approx 1,400$  円と設定した。

### 3.3 評価方法

読みやすさと正確さの 2 軸で評価した。4 つの題材それぞれについて 6 から 7 つの文書を検証対象とした。その内訳は、原文 (1 つ)、協調作業により作成したもの (2 または 3 つ) と単独作業により作成したもの (3 つ) である。

読みやすさに関しては、検証対象の文書を読みやすい順に並べる作業を 10 人のワーカーに依頼し、各文書が何位に何回ランクされたかを調べた。

正確さは文章読解問題の枠組みで評価する。平易化された文書でワーカーに文章読解問題を解かせ、正答率を調べる。正答率が高い文書は原文の情報を読者に正確に伝えていると考えられる。

問題作成のために、まずワーカーに各文書の内容を問う問題を作成してもらい、著者らがその中から妥当と思われるものを各文書 5 題選択した。このとき、問題の解釈に曖昧性があるもの、用語を抜き出すような平易化前後で解答の難度が大きく異なるような問題は避けるようにした。またワーカーの得意不得意の差が現れないように、文を答える問題、語句を答える問題など、複数の解答形式の問題が含まれるようにした。次に、作成した文章読解問題を解くタスクを複数のワーカーに対してクラウドソーシングで実施し、得点を比較した。なお、このように文の正確さを文章読解問題で測る方法は、機械翻訳の評価方法として用いられたことがある [Callison-Burch 09]。

### 3.4 結果: 読みやすさ

読みやすさを評価した結果を表 2 に示す。表は最右列に示す上位にランクされた回数の合計が多かったものから順に並べられている。全体的な傾向として、単独作業による平易化結果は上位に位置づけられる回数が多く、原文は少なかったことがわかる。協調作業では、表 3 に示す例のように、用語や表現の言い換えが頻繁に行われた。また、複数のワーカーの書

図 2: 作業用インターフェース。左中央に表示された原文の一部分を読み、その下のフォームに書き換えた結果を入力する。右に表示されているのは、他のワーカーが同じ文を書き換えた結果である。

ワーカーに分割場所を提示させ、過半数のワーカーが区切った場所から原文を分割する。

## 2.2 文の書き換え

作業用のインターフェース上 (図 2) で、分割された各部分に対して、複数のワーカーが原文を平易化した文を入力する。このとき、他のワーカーが平易化した文の結果すべてを画面に示し、必要に応じて文を複製したうえで修正を加えてもよいことにする。他人の作業結果をもとに作業を行った場合、平均的に作業品質が向上することが知られている [Little 10]。

## 2.3 文の組み合わせ

複数の平易化された文を表示し、分割された各部分に対して最も適する文を候補の中からワーカーが選ぶ。文を組み合わせた系列の候補は複数のワーカーから収集し、その中から他の系列候補との平均ハミング距離が最小の系列を選ぶ。例えば、次のうち平均ハミング距離が最小なのは 2 番目の系列である。

- (文<sub>11</sub>, 文<sub>22</sub>, 文<sub>32</sub>)  $\cdots (2 + 3)/2 = 2.5$
- (文<sub>13</sub>, 文<sub>22</sub>, 文<sub>31</sub>)  $\cdots (2 + 2)/2 = 2.0$
- (文<sub>13</sub>, 文<sub>23</sub>, 文<sub>33</sub>)  $\cdots (3 + 2)/2 = 2.5$

このような系列データに対する統計的な手法が近年研究されており (例えば [Wu 12])、データ数が多いときには有効だと考えられる。

## 3. 実験

### 3.1 平易化する文書

日常で読む必要が生じうる行政文書を想定し、以下の 4 文書を選択した。

- A マイナンバー \*1
- B 健康保険 \*2
- C 酒税法 \*3
- D 雇用保険法 \*4

\*1 Wikipedia, 国民総背番号制,

<https://ja.wikipedia.org/wiki/国民総背番号制>

\*2 Wikipedia, 健康保険,

<https://ja.wikipedia.org/wiki/健康保険>

\*3 国税庁, 酒税法及び酒類行政関係法令等解釈通達, <https://www.nta.go.jp/shiraberu/zeiho-kaisaku/tsutatsu/kihon/sake/2-01.htm#a-04>

\*4 厚生労働省, 雇用保険法等の一部を改正する法律案要綱, <http://www.mhlw.go.jp/topics/bukyoku/soumu/houritu/dl/190-02.pdf>

表 2: 読みやすさの評価結果. 10 人のワーカーがランキングを作成した. 表中の数字は左に示す文書が上に示す文書よりも高いランクを得た回数を示す (最大 10). 文書はその回数の合計 (最右列) の降順に並べられている.

(a) マイナンバー								(b) 健康保険							
>	単独 3	単独 2	単独 1	協業 2	協業 1	原文	計	>	単独 3	協業 2	協業 1	原文	単独 1	単独 2	計
単独 3	-	9	8	8	9	10	44	単独 3	-	8	7	8	7	8	38
単独 2	1	-	7	7	6	8	29	協業 2	2	-	6	5	7	7	27
単独 1	2	3	-	7	8	9	29	協業 1	3	4	-	5	6	7	25
協業 2	2	3	3	-	7	8	23	原文	2	5	5	-	5	6	23
協業 1	1	4	2	3	-	9	19	単独 1	3	3	4	5	-	5	20
原文	0	2	1	2	1	-	6	単独 2	2	3	3	4	5	-	17

(c) 酒税法								(d) 雇用保険法等								
>	単独 3	単独 1	単独 2	原文	協業 1	協業 2	計	>	協業 3	単独 3	単独 2	単独 1	協業 2	協業 1	原文	計
単独 3	-	7	8	8	10	10	43	協業 3	-	6	7	9	9	9	10	50
単独 1	3	-	5	9	9	10	36	単独 3	4	-	5	8	7	8	9	41
単独 2	2	5	-	9	10	10	36	単独 2	3	5	-	8	7	8	8	39
原文	2	1	1	-	6	9	19	単独 1	1	2	2	-	7	8	8	28
協業 1	0	1	0	4	-	9	14	協業 2	1	3	3	3	-	4	6	20
協業 2	0	0	0	1	1	-	2	協業 1	1	2	2	2	6	-	6	19
								原文	0	1	2	2	4	4	-	13

表 3: 健康保険の平易化の例

原文	健康保険制度は、労働者又はその被扶養者の業務災害以外の疾病、負傷若しくは死亡又は出産に関して保険給付を行い、もって国民の生活の安定と福祉の向上に寄与することを目的とする。
協業 1,2	健康保険制度には、働く人やその家族に、病気や怪我、死亡や出産のときに保険金を支払うことにより、人々の生活をより良いものにするという目的がある

表 4: マイナンバーの平易化の例

原文	【概要】 国民総背番号制（こくみんそうせばんごうせい、英: national identification number）は、全ての国民に固有の番号を振り、特定個人を識別し管理しやすくする制度で、共通番号制度ともいう。… 多くの情報を本制度によって管理すれば… メリットがある。一方、国民の基本的な権利が制限されたり、… 懸念する意見がある。
単独?	【マイナンバーとは】 マイナンバーとは、国民一人に対してその人だけの番号をつけて、その番号から個人の情報を知ることができるしくみで、国民総背番号制（こくみんそうせばんごうせい）、英語では「national identification number」という。… 【マイナンバーのメリットとは】 国民一人一人に番号を付けるマイナンバーのメリットは、… 【マイナンバーのデメリット】 国民の情報を管理しやすくなる一方で、… という懸念の声もある。

き換え結果を組み合わせているために、一文書中で文調が変わったり、用語の一貫性が損なわれたりする傾向があった（例えば、「給付」に対して、「給付（言い換え無し）」「支払い」「支払」など）。

一方単独作業は、細部の言い換えは少ないが、文章の構造を変えて読みやすさの向上を図る傾向があった。この傾向は特に法律の条項の解釈（酒税法と雇用保険法）のようにある程度構造化されている文書よりも、構造化されていない文書（マイナンバーと健康保険）で見られた。表 4 にマイナンバーの例を示す。もともと「概説」というひとつの節で書かれていた内容が、単独作業の結果では「マイナンバーとは」「マイナンバーのメリット」のような独自の節が作られて構造化されている。

### 3.5 結果: 正確さ

正確さを検証した結果を図 3 に示す。縦軸はテストの得点（最大 5）を示し、赤の実線は 5 人のワーカーの得点の中央値である。中央値に着目すると、原文を用いた場合の得点がどの文書に関しても高く、協調作業による平易化結果を用いた場合は「健康保険」以外のすべての文書で原文と同等程度の得点が得られていることがわかる。単独作業による平易化結果でも原文と同等程度の得点が得られた場合が多かったが、協調作業と異なり得点が極端に低いものがいくつか存在した。

協業作業での「健康保険」の得点が低下した理由は、言い換えによって解答に必要な語が文から失われてしまったからである。表 5 の例では、原文にあった「福祉の向上」が消えてしまっているために、このキーワードに関する問題で正答が得られなかった。ただし、その言い換え自体は妥当であると考えられ、今回用いられた読解問題の是非について検討が必要である。

協業作業では、単独作業と比べて、作業したワーカーによらずに、成果物の質が安定していた。単独作業では、前節で述べたように単独のワーカーによって構成が大きく変更されたり内容が省略されることがあり、それによって平均点が低くなる事例があった。これは複数人で反復的に作業を行った結果が安定していたという先行研究の報告と一致する [Little 10]。

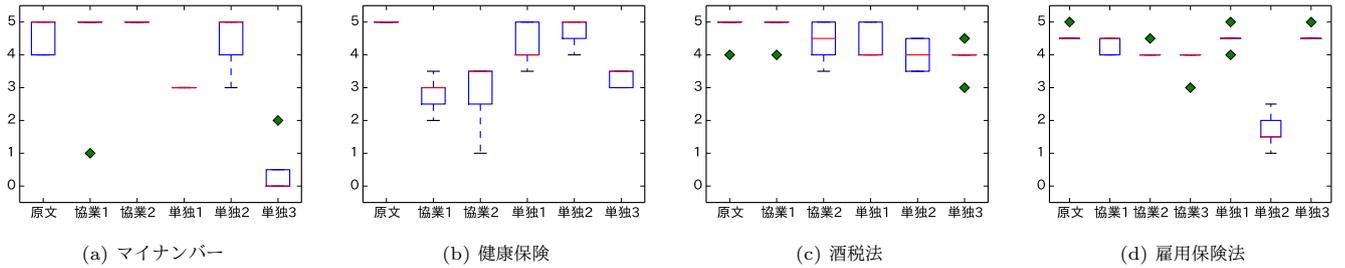


図 3: 正確さの評価結果. 縦軸はテストの得点 (最大 5) を示し, 赤の実線は 5 人の得点の中央値である. 左から, 原文 (1 つ), 協調作業により作成したもの (2 または 3 つ) と単独作業により作成したもの (3 つ) の結果を表す. 青色の箱は第 1・3 四分位点, 点線で結ばれた水平線は最大値または最小値, 緑色の点は外れ値を示す.

表 5: 健康保険の平易化の例

原文	健康保険制度は, … 保険給付を行い, もって国民の生活の安定と福祉の向上に寄与することを目的とする.
協業 1	健康保険制度には, … 保険金を支払うことにより, <u>人々の生活をより良いものにする</u> という目的が…
協業 2	健康保険制度は, … 保険金を支払い, <u>人々の生活をよりよくすることを目的として</u> …

また報酬だけを目的とし求める作業を行わないスパムワーカーは, 今回は検証前に除いたが, 特に単独作業の場合に多く存在した. 今回の報酬設計では, 作業単価 (1,400 円) が平均的なライティング作業よりも高かったために, スпамワーカーを集めやすかったと考えられる.

#### 4. 関連研究

Little らは, 我々がフローに取り入れた反復的に作業結果を改良するものと並列する複数の結果からひとつを選ぶワークフローの比較を行い, 反復型による作業結果は平均的に品質が良く, 並列型による作業結果は品質のばらつきが大きい一方, 最良の結果を生み出す傾向があることを示した [Little 10].

本稿で用いた分割-実行-統合のワークフローは Kittur らによって提案されたもので [Kittur 11], 近年も改良が続けられている. Amazon Mechanical Turk の API を用いる実装がいくつか公開されているが, 我々が用いたプラットフォームには適用できなかったため, 独自に実装した.

文の平易化は自然言語処理の重要なタスクのひとつであり, 自動的に文書を読みやすく言い換える方法が研究されてきた. 梶原らは平易化システムを評価するためのデータセットをクラウドソーシングで収集した [Kajiwara 15]. 彼らは独立した短文を対象としたのに対し我々は大規模文書を対象としている. また, 我々は得られた平易化結果が伝える情報の正確さを文章読解問題の枠組みで評価した.

#### 5. おわりに

我々は研究で蓄積された知見と実際の運用の間にあるギャップを解消するために, 文の平易化を協調型のワークフローによって実践し, 平易化された文書を読みやすさと正確さの観点

で評価した. 実験結果から, 協調作業による文書は語の言い換えによって読みやすさが増すが単独のワーカーによる平易化には劣る傾向があること, ただし正確さの観点では協調作業のほうが品質が安定して良い傾向があることがわかった.

分割-実行-統合型のワークフローでは, ワーカーが全体的な文の構造に修正を加えることができないため, 元の文書の構造が平易化後の読みやすさに影響する. 元の文書に関係なく文を再構成するような平易化を行いたい場合には, ワークフローに修正が必要である. また, 協調作業では語の言い換えが起こりやすい傾向があるため, 重要な固有名詞も別の言葉に置き換えられてしまう可能性がある. それを防ぐためには, 予めワーカーに原形を残したい語を指示する必要がある.

また, 作業コストは実際の運用のためには重要な要素であるが, 本稿では複数のワーカーによる作業と単独のワーカーによる作業で合計の費用を同一に設定し, 議論を行わなかった. コストの比較や最適な報酬の設定などの検討は今後の課題である.

#### 謝辞

本稿は株式会社クラウドワークスの協力を得て京都大学デザインスクールで実施されたデザイン学共通実習科目「問題発見型/解決型学習 (FBL/PBL)」の成果をまとめたものである.

#### 参考文献

[Callison-Burch 09] Callison-Burch, C.: Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon’s Mechanical Turk, in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 286–295 (2009)

[Kajiwara 15] Kajiwara, T. and Yamamoto, K.: Evaluation Dataset and System for Japanese Lexical Simplification, in *Proceedings of the ACL-IJCNLP 2015 Student Research Workshop*, pp. 35–40 (2015)

[Kittur 11] Kittur, A., Smus, B., Khamkar, S., and Kraut, R. E.: CrowdForge: Crowdsourcing Complex Work, in *Proceedings of the 24th annual ACM symposium on User Interface Software and Technology (UIST)*, pp. 43–52 (2011)

[Little 10] Little, G., Chilton, L. B., Goldman, M., and Miller, R. C.: Exploring iterative and parallel human computation processes, in *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP)*, pp. 68–76 (2010)

[Wu 12] Wu, X., Fan, W., and Yu, Y.: Ensembling crowd sequential labeling for improved quality, in *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1713–1719 (2012)