Crowd-Powered Source Searching in Complex Environments

Yong Zhao[1], Zhengqiu Zhu[1], Bin Chen[1,2], and Sihang Qiu[1,2(✉)]

[1] National University of Defense Technology, Changsha, China
[2] Hunan Institute of Advanced Technology, Changsha, China
{zhaoyong15,zhuzhengqiu12,chenbin06,qiusihang11}@nudt.edu.cn

Abstract. Source searching algorithms are widely used in different domains and for various applications, for instance, to find gas or signal sources. As source searching algorithms advance, search problems need to be addressed in increasingly complex environments. Such environments could be high-dimensional and highly dynamic. Therefore, novel search algorithms have been designed, combining heuristic methods and intelligent optimization, to tackle search problems in large and complex search space. However, these intelligent search algorithms usually cannot guarantee completeness and optimality, and therefore commonly suffer from the problems such as local optimum. Recent studies have used crowd-powered systems to address the complex problems that machines cannot solve on their own. While leveraging human rationales in a computer system has been shown to be effective in making a system more reliable, whether using the power of the crowd can improve source searching algorithms remains unanswered. To this end, we propose a crowd-powered sourcing search approach, using human rationales as external supports to improve existing search algorithms, and meanwhile to minimize the human effort using machine predictions. Furthermore, we designed a prototype system, and carried out an experiment with 10 participants (4 experts and 6 non-experts). Quantitative and qualitative analysis showed that the sourcing search algorithm enhanced by crowd could achieve both high effectiveness and efficiency. Our work provides valuable insights in human-computer collaborative system design.

Keywords: Source searching · Crowd-powered system · Crowd computing.

1 Introduction

Source searching problems always exist in nature and in our daily lives, such as animals finding a odor source to acquire foods in the wild and people searching the emission source of air pollution. Traditional search algorithms, such as tree search algorithms and graph search algorithms, work well for limited search space given enough time. However, as technology advances and computing power explodes, people started to expect search algorithms to solve search problems in more complex scenes with a more strict time limit.

Complex search space could be high-dimensional and dynamic, and the computation might be required to be completed in real-time in many applications. Since traditional traverser algorithms can no longer meet those requirements, novel search algorithms have been proposed to tackle source searching problems in large and complex search space [31, 23]. These novel search algorithms integrate the human cognition and animal behaviors into the search rules, and gather information during the search process to dynamically adjust search parameters. While the search efficiency has been dramatically improved, the completeness and optimality can no longer be guaranteed. Therefore, it is possible that these search algorithms return a local optimum or even no solution, instead of the global optimal solution. Researchers and practitioners have noticed this issue [14].

Recent work has focused on crowd-powered systems and human-AI collaborations [1]. These approaches enable humans to take part in the automatic process that is supposed to be completely controlled by machines to solve complex problems. A crowd-powered system provides us with a new perspective that humans could get involved in the automatic problem-solving process, to enhance the effectiveness and efficiency of algorithms [15, 6, 7, 20]. Since human-machine collaboration has been proved to be feasible in a variety of domains, we see an opportunity that combines human rationales with search algorithms, to overcome the difficulties that current source searching approaches usually encounter. However, whether the power of the crowd is really effective and efficient in improving existing source searching algorithms remain unknown. To address this knowledge gap, in this work, we are particularly interested in answering the the research questions: *how can crowd-powered approaches improve the effectiveness and efficiency of existing source searching algorithms?*

To answer the research question, we designed a human-machine collaborative framework that could improve the existing source searching algorithm in a way that crowdsources the problems that occurred during source searching. We implemented a prototype system where a virtual robot autonomously searching a source in complex environments in a simulation setup, to enable a user study. The crowd-powered system is responsible for detecting fatal problems, explaining the algorithm, giving suggestions, and generating tasks for humans to complete. When it comes to humans' turns, humans could either take full control of the robot or aid the robot to address problems. Particularly, to better facilitate effective problem solving, the system predicts the location of the source using Bayesian methods and sequential Monte Carlo methods to further assist humans in making decisions and taking actions. We recruited 10 participants in this study, including 4 domain experts in the field of source searching, and 6 non-experts having no experience in source searching, to evaluate the proposed human-machine collaborative approach in randomly generated complex searching environments.

The experiment shows that the crowd-powered system could improve the performance of the state-of-the-art source searching algorithms, in both effectiveness (success rate 100%, 22% higher than none-human methods) and efficiency (using significantly fewer iterations/steps to find the source). Furthermore, we analyzed system usability scores and cognitive workload scores reported by the participant. Results show that a specific way of interaction could achieve better usability and cognitive workload for a group of participants in the prototype system. Our work provides useful suggestions, valuable insights, and important implications in leveraging human-machine collaboration for improving source searching algorithms.

2 Related Work

We discuss related literature from two perspectives: crowd-powered systems and source searching.

2.1 Crowd-Powered Systems

The goal of designing a crowd-powered system is to leverage human rationales in a way combing with computer systems to collaboratively solve complex problems. A typical crowd-powered system is the ESP game [38], which is an image labeling system developed by Google. This system successfully gamified the image labeling process, and produced a large amount of data while the users were enjoying the game. CrowdDB is another example leveraging human input to process queries that database systems cannot answer [15]. Bozzon et al. proposed Crowdsearcher, a system that is able to answer search queries using the intelligence of crowds [6]. Furthermore, previous work combined human intelligence with machine learning methods to address the problem such as conversational agent learning intents and text classification [39, 3]. Recent studies recruited online users from crowdsourcing platforms, and applied smart task scheduling and output prediction methods to produce city maps [30, 33]. While human-in-the-loop systems have been shown to be effective in many domains, algorithm-in-the-loop systems also started to play a critical role in human decision-making. Previous work introduced this concept, and provided principles for human-AI decision-making and risk analysis [18, 19]. In the domain of robotics and engineering, human-machine collaboration has been used for a long time, to address practical problems that can hardly be considered in theoretical models [17, 25]. For instance, human-machine collaboration was effectively applied to address radiation source search and localization [5], spill finding and perimeter formation [9], and urban search and rescue response [10].

2.2 Source Searching

In general, source searching is a kind of problem that aims to determine the location of the source (of gas or signal) in the shortest possible time, as it is of vital importance for both nature and mankind [12, 24]. For example, the search for preys [21], submarines [11] survivors [36], and pollution sources [41]. As a classical kind of source searching algorithm, the bio-inspired algorithm typically leverages the gradient ascent strategy to approach the source, based on a reasonable assumption that the signal emitted by the source has a greater intensity near the source [29, 22]. However, in the presence of environmental disturbances (e.g., turbulence), the intensity gradient of the emitted signal may be disrupted, undermining the feasibility of the bio-inspired searching algorithm [23]. An alternative kind of source searching algorithm has been developed based on Bayesian theory [12]. Previous works [31, 23] proposed the cognitive searching algorithm that models the source searching process as a Markov Decision Process. To further enhance the performance (i.e., success rate and efficiency) of a searching algorithm, multi-robot collaboration mechanisms [27, 13, 35] were designed and adopted. However, when source searching happens in complex environments, the search process always encounters fatal problems, resulting in wrong outcomes. In this work, we designed a prototype system for source searching in complex environments, and carried out a user study with this system to answer the research questions.

3 Methods

Addressing the problem of source searching in complex environments, the search algorithm needs to navigate a robot to sense the signals emitted by the source, and simultaneously move in the environment. The search process ends when the source is found. There may be many obstacles in the complex environment, which can hinder the movement of the robot and bring challenges to the search task, especially when the entire environment is unknown. Previous work has proposed the Infotaxis search algorithm, featuring a reward function to determine which direction to go for source searching [37]. However, prior studies have shown Infotaxis and its improved versions did not perform well in complex environments and still faced problems occasionally [28, 31, 23, 40]. To this end, we designed a prototype system based on the crowd-powered method to address source searching problems.

3.1 Method Overview

In this section, we show the design of a crowd-powered source searching method, and explain how human rationales can be used during the search process to improve the effectiveness and efficiency of search algorithms. The overview of the method is shown in Figure 1.

In this work, we use the wisdom of the human together with machine intelligence to play important roles in existing source searching algorithms, to address problems that the machine cannot solve on its own. As shown in Figure 1, the human-machine collaboration contains three main parts — problem detection and task generation by machine, task explanation and solution suggestion by machine, and problem solving and task completion by human. In the following sections, we explain the three main parts in detail.

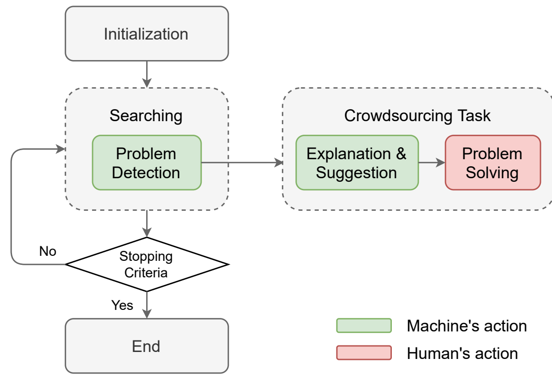

Fig. 1. The crowd-powered method that integrates human-machine collaboration into the search process.

3.2 Human-Machine Collaborative Tasks

The prototype system was designed following the framework shown in Figure 1. The source searching algorithm used in this system is Infotaxis, one of the most popular novel search strategies particularly effective for source searching problems [37, 31].

Problem Detection and Task Generation. Current source searching algorithms (including Infotaxis) usually suffer from local optimum problems, and therefore result in no information gained and infinite loop problems eventually. Therefore, we proposed a simple rule-based mechanism to detect the no information gained and infinite loop problems automatically: if a robot 1) passes by the same spot 5 times within a specific time window, and 2) acquires none information, the system detects a problem and pauses the search process. A task is then generated and crowdsourced, to leverage human intelligence to enable an effective problem solving. The crowdsourcing task features a user interface where crowd workers can view the problem explanation and execute the task. A screenshot of the crowdsourcing task is shown in Figure 2.

Task Explanation and Solution Suggestion. When a problem is detected, on the task interface, we use graphical elements to explain the task as well as the problem. In the prototype system, the goal of explanation is to let users clearly "see" the problem. We showed the direction that the robot wanted to go, and the direction robot had to go (because of obstacles) to help people understand why a problem could happen. We did not further explain the reasons since a problem could be the consequence of many different factors. Future work could focus on deep human understanding on problems. In addition, to enable an effective human-machine collaboration, the task could give a solution suggestion, which helps crowd workers better execute the task. The solution suggestion features

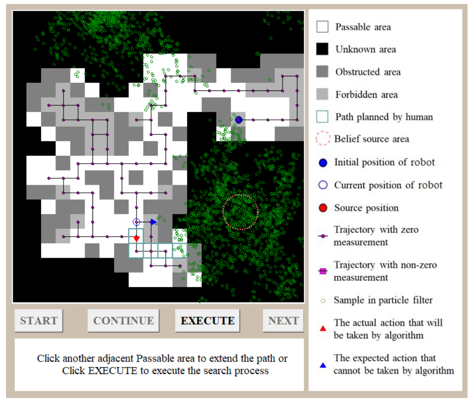

Fig. 2. A screenshot of the crowdsourcing task generated by the crowd-powered source searching prototype system.

a source estimation method that uses Bayesian inference and sequential Monte Carlo methods to show the distribution of the posterior probability of the source location (see green particles in Figure 2) [2, 34]. The machine also suggests an area where the source most likely will be called "belief source area" using DB-SCAN [32] to provide more information to help humans understand and address the problem.

Problem Solving and Task Completion. When the human (crowd worker) starts to execute the task, the prototype system provides two control modes. A full control mode allows the user to take over the search process and control each single step of the robot; An aided control mode allows the user to define a temporary goal (a targeted location), so the robot will pause the current search activities and move to the targeted location set by the user. We did not implement other problem-solving means in the prototype system since they require more expertise and incorrect operations may lead to a high failure rate. Future work could consider implementing more controls modes such as setting forbidden areas and search parameter tuning.

3.3 Task Interface

As shown in Figure 2, the task interface uses graphical elements to explain the source search task and the problem by displaying the searcher (robot), search environment, search route, current search state, and a potential search goal (estimated source). When a problem is found, the system automatically generates a crowdsourcing task and assign it to a human crowd worker, and then the worker canclick on the interface with the button [EXECUTE] to control the robot or plan a path for the robot. The user can click the button [CONTINUE] to resume the automatic search. This search process keeps going until the source is found. The prototype system was developed using Python 3.7 and tkinter packages.

4 User Study

We design a user study to answer our research question. In this section, we introduce experimental conditions, environments, measures, and the procedure of the user study.

4.1 Experimental Conditions

As we have introduced in the previous section, we provided two interaction/-control modes – Full Control (FC) and Aided Control (AC) respectively. A full control interaction mode represents the problem-solving method that requires humans to take over the search process, while an aided control interaction mode represents the problem-solving method that sets a temporary goal (let the robot exits the current search state and then navigate it to a manual defined location).

Furthermore, we used two baseline conditions in our experiment. The baseline 1 condition directly uses the state-of-the-art source search algorithm (Infotaxis), while the baseline 2 condition also uses our proposed automatic problem detection method and then navigates the robot to a random location in order to jump out of the problem. Please note that the baseline 2 condition is also an improvement based on the state-of-the-art source search algorithm.

4.2 Experimental Environments

The source searching activities are performed by a virtual robot in a 2D $20m \times 20m$ squared area in a simulation setup. The 2D search area is divided into 20×20 cells in a grid. Each cell has a probability P_o determining whether this cell contains an obstacle. P_o is set to be 0.75, to give a relatively high difficulty (more obstacles) of tasks, since simple environments (with few obstacles) do not need human assistance that much. In this study, we did not consider the specific types or shapes of obstacles. If there is an obstacle in a cell, the cell is considered to be completely obstructed and cannot be arrived at or passed by the robot.

The prototypes system was deployed on one machine, and all the participants were invited to execute tasks using the same machine to ensure a fair comparison. Participants were invited to a quiet lab, to make sure the experiment would not be interrupted by others.

4.3 Measures

In this study, we measure the effectiveness and efficiency of source search process and outcomes. The effectiveness is measured by the success rate. As the source searching process can forever go on if the source is not found, we define that a successful source searching means the robot finds the actual source within 400 steps (a step means an iteration of updating search states). If the robot cannot find the source (either with or without human involvement) after 400 steps, the source search task is considered to be failed. The efficiency is measured by the number of steps the robot takes to successfully find a source. A fail source search is not taken into account in calculating the efficiency. Furthermore, we measure human execution time per task to see how engaged the participants are during task execution.

Furthermore, we use two standard questionnaires to understand the perceived usability and cognitive workload while using the crowd-powered source searching system. The perceived usability is measured by System Usability Scale (SUS) [8]. Using the ratings of SUS items, we can derive scores of the SUS in two aspects – usability and learnability [26, 4]. Furthermore, we measure cognitive work load using NASA-TLX.

4.4 Procedure

We first asked participants to complete a demographic survey. This survey requires participants to provide their basic background information about their age, gender, education level, and domain knowledge about search algorithms. After the demographic survey, we also briefly explained to the participants our the experimental scenarios (i.e., to find a gas source) and how to use the prototype system.

After demographic surveys, participants were asked to complete source searching tasks. Each participant should complete 20 tasks using 2 control modes, i.e., full control and aided control. To avoid learning biases, the order of the control modes during task execution was pre-scheduled – half of the participants first executed 10 full-control tasks and then aided-control tasks (2 experts + 3 non-experts), and the other half first executed aided-control tasks and then full-control ones. After finishing each control mode (10 tasks), participants were asked to rate their feelings on system usability and cognitive workload using standard questionnaires.

5 Results

We evaluated the effect of using the crowd-powered method in source searching algorithms by measuring the effectiveness (success rate), the efficiency (the number of steps taken to find the source), the human execution time, the self-report SUS scores, and the self-report TLX scores.

5.1 Participants

We asked four experts (academic researchers or engineers), who have been working on the topics related to source searching for at least 1 year, to participate in our study. Furthermore, we recruited 6 non-expert volunteers from our institute who had no experience in source searching. People involved in the prototype system development were not invited to the experiment to avoid potential biases. The experiment was approved by the ethics committee of our institute.

5.2 Source Searching Result

We evaluated source searching from three perspectives, namely the effectiveness (the success rate), the efficiency (the number of steps used to find the source), and the human execution time per task. Results are shown in Table 1. Clearly, the crowd-powered method is proved to be effective, as the success rates can achieve 100% in most cases (except only one case) being approximately 22% higher than baseline 1, and 12% higher than baseline 2. It shows that leveraging human inputs could make the algorithm performance nearly perfect. Furthermore, we observe the improvement of efficiency when the full control mode is used, in comparison with both aided control and the baselines. In general, both experts and non-experts showed good performance while collaborating with the machine to solve the problems of the search algorithm.

Table 1. Results of the source searching experiment.

Groups	**Expertise**	**Effectiveness** (% success rate)	**Efficiency** (# steps per task)	**Human execution time** (seconds per task)
Full control	Expert	100	138.85 ± 79.00	29.59 ± 25.47
	Non-expert	98	144.73 ± 87.62	34.40 ± 30.16
Aided control	Expert	100	175.10 ± 67.67	33.58 ± 27.87
	Non-expert	100	165.67 ± 80.60	29.01 ± 29.51
Baseline 1	-	78.5	154.04 ± 91.32	-
Baseline 2	-	88	179.64 ± 96.45	-

To deeply understand the difference among experimental conditions, we performed statistical analysis for efficiency (the number of steps) and human execution time (per task). Since the numbers of search steps follow normal distributions according to the normality tests, we applied two-way ANOVA to see the effects of two factors considered in this study – expertise (expert vs non-expert) and control mode (full control vs aided control), as well as their interaction effect. Results of statistical tests are shown in Table 2. We found that the efficiency of source searching shows a significant difference in terms of the control mode ($p = 0.026$), meaning the full control mode could achieve a better efficiency regardless of expertise.

Table 2. Results of two-way ANOVA for the efficiency (# of steps) of source searching.

Factors	**Efficiency (# of steps)**		
	dF	F-value	p-value
Expertise (Expert vs Non-expert)	1	0	0.9764
Control mode (Full control vs Aided control)	1	5.01	0.0263*
Expertise × Control mode	1	0.68	0.4089

Note: an asterisk (*) represents significant difference ($p < 0.05$).

Since distributions of the human execution time do not come from a normal distribution according to normality tests (for all the data groups $p < 0.003$), we applied pairwise Mann-Whitney U tests to carry out the significance tests (α value was adjusted by Bonferroni correction). We did not find a significant difference in terms of human execution time ($p > 0.07$ for all the pairs), meaning neither expertise nor control mode could significantly affect execution time.

The source searching result conveys three main messages:

1. *The crowd-powered method is effective and efficient for improving source searching;*
2. *Through our design, non-experts could achieve similar performances as experts could do;*
3. *Taking over the machine during problem-solving could further improve the efficiency of source searching.*

5.3 Usability

We asked all the participants to fill up the system usability scale (SUS) after completing each control mode (i.e., full control and aided control). Therefore, each participant provided 2 SUS responses. Since we only recruited 10 participants (20 SUS responses in total), we did not use statistical tests to perform the analysis. Scores of SUS are reported in Table 3. According to previous studies [8, 26, 4], SUS can measure the usability and learnability of a system.

While we found that the full control mode (meaning the search process is taken over by humans) showed better search efficiency, the participants in general reported that they perceived better usability and learnability from the aided control mode, rather than the full control mode. Interestingly, for all the non-experts, their SUS scores of aided control were not lower than the scores of full control, meaning they all preferred the aided control mode. However, in terms of experts, we observed more diverse opinions, and the difference of experts' overall average SUS scores between aided control and full control was less obvious (Aided Control 86.25 vs Full Control 80.63), in comparison with non-experts (Aided Control 80.42 vs Full Control 65.00).

The SUS result conveys one main message: *The participants, especially the non-experts, generally perceived better usability when they were aiding the machine, in comparison with taking over the machine.*

5.4 Cognitive Workload

To understand cognitive workload during human-machine collaboration, we used NASA-TLX scale after the participants completed each control mode (i.e., full control and aided control). Similarly, each participant also only provided 2 TLX responses, resulting in 20 TLX responses in total. To this end, we did not perform any statistical analysis. Results of NASA-TLX scores on six dimensions, namely physical demand, mental demand, temporal demand, performance, effort, and frustration, as well as the overall TLX score are reported in Table 4.

Table 3. Results of the system usability. FC means the full control mode (humans take over the machine), while AC means the aided control mode (humans aid the machine).

Expertise	**Participant ID**	**Usability**	**Learnability**	**Overall**
Expert	1	AC (90.63) > FC (68.75)	AC (100.0) = FC (100.0)	AC (92.50) > FC (75.00)
	2	FC (93.75) > AC (90.63)	AC (62.50) > FC (50.00)	AC (85.00) = FC (85.00)
	3	AC (84.38) = FC (84.38)	FC (87.50) > AC (75.00)	FC (85.00) > AC (82.50)
	4	AC (81.25) > FC (71.88)	AC (100.0) = FC (100.0)	AC (85.00) > FC (77.50)
	Average	AC (86.72) > FC (79.69)	AC (84.38) = FC (84.38)	**AC (86.25) > FC (80.63)**
Non-Expert	5	AC (78.13) > FC (59.38)	AC (87.50) > FC (50.00)	AC (80.00) > FC (57.50)
	6	AC (87.50) > FC (68.75)	AC (87.50) > FC (62.50)	AC (87.50) > FC (67.50)
	7	AC (87.50) > FC (84.38)	AC (100.0) = FC (100.0)	AC (90.00) > FC (87.50)
	8	AC (71.88) > FC (65.63)	AC (100.0) > FC (87.50)	AC (77.50) > FC (70.00)
	9	AC (87.50) > FC (71.88)	AC (37.50) = FC (37.50)	AC (77.50) > FC (65.00)
	10	AC (78.13) > FC (43.75)	AC (37.50) = FC (37.50)	AC (70.00) > FC (42.50)
	Average	AC (81.77) > FC (65.63)	AC (75.00) > FC (62.50)	**AC (80.42) > FC (65.00)**

We observed that the non-experts in general perceived less cognitive workload in the aided control mode, compared to the full control mode, across all the TLX dimensions. However, in terms of the experts, we found that all the experts thought their performances while using the full control mode were not worse (3 out of 4 reported higher) than the aided control mode. This is an opposite finding compared to non-experts. In terms of other dimensions, again, we observed more diverse opinions from the experts, resulting in tiny differences between the aided control mode and the full control mode.

Table 4. Results of the cognitive workload. FC means the full control mode (humans take over the machine), while AC means the aided control mode (humans aid the machine).

Expertise	**ID**	**Physical demand**	**Mental demand**	**Temporal demand**	**Performance**	**Effort**	**Frustration**	**Overall TLX**
Expert	1	AC (0) < FC (5)	AC (0) < FC (5)	AC (5) < FC (20)	AC (0) = FC (0)	AC (0) < FC (15)	AC (0) = FC (0)	AC (0.8) < FC (7.5)
	2	FC (65) < AC (80)	FC (50) < AC (70)	FC (0) < AC (5)	FC (15) < AC (50)	FC (50) < AC (70)	AC (50) = FC (50)	FC (38.3) < AC (54.2)
	3	AC (10) < FC (20)	AC (5) < FC (30)	AC (0) = FC (0)	FC (0) < AC (10)	AC (10) < FC (15)	AC (0) = FC (0)	AC (5.8) < FC (10.8)
	4	AC (60) = FC (60)	AC (60) < FC (80)	AC (0) = FC (0)	FC (0) < AC (5)	FC (35) < AC (40)	AC (40) < FC (75)	AC (34.2) < FC (41.7)
	Average	AC (37.5) = FC (37.5)	AC (33.8) < FC (41.3)	AC (2.5) < FC (5.0)	FC (3.8) < AC (16.3)	FC (28.8) < AC (30.0)	AC (22.5) < FC (31.3)	**AC (23.8) < FC (24.6)**
Non-Expert	5	AC (5) = FC (5)	AC (0) < FC (5)	AC (0) = FC (0)	AC (5) < FC (50)	AC (10) < FC (15)	FC (5) < AC (10)	AC (5.0) < FC (13.3)
	6	AC (5) < FC (35)	AC (5) < FC (10)	AC (10) < FC (25)	AC (0) < FC (10)	AC (15) < FC (20)	FC (5) < AC (10)	AC (7.5) < FC (17.5)
	7	FC (15) < AC (20)	AC (0) = FC (0)	AC (0) < FC (20)	AC (50) = FC (50)	FC (30)< AC (35)	AC (60) = FC (60)	AC (27.5) < FC (29.2)
	8	AC (0) < FC (10)	AC (0) < FC (10)	AC (0) = FC (0)	AC (0) = FC (0)	AC (0) < FC (50)	AC (0) < FC (50)	AC (0.0) < FC (20.0)
	9	AC (0) = FC (0)	AC (0) < FC (50)	FC (15) < AC (20)	AC (10) = FC (10)	AC (25) < FC (50)	AC (0) < FC (20)	AC (9.2) < FC (24.2)
	10	AC (5) < FC (10)	AC (0) = FC (0)	AC (0) < FC (15)	FC (15) < AC (25)	FC (0) < AC (10)	AC (5) < FC (15)	AC (7.5) < FC (9.2)
	Average	AC (5.8) < FC (12.5)	AC (0.8) < FC (12.5)	AC (5.0) < FC (12.5)	AC (15.0) < FC (22.5)	AC (15.8) < FC (27.5)	AC (14.2) < FC (25.8)	**AC (9.4) < FC (18.9)**

The TLX result conveys two main messages:

1. *The participants, especially the non-experts, generally perceived less cognitive workload when they were aiding the machine, in comparison with taking over the machine;*
2. *The experts reported better performances when they took over the machine during problem-solving, while the non-experts did the opposite.*

6 Discussion

Results have shown that a human-machine collaborative framework could improve the effectiveness and efficiency of a source searching algorithm. We also observed interesting findings in terms of usability and cognitive workload. Clearly, these findings show that experts and non-experts have different needs and preferences.

6.1 Implications for Designing Crowd-Powered Systems

Our study has shown that using human-machine collaboration is effective in improving search algorithms. We have learned important lessons from the experimental result in terms of how to better leverage human rationales. The findings provide important implications in terms of designing general crowd-powered systems.

Personalization. According to the results, personalization is one of the most important factors that should be considered during the design phase. Our experiment shows that the design could achieve good efficiency and effectiveness, but did not satisfy non-experts' needs to some extent. Therefore, we suggest that personalization should play an important role throughout the entire process, since people's needs vary a lot due to different backgrounds, education levels, and personalities.

Learning from Humans. We argue that a crowd-powered system in the future should be able to learn human's (correct) behavior and accordingly adjust its own (problematic) actions. The prototype system has successfully improved the effectiveness and efficiency of the state-of-the-art search algorithms. However, our proposed approach still frames the capability of the machine, meaning the machine could only detect the problem and explain the problem in a way that the task designers decided, and the humans could only help the machine in problem-solving using pre-defined control modes. Human-machine collaboration should be mutually beneficial, as machines could help humans in problem-solving, and simultaneously learn from humans.

Crowd Computing. The experiment has shown that non-experts could achieve comparable output quality, with regard to effectiveness, efficiency, and execution time, in comparison with experts. This is a rather positive finding, implying the potential of massive deployment using crowd computing techniques and the possibility of leveraging swarm intelligence. Researchers and practitioners in the field of human-machine collaboration should focus on lowering the barrier of interaction, to access a larger number of users. Crowd computing has provided numerous new opportunities for multiple disciplines, including AI and HCI communities [16]. It could be connected to prevalent crowdsourcing platforms to access more diverse users, acquire faster responses, and achieve distributed/parallel computing.

6.2 Limitations and Future Work

In this study, we only recruited participants from our institute, and all the participants were either full-time researchers or students. We acknowledge the limitation that the participants in our study are not representative enough. Future work could properly perform a power analysis, consider a larger sample size, and probably recruit participants from online freelancing/crowdsourcing marketplaces, to obtain more general findings.

We also realized that the problem detection and explanation are rather simple in the prototype system design, and only two control modes were implemented in the prototype system. We consider it reasonable since this is a first step studying human-machine collaboration in improving source searching algorithms. As results indeed showed the feasibility of using human-machine collaboration, we suggest that future work could use more recent and advanced AI-based techniques to achieve a more intelligent interaction.

7 Conclusions

In this work, we proposed a research question to investigate the effects of using crowd-powered approaches in source searching algorithms. To answer the research question, we designed a framework enabling human-machine collaboration for improving existing source searching algorithms and carried out an experiment with 4 domain experts and 6 non-experts. The experimental results showed the feasibility of crowd-powered source searching in improving both effectiveness and efficiency. Furthermore, we explicitly asked the participants to report their perceived usability and cognitive workload to deeply understand their needs. Finally, we provided design implications for future human-machine collaboration research.

Acknowledgements This work was supported by the Natural Science Foundation of China (grant number 62202477).

References

1. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., et al.: Guidelines for human-ai interaction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–13 (2019)
2. Andrieu, C., De Freitas, N., Doucet, A., Jordan, M.I.: An introduction to mcmc for machine learning. Machine learning **50**(1), 5–43 (2003)
3. Arous, I., Dolamic, L., Yang, J., Bhardwaj, A., Cuccu, G., Cudré-Mauroux, P.: Marta: Leveraging human rationales for explainable text classification. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 5868–5876 (2021)
4. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the system usability scale. Intl. Journal of Human–Computer Interaction **24**(6), 574–594 (2008)

5. Bashyal, S., Venayagamoorthy, G.K.: Human swarm interaction for radiation source search and localization. In: 2008 IEEE Swarm Intelligence Symposium. pp. 1–8. IEEE (2008)
6. Bozzon, A., Brambilla, M., Ceri, S.: Answering search queries with crowdsearcher. In: Proceedings of the 21st international conference on World Wide Web. pp. 1009–1018 (2012)
7. Bradeško, L., Witbrock, M., Starc, J., Herga, Z., Grobelnik, M., Mladenić, D.: Curious cat–mobile, context-aware conversational crowdsourcing knowledge acquisition. ACM Transactions on Information Systems (TOIS) **35**(4), 1–46 (2017)
8. Brooke, J., et al.: Sus-a quick and dirty usability scale. Usability evaluation in industry **189**(194), 4–7 (1996)
9. Bruemmer, D.J., Dudenhoeffer, D.D., McKay, M.D., Anderson, M.O.: A robotic swarm for spill finding and perimeter formation. Tech. rep., IDAHO NATIONAL ENGINEERING AND ENVIRONMENTAL LAB IDAHO FALLS (2002)
10. Casper, J., Murphy, R.R.: Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **33**(3), 367–385 (2003)
11. Champagne, L., Carl, R.G., Hill, R.: Agent models ii: Search theory, agent-based simulation, and u-boats in the bay of biscay. In: Proceedings of the 35th conference on Winter simulation: driving innovation. pp. 991–998. Citeseer (2003)
12. Chen, X.x., Huang, J.: Odor source localization algorithms on mobile robots: A review and future outlook. Robotics and Autonomous Systems **112**, 123–136 (2019)
13. Dadgar, M., Jafari, S., Hamzeh, A.: A pso-based multi-robot cooperation method for target searching in unknown environments. Neurocomputing **177**, 62–74 (2016)
14. Desale, S., Rasool, A., Andhale, S., Rane, P.: Heuristic and meta-heuristic algorithms and their relevance to the real world: a survey. Int. J. Comput. Eng. Res. Trends **351**(5), 2349–7084 (2015)
15. Franklin, M.J., Kossmann, D., Kraska, T., Ramesh, S., Xin, R.: Crowddb: answering queries with crowdsourcing. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. pp. 61–72 (2011)
16. Gadiraju, U., Yang, J.: What can crowd computing do for the next generation of ai systems? In: CSW@ NeurIPS. pp. 7–13 (2020)
17. Goodrich, M.A., Schultz, A.C.: Human-robot interaction: a survey. Now Publishers Inc (2008)
18. Green, B., Chen, Y.: Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In: Proceedings of the conference on fairness, accountability, and transparency. pp. 90–99 (2019)
19. Green, B., Chen, Y.: The principles and limits of algorithm-in-the-loop decision making. Proceedings of the ACM on Human-Computer Interaction **3**(CSCW), 1–24 (2019)
20. Huang, T.H.K., Chang, J.C., Bigham, J.P.: Evorus: A crowd-powered conversational assistant built to automate itself over time. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. p. 295. ACM (2018)
21. Humphries, N.E., Queiroz, N., Dyer, J.R., Pade, N.G., Musyl, M.K., Schaefer, K.M., Fuller, D.W., Brunnschweiler, J.M., Doyle, T.K., Houghton, J.D., et al.: Environmental context explains lévy and brownian movement patterns of marine predators. Nature **465**(7301), 1066–1069 (2010)
22. Hutchinson, M., Oh, H., Chen, W.H.: A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors. Information Fusion **36**, 130–148 (2017)

23. Hutchinson, M., Oh, H., Chen, W.H.: Entrotaxis as a strategy for autonomous search and source reconstruction in turbulent conditions. Information Fusion **42**, 179–189 (2018)
24. Jing, T., Meng, Q.H., Ishida, H.: Recent progress and trend of robot odor source localization. IEEJ Transactions on Electrical and Electronic Engineering (2021)
25. Kolling, A., Walker, P., Chakraborty, N., Sycara, K., Lewis, M.: Human interaction with robot swarms: A survey. IEEE Transactions on Human-Machine Systems **46**(1), 9–26 (2015)
26. Lewis, J.R., Sauro, J.: The factor structure of the system usability scale. In: International conference on human centered design. pp. 94–103. Springer (2009)
27. Marjovi, A., Nunes, J.G., Marques, L., De Almeida, A.: Multi-robot exploration and fire searching. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 1929–1934. IEEE (2009)
28. Martin Moraud, E., Martinez, D.: Effectiveness and robustness of robot infotaxis for searching in dilute conditions. Frontiers in neurorobotics **4**, 1 (2010)
29. Pang, S., Farrell, J.A.: Chemical plume source localization. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **36**(5), 1068–1080 (2006)
30. Qiu, S., Psyllidis, A., Bozzon, A., Houben, G.J.: Crowd-mapping urban objects from street-level imagery. In: The World Wide Web Conference. pp. 1521–1531 (2019)
31. Ristic, B., Skvortsov, A., Gunatilaka, A.: A study of cognitive strategies for an autonomous search. Information Fusion **28**, 1–9 (2016)
32. Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X.: Dbscan revisited, revisited: why and how you should (still) use dbscan. ACM Transactions on Database Systems (TODS) **42**(3), 1–21 (2017)
33. Sharifi Noorian, S., Qiu, S., Psyllidis, A., Bozzon, A., Houben, G.J.: Detecting, classifying, and mapping retail storefronts using street-level imagery. In: Proceedings of the 2020 International Conference on Multimedia Retrieval. pp. 495–501 (2020)
34. Smith, A.: Sequential Monte Carlo methods in practice. Springer Science & Business Media (2013)
35. Tang, H., Sun, W., Yu, H., Lin, A., Xue, M.: A multirobot target searching method based on bat algorithm in unknown environments. Expert Systems with Applications **141**, 112945 (2020)
36. Unwin, A.R.: Search theory and applications. Journal of the Operational Research Society **33**(5), 494–495 (1982)
37. Vergassola, M., Villermaux, E., Shraiman, B.I.: 'infotaxis' as a strategy for searching without gradients. Nature **445**(7126), 406–409 (2007)
38. Von Ahn, L., Dabbish, L.: Designing games with a purpose. Communications of the ACM **51**(8), 58–67 (2008)
39. Yang, J., Drake, T., Damianou, A., Maarek, Y.: Leveraging crowdsourcing data for deep active learning an application: Learning intents in alexa. In: Proceedings of the 2018 World Wide Web Conference. pp. 23–32 (2018)
40. Zhao, Y., Chen, B., Zhu, Z., Chen, F., Wang, Y., Ji, Y.: Searching the diffusive source in an unknown obstructed environment by cognitive strategies with forbidden areas. Building and Environment **186**, 107349 (2020)
41. Zhao, Y., Chen, B., Zhu, Z., Chen, F., Wang, Y., Ma, D.: Entrotaxis-jump as a hybrid search algorithm for seeking an unknown emission source in a large-scale area with road network constraint. Expert Systems with Applications p. 113484 (2020)