Question-Answer Cards for an Inclusive Micro-tasking Framework for the Elderly

Masatomo Kobayashi, Tatsuya Ishihara, Akihiro Kosugi, Hironobu Takagi, and Chieko Asakawa

IBM Research – Tokyo, 5-6-52 Toyosu, Koto, Tokyo 135-8511, Japan
{mstm,tisihara,a1kosugi,takagih,chie}@jp.ibm.com

Abstract. Micro-tasking (e.g., crowdsourcing) has the potential to help "long-tail" senior workers utilize their knowledge and experience to contribute to their communities. However, their limited ICT skills and their concerns about new technologies can prevent them from participating in emerging work scenarios. We have devised a question-answer card interface to allow the elderly to participate in micro-tasks with minimal ICT skills and learning efforts. Our survey identified a need for skill-based task recommendations, so we also added a probabilistic skill assessment model based on the results of the micro-tasks. We also discuss some scenarios to exploit the question-answer card framework to create new work opportunities for senior citizens. Our experiments showed that untrained seniors performed the micro-tasks effectively with our interface in both controlled and realistic conditions, and the differences in their skills were reliably assessed.

Keywords: Micro-Tasks, Gamification, Skill Assessment, Ageing, Elderly, Senior Workforce.

1 Introduction

In many developed societies, increases in the elderly population and declines in the working-age population are serious social issues. For example, in Japan, which has the oldest population in the world, 24.6% of the population is over 65 as of March 2013 [1]. The retirement of baby-boomers is expected to reduce productivity while increasing the demand for healthcare and economic support. To address these trends, the senior workforce could be more utilized and their participation in work can help them maintain their independent lives. It is known that work benefits people in many ways, for example by reducing the risk of death while improving their well-being [2].

However, the age-related loss of physical, sensory, and cognitive abilities and declining health may prevent senior citizens from participating in full-time jobs. Only a few exceptional seniors are highly active, even though many senior citizens have a desire to work and contribute to their society [3]. As a result, a considerable fraction of the "long-tail" elderly workforce and their valuable knowledge and experience remain unutilized. They need more flexible work opportunities, so they can participate in part-time jobs whenever they want to and at their preferred locations [4].

P. Kotzé et al. (Eds.): INTERACT 2013, Part III, LNCS 8119, pp. 590–607, 2013.

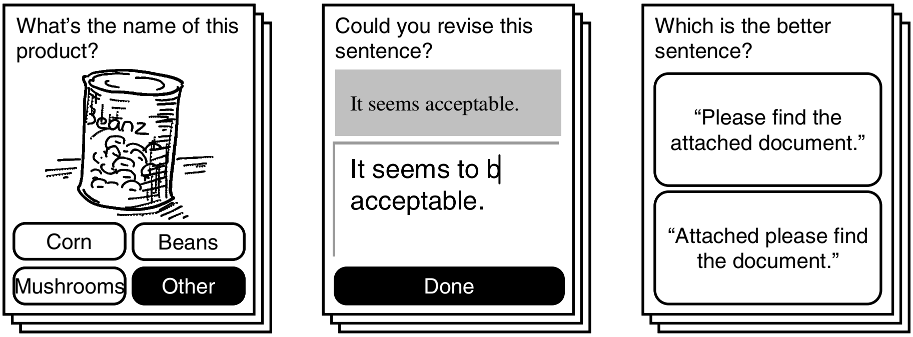
The key to success is how to exploit their abilities that have not declined with age (e.g., linguistic knowledge [2]) while avoiding any risks related to declining abilities.

The information-communication technology (ICT) is a key to such flexible work. Online communication tools allow distant collaborations and outsourcing to remote freelancers. Internet-connected mobile devices remove restrictions on workplaces. Among such ICT capabilities, this study particularly focuses on *micro-tasking* as used in crowdsourcing. It is one of the most suitable ways to work with a long-tail workforce, allowing occasional participation in small tasks at home without requirements for continuous commitment. The examples of micro-tasking range from simple, labor-intensive tasks to complicated, intellectual tasks as sampled in Section 2. Statistics show that the number of crowdsourcing workers has grown by more than 100% each year, reaching 6.3 million as of 2011 [5].

However, most senior citizens are currently excluded from micro-tasking. For example, the participants in the Amazon Mechanical Turk (MTurk) are mainly younger workers [6]. A small study we conducted with local senior citizens in conjunction with [7] found that none of them had ever participated in micro-tasking, such as contributing to MTurk, Wikipedia, or Q&A services, even when they had Internet connections. The literature suggests two reasons: limited ICT skills and concerns about new technologies. Although Internet-connected personal computers (PCs) and mobile devices have become increasingly popular even among senior populations, the elderly are often regarded as passive users of ICT, whose uses of technologies tend to be limited and to favor existing functions [8][9][10].

Our work aims to create an encouraging framework to accelerate elderly participation in micro-tasking. The primary target users involve senior citizens who have ICT devices such as smartphones or PCs but who do not actively use such technologies, who appear to be the majority of the senior population, at least over the next decade. First, we conducted a questionnaire-based survey to identify the elderly's requirements for the framework and to investigate their current usage of ICT. Next, based on the survey results, we introduce two technical foci of the framework: an interface design that minimizes the effort in learning; and a skill assessment mechanism that allows skill-based task recommendation. The interface uses the metaphor of stacked question-answer cards to allow senior workers to participate in micro-tasks with guided, simple, and consistent interactions (Fig. 1). The skill assessment uses a probabilistic model to calculate the skill level based on the results of micro-tasks.

Here is the structure of this paper. In the next section, we discuss related work. We then describe the concept of our inclusive micro-tasking framework, followed by a summary of the questionnaires, descriptions of our interface design and skill assessment mechanism, and some practical scenarios for our framework. Next we describe the prototype implementation. In the first experiment that involved senior and young participants, we tested the usability of our interface and the feasibility of skill assessment with touchscreen tablets. The second experiment examined the framework with home PCs. We conclude by discussing the implications of the experiments.


What's the name of this product?
Corn
Beans
Mushrooms
Other
Could you revise this sentence?
It seems acceptable.
It seems to b acceptable.
Done
Which is the better sentence?
"Please find the attached document."
"Attached please find the document."

(a) (b) (c)

Fig. 1. Examples of micro-task cards, which ask (a) for a product's name (to help a person who is blind), (b) to revise a sentence, and (c) to choose the better sentence

2 Related Work

Micro-Tasking. Crowdsourcing, where workers perform small pieces of work called micro-tasks via the Internet, has been used in many ways such as image recognition to help people who are blind [11], character recognition [12], translation [13], and even user studies [14]. Crowdsourcing is typically regarded as a means to easily access inexpensive labor, but some approaches such as VizWiz [11] have been proposed to exploit social networks for more sophisticated participation. Also, crowdsourcing is useful for enriching the social network information itself. Guy et al. showed crowdsourcing can be used to gather tags for people and social networks [15]. They showed that micro-tasks that ask about social networks can help update obsolete information in social applications.

Crowdsourcing was originally used for solving simple tasks but has expanded into more complex tasks. Kulkarni et al. proposed a framework to decompose complex tasks into micro-tasks [16]. They showed that even the work of decomposing the complex tasks into simple tasks could be crowd-sourced. Soylent decomposes crowd-sourced text-editing into three steps: find, fix, and verify [17]. Noronha et al. showed that nutritional analyses can be crowdsourced by decomposing the tasks into several sub-tasks [18]. We assume that some of decomposed micro-tasks may require simpler ICT skills while still benefiting from the wisdom of age, and are thus especially suitable for senior workers.

Senior-Friendly Systems. What assists and hinders the use of ICT by the elderly? The user interface is a critical factor when designing a system for senior citizens. Leonardi et al. found that seniors preferred intuitive touchscreen interfaces over keyboards [19]. Kobayashi et al. reported that mobile touchscreen interactions are "enjoyable" for seniors and their skills with such interfaces improve rapidly [20]. Those studies motivated us to include touchscreen devices in the primary targets of our framework. The user interface needs of senior citizens have also been discussed in Web accessibility research [21]. There are no standard accessibility guidelines specific to the elderly. However, it is known that seniors have requirements that in many ways are similar to those of people with disabilities, since their physical, sensory, and cognitive abilities tend to decline as they age.

The Office for National Statistics in the U.K. reported that typical senior citizens use their mobile phones only for phone calls and text messaging and avoid using Web browsers and other applications [8]. Kurniawan concluded that senior citizens tend to be fearful of unfamiliar technologies [9]. She also found that seniors are passive users of mobile phones, which means they mostly receive phone calls and text messages rather than originating them. Leung et al. reported that the preference to use trial-and-error decreases with age and senior citizens prefer to have instructions [10]. Hiyama et al. used question-answer interactions through phone calls and text messages as an alternative means for the elderly to participate in social media [7]. Those studies led us to exploit passive interaction styles that can be used without trial-and-error.

Handling Worker Skills. A major problem of crowdsourcing is the difficulty of assuring the quality of the results. Since the tasks are distributed to many people, there are wide variations in workers' skills and risks of vandalism. There are several ways to remove low quality results. CrowdFlower [22], one of the major crowdsourcing services, provides a function to filter noisy workers by distributing some tasks whose ground truths are known. By checking the results of these tasks, the workers who do not have skills can be filtered from the results. Another approach to assure the quality is aggregating multiple workers' output while assessing the skill of each worker and the difficulty of each task at the same time [23]. This approach assumes a correct answer exists for each task and the more skilled workers will agree on the correct answer.

Heimerl et al. showed micro-tasks requiring expertise could be efficiently done by distributing tasks to skilled workers [24]. How to search for the experts is an active area in information retrieval research. For example, Macdonald et al. proposed a method that finds experts by using multiple sources (such as resumes and webpages) to build skill profiles [25]. Guy et al. showed that the content in social applications could help find people related to specific keywords [26]. These projects assumed that the candidates actively used social tools in their daily lives, an assumption that is often strained for senior citizens. As already noted, the elderly tend to be passive users of these technologies. Also, they tend to hesitate in sharing their personal information due to privacy concerns [27]. Thus we decided to devise a new approach to obtain the skill profiles, an approach based on analyzing the results of the micro-tasks.

3 Inclusive Micro-tasking Framework

This section describes our approach for elderly micro-tasking. As noted in Section 2, various types of tasks can be decomposed into micro-tasks and many of them are expected to be suitable for senior workers. For example, among the three steps (find, fix, and verify) in crowdsourced text-editing [17], “verify” seems to be the most suitable task for the elderly. It requires less ICT skills (i.e., simply vote to approve work), but benefits from the wisdom of age (i.e., linguistic knowledge).

In many workplaces it is known that older and younger workers have complementary knowledge, skills, and abilities [28], so younger workers can assist older workers with their ICT skills while the older workers may assist the younger workers with their vocational knowledge. If this type of collaboration is supported in micro-tasking, then elderly participation would be encouraged. Since the people with ICT skills can already use standard desktop interfaces, what we need is an easier gateway for senior citizens who are excluded from micro-tasking opportunities.

3.1 User Requirements Survey

A survey using a paper questionnaire was conducted to identify the elderly's requirements for micro-tasking framework. We gave the questionnaire to 179 senior citizens in a suburban city and 170 of them responded (for a response rate of 95%). The respondents consisted of 99 males and 71 females, ranging in age from 54 to 84 (*mean*=67.7, *SD*=5.3). This survey was conducted as a part of a larger survey regarding ICT and work.

For the usage of ICT, 78% were PC users, with 60% using a PC every day and 88% were mobile phone users, with 61% every-day users. Although most respondents were frequent users of PCs and mobile phones, their usage was limited. The tools they frequently used included word processers, spreadsheets, e-mail, and Web search engines. Meanwhile, they rarely used modern ICT tools such as video chat or a social networking service (SNS). These results indicate they are using specific applications similar to those they used before they retired from work, confirming previous findings that senior citizens tend to avoid active use of new technologies [9]. This suggests that, although the majority of senior citizens now use ICT, we cannot optimistically assume that they will start participating in micro-tasking without additional mechanisms to increase the elderly's easy access to micro-tasking opportunities.

As for requirements for micro-tasking, the majority of the respondents specified: detailed instructions prior to participation (72%) and task recommendations based on their skills and interests (63%). This also confirms previous findings regarding the preference for instructions and the passive usage of ICT [10]. Based on the results, we defined two technical requirements for an inclusive micro-tasking framework. First, the interface should be simple and versatile. The simplicity makes it easy to learn, while the versatility provides consistent interactions for various types of tasks. Second, the system should have skill profiles for task recommendations. As mentioned in Section 2, although skill profiles can be extracted from resumes, webpages, and social tools, these approaches tend not to work well for senior workers. We need a mechanism to assess workers' skills without explicit skill-related sources.

Finally, regarding their interests in micro-tasks, the respondents were particularly interested in the tasks that require linguistic abilities in their native language and tasks that help people with disabilities.

3.2 Question-Answer Cards

To make the interface simple and versatile, we use the metaphor of stacked cards. Each card shows a question that represents a micro-task. Workers perform the task by answering the question. As shown in Fig. 1, the question can be textual or graphical. The answer may be input as free-form text or from multiple choices. Once a task was completed, the card slides out of the screen and the next card appears. This allows the worker to perform multiple tasks in a passive manner without actively searching the task pool for questions to answer. Since a typical micro-task naturally consists of a simple request and response, this fits well with the question-answer interface.

For example, Fig. 1-a shows a micro-task that asks about the name of a product shown in a picture, which simulates a task to help people who are blind [11]. The worker is asked to choose the correct answer by selecting from buttons below the image, where the candidates were generated by a package recognition engine such as used in Yeh et al. [29].

The support of multiple choices is important to allow performing micro-tasks without text entry. It is known that senior people have trouble in using the keypad on mobile phones, particularly on a touchscreen [20]. The problem is more serious in East Asia, where frequent mode-switching is required to access thousands of characters. For tasks involving a question that is by its nature open, the candidate answers need to be generated in advance, so that senior workers can contribute to final decisions based on their deep knowledge in a particular domain. The candidates may be generated by programs (such as the recognition engine in the example above), or by other (possibly less reliable) workers who use a free-form text interface.

3.3 Skill Assessment

To obtain workers' skill profiles without explicit sources, we use a mechanism to estimate the skills from the micro-task results based on a probabilistic model introduced by Whitehill et al. [23]. This mechanism measures a worker's skill by aggregating the results of multiple-choice tasks. The probabilistic approach is needed because, in realistic micro-tasks, the correct answer is often unknown and it is impractical to provide all of the workers with the same sets of tasks to compare their skills. We extended the model in [23], which considers only the completed tasks, to use both completed and uncompleted tasks. This extension is needed because we aim to assess the skills for each worker, whereas the original work aimed to assess the ground truth for each task. Once the skill level is estimated for each task type, the system can recommend appropriate tasks for each worker.

The probability of worker i giving a correct answer for task j is represented as:

$$p(L^c_{ij} = k \mid z^c_j = k, \alpha^c_i, \beta^c_j) = \sigma(\alpha^c_i \cdot \beta^c_j),$$

where α^c_i is the skill of worker i in task type c, β^c_j is the difficulty of task j in c, L^c_{ij} is the answer by i for j in c,σ is a logistic function, and z^c_j represents the true label that cannot be known directly. The probability is always set to 0 when i has not yet completed j. The values of α^c_i and β^c_j are estimated by an Expectation-Maximization (EM) approach. A larger α_i^c means the worker has a higher skill while a larger β_j^c means the task is easier. If a worker has a higher probability of giving correct answers in a task type, then that worker is believed to have a higher skill for that type.

3.4 Elderly Micro-tasking Scenarios

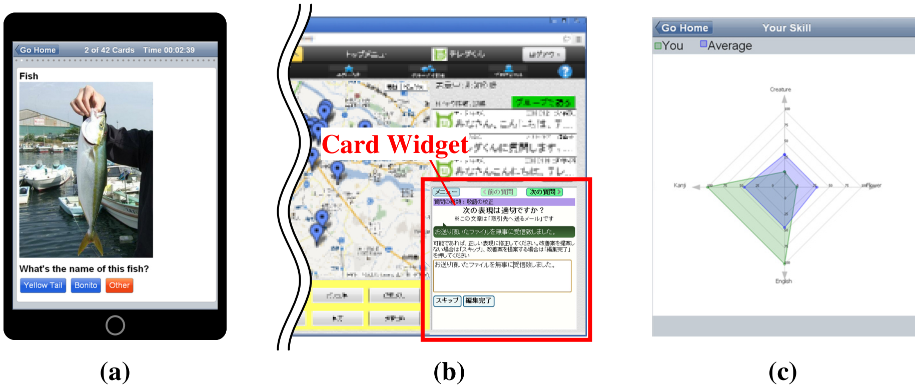
In addition to micro-tasks such as those sampled in Section 2, there are other types of promising scenarios where our framework could be used for elderly micro-tasking. Since the knowledge of the elderly will be most effective in various types of advisory tasks, we believe that the question-answer metaphor is particularly suitable for the senior workforce.

- *Decision Making*: Based on their extensive experience, senior citizens could give advice on difficult topics such as careers and business options. Our card-style interface can be used both for collecting candidate answers and for voting to choose the best answers.
- *Authoring Support*: Senior citizens could contribute to the creation of educational material by using their expert knowledge and linguistic skills. The types of contributions may include translation, proofreading, and other supportive roles as well as writing new material.
- *Crowd Accessibility*: People with disabilities could benefit from working with remote senior workers in many ways. In addition to real-time assistance such as VizWiz [11], examples include captions and audio descriptions for videos, and making accessible books and websites.
- *Product Evaluation*: It is difficult for product designers to meet the usability and accessibility requirements of senior consumers. For example, text information such as menus and instructions must be easy to understand. Cards can show prototype interfaces for products and senior citizens can evaluate them.
- *Marketing Research*: Because of the rapid growth of the senior market, consumer products and service companies are reaching out to the senior population. In the process of developing products and services, enterprises need to research market acceptance and the card-style interface can be a mechanism for such surveys.

In many of these scenarios, the collaboration between the people who create candidate answers and those who make final decisions is a focus. Our framework provides interfaces for those who have weaker ICT skills or who use mobile devices, thus allowing senior citizens to participate in appropriate roles, whenever they want to and at their preferred locations.

4 Prototype Implementation

We implemented a prototype based on this framework as a Web application with support for various types of devices. The user interface written in JavaScript works on the client side while the skill assessment engine written in Java runs on the server side. The Web-based model allows using a standard server-side program for multiple



What's the name of this fish?

質問の種類：敬語の校正

次の表現は適切ですか？

※この文章は「取引先へ送るメール」です

お送り頂いたファイルを無事に受信致しました。

可能であれば、正しい表現に修正してください。改善案を提案しない場合は「スキップ」、改善案を提案する場合は「編集完了」を押してください

お送り頂いたファイルを無事に受信致しました。

(a) (b) (c)

Fig. 2. (a) A card on a mobile touchscreen device, (b) A card seen as a widget in a desktop browser, and (c) Estimated skill visualization

types of micro-tasks, while sharing much of the client-side code for multiple devices. We currently have two variations of the implementation: one for mobile touchscreens and one for desktop browsers.

Fig. 2-a shows a card on a mobile touchscreen device. The card is asking for the name of the object in the picture, which represents a micro-task similar to [11]. The card can be completed by tapping an answer from several options. Once an answer is selected, the next card will appear on the screen with an animation. The worker can also skip cards or go back to previous cards by dragging across the card to the left or right, respectively. Fig. 2-b shows a card in a desktop Web browser. The card is embedded in an SNS interface as a widget, asking to revise a sentence which represents a micro-task similar to the "fix" task in crowdsourced text-editing [17]. The card can be completed by editing the text and clicking on the submit button. The basic design is to the same for the mobile and desktop, except that the desktop interface has "skip" and "back" buttons on top of the card.

The two targets, mobile touchscreens and desktop widgets, were chosen to increase the exposure of senior citizens to micro-tasks. The mobile touchscreen device is known to be easy to use and preferred by the elderly [20]. Although smartphone users are still a minority in the elderly population at this time, it is expected to be an increasingly common device in the next decade. This kind of devices is usually kept switched on and carried on all the day, so it will effectively increase their exposure. The desktop widget can be embedded in any existing interfaces that support HTML-based widgets. The exposure will increase by putting the widget in an interface the users frequently access, such as SNS websites. Although SNS websites are still unfamiliar to most senior citizens, they are becoming popular with them.

Fig. 2-c shows a view that visualizes the estimated skill levels, which we developed to motivate the workers. The skill values for each task type are normalized from 0 to 100 and presented in a radar chart. The average values of all of the workers are also displayed in the chart.

5 Experiment 1: Controlled Study

Our first experiment sought to test the usability of the card-style interface on a touchscreen device and the feasibility of skill assessment, tested in a laboratory setting with simple micro-tasks.

5.1 Settings

Participants. A total of 15 seniors in their 60s to 80s (*mean*=70.1, *SD*=6.8) participated in the experiment. We also involved 16 younger people in their 20s to 40s (*mean*=33.5, *SD*=3.6) as a baseline. A total of 8 of 15 seniors and 14 of 16 younger people had experience with mobile touchscreen devices before the experiment. The senior participants were recruited locally while the younger participants were employees in a global technology company. All of the participants were native Japanese speakers, and the experiment was conducted in Japanese.

Apparatus. The participants used an iPad. As senior citizens rarely have their own tablet devices at this time, the devices were issued by the experimenters and the experiment was conducted in a laboratory. We used the touchscreen interface presented in Fig. 2-a.

Target Micro-Tasks. We tested four types of micro-tasks asking about domain-specific knowledge: Fish, Flowers, Kanji, and English. Fish and Flowers asked for the name of a pictured fish or flower, which simulated crowdsourced image recognition [11]. Kanji asked for the readings of Chinese characters, simulating part of crowdsourced proofreading of digital books [30]. English asked for the meanings of English words, simulating part of crowdsourced translation [13]. We hypothesized that the senior participants would be better at Fish, Flowers, and Kanji since they have deep knowledge about natural things and their native language from their long experience. We also hypothesized that the younger participants would be better at English since they were working at a global company.

The Fish and Flowers types used visual questions while the Kanji and English types used text questions. For all of the four types, we tested multiple-choice answers in the kind of interface shown in Fig. 1-a, where the candidate answers were provided in advance by software or other people. The free-form text input was omitted in this experiment because it is known that text entry on a mobile device is troublesome for the elderly as described in Section 3.2, and our skill assessment model is suitable for multiple-choice tasks as described in Section 3.3. For each of the questions, the experimenters knew the correct answer.

Procedure. Each participant saw 20 different cards for each task type. Thus they each processed (completed or skipped) a total of 80 cards. The presentation order of the cards was balanced. Prior to the tasks, the participants were given a few minutes of

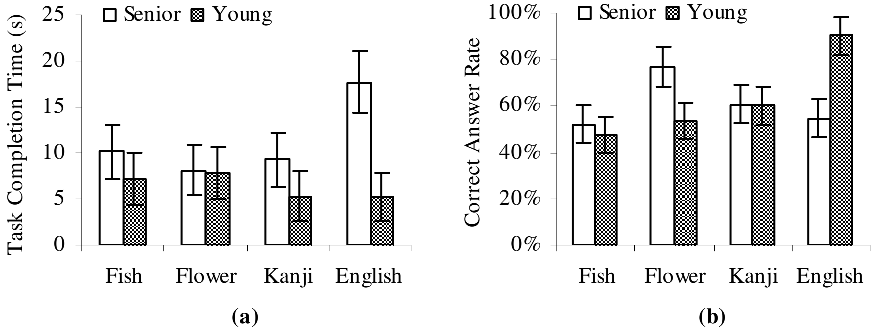


Fig. 3. (a) Task completion times, (b) Actual correct answer rates, by task type with 95% CIs

instructions on how to perform the tasks. After the participants completed all of the micro-tasks, they then answered several multiple-choice questions that asked for feedback on our framework. Each experimental session took less than 30 minutes for most of the participants.

5.2 Results

Task Completion Times. Fig. 3-a shows the average time to complete one task for each task type (skipped cards were excluded from these calculations). We used this measurement to examine whether untrained users could use the card-style interface without confusion. For the senior participants, the overall average values were 10.2, 8.1, 9.3, and 17.7 seconds for Fish, Flowers, Kanji, and English, respectively. For the young participants, the values were 7.2, 7.7, 5.2, and 5.2 seconds. Analysis of variance showed significant main effects for the age ($F_{1,30.54}$=9.59, p<.005) and task type ($F_{3,2162}$=2.96, p<.05). There are also significant interaction effects between the age and task type ($F_{3,2162}$=5.96, p<.001). A post-hoc analysis found that the senior participants were significantly slower than the young participants for Kanji (p<.05) and English (p<.001). There were no significant differences for Fish and Flowers. This indicates that seniors tend to take longer to process text questions, but they are as fast as younger people at visual questions.

Correct Answer Rates. Fig. 3-b shows the percentages of questions answered correctly for each task type. We used this measurement to assess the actual skill levels, which were calculated using knowledge of the correct answers, as a baseline to test the accuracy of our skill estimation mechanism. For the senior participants, the values were 52%, 77%, 61%, and 55% for Fish, Flowers, Kanji, and English, respectively. For the young participants, the values were 48%, 53%, 60%, and 90%. Analysis of variance showed a significant main effect for the task type ($F_{3,87}$=10.64, p<.001). It also showed a significant interaction effect between the age and the task type ($F_{3,87}$=17.77, p<.001). A post-hoc analysis found that the senior participants were significantly better at Flowers than the young participants (p<.001). For English, the young participants significantly outperformed the senior participants (p<.001). The age did not significantly affect the results for Fish and Kanji.

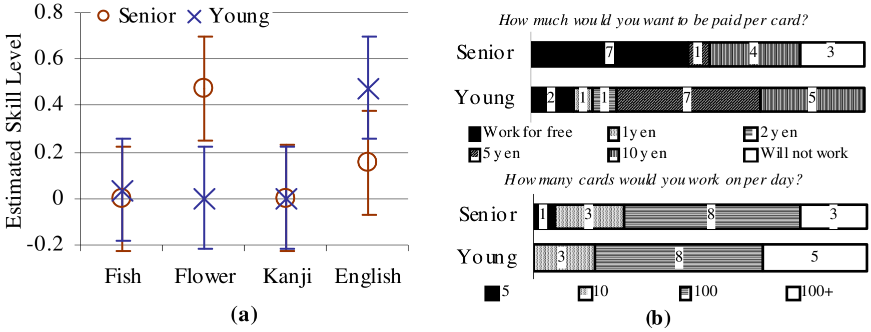


Fig. 4. (a) Estimated skills for each task type with 95% CIs, (b) Results of post-experiment questionnaires (100 yen ≈ 1 USD)

Estimated Skill Levels. To confirm that the skill estimation mechanism works, we compared the participants' estimated skills with their actual correct answer rates. Note that the actual rates are often unknown in realistic micro-tasking situations because the ground truths for the micro-tasks might also be unknown.

Fig. 4-a compares the average estimated skill levels for each task type, which were calculated without using knowledge of the correct answers, as described in Section 3.3. For the senior participants, the values were 2.0×10^{-4}, 4.7×10^{-1}, 1.2×10^{-3}, and 1.5×10^{-1} for Fish, Flowers, Kanji, and English, respectively. For the young participants, the values were 3.4×10^{-2}, 1.5×10^{-3}, 3.8×10^{-4}, and 4.8×10^{-1}. Analysis of variance showed there was a significant main effect of the task type ($F_{3,87}$=3.94, p<.05). There was also a significant interaction effect between the age and task type ($F_{3,87}$=4.29, p<.01). A post-hoc analysis found that the senior participants had significantly higher skills in Flowers (p<.005) while the young participants had significantly higher skills in English (p<.05). For Fish and Kanji, no significant differences were found. This result matches up well with the actual correct answer rates. This confirms that the differences in skills, at least between the senior and young groups, were correctly detected from the results of the micro-tasks. This indicates that the same mechanism could effectively assess the individual workers. Note that, since the skill levels produced by the probabilistic model are relative values compared to other workers, the absolute values of the skill levels are not known.

The probabilistic model estimates the correct answer by aggregating the answers while considering the estimated skills of the workers. The correct answer rates of the aggregated answers are much higher than the average rates of the individual workers for all of the task types (English: 95.0%, Flowers: 85.0%, Kanji: 85.0%, Fish: 80.0%). This also indicates that the skill levels are estimated successfully.

As explained in the Section 3.3, we extended the probabilistic model proposed by Whitehill et al. [23] to support the cases in which workers did not complete, e.g., skipped, some cards. To confirm this extension works, we compared the estimated skills in English for four participants. They completed 20, 11, 9, and 0 English cards, respectively (the first participant skipped no English card while the last one skipped all of the English cards). They did not produce any incorrect answers for the cards they completed. Their estimated English skills were 1.8, 1.9 × 10^{-4}, 1.2 × 10^{-4}, −2.5 × 10^{-4}. The more cards skipped, the lower the estimated skill, as expected.

Subjective Feedback. First we asked "*How much would you want to be paid for each card if you were doing this work as a job?*" The answers were clearly different between the senior and young groups. More people in the senior group said they were willing to do this work even for free. At the same time, more of the seniors answered they would not work on cards regardless of compensation. This result suggests that seniors were less concerned about the rewards for the micro-tasks.

We then asked "*How many cards would you work on in one day?*" For this question, about half of the participants answered they could do about 100 cards in one day regardless of age. Finally we asked "*Was the system enjoyable?*" The majority of participants said that they enjoyed it. In informal post-experiment interviews, several participants commented that they felt the stacked card interface was enjoyable and game-like. The results for the first two questions are shown in Fig. 4-b.

6 Experiment 2: Live Study

The second experiment sought to test the card-style interface and the skill estimation mechanism in a realistic Web-based setting that involved more practical micro-tasks that would benefit more from the wisdom of age.

6.1 Settings

Participants. A total of 28 senior citizens participated in at least one task, with 9 answering the post-experiment questionnaires. For reference, 13 younger people performed at least one task, with 8 answering the questionnaires. The senior participants were members of an experimental local SNS for senior citizens, with no overlap with the participants in the first experiment, but there may have been some overlap with the respondents from the anonymous preliminary survey described in Section 3.1. The younger participants were employees in a global technology company, including four of the earlier participants. None of the participants who answered the questionnaires had prior experience with crowdsourcing, except for one member of the young group. The experiment was conducted in Japanese.

Apparatus. The participants used their own desktop or laptop PC at home or in their workplaces. We used the widget interface shown in Fig. 2-b. For the senior group, the widget was embedded in the portal of the local SNS. For the young group, the widget was embedded in an enterprise SNS for the employees.

Target Micro-Tasks. We tested two types of micro-tasks for proofreading, which is similar to the "fix" and "verify" tasks in Soylent [17]. We chose these tasks since linguistic knowledge is known to be an advantage of the elderly [2]. Also, our preliminary survey (Section 3.1) suggested that senior citizens are motivated to use their native language skills. The Fix task asked participants to revise a sentence to make it more polite and correct. It first asked whether or not the sentence required any revision, and if the participant answered “yes”, then the system asked for a free-form text answer via a card similar to Fig. 1-b. The Verify task asked the participant to choose the more polite and correct sentence from two candidates. It asked for a multiple-choice answer via a card similar to Fig. 1-c. For both types of task, a brief context (e.g., the person who will use the sentence) was included on the card as a note. The scenario assumed that at first some participants would create candidate revisions through the Fix interface, and then other people would make the final decisions through the Verify interface. This experiment is intended to test realistic micro-tasks and the sentences to be proofread were prepared based on actual problems and the correct answers were unknown for each task.

Procedure. Each participant was presented up to 45 and 20 different cards for Fix and Verify tasks, respectively. Thus they completed up to 65 cards at their own pace. The presentation order of the cards was randomized. Prior to the tasks, the senior participants were given a brief explanation of the purpose of this study and instructions on how to perform the tasks. The young participants received an introductory announcement via email or a blog post. After the experiment, they were asked for feedback as in the first experiment. The second experiment ran for approximately one week.

6.2 Results

Participation Status. During the experimental period, a total of 42 senior users were shown one or more cards (counted based on the user ID), and 28 of them completed 906 cards in total (*mean*=32.4, *SD*=27.4). A total of 387 young users were shown cards (counted based on their IP addresses), and 13 of them completed 104 cards (*mean*=8.0, *SD*=11.0). Since 12 of the seniors completed all of the cards for at least in one task type, it seems likely they would have performed more tasks if they were available. None of the young group completed all of the tasks. This might suggest that the card-style interface motivated senior people more than young people, but the relevant experimental conditions were not the same for each group.

Task Completion Times. Fig. 5-a shows the average time to complete one task for each task type (skipped cards and neglected cards, which we defined as cards that took more than 5 minutes to respond to, were excluded from the calculations). For the senior participants, the overall average values were 82.4, 27.5, and 22.1 seconds for Fix-Yes, where they actually fixed the sentence with the free-form text entry card, Fix-No, where they stated that the sentence did not require any revision, and Verify, respectively. For the young participants, the values were 77.9, 45.3, and 29.3 seconds. Note that the value for each trial may be affected by the network delay. Analysis of variance showed significant main effect for the task type ($F_{2,949}$=57.93, p<.001). The age had no significant effect. A post-hoc analysis found that Fix-Yes tasks took significantly longer time than Fix-No and Verify tasks (p<.001).

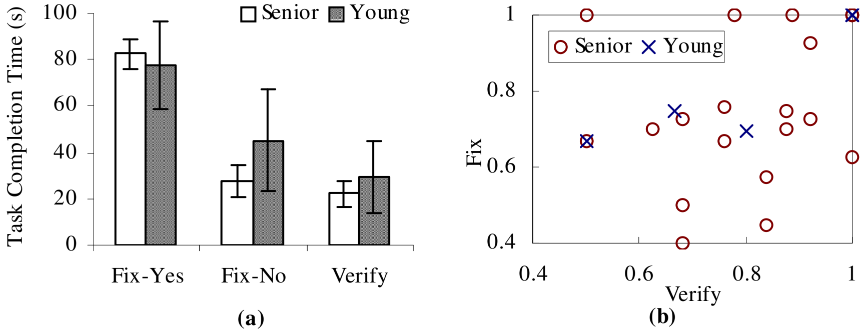


Fig. 5. (a) Task completion times for each task type with 95% CIs, (b) Distribution of estimated correct answer rates for each task type

Estimated Skill Levels. For the senior participants, the average values of the estimated skill levels were 0.11 and 0.23 for Fix and Verify, respectively. For the young participants, the values were 0.29 and 0.58. Note that the skill levels for Fix were calculated based on the results of the yes-no questions that asked the necessity for revisions. The actual content of revised sentences was ignored in this calculation due to a limitation of our skill assessment model, which cannot work well with open questions.

Analysis of variance showed significant main effect for the age ($F_{1,44.77}$=5.19, p<.05) and task type ($F_{1,31}$=5.18, p<.05). Despite the assumption that senior people would be better at linguistic skills, the young group showed higher skills in this experiment, perhaps because all of the young participants were highly educated and can be expected to have better-than-average linguistic skills. Meanwhile, the effect of the task type seems to be caused by the difference in the percentage of the "skipped" trials (17.1% for Fix, 9.3% for Verify). As confirmed in Section 5.2, skipped trials make the estimated skill lower in our skill assessment model.

Fig. 5-b shows the distribution of estimated correct answer rates, calculated by assuming that the aggregated answers are always correct. This indicates that the skill for Fix is not always positively correlated with the skill for Verify (R^2=.18) and the values are vary widely depending on the worker (SD=.28 for Fix, SD=.18 for Verify). This result suggests that we have to carefully choose skilled workers for each task to insure high-quality results.

Subjective Feedback. We asked similar questions as in the first experiment. Although only some of the participants answered the questions (since this experiment was conducted in a real-world setting), the results seemed to generally confirm the previous results. For the question asking about the compensation, the majority of the senior participants were less concerned about the rewards. The young participants tended to want 5 to 10 yen per card. To the question asking for a number, the participants generally responded that they could do around 100 cards per day. In contrast to the first experiment, many of the participants stated that they would not work on Fix and Verify cards regardless of compensation (50% for seniors, 38% for youths). This might indicate that realistic micro-tasks can involve some unenjoyable work, and thus some additional motivational mechanism would be needed.

7 Discussion

Based on the experimental results, we discuss the challenges and future work to invite senior citizens to do micro-tasks.

7.1 Interface Usability and Acceptability

Overall the results indicated that our approach was adequately usable for and accepted by the target users. The senior participants could understand how to perform all of the types of tasks and complete them with only brief instructions. We also confirmed that the proposed framework worked even in a live situation that involves realistic tasks. However the experimental results also indicated that the performance largely depends on the users and the task types. Since we did not control for the participants' ICT experiences or age-related losses of abilities, further research is needed to discuss the realistic usability and accessibility of our interface for specific users and tasks.

We also need to address the range of applications, since the card-style interface basically supports only simple question-answer interactions. Task-decomposition strategies such as discussed by Bernstein et al. [17] are promising ways to handle more complex tasks. They tried to assure the quality of their results by using a voting mechanism. As indicated in our second experiment, the mixed use of free-form text and multiple choice cards can support such an answer-vote strategy. Once the task is split into micro-tasks, the question-answer interactions will work well because each micro-task is typically represented as a pair with a simple request and its response. In such cases, multiple choice cards may be preferred and more effective than free-form text cards, i.e., with fewer skips and quicker responses. Some participants suggested that it would be helpful if voice input is available. As we used Web-based implementations, the voice input functions supported by modern Web standards are available.

7.2 Skill-Based Task Recommendations

Since the elderly are regarded as passive users of ICT, they will not actively explore a pool of tasks to find tasks they like, so our system should offer to the senior workers tasks that fit their skills. The experimental results confirmed that our skill assessment model could assess the differences in skills. This means that once a worker answered several cards for one type of task, the system could assess whether or not the task type is good for that worker. For example, if Alice is better at "fix" and Bob is better at "verify", then the system would provide Alice and Bob with more "fix" and "verify" tasks, respectively. The second experiment indicated that the skills of the elderly vary widely. Even if a worker is better at a type of tasks, it does not necessarily mean the worker is also better at a related, but different type of tasks (e.g., "fix" and "verify" tasks). As the probabilistic approach can estimate workers' skills with a small number of samples, it would allow quickly determining the skills for each specific domain.

A collaborative filtering approach [31] can help recommend task types even when a new worker has never tried those tasks. Our skill estimation mechanism produces a skill profile matrix that can be used for this purpose, where the value of each element represents the skill level of a worker for a task type. We can use a similar approach to Yuen et al.'s work [32], which used a task preference matrix to recommend tasks based on a collaborative filtering mechanism.

7.3 Motivation Management

Our experimental results show that seniors did not care much about the monetary rewards. This indicates that we need special care in the design of the incentive mechanism for senior workers. For example, some participants commented that the skill visualization (Fig. 2-c) allowed them to discover that they had greater knowledge in a particular domain, and that motivated them. The motivation may also be affected by the enjoyability. It is worth noting that many of the participants found that the card-style interface was enjoyable and like a game even though we did not intend to include any gamification mechanisms in our framework. However, we also have to note that, in the second experiment, the participants gave less positive feedback than in the first experiment. At this time we cannot conclude whether this resulted from the differences in the participants' communities, in the interfaces (touchscreen vs. desktop widget), or in the task conditions (laboratory vs. realistic), which means further investigation is needed.

Another key to success would be increasing the exposure of senior citizens to micro-tasking opportunities. We believe that the mobile interface and widget interface will be effective for this purpose, as described in Section 4. Also, the card-style interface could be applied to other situations. Some candidate situations include smart TVs, kiosk terminals, vehicle navigation displays, in-flight entertainment systems, and even when waiting in front of a microwave [33].

8 Conclusion

This paper described a framework that aims to increase the participation of senior citizens in various types of micro-tasks with a simple and versatile card-style interface. It uses a probabilistic model to assess their skill profiles, which is necessary for task recommendations and for quality assessments and improvements. The experimental results showed that even untrained seniors could handle micro-tasks by using the card-style interface in both controlled and real-world settings. We also confirmed that the differences in skills can be assessed by analyzing the results of the micro-tasks. We still have challenges to deploy the framework as a platform for senior citizens who want to work. The scope of tasks should be broadened, but more importantly, enjoyment and motivational factors should be evaluated and improved as well as the economic efficiency. We hope that the proposed framework will help senior citizens participate in the workforce and fully utilize their broad experience and knowledge in the society.

Acknowledgments. This research was partially supported by the Japan Science and Technology Agency (JST) under the Strategic Promotion of Innovative Research and Development Program.

References

1. Statistics Bureau, Monthly Report, http://www.stat.go.jp/english/data/jinsui/tsuki/ (retrieved April 15, 2013)
2. Ouchi, Y., Akiyama, H. (eds.): Gerontology – Overview and Perspectives, 3rd edn. Univ. of Tokyo Press (2010) (in Japanese)
3. United Nations University, Active Ageing, http://wisdom.unu.edu/en/active-aging
4. Leibold, M., Voelpel, S.: Managing the Aging Workforce. John Wiley & Sons (2006)
5. Loten, A.: Small firms, Start-ups Drive Crowdsourcing Growth. Wall Street Journal (February 28, 2012)
6. Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., Tomlinson, B.: Who are the Crowdworkers? Shifting Demographics in Mechanical Turk. In: Proc. CHI EA 2010, pp. 2863–2872. ACM (2010)
7. Hiyama, A., Nagai, Y., Kobayashi, M., Takagi, H., Hirose, M.: Question First: Passive Interaction Model for Gathering Experience and Knowledge from the Elderly. In: Proc. PerCol 2013, pp. 151–156. IEEE (2013)
8. Office for National Statistics: Use of ICT at Home (2007)
9. Kurniawan, S.: Older People and Mobile Phones: A Multi-Method Investigation. Int. J. Hum.-Comput. Stud. 66(12), 889–901 (2008)
10. Leung, R., Tang, C., Haddad, S., McGrenere, J., Graf, P., Ingriany, V.: How Older Adults Learn to Use Mobile Devices: Survey and Field Investigations. ACM Trans. Access. Comput. 4(3), Article 11 (2012)
11. Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R.C., Miller, R., Tatarowicz, A., White, B., White, S., Yeh, T.: VizWiz: Nearly Real-Time Answers to Visual Questions. In: Proc. UIST 2010, pp. 333–342. ACM (2010)
12. von Ahn, L., Maurer, B., McMillen, C., Abraham, D., Blum, M.: reCAPTCHA: Human-based Character Recognition via Web Security Measures. Science 321(5895), 1465–1468 (2008)
13. Callison-Burch, C.: Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk. In: Proc. EMNLP 2009, pp. 286–295. ACL and AFNLP (2009)
14. Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing User Studies with Mechanical Turk. In: Proc. CHI 2008, pp. 453–456. ACM (2008)
15. Guy, I., Perer, A., Daniel, T., Greenshpan, O., Turbahn, I.: Guess Who? Enriching the Social Graph through a Crowdsourcing Game. In: Proc. CHI 2011, pp. 1373–1382. ACM (2011)
16. Kulkarni, A., Can, M., Hartmann, B.: Collaboratively Crowdsourcing Workflows with Turkomatic. In: Proc. CSCW 2012, pp. 1003–1012. ACM (2012)
17. 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: Proc. UIST 2010, pp. 313–322. ACM (2010)

18. Noronha, J., Hysen, E., Zhang, H., Gajos, K.Z.: PlateMate: Crowdsourcing Nutritional Analysis from Food Photographs. In: Proc. UIST 2011, pp. 1–12. ACM (2011)
19. Leonardi, C., Albertini, A., Pianesi, F., Zancanaro, M.: An Exploratory Study of a Touch-based Gestural Interface for Elderly. In: Proc. NordiCHI 2010, pp. 845–850. ACM (2010)
20. Kobayashi, M., Hiyama, A., Miura, T., Asakawa, C., Hirose, M., Ifukube, T.: Elderly User Evaluation of Mobile Touchscreen Interactions. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part I. LNCS, vol. 6946, pp. 83–99. Springer, Heidelberg (2011)
21. Web Accessibility and Older People: Meeting the Needs of Ageing Web Users, http://www.w3.org/WAI/older-users/
22. CrowdFlower, http://crowdflower.com/
23. Whitehill, J., Ruvolo, P., Wu, T., Bergsma, J., Movellan, J.: Whose Vote should Count More? Optimal Integration of Labels from Labelers of Unknown Expertise. In: Proc. NIPS 2009, pp. 2035–2043 (2009)
24. Heimerl, K., Gawalt, B., Chen, K., Parikh, T., Hartmann, B.: CommunitySourcing: Engaging Local Crowds to Perform Expert Work via Physical Kiosks. In: Proc. CHI 2012, pp. 1539–1548. ACM (2012)
25. Macdonald, C., Ounis, I.: Voting for Candidates: Adapting Data Fusion Techniques for an Expert Search Task. In: Proc. CIKM 2006, pp. 387–396. ACM (2006)
26. Guy, I., Jacovi, M., Shahar, E., Meshulam, N., Soroka, V., Farrell, S.: Harvesting with SONAR: The Value of Aggregating Social Network Information. In: Proc. CHI 2008, pp. 1017–1026. ACM (2008)
27. Pfeil, U., Arjan, R., Zaphiris, P.: Age Differences in Online Social Networking – A Study of User Profiles and the Social Capital Divide among Teenagers and Older Users in MySpace. Comput. Hum. Behav. 25(3), 643–654 (2009)
28. Staffing Mature Worker Survey, http://www.goldenworkers.org/images/publication/mature_workers_survey_2012_adecco.pdf
29. Yeh, M.-C., Tai, J.: A Hierarchical Approach to Practical Beverage Package Recognition. In: Ho, Y.-S. (ed.) PSIVT 2011, Part I. LNCS, vol. 7087, pp. 348–357. Springer, Heidelberg (2011)
30. Kobayashi, M., Ishihara, T., Itoko, T., Takagi, H., Asakawa, C.: Age-based Task Specialization for Crowdsourced Proofreading. In: Stephanidis, C., Antona, M. (eds.) UAHCI/HCII 2013, Part II. LNCS, vol. 8010, pp. 104–112. Springer, Heidelberg (2013)
31. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating Collaborative Filtering Recommender Systems. ACM Trans. Inf. Syst. 22(1), 5–53 (2004)
32. Yuen, M.-C., King, I., Leung, K.-S.: TaskRec: Probabilistic Matrix Factorization in Task Recommendation in Crowdsourcing Systems. In: Huang, T., Zeng, Z., Li, C., Leung, C.S. (eds.) ICONIP 2012, Part II. LNCS, vol. 7664, pp. 516–525. Springer, Heidelberg (2012)
33. Watanabe, K., Matsuda, S., Yasumura, M., Inami, M., Igarashi, T.: CastOven: A Microwave Oven with Just-in-Time Video Clips. In: Proc. Ubicomp 2010 Adjunct, pp. 385–386. ACM (2010)