

Crowdsourced Argumentation Feedback for Persuasive Writing

Hiroki Ihoriya¹ and Yusuke Yamamoto²

¹ Shizuoka University, Hamamatsu, Shizuoka, Japan
`ihoriya@design.inf.shizuoka.ac.jp`

² Nagoya City University, Nagoya, Aichi, Japan
`yusuke.yamamoto@acm.org`

Abstract. In this paper, we propose a method for generating feedback to improve an argumentation’s persuasiveness by utilizing multiple crowd workers. This method is aimed at mitigating the typically time-consuming task of providing feedback on argumentation documents, which often falls solely on the author’s supervisor or teacher. The proposed method divides the task argument improvement into three types of micro-tasks: (1) classifying opinions vs. facts, (2) checking that evidence is given to support the facts, and (3) checking that a rationale is given to support the opinions. These micro-tasks are allocated to the crowd workers and executed asynchronously to achieve efficient feedback generation.

Keywords: Writing support · Crowd Sourcing.

1 Introduction

Persuasive argumentation is among the most important skills to be acquired in school education [6]. In order to write a persuasive document, it is essential to present reasonable evidence for claims and to reduce the room for potential objections. We define a persuasive argumentative document as an argumentative document that provides reasonable evidence for its claims and is conscious of the objectivity of its components.

Persuasive argumentation is difficult for those lacking sufficient experience in writing. The first reason is that humans are prone to bias in our opinions, in particular in the form of myside bias, the tendency of humans to value their own opinions [11, 13]. When one collects only those data and references that suit oneself as the evidence for an argument, there is a strong possibility that the resulting argument will lack convincing validity.

The second reason why persuasive argumentation can at first be difficult is the lack of opportunities to acquire and improve skills in persuasive argumentation. Generally, feedback is necessary for acquiring and improving skills [8]. In particular, feedback is considered helpful for development of argumentation skills, a key subset of cognitive skills that are difficult to master. When writing an argumentative document such as a report or an essay, there is generally a

reader, such as an expert in the field or a teacher. Jonassen et al. identified teachers' lesson design and pedagogy skills for developing argumentation skills in students as crucial challenge areas for improving student argumentation skills [9]. According to Jonassen et al., many teachers lack the pedagogical skills to foster argumentation in classrooms. This lack makes it difficult for students to obtain the feedback they need to acquire argumentation skills. Even if teachers possess sufficient teaching skills, it takes substantial time and effort to provide feedback to each student. Furthermore, now that large-scale, remote lectures are common in universities, it is becoming increasingly arduous for teachers to provide students with feedback on their writing skills [4]. Therefore, a system and an environment are needed in which authors can reflect on their writing and improve the persuasiveness of their arguments.

In this paper, we propose a crowdsourcing-based feedback method for improving argumentation so as to foster more persuasive argumentation. Figure 1 shows a schematic diagram of the proposed method. We define argumentation as **something that makes a claim based on a valid argument**'. The proposed method aims to point out poorly reasoned sentences in a written document and provide feedback to the author to improve the argumentation. Given an author's document, the proposed method generates micro-tasks that enable crowd workers to provide feedback for argumentation improvement at a low cost. There are three types of micro-tasks, as described below.

1. Classifying opinion vs. fact
2. Checking that evidence is given to support the facts
3. Checking that a rationale is given to support the opinions

These micro-tasks are executed for each sentence in a document. The proposed method aggregates the results of the micro-tasks executed for each sentence. Moreover, it presents feedback to the writer, as shown in Figure 6, triggering the writer to improve the writing. The micro-tasking allows crowd workers to work asynchronously on the assigned micro-tasks, thus enabling efficient feedback generation. The proposed method is expected to distribute the time and effort load on the person(s) providing feedback compared to the case in which a teacher alone provides feedback. We expect authors to be able to obtain feedback to improve their arguments on short timescales and at a low cost.

2 Related Work

2.1 Writing support system

Arnold et al. proposed a phrase-oriented input assistance system for touchscreen keyboards to improve the text input for written content [2]. Arnold et al. also analyzed how a predictive text input system affects a given writer's written content [1]. The results of user experiments confirmed that while predictive text input systems make written content more concise, the written content itself may be influenced by the text presented by the system. Wambsganss et al. developed

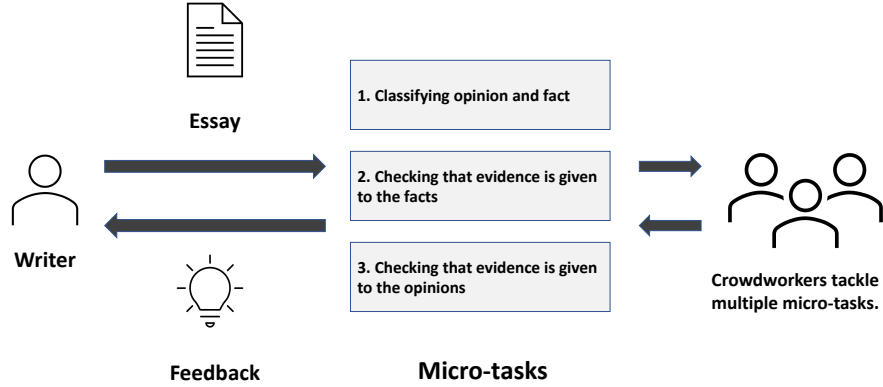


Fig. 1. Overview.

an adaptive system that provides feedback on the argumentative structure of a text [15]. In earlier work, Wambsganss et al. had also proposed a bot-style tutoring system to support analytical writing [14]. In related work, Chen et al. developed Gmail Smart Compose, which assists users in writing email texts by displaying the continuation of phrases that are likely to be typed in Gmail [5].

In this paper, we propose a system that assigns argumentative writing feedback tasks with low implementation costs to crowd workers in order to generate feedback on argumentative documents.

2.2 Crowdsourcing support for content creation and improvement

Crowdsourcing is another tool finding use in the support of text creation and improvement. A study by Bernstein et al. proposed a text-writing system called SOYLENT in which crowd workers reviewed texts based on the Find-Fix-Verify algorithm, which is designed to improve the quality of work [3]. Murukannaiah et al. compared the performance of single and multiple crowd workers (team workers) on an idea generation task [10]. Their analysis suggested that workers were more likely to generate more creative ideas when working in teams than alone. Salehi et al. investigated various strategies for effective communication between crowd workers and clients to perform writing tasks [12]. Their survey revealed that the worker strategies considered helpful by clients varied depending on the writing phase and that Q&A tasks and comments and editing tasks drove client communication costs.

3 Crowdsourcing Feedback for Argumentation Improvement

We propose a crowdsourcing method for providing feedback of a standardized level of quality while reducing the time and effort load of the workers who provide feedback by dividing feedback for argument improvement into three micro-tasks. An argumentation improvement micro-task is a feedback task performed on a single sentence in a document. The micro-tasks here are (1) classifying opinion vs. fact, (2) checking that evidence is given to support the facts, and (3) checking that a rationale is given to support the opinions. A micro-task list is a set of micro-tasks executed for a single paragraph. We describe the details of the proposed method below.

3.1 Workflow for argumentation

First, a writer creates a document draft for a given writing task. Next, the writer submits the draft to the proposed system. At that time, the system generates micro-tasks to improve the draft. The purpose of micro-task generation is to reduce the burden on the crowd workers who perform the feedback by dividing the task of improving the essay, which otherwise might require complex thinking. Following the logic described in section 3.3, the proposed system assigns micro-task lists to crowd workers, and each crowd worker performs multiple micro-tasks. The task of improving the draft is terminated as soon as the specified number of micro-tasks for each sentence has been executed. After the completion of the document-level task, the proposed system aggregates the feedback results and presents the feedback for draft improvement as shown in Figure 6. The aim is for the feedback generated to help writers identify areas for improvement in their documents and write more persuasively.

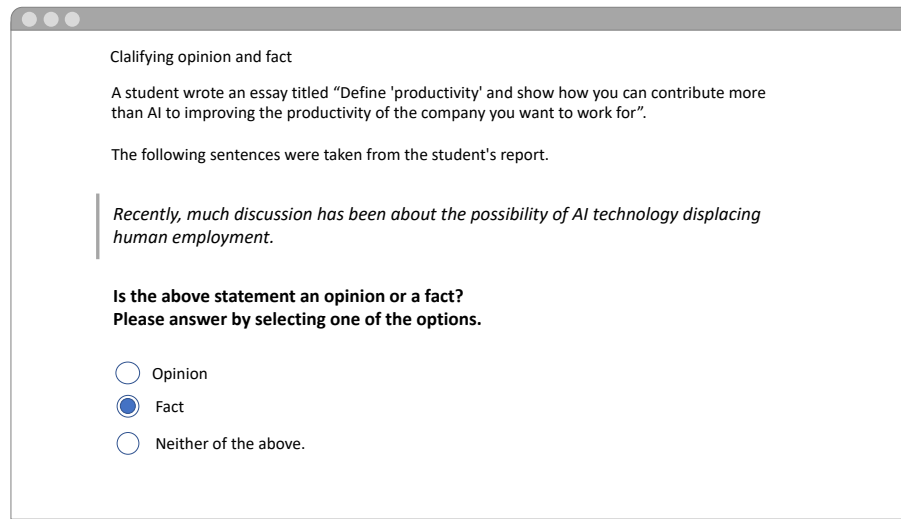
3.2 micro-task-design

As stated above, we designated the following three types of feedback tasks as micro-tasks.

1. Classifying opinions vs. facts
2. Checking that evidence is given to support the facts
3. Checking that a rationale is given to support the opinions

In the classifying opinions vs. facts micro-task, crowd workers classify whether a statement is an opinion or a fact on the screen shown in Figure 2. By performing this micro-task, the crowd worker classifies each sentence in the document as one of three types: (1) opinion, (2) fact, and (3) neither of the above.

When the feedback is presented to the writer, the system highlights the sentences according to the classified type, allowing the writer to objectively observe the ratio of opinions to facts in the document.



Classifying opinion and fact

A student wrote an essay titled "Define 'productivity' and show how you can contribute more than AI to improving the productivity of the company you want to work for".

The following sentences were taken from the student's report.

Recently, much discussion has been about the possibility of AI technology displacing human employment.

Is the above statement an opinion or a fact?
Please answer by selecting one of the options.

☐ Opinion

☒ Fact

☐ Neither of the above.

Fig. 2. Example screen of Micro-task (1): Classifying opinion and fact.

In the micro-task of checking that evidence is given to support the facts, crowd workers determine whether a factual statement is supported by a source of information on the screen shown in Figure 3. crowd workers select from among the following three options.

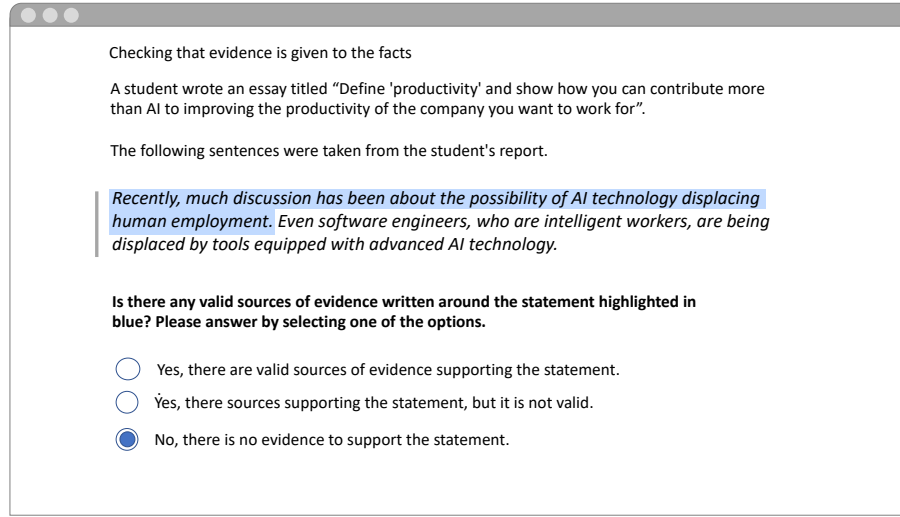
- Yes, there are valid sources of evidence supporting the statement.
- Yes, there are sources supporting the statement, but the statement is not valid.
- No, there is no evidence to support the statement.

When the feedback is presented to the writer, the system highlights any factual statement lacking a valid source of information.

In the micro-task of checking that a rationale is given to support the opinions, crowd workers use the screen shown in Figure 4 to judge whether the writer has offered a rationale supporting each opinion and explain the judgment or offer a proposal for improvement Figure 4. In the descriptive response, crowd workers are asked to choose from among three options.

- Yes, there is a substantial rationale supporting the statement.
- Yes, there is a rationale supporting the statement, but it is limited.
- No, there is no rationale to support the statement.

In the descriptive response, crowd workers are asked to provide reasons for their judgments or suggestions for improvement. Specifically, in the case of “Yes, there is a substantial rationale supporting the statement,” crowd workers are asked to describe the text on which this judgment was based. In other two cases, crowd workers are asked to describe how the text can be improved to provide



Checking that evidence is given to the facts

A student wrote an essay titled "Define 'productivity' and show how you can contribute more than AI to improving the productivity of the company you want to work for".

The following sentences were taken from the student's report.

Recently, much discussion has been about the possibility of AI technology displacing human employment. Even software engineers, who are intelligent workers, are being displaced by tools equipped with advanced AI technology.

Is there any valid sources of evidence written around the statement highlighted in blue? Please answer by selecting one of the options.

- ☐ Yes, there are valid sources of evidence supporting the statement.
- ☐ Yes, there sources supporting the statement, but it is not valid.
- ☒ No, there is no evidence to support the statement.

Fig. 3. Example screen of Micro-task (2): checking that evidence is given to the facts.

a better rationale. When this feedback is presented to the writer, the system highlights opinion sentences lacking a valid rationale, along with reasons for the judgment and suggestions for improvement.

3.3 Micro-task assignment

The proposed method assigns multiple micro-tasks (micro-task list) to a given crowd worker according to the following procedure.

1. Split a document into paragraphs.
2. Create micro-tasks for each sentence in a given paragraph
3. Allocate the paragraph's set of generated micro-tasks (micro-task list) to crowd workers

The flow of the argumentation improvement task is shown in Figure 5. With the creation of a micro-task for each sentence in a segmented paragraph, the scope of text subject to feedback by a single crowd worker can be reduced from the entire document to a paragraph or a few sentences. Assigning a micro-task list defined at paragraph scope to a single crowd worker allows each crowd worker to provide coherent feedback on a single paragraph.

Sequential dependencies exist in the execution of micro-tasks. The micro-task of classifying opinion vs. fact must be completed before the micro-tasks involving checks for support are executed. This sequentiality is due to the fact that support check micro-task requires that each statement is already identified as an opinion, a fact, or neither. Therefore, logistical priority is given to the task of classifying opinion vs. fact.

Checking that evidence is given to the opinions
...

In software development, there are many uncertainties, such as schedules and human resource characteristics, etc. Inference by AI based on training data may result in extremely low productivity in urgent situations. Therefore, the long-term productivity of an organization would be higher if it employs humans, who can respond flexibly to different situations.

**Is there any reasonable evidence written around the statement highlighted in orange?
Please answer by selecting one of the options.**

☐ Yes, there is substantial evidence supporting the statement.

☐ Yes, there is some evidence supporting the statement, but it is limited.

☒ No, there is no evidence to support the statement.

Briefly explain the reason for your answer to the above question.
If you answered "Yes, there is substantial evidence supporting the statement.", please briefly explain what you read in the text that led you to that conclusion. If your answer is other than that, please briefly explain how the text could be improved to provide more evidence.

Fig. 4. Example screen of Micro-task (3): checking that a rationale is given to support the opinions.

After the specified classification micro-tasks for a document have been completed, the allocation of support check micro-tasks begins. When all the allocated micro-tasks are completed, the document-level task is complete, and the process terminates.

3.4 Aggregation and presentation of micro-tasks

The aggregation and presentation methods are different for each type of micro-task. The results of each micro-task may differ among crowd workers when multiple crowd workers are assigned the same micro-task. For example, it is possible that in the classification of opinion vs. fact, crowd worker A, judges a certain statement to be an opinion, while crowd worker B judges the same statement to be a fact. In order to account for such differences, the proposed system aggregates the results for the micro-task according to a plurality rule and assigns the plurality-rule assessment as the final result for the micro-task.

For example, in the micro-task of classifying a particular statement as an opinion vs. a fact, the results are obtained from all crowd workers' judgments. For the set of crowd worker assessments of a given statement, the assessment adopted is the one associated with the highest number of crowd worker assessments. Once a given statement has been thus classified as an opinion, a fact, or neither, this result is offered to the writer on the feedback screen. Specifically, the feedback is presented to the writer by highlighting the opinions and the facts so that each can be distinguished, as shown in Figure 6.

A similar plurality rule is applied in the support check micro-tasks. In the feedback to the writer, each sentence judged to lack support is accompanied by

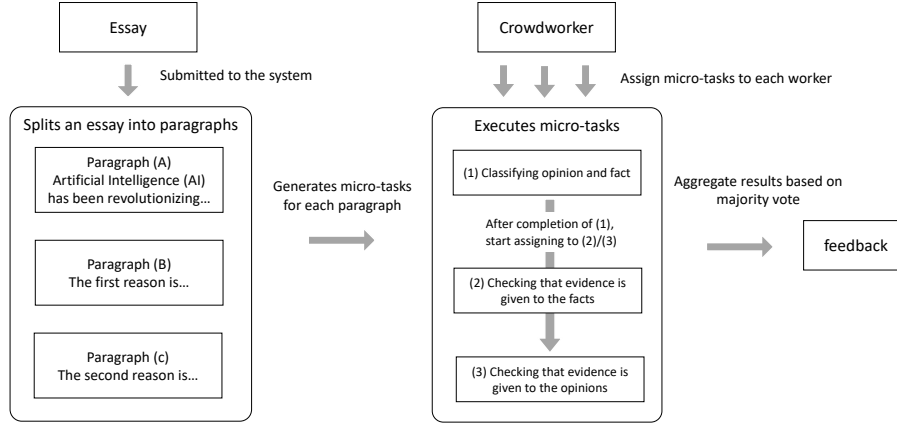


Fig. 5. Workflow overview.

the crowd worker’s stated reasoning explaining why it was judged to be lacking in support. These reasons are not semantically aggregated or summarized; instead, all reasons or hints for improvement provided by the crowd workers are presented to the writer.

As shown in Figure 6, when the writer accessing the system then clicks on an opinion or factual statement, the system presents the adopted plurality judgment regarding support for the statement along with the detailed reasons for that judgment as provided by the crowd workers. The feedback details include the number of crowd workers who completed the support / check task and those who responded “Yes but” or “No.” If an opinion statement is selected by the writer, then crowd worker feedback comments appear onscreen regarding “How support can be improved.”

4 Experiments

To evaluate the effectiveness of the proposed method, we conducted user experiments. We refer to experiment participants who receive feedback on their documents as writers. We further refer to experiment participants recruited through a crowdsourcing website to perform an essay improvement task as crowd workers. Hereafter, the term “participants” includes both writers and crowd workers. In user experiments, writers are assigned an essay task, while crowd workers are assigned micro-tasks to improve the essay .



Fig. 6. An example image of feedback after the argumentation improvement task is executed. The left column displays the written document. Orange highlighting indicates statements identified as opinions. Blue highlighting indicates statements identified as facts. The right column displays the feedback decisions and suggestions for improvement obtained from the crowd workers. The system currently supports only Japanese language.

4.1 Participants

Four students affiliated with the Faculty of Informatics at Shizuoka University were recruited as writers. 150 crowd workers were recruited using the Japanese crowdsourcing service CrowdWorks³. We explained the data collection policy to each crowd worker before starting the task, and they each crowd worker was allowed to proceed only once agreement was reached. crowd workers who did not complete the task properly were excluded from the analysis. Finally, data from 134 participants were used for the analysis of the system. Of the 134, 78 were male, while 56 were female. 85 crowd workers had completed their undergraduate or graduate education. We paid 100 Japanese yen (about 1 USD) to each participant for completion of the document-level task.

4.2 System

We built a system for participants to perform the task, hereafter referred to as “the experimental system.” Writers viewed an explanation page for the writing

³ <https://crowdworks.jp>

task on the experimental system, and later they viewed the feedback generated by the crowd workers on the same experimental system. crowd workers browsed a page explaining the content of the argumentation improvement task on the experimental system, and later they executed the task on the same experimental system.

4.3 Experimental procedure for writer

The writer moved through the experimental system according to the following procedure: (1) submission of argumentation document, (2) item Viewing of feedback, and (3) responding to questionnaires and interviews.

Each writer was asked to submit an essay for the argumentation task in response to the “Artificial Intelligence and Productivity” prompt below. **Artificial**

Intelligence and Productivity

Assume that you are currently seeking employment and that you have applied for a job at a company you want to work for. That company is hiring only one person, and you have been invited for a final interview at that company. According to preliminary information, all applicants for the final interview except you are AIs (artificial intelligences). In order to win a job offer from the company, you need to appeal to the hiring manager to demonstrate that hiring you is more beneficial to the company than hiring an AI. Define “productivity” and explain how you can contribute more than an AI to improving the productivity of your company of interest in about 0.5~1 page of A4 paper. Note that you may freely set your “company of interest,” though you do not need to name the company you are interested in.

The resulting essays were the ones on which the crowd workers offered feedback. After viewing the crowdsourced feedback, each writer answered a questionnaire. The questionnaire asked about the system’s usefulness for viewing feedback. Specifically, each writer answered the following questions.

- Q1** Did viewing the feedback make you feel a desire to make your document more persuasive?
- Q2** Did the feedback help you understand what you need to do to make your document more persuasive?
- Q3** Did the feedback help you find what you were missing to write a more persuasive document?
- Q4** Did you find the feedback easy to understand overall?
- Q5** Did you find the content of the feedback to be reasonable overall?
- Q6** If there were a system that would allow you to get feedback this way when writing your next report or paper, would you want to use it?

In the interview, we asked each writer which characteristics of the feedback were useful. We further asked each writer to name any problems with the system or improvements that could be made.

4.4 Experimental procedure for crowd worker

crowd workers moved through the experimental system according to the following procedure: (1) user registration, (2) execution of the argumentation improvement task, and (3) post-event questionnaire.

After agreeing to the experiment participation agreement on CrowdWorks, each crowd worker was directed to the experimental website. Next, after entering the experimental website, each crowd worker registered as a user and read the descriptions of specific tasks. Then three micro-task lists, one for each paragraph of the essay, were assigned to each crowd worker. The number of crowd workers assigned to each sentence was set to a predetermined number of completions ($n = 5$) for each micro-task.

For the argumentation improvement task, each crowd worker was asked to perform all the micro-tasks in each of the three assigned micro-task lists. As described in section 3.2, there are three types of micro-tasks. For micro-task (1), a single statement was displayed on the task execution screen, since no broader context is necessary for successful completion of this micro-task. For the statement support check tasks, a full paragraph was displayed on the task execution screen, since for any given statement in a paragraph, the paragraph defines the reasonable search scope for support.

Finally, after completing all the assigned micro-tasks, each participant was asked to answer a questionnaire regarding demographic characteristics (gender, age, and highest level of education) and task type. The questionnaire used a five-point Likert scale to rate each question item. Details of the questions presented to the crowd workers are provided below.

- Q1** Did you understand the task instructions immediately? (1: Did not understand at all immediately ~ 5: immediately understood perfectly)
- Q2** Did you enjoy doing the task? (1: Not at all enjoyable ~ 5: Very enjoyable)
- Q3** Did you find the task easy? (1: Not at all easy ~ 5: Very easy)
- Q4** How tired did you feel after doing the task? (1: not at all tired ~ 5: very tired)
- Q5** Would you like to do the task again? (1: I don't want to do it at all ~ 5: I want to do it very much)

5 Results

We analyzed the questionnaire results obtained through the crowd worker experiment for all 134 successful crowd workers. We also analyzed the results of the questionnaires obtained through the writer experiment for all four writers. Below we analyze the Likert scale results of the questionnaires given to the crowd workers and the writers.

5.1 Crowd workers questionnaire

Each crowd worker was asked to complete a questionnaire regarding her or his experience with micro-task execution for each micro-task type. Table 1 summarizes the questionnaire results. For Q1, regarding the ease of understanding

Table 1. Results of a survey of crowd workers using a 5-point scale. Values outside the parentheses represent the mean, and values inside the parentheses represent the standard deviation. The **Task_{divide}**, **Task_{fact}**, and **Task_{opinion}** in the figure represent the task (1) classifying opinion and fact, (2) checking that evidence is given to support the facts, and (3) checking that a rationale is given to support the opinions, respectively.

Question	Micro-task kind		
	Task _{divide}	Task _{fact}	Task _{opinion}
Q1 Did you immediately understand the task instructions?	3.77 (1.05)	3.21 (1.10)	3.21 (1.04)
Q2 Did you find the task enjoyable?	3.62 (0.78)	2.92 (0.93)	2.79 (0.94)
Q3 Did you find the task easy when you tried it?	2.46 (0.89)	2.06 (0.77)	2.21 (0.67)
Q4 How tired did you feel after doing the task?	2.58 (1.00)	3.19 (1.07)	3.55 (0.91)
Q5 Would you like to do the task you performed again?	4.23 (0.75)	3.52 (1.02)	3.26 (1.11)

the task instructions, all micro-tasks received positive evaluations, each with an average score above 3.0. Here the average value was higher for the micro-task of separating opinions from facts than for the other two micro-tasks. As for Q2, regarding task enjoyment, the mean value for the opinion/fact classification micro-task indicated a positive evaluation, again above 3.0. On the other hand, the other two micro-tasks received negative evaluations, with mean scores below 3.0. The opinion support check micro-task had the lowest mean score, suggesting that it was the least enjoyable of the three micro-task types. For Q3, regarding the ease of the task, evaluations were negative, with all mean values below 3.0. This finding suggests that the crowd workers did not feel that the micro-tasks they performed were easy. For Q4, regarding task fatigue, a higher values indicated higher degrees of fatigue, so the mean score below 3.0 for the opinions/fact separation task suggests that this micro-task was the least fatiguing. Meanwhile, the opinion support check micro-task was the most fatiguing. The results for Q5, regarding the willingness to perform the task, showed that opinion/fact separation received a positive evaluation, with a mean score of 4.23. The other two micro-tasks also received positive evaluations, with mean scores exceeding 3.0. These results suggest that crowd workers remained motivated to conduct the next micro-task.

5.2 Feedback Evaluation by Writers

Each writer was asked to view the feedback generated by the system and answer a questionnaire about the system’s usefulness. Table 2 shows the results.

For Q1, regarding the feedback’s motivating effect on writing persuasive documents and understanding the next action to take, the mean score was higher than 3.0, indicating a favorable evaluation of the system. Viewing sentence-level feedback by the proposed system may thus have a positive impact on improving the persuasiveness of documents. The mean value was also higher than 3.0, again indicating positive results, for Q6 and Q7, regarding the ease of understanding the feedback content of the proposed system and the willingness to use the system, respectively. On the other hand, the mean score for Q5 was lower than 3.0, suggesting dissatisfaction regarding the appropriateness of the feedback content.

Table 2. Results of a questionnaire survey of writers using a 5-point scale. (Values outside the parentheses indicate the mean, and values inside the parentheses indicate the standard deviation.)

Question	Answer
Q1 Did the feedback make you want to make your document more persuasive?	3.77 (1.05)
Q2 Does the feedback give you an idea of what you need to do to make your document more persuasive?	4.00 (1.15)
Q3 Did the feedback help you find what you were missing to write a persuasive document?	3.75 (0.5)
Q4 Did you find the screen on which you viewed the feedback easy to use?	3.00 (0.82)
Q5 Did you feel that the feedback was generally appropriate?	2.25 (0.96)
Q6 Did you find the feedback generally easy to understand?	3.25 (0.5)
Q7 When you write your next essay or paper, would you be willing to use a system such as the one you used?	3.5 (1.7)

In order to clarify the problems with the proposed system in greater detail, the writers were asked about the characteristics of the feedback that they found useful, about problems with the system, and about potential improvements to the system. In response, the writer noted the unsatisfactory quality of the feedback. Specifically, the following comments were obtained.

Much of the feedback was of low quality, including the suggestions for improvement to provide additional rationale for opinions. Much of the feedback text was similarly written, with stock phrases such as “...with my experience.”

The execution of the opinion support check micro-task was reported by crowd workers to be the most tiring in Q4 of the questionnaire, as the average work time per sentence for this micro-task was longer than for the other two micro-tasks. Based on the crowd workers opinions obtained, it is possible that the crowd workers felt tired in the middle of the task and worked on the task appropriately. The following comment was obtained on the usefulness of the feedback.

I reviewed my document objectively with an overall suspicion of the feedback indicating that there might not be enough evidence for some statements. Even when the accuracy of the feedback was low, I found value in that it helped me locate areas where the rationale was indeed lacking...

Even when the quality of the feedback obtained is low, it is necessary for the writer to objectively reflect on its relevance in order to choose not to adopt it. This process may have facilitated metacognitive monitoring (observation of how thoughts are produced) during the writing process [7], motivating writers to improve their persuasiveness and helping them to know what to do next (5.2 clause).

6 Discussion for Improvements

Feedback Quality In order to ensure the quality of the feedback, we set a specified number of completions ($n = 5$) for each micro-task. The system performed a plurality rule-based aggregation of the results for each micro-task to determine

the feedback to be displayed to the writer. Questionnaire interviews with writers about the system suggested that the quality of feedback was not sufficient (see section 5.2). We discussed the possibility of positive impacts even when feedback quality was low, but they responded that the higher the feedback quality, the more desirable it is. When we asked the writers who noted inadequate feedback for their reasons for their assessment of inadequacy, they offered the following responses.

Some feedback did not consider the context of what came before and after it. Much of the feedback contained similar phrases and phrasing...

We attempted to ensure quality by applying a plurality rule to the judgments of the crowd workers. However, we found that the feedback we presented was insufficient for the writers, as shown in the above responses. To improve the quality of feedback, the adoption of the Find-Fix-Verify algorithm [3] is a possible direction. Incorporating a micro-task in which one crowd worker’s micro-task results are evaluated by another crowd worker would be a possible application in this paper, with the potential to help the crowd workers realize more rigorous evidence or rationale.

Crowd worker’s prior knowledge as a basis for assignment logic To test the proposed system, we selected the argumentation task “Artificial Intelligence and Productivity” (4.3 section). We chose this topic because artificial intelligence is among today’s most hotly debated issues and because we believe the content did not require substantial specialized knowledge. However, some crowd workers were unfamiliar with the theme and may therefore have had difficulty performing the task. The impact of the prior knowledge of the crowd workers on the quality and experience (e.g., fatigue and ease) of the feedback needs to be investigated in detail. It would also be necessary for the system to implement allocation logic that considers the prior knowledge of the crowd workers in order to handle multiple subject fields of argumentation effectively.

7 Conclusion

We proposed a crowd worker-based method for generating argumentation improvement feedback to improve the persuasiveness of argumentation. The proposed method divides a complex argumentation improvement task into three types of micro-tasks: (1) classifying opinion vs. fact, (2) checking that evidence is given to support the facts, and (3) checking that a rationale is given to support the opinions. The results of user experiments showed that, writers who viewed the feedback in the system gave promising feedback, as measured by their increased motivation to persuade and their understanding of what to do next.

Acknowledgment

The work was supported in part by the Grants-in-Aid for Scientific Research (18H03244, 18H03554, 18H03775, 22H03905) from the MEXT of Japan.

References

1. Arnold, K.C., Chauncey, K., Gajos, K.Z.: Predictive text encourages predictable writing. In: Proceedings of the 25th International Conference on Intelligent User Interfaces. pp. 128–138 (2020)
2. Arnold, K.C., Gajos, K.Z., Kalai, A.T.: 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)
3. Bernstein, M.S., Little, G., Miller, R.C., Hartmann, B., Ackerman, M.S., Karger, D.R., Crowell, D., Panovich, K.: 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)
4. Black, P., Wiliam, D.: Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education)* **21**, 5–31 (2009)
5. Chen, M.X., Lee, B.N., Bansal, G., Cao, Y., Zhang, S., Lu, J., Tsay, J., Wang, Y., Dai, A.M., Chen, Z., 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)
6. Ferretti, R.P., Lewis, W.E., Andrews-Weckerly, S.: Do goals affect the structure of students’ argumentative writing strategies? *Journal of Educational Psychology* **101**(3), 577 (2009)
7. Hacker, D.J.: A metacognitive model of writing: An update from a developmental perspective. *Educational Psychologist* **53**(4), 220–237 (2018)
8. Hattie, J., Timperley, H.: The power of feedback. *Review of educational research* **77**(1), 81–112 (2007)
9. Jonassen, D.H., Kim, B.: Arguing to learn and learning to argue: Design justifications and guidelines. *Educational Technology Research and Development* **58**(4), 439–457 (2010)
10. Murukannaiah, P.K., Ajmeri, N., Singh, M.P.: Enhancing creativity as innovation via asynchronous crowdwork. In: 14th ACM Web Science Conference 2022. pp. 66–74 (2022)
11. Nickerson, R.S.: Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology* **2**(2), 175–220 (1998)
12. Salehi, N., Teevan, J., Iqbal, S., Kamar, E.: 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)
13. Toplak, M.E., Stanovich, K.E.: Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* **17**(7), 851–860 (2003)
14. Wambsganss, T., Kueng, T., Soellner, M., Leimeister, J.M.: 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)
15. Wambsganss, T., Niklaus, C., Cetto, M., Söllner, M., Handschuh, S., Leimeister, J.M.: 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)