



An Agent that Facilitates Crowd Discussion

A Crowd Discussion Support System based on an Automated Facilitation Agent

Takayuki Ito¹ · Rafik Hadfi¹ · Shota Suzuki²

Accepted: 10 October 2021 / Published online: 6 November 2021
© The Author(s) 2021

Abstract

Online discussion platforms are perceived as the next-generation method of citizen involvement. Such platforms can collect, integrate, and synthesize opinions to achieve social good. Crowd-scale platforms are being developed and deployed in social experiments that involve citizens and local governments. In such platforms, human facilitation is often used to preserve the quality of the discussions. Human facilitators often face difficulties when the discussions grow in size. In this paper, we present “D-agree,” a crowd-scale discussion support system based on an automated facilitation agent. The agent extracts discussion structures from online discussions, analyzes them, and posts facilitation messages. We conducted small- and large-scale social experiments in Japan to assess the social impact of the platform. The results showcase the success of our automated facilitation agents in gathering valuable opinions from citizens. In addition, our experiments highlight the effect of an automated facilitation agent on online discussions. In particular, we find that combining the agent facilitator with human facilitators leads to higher user satisfaction.

Keywords Crowd discussion support · Automated facilitation agent · Crowd decision support · Collective intelligence

1 Introduction

Crowd-scale discussion platforms are posed to be the next next-generation platforms for democratic citizen involvement. Such platforms require support functions that can integrate ideas, opinions, and arguments, discourage the publication of toxic content, and even achieve consensus (Malone and Klein 2007; Malone 2018a; Ito et al. 2019).

✉ Takayuki Ito
ito@i.kyoto-u.ac.jp

¹ Kyoto University, Yoshida-Honmachi, Kyoto 606-8501, Japan

² Nagoya Institute of Technology, Gokiso, Showa-ku 466-8555, Nagoya, Japan

An example of such platforms is the “COLLAGREE” system, with its ability to work jointly with human facilitators to promote crowd-scale online discussions (Sengoku et al. 2016; Ito et al. 2015, 2014; Ito 2018).

Despite their ability to promote citizen participation, human facilitators face cognitive challenges due to the possible scale of discussions and the complexity of the themes discussed (Kawase et al. 2018; Nishida et al. 2018, 2017). For instance, in the case of “COLLAGREE” discussions, some threads had over one thousand opinions that were posted simultaneously by the users of the system. In this paper, we propose to address such facilitation challenges by building an automated facilitation agent that can manage online discussions in a new crowd-scale discussion support system called “D-agree”. The automated facilitation agent extracts the structure of the discussion, analyzes it, and posts targeted messages to effectively facilitate the discussion.

To evaluate our system, we conducted small- and large-scale social experiments within the city of Nagoya (Japan) with the collaboration of the local municipal government. We initially posited the following three hypotheses:

Hypothesis 1 The agent can incentivize the participants to submit more postings and to diversify these postings.

Hypothesis 2 When the agent works collaboratively with the human facilitator, the overall performance of the facilitation increases.

Hypothesis 3 The satisfaction of the participants in the discussions facilitated by the agent is more than average. This means that the participants were at least not dissatisfied with the discussion facilitated by the agent.

The results of our experiments verify the above three hypotheses. Moreover, in the experiment with the collaboration of the municipal government of Nagoya, the collected insights were later analyzed and used to elaborate upon social decisions and policies.

The contribution of the paper is twofold. First, we propose an agent platform that can intelligently interact with humans and extract insights from their discussions. Second, the platform successfully guided humans in their discussions using facilitation mechanisms that were evaluated in real social experiments.

The paper is structured as follows. In Sect. 2, we cover the relevant literature on crowd-scale platforms and the underlying technologies. In Sect. 3, we present an outline of our system, including the automated facilitation agent. In Sect. 4, we review the large-scale experiment and its results. In Sect. 5, we cover the small-scale controlled experiment. Finally, we summarize our work and highlight the future directions.

2 Related Work

2.1 Online Platforms

Online platforms are becoming crucial for the empowerment of citizens and in implementing sustainable goals Savaget et al. (2019). They can now collect opinions and even lead to advanced forms of social agreements (Malone 2018b; Malone and Klein 2007). For instance, the Climate CoLab system (Malone and Klein 2007) was used to integrate the collective intelligence of thousands of people worldwide to address climate change. The Deliberatorium Iandoli et al. 2007 is another system where people submit ideas by following an argumentation map through which participants frame their ideas. The first difference between our system and the Deliberatorium is that our discussions are structured around issues, or critical questions, to be addressed based on the Issue-Based Information System (IBIS) (Conklin and Begeman 1989). The second difference is that participants in the Deliberatorium create their discussions according to a predefined argumentation map, while in our system we do not constrain the participants to using such a map. Instead, the system builds the argumentation structure automatically from their posts after classifying them into IBIS elements.

Another similar system that shares many aspects with our proposal is called “COLLAGREE” (Sengoku et al. 2016; Ito et al. 2015, 2014; Ito 2018). COLLAGREE has been employed for large-scale social experiments in Japan. The system was used in the context of a collaboration with the local government to gather opinions from the public about next-generation planning. The real social impact of the system was that it succeeded in gathering opinions from younger people at a lower cost. The main difference between our current platform and COLLAGREE is that the latter used human facilitators. Other platforms for citizen participation rely on decision theory with insights from social sciences. For instance, the work Mkude et al. (2014) focused on participatory budgeting and the assessment of added public value. The Participatory On-Line Interactive System (POLIS) is another platform that allows multi-method, multi-stage, and semi-structured electronic public participation for citizens (Williams 2010). Our proposed system shares the same motivation with respect to the future of deliberative democracy and public sociology.

In practice, intelligent discussion platforms combine algorithmic and statistical techniques to harness the intelligence of the crowds. In our work, we focus on the use of artificial agents for their ability to adapt to complex human behavior, particularly in argumentative domains. These complex problem domains will first be addressed using argumentation mining, and then a facilitation agent will be implemented to handle them.

2.2 Argumentation Mining

A crucial component to the development of our platform is the ability to manipulate argumentative text in online discussions. This task is performed using argumentation mining, which also refers to the research area that is closely related to our study.

Argumentation mining aims at identifying the structure of different arguments in natural language texts. For instance, many studies in the field of argumentation mining extract structures from essays (Stab and Gurevych 2014b; Nguyen and Litman 2016, reviews Kim 2014, and legal texts Palau and Moens 2009) in the same way as we propose for extracting structures from online discussions. These essays, reviews, and legal texts are represented according to different data models such as the Issue-Based Information System (IBIS) model Conklin and Begeman 1989. Among the studies in the field of argumentation mining, the subtasks of the component classification and the structure identification Stab and Gurevych (2017) are particularly related to our subtasks of node and link extractions. The difference between classical argumentation mining and our approach is that we perform the mining in real time as people discuss and alter the mined text. In addition, our agent dynamically posts its facilitation messages so that the entire discussion grows according to the IBIS model. In our mining approach, we extract the IBIS nodes from the text and then add the links that connect these nodes in the original text. The links are crucial for obtaining the final IBIS hierarchy that represents the argumentative structure of the discussion as illustrated, for instance, in Fig. 7. In the results of our extractions, we found that the F scores of extracting issues exceeded 0.80, and the precision of identifying the links among the IBIS elements was around 0.88; consequently, these scores are higher than state-of-the-art argumentation mining. Furthermore, our results greatly depended on the manual annotation efforts on over 38 discussion datasets, that is, after carefully defining our annotation scheme, our annotation results had a Fleiss Kappa value of around 0.66 (Yamaguchi et al. 2019). Here, the Fleiss Kappa value is a statistical quantity that measures inter-rater agreement for qualitative items. One of the main limitations of our method is that we focus on simplified discussion structures using the IBIS model. This model assumes that there are only four clearly classifiable components: issue, idea, pros, and cons. This assumption is the reason why we obtained higher accuracy. This being said, our major goal in this paper is not the classification of arguments but a clarification of the effect that a facilitation agent could have on online “argumentative” discussions. In the field of argumentation mining, more general components are often considered such as the major claims, minor claims, and premises (Stab and Gurevych 2014a). Another alternative to IBIS is to use coarse discourse acts and their richer set of argumentative types (Zhang et al. 2017). In the end, maintaining a limited set of argumentative utterances made the extraction more tractable and allowed the agent to interact in real time with the participants. In addition, it allowed us to create around 200 tractable facilitation rules that were carefully assembled after consultation with professional human facilitators. By combining these rules and the obtained IBIS structures, we could generate and use the facilitation messages in real time.

2.3 Chatbots

The key component in our platform is the facilitation agent and its ability to interact with humans in online discussions. Such an agent is identified as a conversational agent, a chatbot, or a social bot, and it is defined as a computer program that is

designed to converse using natural language (Almansor and Hussain 2020; Tavanapour and Bittner 2018; Tavanapour et al. 2019). Such agents could generally be classified as task-oriented agents and non-task-oriented agents Chen et al. 2017; Yan et al. 2017. Task-oriented agents are designed for a particular task and are set up to have short conversations, usually within a closed domain such as online shopping, customer support, or medical expertise. Many techniques can be adopted to build this type of agent, such as parsing Weizenbaum (1966), pattern matching Wallace (2009), and more recently with the use of neural networks (Nuez Ezquerro 2018; Csaky 2019). The approach we adopt is rule-based and relies on deep learning classification, which gives the agent the ability to respond to a given message with the purpose of facilitating argumentative discussions. That is, our facilitation agent can identify argumentative utterances, build the corresponding semantic structure, and post adequate facilitation messages based on this structure.

2.4 Evaluation Methodology of Crowd-Scale Systems

The evaluation of crowd-scale systems requires the use of appropriate methodology when looking at the usefulness of the system and its acceptance among the crowds. Examples of such methodologies include the Technology Acceptance Model (TAM) (Davis 1985; Dasgupta et al. 2002; Venkatesh and Davis 2000), user satisfaction Zviran and Erlich (2003), usability evaluation Lewis (2018), and so forth. Due to the large scale of our studies and the ill-defined nature of the discussions, we relied on a quantitative method that combines questionnaires, annotated data, and statistical analyses of the argumentative data generated from the discussions. We particularly looked at how many IBIS elements are generated in a discussion and how many of these are generated as a result of the facilitation messages. We then combined such measures with the satisfaction levels of the users (Joshi et al. 2015). To this end, we used questionnaires created by experts in social psychology as well as psychological measurement scales (Hori and Yamamoto 2001; Hori and Yoshida 2001; Hori and Matsui 2001; Hori et al. 2007, 2011). A detailed investigation based on TAM will be one of our future works.

Finally, we looked at the interaction among the types of replies; i.e., those from participants to other participants, from participants to facilitator, and from facilitator to participants.

3 The D-Agree Platform

The D-agree system is composed of our artificial agent and the Web platform that hosts the participants and allows them to exchange messages with each other and with the agent. An example of such an exchange is shown on the left side of Fig. 1, where the first person submits a question in the form “How can we solve congestion in Nagoya city?”. Then, the automated facilitation agent identifies this post as an issue, labels it as an issue, and stores it in the database. The second person submits his/her post “How about introducing a traffic tax mechanism?”. Our agent

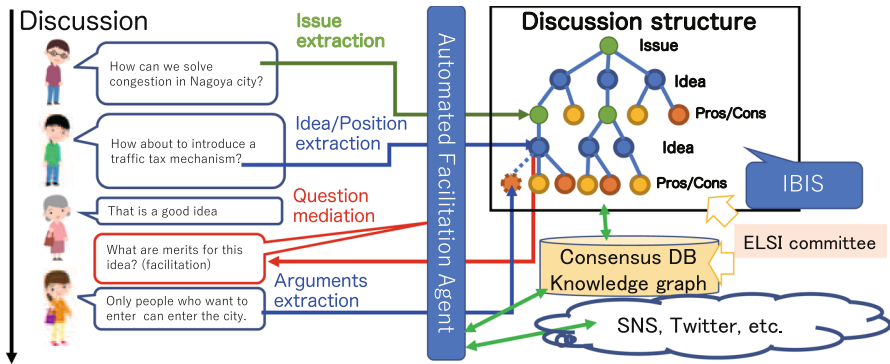


Fig. 1 D-agree: Web interface and automated facilitation agent

identifies this as an idea corresponding to the issue submitted by the first person. This new post is labeled as an idea, which is stored in the database with a link to the corresponding issue submitted by the first person. By following this process, the agent constructs a typed hierarchic structure of the discussion. Finally, given predefined facilitation rules, the agent posts a facilitation message such as “What are the merits of this idea?” whenever the number of the pros is small for the idea under discussion.

For the extraction of the discussion structure, we adopted the Issue-Based Information System (IBIS) Kunz and Rittel (1970), shown in the right side of Fig. 1. This choice is justified by the need to lead the discussions while allowing people to clarify issues and ideas and then debate their merits and demerits. The IBIS model can comprehensively distinguish between such elements as well as any argumentative text (Lawrence and Reed 2017). Once the IBIS structure is automatically extracted, the facilitation agent posts facilitation messages in relation to the discussion to incentivize the users to cover more issues, ideas, pros, and cons. The resulting structured discussion is stored in the discussion database and later solicited in future discussions as reference.

The system’s Web interface is shown in Fig. 2. The example is taken from an experiment conducted during an official governmental meeting in Afghanistan Haqbeen et al. (2020). The features of the interface are described as follows.

1. The phase of the discussion.
2. Discussion topic posted by the moderator.
3. Human-based facilitated message.
4. Facilitation message of the agent.
5. Ranking that includes user aspects of performance such as the number of posted items and the activity-based points.
6. Summary of agent activities such as classification, analysis, and visualization.
7. The post form used to post discussion topics.
8. The reply function used by users to post opinions.
9. Search function used to refer to current and past discussions using keywords.

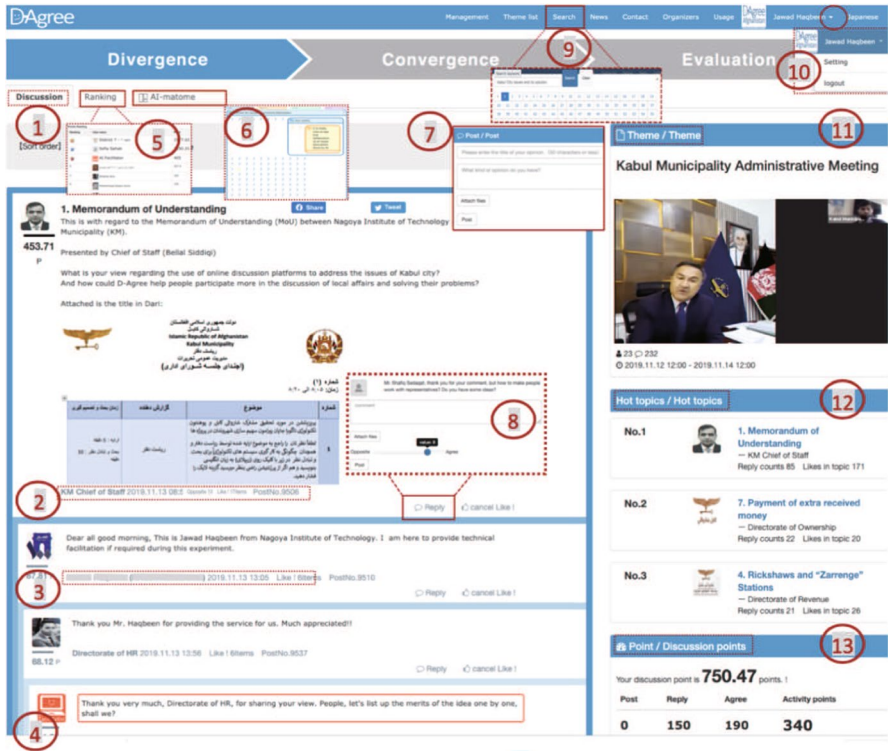


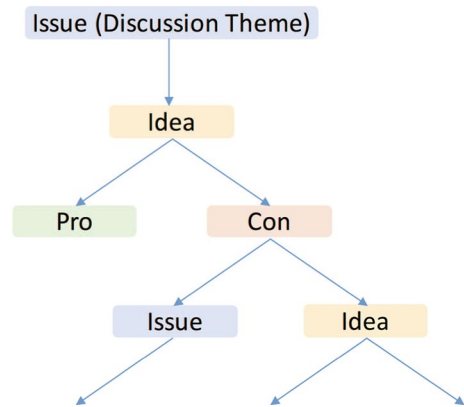
Fig. 2 User interface of proposed system

- 10. Menu bar that includes account settings and logout button.
- 11. Discussion theme and media. Users can see the total number of discussants, posted items, discussion time, and live discussion videos.
- 12. Ranking of the posted topics.
- 13. Discussion points earned through participation.

3.1 Issue-Based Information System (IBIS)

The Issue-Based Information System (IBIS) is a practical model to structure arguments in textual discourse (Noble and Rittel 1989). This is done by categorizing sentences into issues, positions, and arguments in a graphical manner. There were previous attempts to use the IBIS model in the context of face-to-face meetings (Noble and Rittel 1989). Similarly, another approach Conklin (2003) proposed a system called Dialog Mapping, where an idea is used instead of “position” and arguments are set to “pros” and “cons.” Here, we use a similar formalism as illustrated in the example of Fig. 3. The root node is often the main question to be addressed by adding new ideas or arguments.

Fig. 3 Example of IBIS-based discussion structure



3.2 Automated Facilitation Agent

We developed an automated facilitation agent software that performs the following tasks:

1. Observing the textual content posted by the users,
2. Extracting the argumentative utterances from the content,
3. Generating facilitation messages according to predefined rules, and
4. Posting the messages to the discussion board in response to other posts.

The agent has additional functions such as filtering inappropriate posts and visualizing the IBIS elements as a tree. The agent consists of two main modules:

1. Observation and posting module, and
2. Data extraction module.

The observation and posting module was implemented using Amazon CloudWatch Wittig and Wittig (2018) and AWS Lambda functions to enable scalable observation and posting functions Varia and Mathew (2014). Accordingly, the agent can be activated when events happen within the discussion such as detecting certain utterances or receiving events from external triggers (Cloud Watch). The posting function is activated when a particular clue is detected, which allows the agent to post a message based on predefined rules. For instance, if three posts are added to the discussion and the last post is an issue, then the agent could post a message that asks the user to elaborate on the issue or propose a solution.

To detect the types of the posts, the agent needs to classify the text according to the IBIS types. To this end, we implemented the data extraction module using a Bidirectional Long Short-Term Memory (BiLSTM) classifier (Suzuki et al. 2019; Lample et al. 2016). The module captures the sentences and their IBIS word constituents (issues, ideas, pros, and cons). Then, it identifies the links that connect these nodes within the textual data. Finally, the module adds these relationships to the

IBIS data model of the agent. Our proposed extraction method relies on previous works in argumentation mining Suzuki et al. (2019; Lawrence and Reed 2017; Stab and Gurevych 2017, 2014b, a) while remaining better suited to the IBIS data types. Such types often include an issue component, which is different from conventional argumentation structures. Furthermore, most of the literature on argumentation mining focuses on the use of claims and premises, thus lacking issue components in their structure (Cabrio and Villata 2018). Issues are critical for ill-defined discussions and wicked problems (Churchman 1967). To overcome this limitation, we adopt the conception where claims are decomposable into issues and ideas, while premises are equivalent to arguments (pros and cons). Adopting this mapping in the IBIS model provides a richer data model for argumentative discussions. More details on our implementation of the extraction method can be found in a previous study (Suzuki et al. 2019).

The generation of the facilitation messages is controlled with two parameters: a period of 1 minute specific to Amazon CloudWatch Wittig and Wittig (2018) and a threshold of 3 messages. This threshold sets the number of messages that the agent should count before taking part in the discussion. That is, the agent will wait before extracting the node types of the last message and then selecting an adequate message. The messages are selected based on rules that map each IBIS type to a random sentence. For example, a message to an idea would look like “That is a good perspective. Anybody else agree with your idea?” or “You are absolutely right. Anyone else support {user}’s idea?”. The variable “{user}” is the name of the participant to whom the agent is replying to Hadfi et al. (2020).

3.3 Architecture of System

Figure 4 illustrates the architecture of our system and its user interface. The system operates on Amazon Web Services (AWS) to manage the scale of the discussions (Varia and Mathew 2014). Here, the discussions are conducted in Japanese or English. The Web server component manages discussion boards and all of the data stored in the database. Users can access our system using Web browsers or iPhone

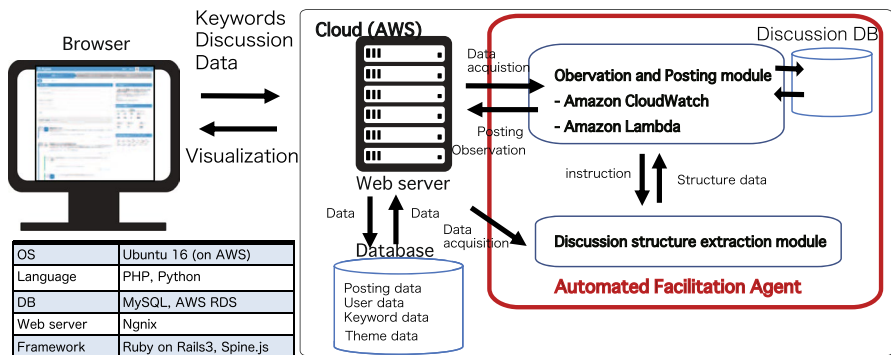


Fig. 4 System Architecture and User Interface

and Android applications. The red boxed area in Fig. 4 shows the automated facilitation agent and its constituents.

4 Large-Scale Societal Experiment

4.1 Setting

The objective of this experiment was to gather opinions on the next-generation planning in the city of Nagoya (Japan). The resulting comprehensive plan will be the basis for the administrative decisions within the next few decades in the city of Nagoya. The D-agree system was used for this task and allowed Nagoya citizens to discuss five themes about the future of their city. As a result, we received 15,199 page views, 157 registered participants, and 432 submitted opinions, and the system was visited by 798 participants. These discussions were also held in the context of more than 10 face-to-face meetings with the city's administrative staff. In a typical town meeting, there were more than 100 people gathered, and each person had an opportunity to provide opinions to the city administrators during the two-hour session. People who attend such town meetings are generally senior citizens, since such meetings are held in daytime. In contrast, our online platform attracted younger people at lower participation cost.

The experiment was conducted from November 1 to December 7, 2018. The whole campaign was advertised on Google Ads, on the homepage of the Nagoya municipal government, on the town meetings announcement of the Nagoya municipal government, and on various social media (Facebook, Twitter, Line, etc.).

The plan has five main themes in total.

Theme1 : Human rights and diversity.

Theme2 : Secure childcare.

Theme3 : Disaster prevention.

Theme4 : City environment.

Theme5 : Attractiveness to industry and the world.

Themes 1 and 2 were facilitated by expert human facilitators. In particular, for theme 1, the facilitators used their own facilitation methodologies, while for theme 2, their facilitation was based on the IBIS model. Themes 3 and 4 were facilitated by automated facilitation agents only. Theme 5 was facilitated through the cooperation of humans and agents, and here human facilitators used IBIS.

The choice of themes and the differences between them are paramount to conducting significant evaluations of the system's output. In our case, theme differences could in fact give rise to distinct distributions of ideas, issues, and arguments, depending on the initial questions and the populations. Here, we were not mainly focusing on comparing the discussions and the resulting IBIS data, since they revolve around completely different themes. For example, some topics could naturally lead to more questions (for example, unresolved social problems), while other

topics might lead to more ideas (for example, well-understood topics). Our main goal was to globally assess the behaviors of the agent facilitator, human facilitator, and participants within their discussions.

We established two phases with the goal of summarizing the discussed ideas. The first phase had 30-day discussions, and the second phase was conducted during 7-day discussions. In the first phase, people discussed issues using the D-agree system. In the second phase, administrative staff members summarized the discussions into several concrete ideas on which the citizens voted.

In the first phase, we launched the D-agree system on the internet. Anyone could register with the platform and post comments on the discussion threads. To register, users provided their email address, nickname, gender information, and home region (town-level). We did not gather actual names and exact home addresses. The collected information was carefully secured by the administrative staff to protect the privacy of the participants who only know each other by the registered nicknames.

4.2 Results of Large-scale Experiment

As mentioned in the Introduction section, we proposed to study the following three hypotheses:

Hypothesis 1 The agent can incentivize the participants to submit more postings and to diversify these postings.

Hypothesis 2 When the agent works collaboratively with the human facilitator, the overall performance of the facilitation increases.

Hypothesis 3 The satisfaction of the participants in the discussions facilitated by the agent is more than average. This means that the participants were at least not dissatisfied with the discussion facilitated by the agent.

4.2.1 Example

Figure 5 shows an example of a discussion where the facilitation agent is responding to citizens' posts in Japanese. Here, Issue 1 was raised by the participant, and the automated facilitation agent identified this post as an issue. Then, the agent asked "What can we do to solve it?" and a participant posted Idea 1. The facilitation agent identified this post as an idea and raised an issue to deepen the idea. Then, a participant posted Idea 2. This behavior is perceived as a successful facilitation behavior. However, there are cases where the agent could misidentify the IBIS nodes and links. This is due to the non-determinism of the classification method and the possibility that it encounters elements that do not fall within the IBIS taxonomy. For instance, if a participant submits a generic text that is neither an idea, issue, pro, nor con, then the agent cannot identify the post correctly. The solution we found to this problem is to use generic facilitation messages that do not alter the ongoing discourse. This situation is common given the complexity of the linguistic domains



Fig. 5 Example of automated facilitation from discussions among citizens of Nagoya City

adopted in online discussions, and our proposed method is a practical way to control and bound the discussions within the limits of argumentative discourse. Another solution to address this limitation would be to generate messages only when the classification is deterministic, or to adopt an extended discussion model to account for more node types (Zhang et al. 2017).

4.2.2 Number of Postings

We started by looking at the total number of posts resulting from the experiments as illustrated in Table 1. The analysis was done according to a student’s t-test with $N > 700$ in a setting similar to that of an earlier work Woolley et al. (2010). The “Human” rows show the number of postings by human facilitators, and the “Agent” rows show the number of postings by the automated facilitation agent. The column “Participants” shows the number of postings by the participants. Themes 3, 4, and 5 were facilitated by the automated facilitation agent, and they obtained more posts than in the case of the themes facilitated by participants. This implies that the automated facilitation agent incentivized participants to submit more postings. This result supports our Hypothesis 1.

Table 1 Number of postings depending on the themes and the facilitator

Theme	Number of postings			
	Total	Human	Agent	Participants
1: Human	81	43	0	38
2: Human	56	21	0	35
3: Agent	88	0	24	64
4: Agent	70	0	18	52
5: Agent & Human	137	17	21	99

In Theme 5, the agent and humans worked together and thus generated more postings from the participants. This is a complementary result on the interaction between human and agent facilitators. In this sense, humans can understand complex linguistic content while an agent could work with large textual content over long periods of time. This leads to human facilitators interacting more with the participants and deepening the discussion. Note that the number of posts alone is not sufficient to assess the quality of a discussion and the performance of the crowd Hong et al. (2016). We propose looking at the satisfaction levels of the users as well as the types of generated argumentative content.

4.2.3 User Satisfaction

We used questionnaires to evaluate the satisfaction of the system's users (Joshi et al. (2015). Typical questions were of the form "Are you satisfied with the discussion of the city plan?". The participants had to select their level of satisfaction from strongly satisfied (5), satisfied (4), neutral (3), dissatisfied (2), and strongly dissatisfied (1). In the results, illustrated in Fig. 6, the satisfaction scores had nearly the same scores, 3.2 to 3.7, across all themes. This suggests that users experienced satisfying discussions even if they were managed by the automated facilitation agent (Agent). We also noted that the collaboration between the automated facilitation agent and the humans (Agent and Human) achieved the highest satisfaction score. We believe this is due to the complementarity effect in which the agents respond efficiently to multiple users' postings while humans can post well-thought-out comments. These results support Hypotheses 2 and 3.

4.2.4 Diversity of IBIS Nodes

We analyzed the discussion data and labeled all of the postings in the discussions based on their IBIS structure. Table 2 shows the number of IBIS nodes obtained in each discussion theme. Here, issue, idea, pros and cons are the IBIS nodes. "Facilitation" represents the number of facilitation messages produced by the agent. The label N/A refers to unclassified nodes. As explained above, human facilitators for

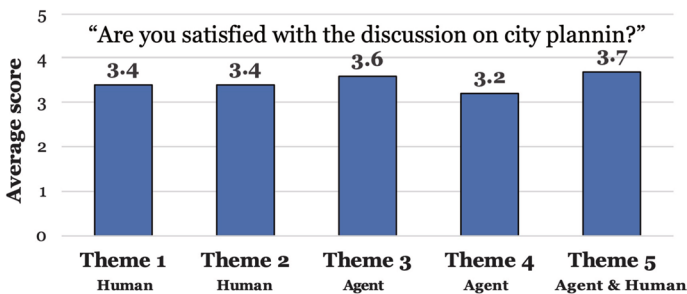


Fig. 6 User satisfaction scores for the five different themes. The score given answers the question "Are you satisfied with the discussion on city planning?" for each of the five themes.

Table 2 Number of IBIS nodes

Nodes	Themes					Total
	Theme 1	Theme 2	Theme 3	Theme 4	Theme 5	
Issue	15	10	8	3	9	45
Idea	37	29	39	27	101	233
Pros	7	8	6	13	47	81
Cons	6	6	6	11	11	40
Facilitation	137	33	36	26	36	269
N/A	51	46	51	68	133	349

Chi-test*: $p < 0.05$

Theme 1 did not follow the IBIS structure, while the human facilitators of Theme 2 did follow it. Consequently, it is clear that the human facilitators of Theme 1 posted more than the facilitators of the other themes. Due to the lack of clear semantics in the facilitation, there was difficulty in achieving adequate facilitation. This suggests that a facilitation structure would reduce the facilitation messages. The case of collaboration between an automated facilitation agent and human facilitation can be viewed as working well because it obtained 101 ideas and 47 pros, which is large compared to the results for the other themes. The diversification of the IBIS nodes and the successful combination of human and agent facilitation support our Hypotheses 1 and 2.

4.2.5 Agent Performance in Terms of Post-Type Generation

As a general measure of performance of the automated facilitation agent, we investigated how many nodes are generated from a single facilitation message. This is computed as the ratio of the number of nodes versus the number of facilitation messages. The results are illustrated in Table 3. We can see that the number of postings increases by the automated facilitation agent. The performance of the human facilitator in Theme 1 is clearly lower than that of facilitators in the other themes. The performances of our agent for Theme 3 and Theme 4 are 41.0% and

Table 3 Facilitation performance in terms of node-type count (%)

Theme	Discussion theme				
	Theme 1	Theme 2	Theme 3	Theme 4	Theme 5
Facilitator	Human	Human	Agent	Agent	Human & Agent
Issue	10.9	30.3	22.2	11.1	25.0
Idea	27.0	87.9	108.3	100.0	280.6
Pro	5.1	24.2	16.7	48.1	130.6
Con	4.4	18.2	16.7	40.7	30.6
Average	11.9	40.2	41.0	50.0	116.7

Chi-test*: $p < 0.05$

50.0%, respectively, which are competitive with the performance of 40.2% for the human facilitator who follows the IBIS structure in Theme 2. Accordingly, the performance of our agent is at the same level as that of the human facilitator who follows IBIS. Additionally, in the case of collaborative facilitation by a human and an agent in Theme 5, the performances in getting ideas and pros are 280.6% and 130.6%, respectively, which are dramatically better than the other cases. From the viewpoint of facilitation performance, collaborative facilitation between humans and agents worked well. The performance of the agent when alone and when coupled with humans supports Hypotheses 1 and 2, respectively.

4.2.6 Qualitative Analysis of the Discussions

We qualitatively analyzed the discussions with the goal of identifying the most common IBIS topologies. Figure 7 illustrates the discussion trees labeled manually with the IBIS types. In each figure, the boxes colored in blue, orange, green, purple, and gray are issues, ideas, pros, cons, and N/A nodes. The obtained tree in Theme 5 is the largest, widest, and deepest. In this sense, collaboration between humans and agents would help crowd discussions. On the other hand, we cannot see a big difference among Themes 1 to 4. This suggests that our automated facilitation agent worked in the same way as the human facilitators. These results support Hypotheses 1 and 2.

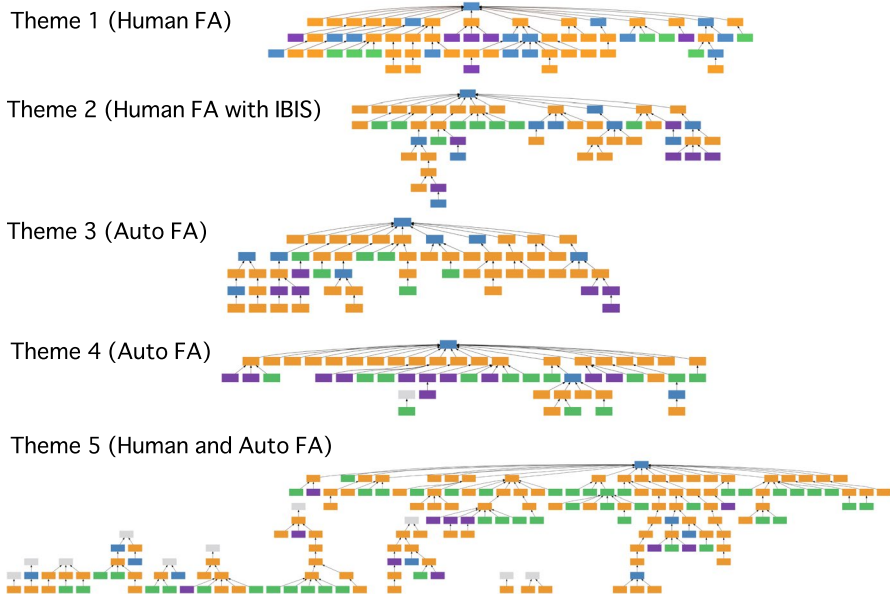


Fig. 7 Obtained Discussion Trees

5 Small-Scale Controlled Experiment

In order to verify the direct effect of the automated facilitation on online discussion, we conducted small-scale controlled experiments. The discussion theme was set to “city development.” The subjects were randomly selected graduate students from 22 Japanese universities. We randomly separated them into 2 isolated groups. The discussion time was set to durations of 45 and 60 minutes. We chose six themes for discussion. Each of the two groups conducted six online discussions with these six themes.

Human and agent facilitators participated in the discussion, and both conducted their facilitation according to the IBIS model. We informed the subjects that a facilitator was involved in the discussions but we did not announce the facilitator type (human or agent). In the discussion experiment, the human facilitators lead the discussion based on the IBIS structure, and they were aware of the presence of the agent facilitation.

In the experiment, the subjects were gathered in one venue as shown in Fig. 8. We prohibited spoken communication between subjects so that online discussions would not be affected. Subjects conducted anonymous discussion using an arbitrarily assigned nickname so that opinions could be expressed freely. Table 4 shows an outline of the discussion experiment.

The setting of the timing and frequency of the automated facilitation is critical for the comfort of the participants. The automated facilitation agent was set to reply once every five participant posts in discussions B and C, and once every three participant posts in discussions F, G, J and K. In addition, the facilitation agent had access to the new postings from the discussion board once every one minute. This particular setting depended on our trial experiments and agent configuration.



Fig. 8 The small-scale controlled experiments

Table 4 Outline of Discussion Experiment

Day 1 16:20 to 18:50	1st 60min, Disc. A: 11 subjects & Human FA, Disc. B: 11 subjects & Auto FA <i>Theme: How to support lifelong mental and physical health promotion?</i>
Day 2 16:20 to 18:00	2nd 60min, Disc. C: 11 subjects & Auto FA, Disc. D: 11 subjects & Human FA <i>Theme: How can I make an environment where children grow up safely?</i>
Day 3 16:20 to 18:00	3rd 45min, Disc. E: 11 subjects & Human FA, Disc. F: 10 subjects & Auto FA <i>Theme: How to realize a safe area without crime or traffic accidents?</i>
Day 3 16:20 to 18:00	4th 45min, Disc. G: 11 subjects & Auto FA, Disc. H: 10 subjects & Human FA <i>Theme: How can we create an environment that is natural and agricultural?</i>
Day 3 16:20 to 18:00	5th 45min, Disc. I: 10 subjects & Human FA, Disc. J: 10 subjects & Auto FA <i>Theme: How can we enhance Nagoya city as a major city in the world?</i>
Day 3 16:20 to 18:00	6th 45min, Disc. K: 10 subjects & Auto FA, Disc. L: 10 subjects & Human FA <i>Theme: How can we support improvement of regional disaster management?</i>

5.1 Results

In order to identify the differences between the facilitation agent and the human facilitator, we analyzed the number of responses, the number of nodes in the IBIS data, the number of replies for each node, and the questionnaire data. The obtained nodes were labelled, classified, and validated by six researchers. An outline of the experiment is shown in Table 4. The format “FA X” means that node X was targeted by the facilitator who requested the participants to address such a node. For example, “FA Issue” is a facilitator message that targets an issue, and “FA Response” is a facilitation message that requests an opinion. “Issue (FA)” is an issue submitted by a facilitator. The general discussion theme was treated as “Issue (FA),” which is the top level of IBIS discussion structure.

5.1.1 Number of Posts and Interval Time

In order to clarify the differences between human and agent facilitation, we summarized the number of posts by participants within certain interval times. Table 5

Table 5 Number of postings by human facilitator

Discussion	Posting (All)	Posting (FA)	Posting (PA)	Posting (FA/PA)	Average Interval (PA → FA)
No. 1 A	194	32	162	16.5%	350.6
No. 2 D	169	27	142	16%	339.2
No. 3 E	136	21	115	15.4%	497.1
No. 4 H	161	20	141	12.4%	396.7
No. 5 I	98	27	71	27.6%	320.8
No. 6 L	97	14	83	14.4%	415.7
Average	142.5	23.5	119	16.5%	377.5

shows the case where the facilitator is a human, and Table 6 shows the case where the facilitator is an artificial agent.

In Tables 5 and 6, the quantity “Posting (FA)” refers to the number of postings by the facilitator and “Posting (PA)” refers to the number of postings by the participants. The quantity “Avg. Interval (PA→FA)” refers to the time in seconds that has passed from the final participant posting up to the posting of the facilitator. From Table 5, the average number of posts by the human facilitators is 23.5, while Table 6 shows that the average number of posts made by the automated facilitation agent was 33. We found a significant difference at the 1% level by a t-test, where the automated facilitation agent can reply about 1.4 times more than the human facilitator in this experiment. As the number of participants in the discussion increases, the number of posts will increase too, so the difference is likely to widen. This result is obvious because the agent is automated. However, it is not easy to post facilitation messages at an optimally chosen timing. Our agent analyzes all of the postings in order to submit facilitation messages at a meaningful timing while trying to maintain the user’s satisfaction at the same level as that of human facilitation. On the other hand, human facilitators cannot post facilitation messages frequently. This is because a human facilitator cannot perceive all of the ongoing sub-threads and he/she has limited time to post elaborate facilitation messages.

Table 5 shows that the average response interval of the human facilitators was 377.5 seconds. Table 6 shows that the average response interval of the automated facilitation agent was 57.5 seconds. There is a significant difference at the 1% level by t-test. The automated facilitation agent can reply at the 1/6 short interval compared to the human facilitator in this experiment. Of course, this result is due to the fact that the agent facilitator has a quicker reaction time when compared to the human facilitator. From these results, we can say that our agent can analyze more postings. By analyzing more postings, the agent can post adequate facilitation messages, which corresponds to substantial support of our hypotheses.

5.1.2 Number of Nodes

Here, we summarize the number of nodes obtained at each layer of the discussion threads. A layer refers to the depth in the IBIS tree starting from the root node of

Table 6 Number of postings by facilitation agent

Discussion	Posting (All)	Posting (FA)	Posting (PA)	Posting (FA/PA)	Average Interval (PA → FA)
No. 1 B	154	23	131	14.90%	77.5
No. 2 C	241	36	205	14.90%	59.3
No. 3 F	167	43	124	25.70%	65.7
No. 4 G	153	39	114	25.50%	47.4
No. 5 J	109	28	81	25.70%	44.4
No. 6 K	113	29	84	25.70%	53.6
Average	156.17	33	123.17	21.10%	57.5

the discussion, or root issue. Table 7 illustrates the number of nodes for each level. When a ratio test is performed on the number of nodes for each node type among the participants, there was no significant difference at the 5% level. Thus, it became clear that both the human facilitator and the automated facilitation agent could incentivize the participants to post at a similar rate. In addition, when looking at the ratio for each node type, there was no significant difference at the 5% level. From this result, it became clear that the number of node types posted by the participants is similar to that posted by the human and the agent. This result supports our Hypothesis 1.

In the following, we look at the interactions between the types of posts by the participants and facilitators.

5.1.3 Types of Participant Replies given the Types of Facilitator Posts

We are particularly interested in the cases where the facilitation agent receives more ideas after requesting posts from the participants. Therefore, we looked at the types of replies for each posting type by facilitators, as illustrated in Table 8.

Table 7 Number of posts by facilitator

Node	Facilitator	Level									
		1	2	3	4	5	6	7	8	9	10
Issue	Human	0	14	9	2	1	0	0	0	0	0
	Agent	0	7	16	2	1	0	0	0	0	0
Idea	Human	0	93	131	82	32	9	1	0	0	0
	Agent	0	113	120	71	31	11	0	0	0	0
Pros	Human	0	0	60	27	15	3	2	1	0	0
	Agent	0	1	48	26	16	3	1	0	0	0
Cons	Human	0	24	57	35	12	2	1	0	0	0
	Agent	0	21	64	34	9	4	2	0	0	0
N/A	Human	0	8	48	26	13	0	1	0	0	0
	Agent	0	9	69	38	20	6	1	0	0	0
Issue (FA)	Human	5	0	24	20	6	2	1	0	0	0
	Agent	5	0	29	19	22	1	1	1	0	0
FA Response	Human	1	19	4	11	1	0	0	0	0	0
	Agent	4	14	12	6	1	0	0	0	0	0
FA Issue	Human	0	3	0	0	1	0	0	0	0	0
	Agent	0	3	0	0	0	0	0	0	0	0
FA Pros	Human	0	6	3	1	1	0	0	0	0	0
	Agent	0	8	2	3	1	0	0	0	0	0
FA Cons	Human	0	7	6	1	0	0	0	0	0	0
	Agent	0	15	9	3	4	0	0	0	0	0
FA Pros or Cons	Human	0	3	0	2	0	0	0	0	0	0
	Agent	0	0	2	1	0	0	0	0	0	0

Chi-test*: $p < 0.05$

Table 8 Participants' reply types for each facilitator node type

Node	Facilitator	Node Count	Issue	Idea	Pros	Cons	N/A	Total
Issue (FA)	Human	62	11	116	3	1	5	136
	Agent	90	7	102	0	0	10	119
FA Response*	Human	43	0	8	8	6	3	25
	Agent	48	0	22	1	2	2	27
FA Issue	Human	4	0	1	0	2	0	3
	Agent	3	0	0	0	0	1	1
FA Pros	Human	11	0	2	3	0	0	5
	Agent	20	0	0	7	1	2	10
FA Cons	Human	17	0	2	2	15	0	19
	Agent	41	2	1	1	17	1	22
FA Pros or Cons	Human	6	0	0	1	3	0	4
	Agent	3	0	0	0	0	0	0

Chi-test*: $p < 0.05$

For instance, the responses from participants in response to the agent message “FA Response” are 0 for issues, 8 for ideas, 8 for pros, 6 for cons, and 3 for N/A. Here, “FA Response” means that the facilitator is requesting a response from participants. In addition, the reply post from the agent is 0 for issues, 22 for ideas, 1 for pros, 2 for cons, and 2 for N/A. Hence, the facilitation agent has brought out more ideas and fewer pros and cons compared to the human facilitator. Here, there was a significant difference in “FA Response” at the 5% level by a chi-test. This proves that the agent can incentivize the participants to post more messages and to diversify their types. This result also supports our Hypothesis 1.

5.1.4 Types of Facilitator Replies given the Types of Participant Posts

Similarly, the following analysis found that the human and the agent facilitator elicited replies at the same ratio. Table 9 shows the number of the facilitator's replying node types for each posting node type by participants. For instance, there is 1 issue node in reply to the “FA Response” nodes in the case of the human facilitator. Again, “FA Response” means that the facilitator is requesting a response from participants. There was no significant difference at the 5% level for all posting nodes, which means that the human and the agent facilitator could elicit replies from the participant at the same ratio. It is clear that the agent and humans could incentivize the participants to submit posts at the same ratio, which again supports our Hypothesis 1.

5.1.5 Types of Participant Replies given the Types of Participant Posts

Here, the automated facilitation agent obtained more ideas to the issues raised by the participants. Table 10 shows the number of participants' replying node types for each posting node type by participants. In Table 10, we can see, for example, that

Table 9 Facilitator reply type for each participant node type

Node	Facilitator	Node Count	FA Response	FA Issue	Issue (FA)	FA Pros	FA Cons	FA Pros or Cons	Total
Issue	Human	28	1	0	2	1	1	0	5
	Agent	32	3	0	4	1	0	0	8
Idea	Human	391	25	4	22	4	10	2	67
	Agent	393	14	2	26	11	25	1	79
Pros	Human	146	5	0	1	1	1	0	8
	Agent	132	0	0	3	1	2	0	6
Cons	Human	143	3	0	14	1	1	0	19
	Agent	160	2	0	9	1	0	0	12
N/A	Human	102	0	0	2	0	0	1	3
	Agent	168	5	0	3	1	2	0	11

Chi-test: $p < 0.05$

Table 10 Number of participant reply types for each participant posting type

Node	Facilitator	Node Count	Issue	Idea	Pros	Cons	N/A	Total
Issue	Human	28	0	34	1	2	4	41
	Agent	32	0	20	0	6	4	30
Idea*	Human	391	5	66	74	53	24	222
	Agent	393	15	77	63	41	52	248
Pros	Human	146	1	4	7	0	0	12
	Agent	132	0	0	6	0	1	7
Cons	Human	143	2	30	2	1	7	42
	Agent	160	1	29	0	3	4	37
N/A*	Human	102	0	5	1	4	27	37
	Agent	168	2	0	4	4	21	31

Chi-test*: $p < 0.05$

there are 34 issue nodes in reply to the idea nodes in the case of the human facilitator. In the case of the facilitation agent, we obtained three times more idea nodes (15) than in the case of the human facilitation (5). There was a significant difference at the 5% level, using a chi-test, in the cases of idea and N/A. Consequently, the considerable increase in the number of proposed ideas supports our Hypothesis 1.

5.1.6 Measuring Facilitation Effect by Questionnaires

In order to understand the differences in the psychological effects of facilitation, we conducted a questionnaire evaluation using a five-level Likert scale for the discussions and facilitators (Joshi et al. 2015). A questionnaire was sent to all participants after the discussions. We received 62 answers about the agent's facilitation and 63 answers about the human's facilitation. The averages of the answers were calculated.

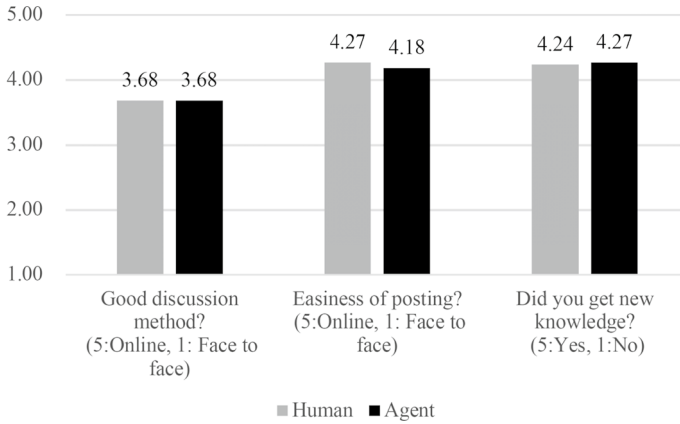


Fig. 9 Evaluation of Discussion by Questionnaire (each N=63, t-test *: $p < 0.05$)

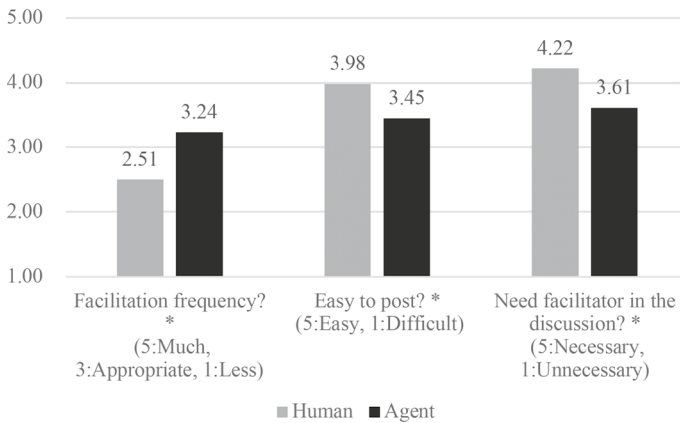


Fig. 10 Evaluation of the facilitation using questionnaires (N=63, t-test* $p < 0.05$)

Figure 9 shows the evaluation value for the discussion, and Fig. 10 shows the evaluation value for the facilitator. The content of the questions, the rating scale, and the t-test results are also shown in the figures. Here, the Likert scale is used to precisely assess the level of satisfaction of the users with respect to agent facilitation and human facilitation, each evaluated separately. The users rate each method using the same Likert scale. The rating is also used to compare user satisfaction between the two facilitation methods as shown in Fig. 9.

In Fig. 9, the evaluation of the criterion “Good discussion method?” in both cases gave 3.68. In addition, the evaluations on “Easiness of posting?” are 4.27 for the case of human facilitation and 4.18 for the agent. The evaluations on “Did you get new knowledge?” are 4.24 for the human case and 4.27 for the agent case. Results of these questions are not significantly different at the 5% level.

The evaluations for “Facilitation frequency?” shown in Fig. 10 are 2.51 for the human case and 3.24 for the agent case. This suggests that the agent adopted an appropriate frequency of facilitation when the human facilitation frequency was smaller. This means that participants were comfortable with the frequency of facilitation even if our agent responded more often than the human facilitator. The ease of posting criterion gave 3.98 for the case of human facilitation and 3.45 for the agent. The subjects found it easier to post when the human facilitator was present. Concerning the need for the facilitator in the discussion, the human facilitator obtained 4.22 and the facilitation agent obtained 3.61. This difference could be caused by the short duration (45 and 60 minutes) of the discussions and the inability to see the effect of the agent over prolonged periods of time, as in the case of the large-scale social experiment. That is, the effect of humans in brief discussions is stronger than that of artificial agents. This is due in part to the linguistic abilities of humans when compared to the limited, rule-based abilities of the artificial agent. But overall, the scores are higher than 3, which suggests that the participants expressed their satisfaction with the facilitators. This result supports our Hypothesis 3.

6 Conclusion

We presented our implementation of a crowd-scale discussion support system based on an automated facilitation agent that can extract argumentative content, analyze it, and post facilitation messages. We performed large-scale experiments within the city of Nagoya in the context of local governmental meetings. We analyzed several aspects of the experiments, including the number of postings, the satisfaction of the participants, the facilitation agent performance, and the structure of the discussion trees. We found that the automated facilitation agent worked in a similar manner to human facilitators. Moreover, we found that the results were optimal when human and agent facilitators worked together. In addition to the large-scale experiment, we conducted small-scale controlled experiments to verify the direct effects of the automated facilitation.

We verified the following hypotheses through the large-scale and small-scale experiments:

Hypothesis 1 The agent can incentivize the participants to submit more postings and to diversify these postings.

Hypothesis 2 When the agent works collaboratively with the human facilitator, the overall performance of the facilitation increases.

Hypothesis 3 The satisfaction of the participants in the discussions facilitated by the agent is more than average. This means that the participants were at least not dissatisfied with the discussion facilitated by the agent.

We consequently obtained the following conclusions.

- The agent had 1.4 times more replies than in the case of a human and 6 times shorter response intervals.
- There was no difference in the ratio of the number of nodes in the IBIS structure created by the discussion. This means that our automated agent could facilitate discussions as effectively as a human facilitator does in order to extract the IBIS elements from the discussion content.
- When replying to ideas, in the case of the agent, there were about three times more issues than in the case of humans.
- There was no difference in the ratio of the number of reply nodes between automated facilitation and human facilitation.
- The use of the agent produced more ideas for any given issue.
- From the results of the questionnaires, and from the frequency of the facilitation, we found that the frequency of the agent facilitation suited the participants' reflection and comfort time, despite the fact that the agent responded more often than the human facilitator as shown in Fig. 10. The chosen timing of the automated facilitation could be optimized in future experiments to account for cognitive differences in the participant pool, or the difficulty of the discussed topics.

As a future direction, we would like to investigate the extent to which the agent can polarize a discussion, thus leading to ethical issues. Although our current agent is implemented to generate fair facilitation, such considerations need to be scrutinized thoroughly within controlled experiments. To this end, we would like to investigate how much bias people can accept in a given discussion. We are currently investigating such questions in similar social experiments (Haqbeen et al. 2020; Hadfi et al. 2020, 2021). As a second direction, we would like to investigate how the platform could be used to automatically foster inclusion by focusing on subsets of participants such as social minorities. Finally, we would like to devise mechanisms that allow the agent to keep track of the effects of its messages and adaptively adjust its facilitation policies.

Acknowledgements This work was supported by JST CREST Grant Numbers JPMJCR15E1 and JPMJCR20D1, Japan.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Almansor EH, Hussain FK (2020) Survey on intelligent chatbots: State-of-the-art and future research directions. In: Barolli L, Hussain FK, Ikeda M (eds) *Complex, Intelligent, and Software Intensive Systems*, Springer International Publishing, pp 534–543
- Cabrio E, Villata S (2018) Five years of argument mining: A data-driven analysis. *IJCAI* 18:5427–5433
- Chen H, Liu X, Yin D, Tang J (2017) A survey on dialogue systems: recent advances and new frontiers. *ACM SIGKDD Explorations News* 19(2):25–35
- Churchman CW (1967) *Wicked problems*. Management Science
- Conklin J (2003) *Dialog Mapping: Reflections on an industrial strength case study*, Springer, London, London, pp 117–136
- Conklin J, Begeman ML (1989) gibis: A tool for all reasons. *J Am Soc Inf Sci* 40(3):200–213
- Csaky R (2019) Deep learning based chatbot models. arXiv preprint [arXiv:190808835](https://arxiv.org/abs/190808835)
- Dasgupta S, Granger M, McGarry N (2002) User acceptance of e-collaboration technology: An extension of the technology acceptance model. *Group Decis Negot* 11(2):87–100
- Davis FD (1985) A technology acceptance model for empirically testing new end-user information systems: Theory and results. PhD thesis, Massachusetts Institute of Technology
- Hadfi R, Kawamura N, Sakai A, Yamaguchi N, Ito T (2020) A study on the polarisation effects of biased conversational agents in online debates. In: *Proceedings of the Annual Conference of JSAI JSAI2020*
- Hadfi R, Haqbeen J, Sahab S, Ito T (2021) Argumentative conversational agents for online discussions. *J Syst Sci Syst Eng*. <https://doi.org/10.1007/s11518-021-5497-1>
- Haqbeen J, Ito T, Hadifi R, Nishida T, Sahab Z, Sahab S, Roghmal S, Amiryar R (2020) Promoting discussion with AI-based facilitation: Urban dialogue with Kabul City. In: *ACM Collective Intelligence*
- Hong H, Du Q, Wang G, Fan W, Xu D (2016) Crowd wisdom: The impact of opinion diversity and participant independence on crowd performance
- Hori H, Matsui Y (2001) The psychological measurement scales III. *Saiensu-sya*
- Hori H, Yamamoto M (2001) The psychological measurement scales I. *Saiensu-sya*
- Hori H, Yoshida F (2001) The psychological measurement scales II. *Saiensu-sya*
- Hori H, Sakurai S, Matsui Y (2007) The psychological measurement scales IV. *Saiensu-sya*
- Hori H, Yoshida F, Miyamoto S (2011) The psychological measurement scales V. *Saiensu-sya*
- Iandoli L, Klein M, Zollo G (2007) Can we exploit collective intelligence for collaborative deliberation? the case of the climate change collaboratorium
- Ito T (2018) Towards agent-based large-scale decision support system: The effect of facilitator. In: *The 51st Hawaii International Conference on System Sciences (HICSS2018)*
- Ito T, Imi Y, Ito T, Hideshima E (2014) COLLAGREE: A facilitator-mediated large-scale consensus support system. *Collective Intelligence* 2014
- Ito T, Imi Y, Sato M, Ito T, Hideshima E (2015) Incentive mechanism for managing large-scale internet-based discussions on collagree. In: *Collective Intelligence*
- Ito T, Shibata D, Suzuki S, Yamaguchi N, Nishida T, Hiraishi K, Yoshino K (2019) Agent that facilitates crowd discussion. In: *Collective Intelligence*
- Joshi A, Kale S, Chandel S, Pal DK (2015) Likert scale: Explored and explained. *Current J Appl Sci Technol* pp 396–403
- Kawase S, Ito T, Otsuka T, Sengoku A, Shiramatsu S, Matsuo T, Oishi T, Fujita R, Fukuta N, Fujita K (2018) Cyber-physical hybrid environment using a largescale discussion system enhances audiences' participation and satisfaction in the panel discussion. *The IEICE Trans Inf Syst* E101.D(4):847–855
- Kim Y (2014) Convolutional neural networks for sentence classification. arXiv preprint [arXiv:14085882](https://arxiv.org/abs/14085882)
- Kunz W, Rittel HW (1970) Issues as elements of information systems. *Tech. rep.*, CiteSeerX 10.1.1.134.1741
- Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C (2016) Neural architectures for named entity recognition. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, pp 260–270, 10.18653/v1/N16-1030, <http://aclweb.org/anthology/N16-1030>
- Lawrence J, Reed C (2017) Mining argumentative structure from natural language text using automatically generated premise-conclusion topic models. In: *Proceedings of the 4th Workshop on Argument Mining*, pp 39–48

- Lewis JR (2018) Measuring perceived usability: the CSUQ, SUS, and UMUX. *Int J Human-Comput Interact* 34(12):1148–1156
- Malone TW (2018a) *Superminds: The surprising power of people and computers thinking together*. Little, Brown and Company
- Malone TW (2018b) *Superminds: The surprising power of people and computers thinking together*. Little, Brown Spark
- Malone TW, Klein M (2007) Harnessing collective intelligence to address global climate change. *Innovations: Technology, Governance, Globalization* 2(3):15–26
- Mkude CG, Pérez-Espés C, Wimmer MA (2014) Participatory budgeting: A framework to analyze the value-add of citizen participation. In: 2014 47th Hawaii International Conference on System Sciences, IEEE, pp 2054–2062
- Nguyen H, Litman D (2016) Context-aware argumentative relation mining. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol 1, pp 1127–1137
- Nishida T, Ito T, Ito T, Hideshima E, Fukamachi S, Sengoku A, Sugiyama Y (2017) Core time mechanism for managing large-scale internet-based discussions on COLLAGREE. In: *the Proceedings of the 2nd IEEE International Conference on Agents (IEEE ICA2017)*
- Nishida T, Ito T, Ito T (2018) Verification of effects using consensus-building support system in continuous workshops for city development. *J Sci Design*
- Noble D, Rittel HWJ (1989) Issue-based information systems for design. Working paper / Institute of Urban and Regional Development, University of California, Berkeley (492)
- Nuez Ezquerro A (2018) *Implementing chatbots using neural machine translation techniques*. B.S. thesis, Universitat Politècnica de Catalunya
- Palau RM, Moens MF (2009) Argumentation mining: the detection, classification and structure of arguments in text. In: *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, ACM, pp 98–107
- Savaget P, Chiarini T, Evans S (2019) Empowering political participation through artificial intelligence. *Sci Public Policy* 46(3):369–380
- Sengoku A, Ito T, Takahashi K, Shiramatsu S, Ito T, Hideshima E, Fujita K (2016) Discussion tree for managing large-scale internet-based discussions. In: *Collective Intelligence*
- Stab C, Gurevych I (2014a) Annotating argument components and relations in persuasive essays. In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp 1501–1510
- Stab C, Gurevych I (2014b) Identifying argumentative discourse structures in persuasive essays. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp 46–56
- Stab C, Gurevych I (2017) Parsing argumentation structures in persuasive essays. *Comput Linguist* 43(3):619–659
- Suzuki S, Yamaguchi N, Nishida T, Moustafa A, Shibata D, Yoshino K, Hiraishi K, Ito T (2019) Extraction of online discussion structures for automated facilitation agent. *JSAI Int Sessions Proc JSAI2019:2F1E302–2F1E302*
- Tavanapour N, Bittner E (2018) Automated facilitation for idea platforms: Design and evaluation of a chatbot prototype. In: *Proceedings of Thirty Ninth International Conference on Information Systems (ICIS2018)*
- Tavanapour N, Poser M, Bittner E (2019) Supporting the idea generation process in citizen participation - Toward an interactive system with a conversational agent as facilitator. In: *Proceedings of Twenty-Seventh European Conference on Information Systems (ECIS2019)*
- Varia J, Mathew S, et al. (2014) Overview of Amazon Web Services. *Amazon Web Services* 105
- Venkatesh V, Davis FD (2000) A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Manage Sci* 46(2):186–204
- Wallace RS (2009) The anatomy of ALICE. In: *Parsing the Turing Test*, Springer, pp 181–210
- Weizenbaum J (1966) ELIZA-a computer program for the study of natural language communication between man and machine. *Commun ACM* 9(1):36–45
- Williams SN (2010) A twenty-first century citizens POLIS: Introducing a democratic experiment in electronic citizen participation in science and technology decision-making. *Public Underst Sci* 19(5):528–544
- Wittig M, Wittig A (2018) *Amazon Web Services in action*. Simon and Schuster

- Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW (2010) Evidence for a collective intelligence factor in the performance of human groups. *science* 330(6004):686–688
- Yamaguchi N, Nishida T, Ito T (2019) An annotation scheme for developing prototype automated facilitation software. In: 2019 Portland International Conference on Management of Engineering and Technology (PICMET), IEEE, pp 1–7
- Yan Z, Duan N, Chen P, Zhou M, Zhou J, Li Z (2017) Building task-oriented dialogue systems for online shopping. In: Thirty-First AAAI Conference on Artificial Intelligence
- Zhang A, Culbertson B, Paritosh P (2017) Characterizing online discussion using coarse discourse sequences. In: Proceedings of the International AAAI Conference on Web and Social Media, vol 11
- Zviran M, Erlich Z (2003) Measuring IS user satisfaction: Review and implications. *Commun Assoc Inf Syst* 12(1):5

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.