

# 説得力のある文書執筆のための クラウドソーシングを用いた論述改善フィードバック

庵谷 拓輝<sup>†</sup> 山本 祐輔<sup>†</sup>

<sup>†</sup> 静岡大学大学院総合科学技術研究科 〒432-8011 静岡県浜松市中区城北 3-5-1

E-mail: <sup>†</sup>ihoriya@design.inf.shizuoka.ac.jp, <sup>††</sup>yamamoto@inf.shizuoka.ac.jp

**あらまし** 本稿では、多数のクラウドワーカーを活用し、論述の説得力を向上させるためのフィードバックを生成する手法を提案する。本手法の目的は、論述文書に対するフィードバックという一般的に時間のかかるタスクが、教員などの専門家のみに委ねられていることによる負荷の軽減である。提案手法は、論述改善タスクを (1) 意見と事実の切り分け、(2) 事実の根拠の有無の確認、(3) 意見の根拠の有無の確認の 3 種類のマイクロタスクに分割する。これらマイクロタスクをクラウドワーカーに割り振り、非同期的に実施することで、論述改善のための効率的なフィードバック生成を実現する。150 名のクラウドワーカーにマイクロタスクを実施してもらい、4 名の執筆者に提案手法を評価してもらうユーザ実験を行った。システムにてフィードバックを閲覧した執筆者からは、説得力を高める意欲の向上・次にやるべきことの理解度に関する測定結果にて、有望な評価が得られた。システムを用いることで、教員が単独でフィードバックを行う場合に比べて、クラウドワーカーは低い負荷でフィードバックを生成できる可能性を示した。

**キーワード** 文章執筆、クラウドソーシング

## 1 はじめに

説得力のある論述を行うスキルは、学校教育において獲得が望まれる重要なスキルの 1 つである [1]。レポートや小論文などの論述文書を執筆するとき、一般的にはその分野の専門家や教員などの読者が存在する。説得力のある文書を書くためには、主張に対して妥当な根拠を提示し、想定される反論の余地を減らすことが重要である。本稿では、説得力のある論述文書を「主張に対して妥当な根拠が与えられており、客観性が意識された論述文書」と定義する。

文章執筆に慣れていない者にとって、説得力のある論述を行うことは容易ではない。第 1 の理由には、マイサイドバイアスの影響によって、偏った主張を行ってしまいやすいことが挙げられる。マイサイドバイアスとは、人間には自身の意見を重視する傾向があることを説明するバイアスである [2], [?]. 主張の根拠として自身に都合の良いデータや文献ばかりを集める場合、妥当性を欠いた論述となる可能性が高い。

第 2 の理由には、説得力のある論述を行うスキルを修得・改善する機会が乏しいことが挙げられる。一般に、スキル獲得・改善には学習者に対するフィードバックが重要とされている [3]。学習が難しい認知的なスキルである論述スキルの向上には、文書執筆に対するフィードバックは特に有用と考えられる。Jonassen らは、論述スキル向上に向けた課題の 1 つとして、論述スキル開発のための授業設計や教育方法論に関する教師のスキルを挙げている [4]。Jonassen らによると、教師はこの種のスキルを十分に有していないことが多い。そのため、学生は論述スキルを習得するために必要となるフィードバックを得ることが難しい。また、仮に教師側が十分な教育スキルを持ってい

たとしても、学生ごとにフィードバックを行うには膨大な時間と労力がかかる。さらに、大学などにおいての大規模・遠隔講義が一般的となった現在、学生は教員から論述スキル獲得のためのフィードバックを受ける機会を確保することが難しくなっている [5]。そのため、執筆者自らが自分の書いた文章を振り返り、より論述の説得力を向上させられる仕組みや環境が必要となる。

本稿では、より説得力のある論述を可能にするために、クラウドソーシングを用いた論述改善フィードバック手法を提案する。図 1 に、提案手法の概略図を示す。本稿では、論述を**何らかの主張を妥当な根拠に基づいて行うこと**と定義する。提案手法は、執筆中の文書において根拠が乏しい文を指摘し、執筆者に論述改善のためのフィードバックを与えることを目的とする。執筆者の文書が与えられると、提案手法はクラウドワーカーが低コストで論述改善フィードバックを行えるようなマイクロタスクを生成する。提案手法によって生成されるマイクロタスクは、以下の 3 種類である。

- (1) 意見と事実の切り分け
- (2) 事実の根拠の確認
- (3) 意見の根拠の確認

上記 3 種類のマイクロタスクは、執筆中の文書内の 1 文ごとに実行される。提案手法は、文ごとに実行された 3 種類のマイクロタスクの結果を集約し、図 6 に示すようなフィードバックを執筆者に提示することで、論述改善のきっかけを与える。マイクロタスク化により、クラウドワーカーは割り当てられたマイクロタスクに非同期的に取り組むことができ、効率的なフィードバック生成が可能となる。提案手法によって、教員が単独でフィードバックを行う場合に比べて、フィードバックを行う者にかかる時間的・労力的負荷の分散が期待できる。また、執筆

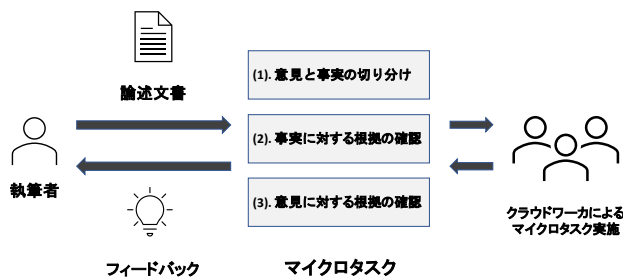


図 1 提案手法の概略図

者は、短時間かつ低コストで論述改善のフィードバックを得られるようになると期待される。

## 2 関連研究

### 2.1 文章執筆支援システム

Arnold らは、テキスト入力が文章内容に与える影響についてタッチスクリーンキーボードのフレーズ指向の入力補助システムを提案している [6]。また Arnold らは、テキストの予測入力システムが人々の執筆内容にどのような影響を及ぼすかを分析している [7]。ユーザ実験の結果、予測入力システムの利用は文章の内容をより簡潔する一方、執筆内容そのものがシステムが提示したテキストに影響されてしまう可能性があることが確認された。Wambsganss らは、文章の論証構造に対するフィードバックを提供する適応的なシステムを開発している [8]。また Wambsganss らは、論理的な文章の執筆を支援するためのボット形式のチュータリングシステムを提案している [9]。Chen らは、Gmail において入力可能性の高いフレーズの続きを表示することで、ユーザのメール文章の執筆支援を行う Gmail Smart Compose を開発している [10]。

本稿では、クラウドワーカーに実施コストの低いタスクを割り当て、論述文書へのフィードバックを生成するシステムを提案する。

### 2.2 クラウドソーシングによるコンテンツ作成・改善支援

クラウドソーシングを活用した文章作成、改善支援の研究も行われている。Bernstein らは、作業品質を向上するために考案された Find-Fix-Verify パターンに基づいて、クラウドワーカーが文章の校閲を行う文章執筆システム SOYLENT を提案した [11]。Murukannaiah らは、単独のクラウドワーカーと複数人のクラウドワーカー（チームワーカー）のアイデア発想タスクにおけるパフォーマンスを比較分析した [12]。分析の結果、ワーカーが単独で作業するよりもチームで作業した方がより創造的なアイデアを生成しやすいことが示唆された。Salehi らは、クラウドワーカーと依頼者がコミュニケーションを取りながら、効果的に執筆タスクを行うための方略について調査した [13]。調査の結果、執筆フェーズによってワーカーに有用な方略が異なり、依頼者にとってはコメントと編集、Q&A にコミュニケーション

コストがかかることが明らかになった。

本稿では、論述改善フィードバックを (1) 意見と事実の切り分け、(2) 事実の根拠の有無の確認、(3) 意見の根拠の有無の確認、の 3 つのマイクロタスクに分割したクラウドソーシングを行うことで、フィードバックを行うワーカーの時間的、労力的負荷を低くしつつ、ある程度の質のフィードバックを行う手法を提案する。

## 3 クラウドソーシングによる論述改善フィードバック

本稿では、クラウドソーシングを用いて、論述改善のためのフィードバックを効率的に生成する手法を提案する。AI ではなくクラウドソーシングを用いて人間に文書を読んでもらうことにより、他者を意識した客観視機会を創出を狙う。本稿において、論述改善のマイクロタスクとは、文書内の 1 文に対して実行されるフィードバック作業のことを指す。本稿で設計するマイクロタスクは、(1) 意見と事実の切り分け、(2) 事実に対する根拠の有無の確認、(3) 意見に対する根拠の有無の確認、の 3 種類とする。また、マイクロタスクリストとは、単一のパラグラフに対して実行されるマイクロタスクの集合と定義する。

以下、提案手法の詳細について述べる。

### 3.1 論述改善のワークフロー

まず執筆者は、与えられた論述課題に関する文書ドラフトを作成する。次に、執筆者はドラフトを提案システムに提出する。そのタイミングで、システムはドラフトを改善するためのマイクロタスクを生成する。マイクロタスク生成の目的は、複雑な思考が求められる論述改善タスクを分割し、フィードバックを実施するクラウドワーカーの負担を減らすことにある。3.3 節に述べるロジックに従い、提案システムはクラウドワーカーにマイクロタスクリストを割り当て、複数のマイクロタスクを実施してもらう。ドラフトに対する論述改善タスクは、各文に対するマイクロタスクが、規定完了数分実行され次第終了する。論述改善タスクの終了後、提案システムはフィードバック結果を集約し、図 6 に示すような論述改善のフィードバックを提示する。執筆者が生成されたフィードバックを閲覧することで、文書の改善箇所を特定して説得力向上のための執筆が可能になることが期待される。

### 3.2 マイクロタスクの設計

本稿では、以下の 3 種類のフィードバック作業をマイクロタスクとして設計する。

- (1) 意見と事実の切り分け
- (2) 事実に対する根拠の有無の確認
- (3) 意見に対する根拠の有無の確認

意見と事実の切り分けタスクでは、クラウドワーカーは図 2 に示す画面にて、ある文が意見であるか事実であるかの判断を行う。クラウドワーカーが本マイクロタスクを実施することで、文書の各文が、「意見」「事実」「どちらでもない」の 3 種のいずれかに区別される。システムが区別された種類に応じて文をハイ

評価タスクページ（意見と事実の切り分け）

ある学生が【「生産性」を定義し、志望企業の生産性向上に対して、AIよりもあなたが貢献できる点をアピールせよ】というレポート課題に取り組み、レポートを執筆しました。次の文は、学生のレポートから切り出された文です。

「**昨今、AI技術が人間の雇用を奪う可能性について議論されている。**」

上の文は、意見と事実のどちらですか？  
選択肢の中から1つ選んで回答してください。

☐ 意見

☒ 事実

☐ どちらでもない

回答して次のタスクへ

図 2 意見と事実の切り分けタスクの例

評価タスクページ（意見に対する根拠の有無の確認）

「**昨今、AI技術が人間の雇用を奪う可能性について議論されている。**」

ソフトウェア開発の現場には、スケジュールや人材の特性など多くの不確実性が存在する。AIによる、学習データに基づいた推論では、緊急度の高い場面において、生産性が極端に低いケースが考えられる。したがって、場面に応じた柔軟な対応が可能である人間を採用する場合のほうが、組織にとって、長期的な生産性は高いであろう。

上のオレンジ色でハイライトされた文について、妥当な根拠が周囲に書かれていますか？  
選択肢の中から1つ選んで回答してください。

☐ 妥当な根拠が書かれている。

☐ 根拠は書かれているが、妥当とはいえない

☒ 根拠は書かれていない

上記の回答理由を簡単に述べてください。  
「妥当な根拠が書かれている」と回答した場合、「どの文章を読んでそう判断したのか」を簡単に述べてください。それ以外を回答した場合「どのように文章を改善すれば、根拠をもたせられるか」を簡単に述べてください。

図 4 意見に対する根拠の有無の確認の例

評価タスクページ（事実に対する根拠の有無の確認）

ある学生が【「生産性」を定義し、志望企業の生産性向上に対して、AIよりもあなたが貢献できる点をアピールせよ】というレポート課題に取り組み、レポートを執筆しました。次の文は、学生のレポートから切り出された文です。

「**昨今、AI技術が人間の雇用を奪う可能性について議論されている。**」

上の青色でハイライトされた文について、根拠として妥当な情報源が周囲に書かれていますか？選択肢の中から1つ選んで回答してください。

☐ 根拠として妥当な情報源が書かれている。

☐ 情報源は書かれているが、妥当とはいえない

☒ 情報源は書かれていない

回答して次のタスクへ

図 3 事実に対する根拠の有無の確認の例

ライトすることで、執筆者が文書の意見と事実の比率を客観的に観察できるようになることを狙う。

事実に対する根拠の有無の確認タスクでは、クラウドワーカは図 3 に示す画面にて、ある事実文に対して根拠となる情報源が与えられているかどうかの判断を行う。回答は選択式であり、「根拠として妥当な情報源が書かれている」「情報源は書かれているが妥当とはいえない」「情報源は書かれていない」の三択から回答してもらう。本マイクロタスクが実施された後、システムは妥当な根拠となる情報源が不足している事実文の存在を示す。

意見に対する根拠の有無の確認タスクでは、クラウドワーカは図 4 に示す画面にて、ある意見文に対して意見を裏付ける事実が書かれているかどうかの判断と、判断の理由もしくは改善案の記述を行う。根拠の有無の判断については選択式、判断理由・改善案については記述式で回答を行う。選択回答では「妥当な根拠が書かれている」「根拠は書かれているが妥当とはいえない」「根拠が書かれていない」の三択から回答してもらう。記述式回答では、前述の通り、判断の理由もしくは改善案を記述してもらう。具体的には、「妥当な根拠が書かれている」と回答した場合、その判断元となった文章を述べてもらい、「根拠は書かれているが妥当とはいえない」もしくは「根拠が書かれていない」と回答した場合、どのように文章を改善すれば、根拠を持たせられるか述べてもらう。本マイクロタスクが実施された後、システムは、妥当な根拠となる事実が書かれていない意見文の存在とその判断理由・改善案を執筆者に提示する。

### 3.3 マイクロタスクの割り当て

本稿では、以下の手順でクラウドワーカに複数のマイクロタスクを割り当てる。

- (1) ある文書を段落（パラグラフ）単位に分割する
- (2) 分割されたパラグラフ内の各文に対してマイクロタスクを作成する
- (3) 生成されたマイクロタスクの集合（マイクロタスクリスト）をクラウドワーカに割り当てる

論述改善タスクのフローを図 5 に示す。分割したパラグラフ内の各文に対してマイクロタスクを作成することで、1 人のクラウドワーカがフィードバックの対象とする文章量を、文書全体からパラグラフ・文単位に削減できる。クラウドワーカに対して、パラグラフに定義されたマイクロタスクリストを割り当てることで、そのパラグラフに焦点を当てたフィードバックの実施が可能となる。

マイクロタスクの実行には、順序の依存関係が存在する。意見と事実の切り分けタスクは、事実/意見に対する根拠の有無の確認タスクが実施される前に完了されていることが望ましい。なぜなら、事実/意見に対する根拠の有無の確認タスクの実施には、各文が意見であるか事実であるかそのどちらでもないかが確定されていることが前提となるからである。したがって、まずクラウドワーカには、意見と事実の切り分けタスクを優先的に割り当てる。

ある文書に対する意見と事実の切り分けタスクが規定完了数を満たした後に、事実/意見に対する根拠の有無の確認タスクの割り当てを開始する。各種マイクロタスクの規定完了数が満たされたとき、対象文書に対する論述改善タスクを終了する。

### 3.4 マイクロタスクの集約・提示

マイクロタスクの種類ごとに、集約方法と提示方法は異なる。各マイクロタスクの結果は、クラウドワーカによって異なる可能性がある。例えば、意見と事実の切り分けタスクを実施したクラウドワーカ A は、ある文章を意見と判断したが、クラウドワーカ B はその文章を事実として判断したというケースが考えられる。マイクロタスクの結果の違いを考慮するために、提案システムではマイクロタスク結果を多数決によって集約することで、各マイクロタスクの最終的な結果とする。

意見と事実の切り分けタスクでは、クラウドワーカが、ある

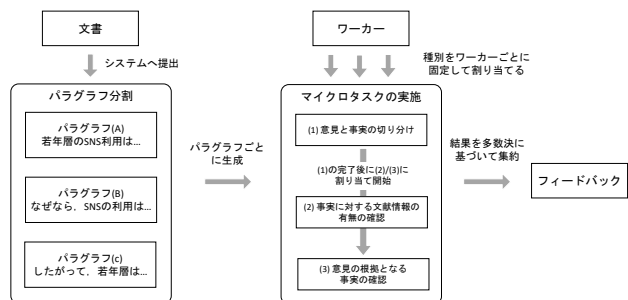


図 5 論述改善タスクのフロー

文を意見であるか事実であるか判断した結果が得られる。ここで得られた各クラウドワーカーのマイクロタスク結果に対して多数決を適用し、文が意見であるか、事実であるか、どちらでもないかの判断のうち、最も多かった判断結果を採用する。ある文が意見か事実か、どちらでもないかが一意に決められた後、この結果をフィードバック画面に反映する。具体的には、図6のように、意見と事実をそれぞれが区別できるようにハイライトし、フィードバックを執筆者に提示する。

事実/意見に対する根拠の有無の確認タスクでは、事実文/意見文に対して根拠が不足しているかどうかの判断結果が得られる。この判断結果に多数決を適用し、一定数のワーカが根拠が不足していると判断した文について、それを執筆者にフィードバックする。意見に対する根拠の有無の確認タスクにおいて、根拠が不足していると判断された文に対してはクラウドワーカがそのように判断した理由および改善案が添えられている。これらの理由については意味的な集約や要約は行わず、ワーカから提供された理由および改善案をすべて執筆者に提示する。

図6に示すように、事実/意見に対する根拠の有無の確認タスクのフィードバックは、ある意見文または事実文がクリックされたタイミングで、その文に関連する事実/意見に対する根拠の有無の判断結果、およびその判断理由の詳細を提示する形で行われる。フィードバックの詳細には、事実/意見に対する根拠の有無の確認タスクの判断を行ったクラウドワーカーの人数およびそれらの中で「根拠がある」と回答しなかった人数を示す。意見文が選択された場合には、「根拠を持たせるためにどのように改善すればよいか」を提示する。

## 4 ユーザ実験

提案手法の有効性を評価するために、ユーザ実験を実施した。本稿では、文書のフィードバックを受ける実験協力者を執筆者、論述改善タスク実施のためにクラウドソーシングサイトにて募集する実験協力者をクラウドワーカーと呼ぶ。以降で実験協力者と呼ぶ場合には、執筆者とクラウドワーカーの両方を含む。ユーザ実験では、執筆者に論述タスクを課し、クラウドワーカーに論述改善タスクを課す。



図 6 論述改善タスク実施後のフィードバックのイメージ例。  
左カラムは執筆された文書を表示する。オレンジ色でハイライトされた文章は意見、青色でハイライトされた文章は事実として切り分けられたことを示す。右カラムには、得られたフィードバックの判断と改善案を表示する。

#### 4.1 実験協力者

静岡大学情報学部 に所属する 4 名の学生を執筆者として採用した。また、クラウドソーシングサービスの CrowdWorks<sup>1</sup> を利用して、150 名のクラウドワーカーを募集した。各クラウドワーカーには、タスク開始前にデータ収集の方針を説明し、同意を得られた場合のみタスクに進んでもらう。タスクを適切に完了していないクラウドワーカーを分析対象から除外し、最終的に 134 名のデータを利用して分析を行った。134 名の内、78 名が男性、56 名が女性であった。最終学歴が大学・大学院であるクラウドワーカーは 85 名であった。タスクを完了した実験協力者には 100 円の報酬を支払った。

## 4.2 実験システム

実験協力者がタスクを行うためのシステム（以下、実験システム）を構築した。執筆者は実験システム上で、論述タスク内容の説明ページの閲覧、およびクラウドワークによって生成されたフィードバックの閲覧を行った。クラウドワークは実験システム上で、論述改善タスク内容の説明ページの閲覧、および論述改善タスクの実施を行った。

### 4.3 実験手順：執筆者

執筆者は、以下の手順で実験を実施した。

- (1) 論述文書の提出
- (2) フィードバックの閲覧
- (3) アンケートとインタビューへの回答

まず、執筆者には、「人工知能と生産性」についての論述課題の回答文書を提出してもらった。ここで提出された文書をフィードバック対象の文書とした。執筆者が取り組んだ課題の内容を以下に示す。

## 人工知能と生産性



あなたは現在就職活動中であるとする。あなたは、採用人数が1名の企業を志望しており、この度その企業から最終面接に呼ばれたとする。事前情報によると、最終面接に残った応募者のうち、あなた以外はすべてAI（人工知能）であることが分かった。志望企業から内定を勝ち取るためには、「AIよりもあなたを採用した方が企業にとってメリットがある」ことを採用担当にアピールする必要がある。「生産性」を定義し、志望企業の生産性向上に対して、AIよりもあなたが貢献できる点をA4用紙0.5～1ページ程度でアピールせよ。なお、「志望企業」は自由に設定してよい（志望企業名を出す必要はない）。

提出した文書に対するフィードバックタスクが完了された後、執筆者にはアンケートに回答してもらった。アンケートでは、フィードバックを閲覧するシステムの有用性についての質問に回答してもらった。具体的には、以下に関する質問に回答してもらった。

**Q1** 文書の説得力の向上意欲

**Q2** 文書の説得力を向上させるためにやるべきことが分かるようになったか

**Q3** 文書に説得力を持たせるために不足していることが見つかったか

**Q4** フィードバックの理解容易性

**Q5** フィードバックの妥当性

**Q6** 提案システムの利用意欲

事後インタビューでは、有用だと感じられたフィードバックの特性やシステムの問題点・改善点について尋ねた。

#### 4.4 実験手順：クラウドワーカ

クラウドワーカは、以下の手順で実験を実施した。

- (1) ユーザ登録
- (2) 論述改善タスクの実施
- (3) 事後アンケート

まず、クラウドワーカはCrowdWorksにて実験参加のための同意事項に同意した後、実験用ウェブサイトへ遷移した。次に、実験用ウェブサイトへ遷移した後、ユーザ登録を行い、具体的なタスクの説明を読んでもらった。タスクの説明を読み、タスクを開始した時、クラウドワーカにはパラグラフごとに生成されたマイクロタスクリストが3つ割り当てられた。なお、マイクロタスクリストはパラグラフに対応し、マイクロタスクは文に対応する。ある文に対するマイクロタスクは、種別ごとに規定完了数( $n=5$ )を設けた。

論述改善タスクでは、割り当てられた3つのマイクロタスクリストに含まれるマイクロタスクをすべて実施してもらった。3.2節で述べたように、マイクロタスクは3種類である。意見と事実の切り分けタスクは、単一の文に対して実行可能であるため、タスク実施画面では単一の文を表示した。事実/意見に対する根拠の有無の確認タスクは、パラグラフに対して実行可能であるため、タスク実施画面ではパラグラフを表示した。

最後に、割り当てられたマイクロタスクをすべて完了した後、人口統計学的属性（性別、年齢、最終学歴）と実施したマイクロタスクの種類に応じたアンケートに回答してもらった。アンケートでは、各質問項目の評価尺度を5段階のリッカート尺度として作成した。具体的には、以下に関する質問に回答してもらった。

**Q1** タスク指示内容の理解容易性

**Q2** タスク実施の楽しさ

**Q3** タスク実施の容易性

**Q4** タスク実施の疲労度

**Q5** タスク実施の意欲

## 5 結 果

クラウドワーカ実験によって得られた134名のクラウドワーカの行動ログとアンケート結果を分析した。また、執筆者実験によって得られた4名の執筆者のアンケート結果を分析した。本稿では、クラウドワーカと執筆者に実施してもらったアンケートの回答結果を間隔尺度と見なして分析を行った。

### 5.1 クラウドワーカによるマイクロタスク実施コスト

クラウドワーカによるフィードバック実施に対して、どれほどの時間的コストが必要かを調べるために、マイクロタスクの作業時間について分析した。

#### 5.1.1 文に対するマイクロタスク平均作業時間

クラウドワーカが、ある文に対するマイクロタスクの実施に、どれほどの時間がかかっているかを算出した。表1に、クラウドワーカによる文に対するマイクロタスク平均作業時間を記す。対象とした4文書すべてにおいて、各種マイクロタスクの中では意見に対する根拠の有無の確認タスクの平均作業時間が最も長かった。この要因として、意見に対する根拠の有無の確認タスクでは、他の2種とは異なり、記述回答が存在することが考えられる。クラウドワーカは、文章の改善が必要な理由・不要な理由を考えるために、時間を長く費やしてしまったと推測される。

#### 5.1.2 文書に対するフィードバックの開始から完了までの所要時間

クラウドワーカが実際にフィードバックを行うのにかかる時間を調べるために、各文書に対する各種フィードバックの開始から完了までの所要時間を算出した。ここで、各種フィードバックの完了とは、ある文に対するマイクロタスクの結果が規定回数分だけ生成された状態を示す。表1に、各種フィードバックの開始から完了までの所要時間を記す。各種フィードバックの開始から完了までの所要時間は、文書1では合計で2804秒（約47分）、文書2では合計で4173秒（約1時間10分）、文書3では合計で2311秒（約39分）、文書4では合計で2412秒（約40分）であった。最短で約39分、最長で約1時間10分でフィードバック作業が終了している。

#### 5.1.3 クラウドワーカ1人あたりの平均作業時間

クラウドワーカ個人が、割り当てられたタスクをすべて完了するまでの平均作業時間を算出した。表1が示すとおり、クラ

ウドワーカは平均的に 361.9 秒（約 6 分）から 508.9 秒（約 9 分）の間で割り当てられたタスクが完了していた。中央値は、文書 1 では 370.0 秒、文書 2 では 206.0 秒、文書 3 では 174.0 秒、文書 4 では 229.5 秒であった。

## 5.2 クラウドワーカへのアンケート

クラウドワーカに、マイクロタスク実施の体験に関するアンケートをマイクロタスクの種類ごとに回答してもらった。表 2 に、その分析結果を表す。タスク指示内容の理解容易性に関する Q1 の回答結果において、すべてのマイクロタスク種別で 3.00 を超えるポジティブな評価が得られた。そのうち、意見と事実の切り分けタスクの平均値は他の 2 種類よりも高かった。タスクの楽しさに関する Q2 の回答結果において、意見と事実の切り分けタスクの平均値は 3.0 を超えるポジティブな評価が得られた。一方、他の 2 種類では、平均値が 3.0 を下回るネガティブな評価が得られた。意見に対する根拠の有無の確認タスクの平均値は最も低く、3 種の中では最も楽しいと感じられていないことが示唆された。タスクの容易性に関する Q3 の回答結果では、マイクロタスク種別に関わらず平均値が 3.0 を下回るネガティブな評価が得られた。クラウドワーカは、実施したマイクロタスクについて簡単だと感じていないことが示唆された。タスクの疲労度に関する Q4 の回答結果では、意見と事実の切り分けタスクの平均値が 3.0 を下回った。Q4 では数値が高いほど疲労度が高いことを示す。意見と事実の切り分けタスクは最も疲労度が少なく、意見に対する根拠の有無の確認タスクは最も疲労度が高いことが示唆された。タスクの実施意欲に関する Q5 の回答結果では、意見と事実の切り分けタスクにて平均値が 4.23 とポジティブな評価が得られた。また他の 2 種類の平均値においても、3.0 を超えるポジティブな評価が得られた。クラウドワーカが、次のマイクロタスク実施に意欲的であることが示唆された。

## 5.3 執筆者によるフィードバック結果の評価

執筆者に、生成されたフィードバックを閲覧し、システムの有用性に関するアンケートに回答してもらった。表 3 に、その結果を示す。

フィードバックを見ることによる説得力のある文書執筆への意欲向上と次取るべき行動の理解に関する Q1~Q3 の回答結果の平均値は 3.0 よりも高く、好意的な評価が得られた。提案システムによる文レベルのフィードバックの閲覧は、文書の説得力向上に対してポジティブな影響を与えられる可能性がある。また、Q6~Q7 に関しても平均値は 3.0 よりも高い評価が得られた。提案システムのフィードバック内容の理解容易性とシステム利用の意欲についてポジティブな結果となった。一方で、Q5 の回答結果の平均値は 3.0 よりも低いネガティブな評価が得られた。フィードバック内容の妥当性について課題があることが示唆された。

提案システムの課題をより詳細に明らかにするために、有用だと感じられたフィードバックの特性やシステムの問題点・改善点について、文書執筆者に対して尋ねた。システムの問題点

としては、フィードバックの品質が挙げられた。具体的には、以下のような意見が得られた。

**フィードバック（意見に対する根拠の有無の確認タスクによる根拠を持たせるための改善案）の質が低いものが多い。「自分の経験を交えて」のように似たフィードバックばかりが書かれていた。**

意見に対する根拠の有無の確認タスクの実施は、他の 2 種類と比べて、ある文に対する平均作業時間が長く、クラウドワーカへのアンケートの Q4 において最も疲れを感じさせるものになっていた。得られた意見から、クラウドワーカがタスクの途中から疲れを感じて適当にタスクに取り組んでしまった可能性が考えられる。

フィードバックの有用性については以下のような意見が得られた。

**自分が書いた文書に対して、全体的に「根拠が無いのではないか」と疑った状態で客観的にレビューできた。フィードバックの正確性が低い場合でも、実際に根拠が足りていない部分が見つかった点において価値を感じた。**

得られたフィードバックの品質が低い場合でも、それを採用しないという選択を取るためには、執筆中の文章との関連を客観的に見極める必要がある。この過程によって、文章執筆中のメタ認知的モニタリング（思考が生み出される様子の観察）[14]が促進され、説得力を向上させる意欲が湧き、次にやるべきことが分かるようになったのではないかと考えられる（5.3 節）。

## 6 考察

本稿では、執筆者の説得力のある論述を可能にするために、クラウドソーシングを用いた論述改善フィードバック手法を提案した。提案手法は、クラウドワーカが過度な負荷なく効果的な論証改善フィードバック作業を実施できるよう、フィードバックタスクを 3 種類のマイクロタスクとして分割する。これによって、クラウドワーカが割り当てられたマイクロタスクに対して非同期的に取り組めるようになり、執筆者が短時間かつ低コストでフィードバックを得られる環境を実現することを狙った。

クラウドワーカ 1 人あたりのフィードバック作業時間は、平均的に 361.9 秒（約 6 分）から 508.9 秒（約 9 分）であった。本稿におけるユーザ実験と同じ状況を想定すると、ある文書に対するフィードバックを完了するために、少なくとも約 39 分のフィードバック作業時間と 1500 円の費用（クラウドワーカへの報酬）が必要となる。大学教員の時給を 5000 円として考えた時、ある 1 人の学生に対するフィードバックを 10 分で完了できるとして最低 833 円が費やされる。1 人の学生に対するフィードバックにおいては、本稿の提案手法を利用するよりも教員が単独で行ったほうが時間的・品質的に勝っていると考えられるかもしれない。しかし、1 章において述べたように、学

表 1 論述改善タスクにおけるクラウドワーカの行動ログの統計（括弧外の数値は平均値，括弧内の数値は標準偏差を表す）.

文書	文章要素のサイズ	クラウドワーカ数	1 文あたりの 平均作業時間			1 文書あたりの フィードバック所要時間			ワーカ 1 人あたりの 平均作業時間
			(1)	(2)	(3)	(1)	(2)	(3)	
1 志望企業: 広告代理店	文字数: 711	71	22.57	57.1	126.58	456	496	1852	508.90 (462.96)
	パラグラフ数: 3 文数: 18		(43.35)	(65.80)	(108.70)				
2 志望企業: ウェブ企業	文字数: 566	27	26.08	-	128.16	450	-	3723	405.30 (463.46)
	パラグラフ数: 2 文数: 11		(35.13)	(168.76)					
3 志望企業: カウンセリング企業	文字数: 651	55	28.45	59.09	171.33	286	586	1439	361.93 (531.24)
	パラグラフ数: 3 文数: 12		(32.16)	(43.36)	(290.96)				
4 志望企業: 自動車メーカー	文字数: 825	54	37.30	77.76	164.10	388	538	1486	414.43 (474.25)
	パラグラフ数: 3 文数: 12		(40.34)	(43.30)	(163.24)				
全文書の平均値			25.86 (40.86)	60.86 (56.13)	146.31 (187.24)	394.75 (78.64)	533.33 (35.23)	2125.00 (1081.21)	663.07 (688.25)

表 2 5 段階評価を用いたクラウドワーカに対するアンケート調査結果. 括弧外の数値は平均値，括弧内の数値は標準偏差を表す. また，図中の **Task<sub>divide</sub>**，**Task<sub>fact</sub>**，**Task<sub>opinion</sub>** はそれぞれ，意見と事実の切り分けタスク，事実に対する根拠の有無の確認タスク，意見に対する根拠の有無の確認タスクを表す.

マイクロタスク種別			
質問項目	Task <sub>divide</sub> Task <sub>fact</sub> Task <sub>opinion</sub>		
	Task <sub>divide</sub>	Task <sub>fact</sub>	Task <sub>opinion</sub>
Q1 タスクの指示内容はすぐに理解できましたか？	3.77 (1.05)	3.21 (1.10)	3.21 (1.04)
Q2 タスクをやってみて，楽しいと感じましたか？	3.62 (0.78)	2.92 (0.93)	2.79 (0.94)
Q3 タスクをやってみて，簡単だと感じましたか？	2.46 (0.89)	2.06 (0.77)	2.21 (0.67)
Q4 タスクをやってみて，どの程度疲れを感じましたか？	2.58 (1.00)	3.19 (1.07)	3.55 (0.91)
Q5 実施したタスクをまたやりたいと思いますか？	4.23 (0.75)	3.52 (1.02)	3.26 (1.11)

表 3 5 段階評価を用いた執筆者に対するアンケート調査結果. (括弧外の数値は平均値，括弧内の数値は標準偏差を表す.)

質問項目	回答
Q1 フィードバックを見ることで，文書の説得力をより向上させたいと感じましたか？	3.77 (1.05)
Q2 フィードバックを見ることで，文書の説得力を向上させるためにやるべきことが分かるようになりましたか？	4.00 (1.15)
Q3 フィードバックを見ることで，説得力のある文書を書くために足りていなかったことが見つかりましたか？	3.75 (0.5)
Q4 フィードバックを閲覧した画面は，使いやすいと感じましたか？	3.00 (0.82)
Q5 フィードバックの内容は，全体的に妥当なものが多いと感じましたか？	2.25 (0.96)
Q6 フィードバックの内容は，全体的に理解しやすいと感じましたか？	3.25 (0.5)
Q7 次回のレポートや論文の執筆の際に，使ってもらったようなシステムがあるとすると利用したいと思いますか？	3.5 (1.7)

生ごとにフィードバックを行うとすれば，教員には膨大な時間と労力が求められる．例えば，フィードバックの対象を 80 人として，各学生のフィードバックを 10 分で完了できる場合，13 時間の作業が必要となる．現実的には，研究や講義の準備やその他の業務がある中で，講義のレポートのフィードバックの実施優先度を高めて即座に行くことは難しいのではないだろうか．執筆者が，より早く論述を改善するサイクルを得るには，本稿の提案手法を採用することが望ましいと考えられる．

6.1 改善点

本稿における主要な改善点について述べる．

6.1.1 フィードバックの品質

本稿では，クラウドワーカによる論述改善フィードバックの品質を担保するために，各マイクロタスクの規定完了数 ( $n = 5$ ) を設けた．システムは，規定完了数以上得られたマイクロタスク実施結果に対して多数決的な集約を行い，執筆者に表示するフィードバック内容を決定した．執筆者に対するシステムについてのアンケート・インタビューからは，フィードバックの品質が十分でないことが示唆された (5.3 節)．フィードバックの品質が低い場合においてもポジティブな影響がある可能性について議論したが，フィードバック品質は高ければ高いほど望

ましい．妥当なフィードバックが少ないと感じていた執筆者に対してその理由を聞いたところ，以下のような回答が得られた．

前後の文脈を考慮できていないフィードバックがあった．

似たような内容のフィードバックが多くあった．

本稿ではクラウドワーカの多数決によって品質を担保しようと試みたが，上記の回答のように，提示したフィードバックが執筆者にとって不十分であることが分かった．フィードバックの品質向上のために，Bernstein らの Find-Fix-Verify パターン [11] の採用が改善の方向性として考えられる．あるクラウドワーカのマイクロタスク結果を他のクラウドワーカが評価するマイクロタスクを組み込むことで，本稿での応用が可能となるであろう．事実/意見に対する根拠の有無の確認タスクの判断に対して評価を行うことで，より厳密に根拠の有無が実現できると思われる．また，意見に対する根拠の有無の確認タスクの記述回答の結果に対してスコアを付け，その内容を評価して，教員の結果と比較することも課題として残されている．

加えて，フィードバック品質向上のためには，機械学習モデルを用いたマイクロタスク実施の支援も改善の方向性として考

えられる。機械学習モデルに、ある文の意見と事実の切り分けや根拠の有無の確認の判定結果を予測させることが可能となれば、クラウドワークはその予測結果を確認するのみでマイクロタスクを完了できるようになる。クラウドワークが文の判定タスクを実施するための時間的・労力的負荷が下がることで、人間の能力が必要なマイクロタスクにリソースを費やすことが可能となる。機械学習モデルの支援のもとで、多くのクラウドワークが時間的・労力的な負荷の高かった意見に対する根拠の有無の確認タスクの実施が可能になることは、フィードバック品質の向上につながるであろう。

### 6.1.2 クラウドワークの事前知識を考慮した割り当てロジック

クラウドワークは、論述課題「人工知能と生産性」の回答文書に対してフィードバックを行った(4.3節)。昨今、人工知能は社会的に取り扱われることが多く、専門的な知識が過剰に求められない内容であることを理由に、本テーマを採用した。しかし、一部のクラウドワークによっては馴染みのないテーマであり、タスクが実施しづらかった可能性がある。フィードバックを行うクラウドワークの事前知識が、フィードバックの品質や体験(疲労度や容易性など)に与える影響を詳細に調査する必要がある。またシステム側では、複数の分野の論述文書に効果的に対応するために、クラウドワークの事前知識を考慮した割り当てロジックなどの検討が求められる。

## 7 おわりに

本稿では、論述の説得力を向上させるために、クラウドワークによる論述改善フィードバック生成のための手法を提案した。提案手法では、複雑な論述改善タスクを、(1)意見と事実の切り分け、(2)事実の根拠の有無の確認、(3)意見の根拠の有無の確認の3種類のマイクロタスクに分割した。クラウドワークを対象としたユーザ実験の結果、実験対象とした文書内において、ある文書のフィードバックは作業開始から最短で約39分で完了できること、クラウドワークは割り当てられたフィードバック作業を平均的に約6分から9分の間で完了できることが分かった。システムを用いてフィードバックを閲覧した執筆者からは、説得力を高める意欲の向上・次にやるべきことの理解度に関する測定結果にて有望な評価が得られた。

今後は、フィードバック品質の向上やクラウドワークの事前知識を考慮した割り当てロジックの実装が求められる。

## 謝 辞

本研究はJSPS科研費JP18H03244, 21H03554, 21H03775, 22H03905の助成を受けたものです。ここに記して謝意を表します。

## 文 献

- [1] Ralph P Ferretti, William E Lewis, and Scott Andrews-Weckerly. Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology*, Vol. 101, No. 3, p. 577, 2009.

- [2] Raymond S Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, Vol. 2, No. 2, pp. 175–220, 1998.
- [3] John Hattie and Helen Timperley. The power of feedback. *Review of educational research*, Vol. 77, No. 1, pp. 81–112, 2007.
- [4] David H Jonassen and Bosung Kim. Arguing to learn and learning to argue: Design justifications and guidelines. *Educational Technology Research and Development*, Vol. 58, No. 4, pp. 439–457, 2010.
- [5] Paul Black and Dylan Wiliam. Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education)*, Vol. 21, pp. 5–31, 2009.
- [6] Kenneth C Arnold, Krzysztof Z Gajos, and Adam T Kalai. On suggesting phrases vs. predicting words for mobile text composition. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 603–608, 2016.
- [7] Kenneth C Arnold, Krysta Chauncey, and Krzysztof Z Gajos. Predictive text encourages predictable writing. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pp. 128–138, 2020.
- [8] Thiemo Wamborg, Christina Niklaus, Matthias Cetto, Matthias Söllner, Siegfried Handschuh, and Jan Marco Leimeister. Al: An adaptive learning support system for argumentation skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020.
- [9] Thiemo Wamborg, Tobias Kueng, Matthias Soellner, and Jan Marco Leimeister. Arguetutor: an adaptive dialog-based learning system for argumentation skills. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–13, 2021.
- [10] Mia Xu Chen, Benjamin N Lee, Gagan Bansal, Yuan Cao, Shuyuan Zhang, Justin Lu, Jackie Tsay, Yinan Wang, Andrew M Dai, Zhifeng Chen, et al. Gmail smart compose: Real-time assisted writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2287–2295, 2019.
- [11] Michael S Bernstein, Greg Little, Robert C Miller, Björn Hartmann, Mark S Ackerman, David R Karger, David Crowell, and Katrina Panovich. Soylent: a word processor with a crowd inside. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology*, pp. 313–322, 2010.
- [12] Pradeep Kumar Murukannaiah, Nirav Ajmeri, and Munindar P Singh. Enhancing creativity as innovation via asynchronous crowdwork. In *14th ACM Web Science Conference 2022*, pp. 66–74, 2022.
- [13] Niloufar Salehi, Jaime Teevan, Shamsi Iqbal, and Ece Kamar. Communicating context to the crowd for complex writing tasks. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 1890–1901, 2017.
- [14] Douglas J Hacker. A metacognitive model of writing: An update from a developmental perspective. *Educational Psychologist*, Vol. 53, No. 4, pp. 220–237, 2018.