Toward Agent-based Large-scale Decision Support System: The Effect of Facilitators

Takayuki Ito
Department of Computer Science,
Frontier Research Institute of Information Science,
Nagoya Institute of Technology, Japan
ito.takayuki@nitech.ac.jp

Abstract

Good discussions are essential for group decisions, especially when a group has many people. Providing good support is critical for establishing and maintaining coherent discussions that avoid such anti-social behaviors as flaming, which has been observed in some large discussion groups. We have developed a large-scale online decision support system that has facilitator support functions and deployed it in case studies for several real-world online discussion supports. In this paper, we propose a facilitator-mediated online discussion model to lead discussions in a better direction to reach decisions. Our ultimate goal is an automated facilitator agent that can help participants reach reasonable decisions. In reality, online discussion is often plagued by "flaming," which is posting or sending offensive messages during a discussion. Such flaming phenomena have been focused on as anti-social behavior in online discussion forums. After several cases studies, we learned several lessons. Critically, in all of our social experiments, no flaming has been observed in our facilitator-mediated decision support system. Our insights also suggest that the social presence of a facilitator would greatly affect participant behavior.

1. Introduction

Online discussion forums are receiving much attention because they are likely to be one of the next generation methods for open and public democratic citizen forums. Such forums require systematic methodologies that can efficiently achieve a consensus, reasonably integrate ideas, and discourage flaming. We developed an intelligent crowd decision-making support system that has facilitator support functions and deployed it for several real-world online discussion supports as case studies.

We were inspired to enter this area by several ongoing intriguing projects, of which the following are representative. The goal of the Climate CoLab [2][5][6], which is one of the most famous web-based collective intelligence projects, is to harness the collective intelligence of thousands of people worldwide to address global climate change. Like Wikipedia and Linux, MIT CCI developed a crowdsourcing platform where citizens work with experts to create, analyze, and select detailed proposals that tackle climate change. This system defined several steps, including "proposal creation," "finalist selection," "proposal revisions," "voting," and "presentations to potential implementers" to integrate innovative opinions with crystalized ideas that are implementable. Deliveratorium [1][3][4] is another project where people submit ideas by following an argumentation map, which is a kind of discussion structure through which people frame their ideas. With structured argumentation maps, Deliveratorium makes it possible to clearly show the entire relations among ideas and opinions. Such structuring can be done even if the opinions are completely divided.

We propose a facilitator-mediated online discussion model to take discussions in better directions. Online discussion often degenerates into flaming, which is posting or sending offensive messages during a discussion. Such flaming phenomena have been criticized in online discussion forums because they discourage people from engaging in online discussion forums. Such forums need more effective ways to avoid flaming.

In fact, real-world workshops or town meetings among citizens are usually coordinated by a facilitator who coordinates, leads, integrates, classifies, and summarizes discussions that might reach an acceptable consensus or an alternative. The main issue is that facilitators must be supported so that they can manage large-scale discussions. Even though professional facilitators joined our experiment's project, this was their first experience to harness discussions on the internet that involved over

URI: http://hdl.handle.net/10125/49934
ISBN: 978-0-9981331-1-9

100 people. To aid these facilitators, our system provides support functions for them.

We learned several lessons from our previous cases studies. The most important achievement in our social experiments is that no flaming was observed. Also, we obtained insights that suggest that the social presence of a facilitator might greatly affect participant behavior. Social presence refers to the feeling of being socially present with another person at a remote location.

The remainder of this paper is organized as follows. Section 2 presents the importance of facilitators who mediate large-scale internet discussions. Section 3 introduces the current implementation of our system, and Section 4 presents our case studies of social experiments of online discussions. Section 5 discusses the lessons learned from the case studies and provides further discussion on automated facilitators. Finally, Section 6 makes concluding remarks.

2. Facilitator-mediated Online Discussion

We propose a facilitator-mediated online discussion model to lead discussions to better results. Online discussion often degenerates into flaming, which is posting or sending offensive messages during a discussion. Such flaming phenomena have been criticized in online discussion forums because they discourage participants from joining online discussion forums. Online discussion forums need more effective ways to avoid flaming.

T. W. Malone et al. [7] described the importance of a hierarchy for harnessing crowds to produce collective intelligence and classified the genomes of collective intelligence into several types based on four categories: Who, Why, What, and How. They described the Crowd and Hierarchy Genes as the foundation for the Who category for crowd-based intelligence (collective intelligence).

Crowd Gene: "Using the Crowd gene, activities can be undertaken by anyone in a large group who chooses to do so, without being assigned by someone in a position of authority." "Reliance on the Crowd gene is a central feature of web-enabled collective intelligence systems. In fact, all of the examples we studied include at least one instance of the Crowd gene - at least one task where anyone who chooses to can participate".

Hierarchy Gene: "When the conditions for using a Crowd aren't met, you can use a Hierarchy (often meaning: "management")". "For instance, if only a few people have the skills you need, and you already know who they are, you can assign the task to them directly. Or if you can't figure out how to prevent people in a Crowd from sabotaging your goals, you many need to use a Hierarchy instead. In this sense, you can think of the traditional Hierarchy gene as the "default" gene, the one to use when you can't figure out how to get a Crowd gene to work."

For example, in the Wikipedia project, since anyone can edit or add/delete articles, this situation resembles the Crowd Gene. On the other hand, these activities are all monitored and overseen by moderators, whose actions reflect the Hierarchy Gene. In the Linux project, anyone can generally post and edit source codes, like the Crowd Gene. Linux Torvalds et al. decided which of the many modules submitted by people to actually include in the next release, which is the Hierarchy Gene.

In this paper, we propose using a facilitator as a hierarchy gene for large-scale online discussions to discourage flaming. A facilitator usually leads collaborative discussions so that members can achieve effective results after discussions. A facilitator [8] is defined as a process guide, someone who simplifies a process or makes it more convenient. Facilitation enables a group of people to achieve its own purpose in its own agreeded-upon way. A facilitator is especially critical for collaborative discussion in the world. For example, local governments often hold facilitator-mediated workshops to gather opinions from citizens.

Online discussion should also be mediated by a facilitator and taken in an acceptable direction to obtain effective results after discussion. However, to the best of our knowledge, no such systems currently exist. This is because the nature of online discussion is completely different from physical (face-to-face) discussion.

In online discussions, the amount of participants is usually large and they are often located remotely and cannot see each other. Online discussions often become dispersed, multi-threads, and asynchronous and might branch into many sub-discussions. The response times between posts might be very long, too. On the other hand, physical discussions are continuous, single-threaded, and synchronous. It is very difficult to simultaneously have several threads in real discussions; they must be synchronous because all participants are attending a single discussion thread.

The existing online discussion systems are usually based on the Crowd Gene. Their characteristics, i.e., dispersive, multi-threaded, and asynchronous, are the features of Crowd Gene. Current online discussion systems often fail to avoid flaming because the Crowd Gene does not diligently discourage it.

One obvious way to avoid flaming is to observe and manage discussions from a higher level: by introducing the Hierarchical Gene. But current online discussion systems have no such observation or management, e.g., a Hierarchical Gene, although Wikipedia and other successful social computing systems do have their own versions of it. Thus, introducing a Hierarchy Gene into online discussions is promising and reasonable.

In this research, we introduce a facilitator as a Hierarchy Gene into online discussions. Facilitators can manage online discussions and lead and motivate participants to have productive and fruitful discussions. They also observe postings, replies, and other actions by participants and identify individuals who are engaging in anti-social behavior. Since installing facilitators might enhance the possibility of online discussions, we have been introducing them into our social experiments on online discussions. We have not yet experienced any flaming in more than ten social experiments.

On the other hand, several real problems have surfaced for introducing facilitators into online discussions. Because of the characteristics of online discussions, human facilitators have difficulty facilitating them. Since no actual expert facilitators exist for managing such online discussions, we have developed several facilitator support functions. For example, incentive mechanisms motivate participants to post opinions by assigning points (virtual money) to their actions. Such facilitation support functions have successfully helped facilitators. We are also trying to include intelligent algorithms, e.g., NLP-based analysis of discussion contexts, which can intelligently support facilitators.

3. An Intelligent Crowd Decision-Making Support System

3.1. Facilitator Support Functions

We implemented an intelligent crowd decision-making support system called COLLAGREE: COLLective, COLLaborative and AGREEment. Its first version was implemented in 2013, and it has also been upgraded and branched into a couple of slightly different versions. Fig. 1 shows a typical user-interface employed by both facilitators and participants. Numbers (1)-(5) correspond to numbers 1-5 in red circles. The following are its typical functions, and we adopted (1), (2), and (3) to support facilitators. (1) Agreement or disagreement analysis for a comment is shown so that facilitators can understand whether a discussion thread is positive or negative. (2) Keywords are highlighted so that facilitators can quickly understand what words are being focused on and which are important. (3) With facilitation tabs, facilitators can input their instructions to participants. (4) Opinions and discussions can be searched for and reordered. (5) Issue tags allow participants to add to each opinion and comment so that they can search for them afterwards. (6) E-mail reminders are given to participants about related future events.

Fig. 1 Basic user-interface: each number shows characteristic functions explained in 3.1. Users can post their opinions in (3). (2) is a keyword extractor, and (1) is an automatic sentiment analyzer.

3.2 Incentive Mechanisms

Incentive mechanisms have been greatly focused on in the field of social computing. Incentives in social network are very effective for efficient information gathering and finding. One of the most well-known success stories about incentives is the 2009 DARPA Network Challenge, where competing teams were asked to locate ten red weather balloons placed around the continental United States. Using a recursive incentive mechanism that both spread information about the task and incentivized individuals to act, the MIT team won the competition by finding all ten balloons in less than 9 hours [18].

We developed incentive mechanisms [10][12] for participants and employed both incentives and facilitators to harness collective intelligence. While facilitators, who are one element of a hierarchical management, can be seen as a top-down approach to produce collective discussions, incentive is a bottom-up approach.

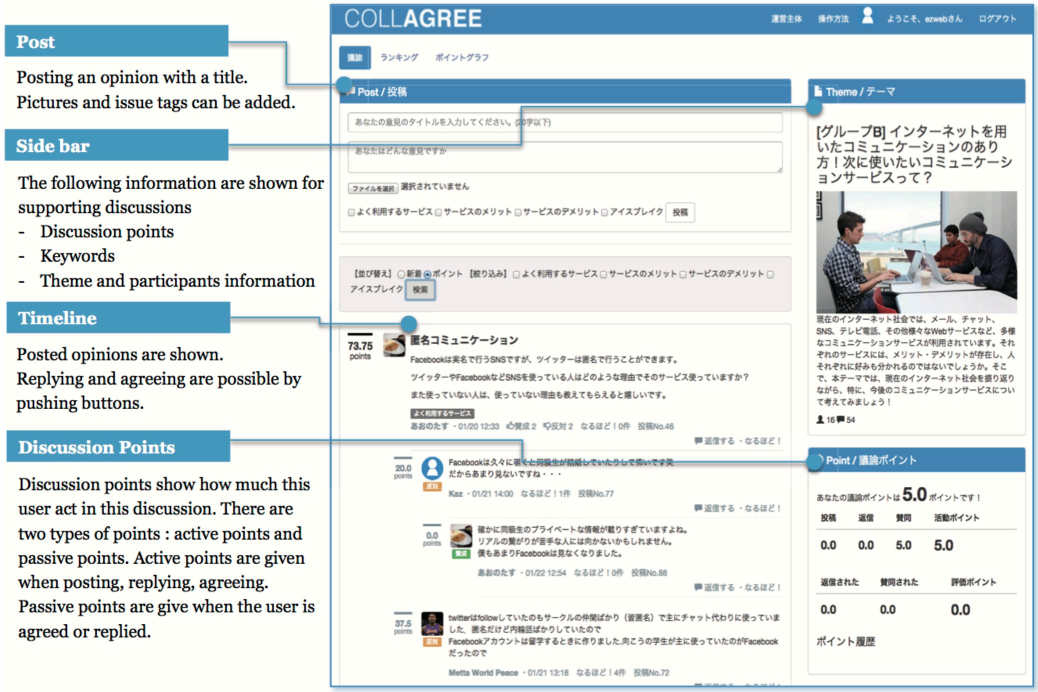


Fig. 2 UI with discussion points: another user-interface with discussion points obtained by this user.

We also identified what facilitators require from participants who want to contribute to online discussions. This is because the information gained by facilitators in online discussions is drastically less than in face-to-face discussions. After several social experiments, such requirements will become more important. Thus, we implemented several functions that incentivize participants to post comments/opinions to our system.

As our first incentive mechanism [12], we adopted discussion points. Fig. 2 shows a user-interface of our system which has an initial incentive mechanism. Users can post opinions/comments through the top boxes. The side bar has functions for showing discussion points, user rankings of discussion points, highlighted keywords, themes, and participant information. The timeline shows the sequence of opinions and replies to them. Users can re-order the sequence by points, keywords, etc. By re-ordering the points, users can easily find the focused on and noteworthy discussions from the timeline. Fig. 2 gives a detailed description of the discussion points as an incentive mechanism in our system. We have two types of discussion points: action (active) and evaluated (passive). Action points include posts, replies, and agreements, all of which are obtained when a user posts, replies, and agrees. We expect these points to encourage users to actively post, reply, and agree.

Evaluated points are those to which others replied and with which they agreed. When posted comments are replied to or agreed to, they have been evaluated, suggesting that they have discussion value. Thus, we give discussion points to these comments. We expect that evaluated points will encourage participants to submit more thoughtful comments to get replies or agreements. We adopted a recursive (or propagating) pointing idea for the agreed points; if comment X is agreed with, then its ancestor (parents) comments are also evaluated because these ancestor comments might have produced comment X that was agreed with. This incentivizes the participants to solicit agreements and replies.

3.3 Quality of Opinions

The initial incentive mechanism described in the previous section did not use the quality of opinions. We observed that facilitators want different opinions for different phases in a discussion. For example, in the beginning (divergence) phase, they want to identify the variety of different and diverse opinions, while in the final (convergence) phase they want to summarize the discussion. Thus, they prefer concentrated and similar opinions.

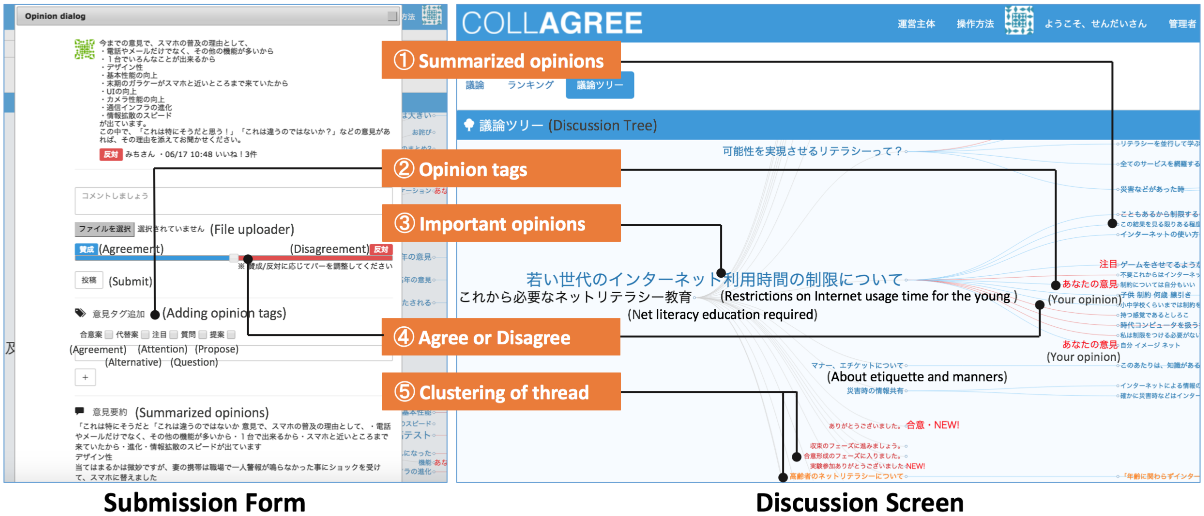


Fig. 3 Right window shows a discussion graph, which is semi-automatically generated by discussion forum in left window.

We previously proposed a discussion point function based on the quality of opinions [10], as judged by content and posted timing. Five members of the Japan Facilitation Association defined these two elements. Since they also participated as facilitators in past experiments, they were familiar with our system.

Criteria for quality of posted opinions:

- ✓ Content: opinions that fit the particular phase (divergence, convergence, and agreement) are highly evaluated.
- ✓ Posted timing: quick replies and posts when the discussion is stagnating are highly evaluated.

Evaluation method of opinion content:

The features of new posts are judged as either divergence or convergence. The system uses the word weighting algorithm BM25 to judge them [19]. First, it extracts all the nouns from the new post. Then it extracts keywords using BM25 from the previous discussion. After that, it determines whether noun w_i and keyword e_j match. When w_i is not equal to e_j, M_1 points are given to the user. When w_i equals e_j, points are given to user:

$$P_{w_i} = \begin{cases} M_1 & (w_i \neq e_j) \\ \sum_{j=1}^{n} score(e_j, D) \times M_2 & (w_i = e_j) \end{cases}$$

w_i is the noun that was extracted from the new posts. e_j is the keyword that was extracted from document set $D=\{ d_1, d_2, \ldots, d_n\}$. $score(e_j,D)$ is the importance of e_j that was calculated by BM25. The sum adopted in this process for all the nouns is P_d, which is an additional discussion point given for opinion content. M_1 and M_2 in the formula are parameters that can be freely set. M_1 corresponds to the divergence, and M_2 corresponds to the convergence.

When divergence should be emphasized, M_1 needs to be higher. In contrast, when convergence should be emphasized, M_2 needs to be higher. In our system, these parameters are changed in accordance with the discussion phase as follows:

- Divergence phase: M_1=0.7, M_2=20.
- Convergence phase: M_1=0.5, M_2=25.
- Agreement phase: M_1=0.3, M_2=30.

P_d is the sum adapted for this process for all the nouns. It is an additional discussion point given for opinion content:

$$P_d = \sum_{i=1}^{n} P_{w_i}$$

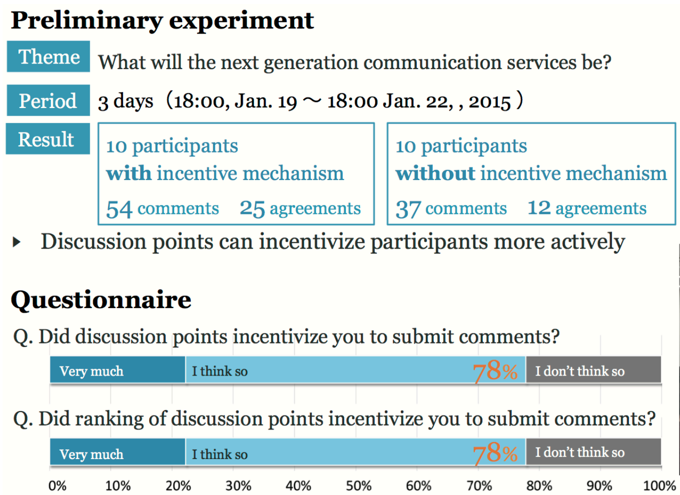
As a result, an opinion that fits the discussion phase will obtain a high rating.

Evaluation method of posted timing:

A reply within 30 minutes will also be given five discussion points. When there are no new posts for more than three hours, a new post will be given ten discussion points. The sum of the additional points on the posted timing is P_t.

Table 1 Point setting

Action points	"Posting" points	When I post a comment, I can get posting points.	10 + P_s + P_t points
	"Replying" points	When I post a reply, I can get replying points.	5 + P_s + P_t points
	"Agreeing" points	When I agree once, I can get agreeing points.	5 points
Evaluated points	"Replied-to" points	When my opinion gets one reply, I can get "replied-to" point.	15 points
	"Agreed-to" points	When my opinion gets one agreement, I can get "Agreed-to" points.	5 points



Preliminary experiment and it results

Large-scale Aichi-prefecture City-design discussion

Fig. 4 Aichi Design League: we conducted a large-scale Aichi Prefecture city-design discussion with our system (pictures on right). Before this social experiment, we conducted a lab-scale preliminary experiment (on left) that shows discussion points that incentivized people to join discussions.

By using P_d and P_t, we can now set discussion points adaptively, as shown in Table 1. A previous work [10] reported that this discussion point mechanism works well in experiments.

3.3 Discussion Graphs

The initial implementation of our system faced problems, including "High viewing cost" and "Creating a draft agreement," according to the previous advanced research. "High viewing cost" means the number of posted opinions is too unwieldy when a discussion becomes large. Therefore, participants have difficulty grasping the discussion contents. "Creating a draft agreement" means that the opinions posted by participants weren't integrated because of anonymity and asynchronicity, which are features of online discussions. We solved these two problems by introducing a **Discussion Tree** to support large-scale opinion gathering.

We proposed a new Discussion Tree-based discussion method for opinion gathering in large-scale discussions on the web [9]. A Discussion Tree is a tree diagram that visualizes a discussion's flow on the basis of the reply relationships in the conversations to make discussions more efficient. Discussion Trees are commonly used as a facilitation tool for face-to-face workshops.

We introduced our Discussion Tree method in online discussions and extended it so that it can support such discussions. Participants can use it to grasp a discussion's flow and issues and to help them grasp the discussion contents. In addition, it can provide the positioning and the mutual relationships of the opinions to participants so that they can easily create a draft agreement.

A Discussion Tree is a tree diagram that visualizes a discussion's flow on the basis of the reply relationships in conversations to make the discussion more efficient. A major difference of Discussion Trees from argumentation map used in Deliberatrium [1][3][4] is that a Discussion Tree is generated automatically from chunks of texts submitted freely by participants on discussion forums. In addition, our Discussion Tree uses text-mining techniques to present critical keywords in the discussion contents. These features avoid imposing a load on participants when the argumentation map requests them to manually create a logical argumentation structure. An automatically created Discussion Tree edited by facilitators can create an accurate Discussion Tree. Therefore, participants can smoothly discuss by viewing a Discussion Tree.

Figure 3 shows a Discussion Tree created for each discussion theme in COLLAGREE as well the following functions. The numbers below correspond to the numbers with red circles in Fig. 3:

(1) summarizing opinions display function, (2) opinion tag adding function, (3) important opinions display function for helping readers grasp discussion content, (4) agree or disagree display function, and (5) clustering of thread functions for creating draft agreements. We implemented each function on the basis of the results of a preliminary experiment that has a Discussion Tree.

The nodes of the Discussion Tree in Fig. 3 are each opinion, and the links show the reply relationships. The size of each node denotes the opinion's significance. The text displayed in the node is a summarized opinion. The node's color is different for each classification. Blue links mean agreement with an opinion, and red links denote disagreement. Another work [9] demonstrated experimental results that show the effectiveness of Discussion Trees.

3.4 Toward Intelligent Facilitator

Implementations on multiple new functions remain on-going. We are implementing an intelligent software agent that can facilitate human discussion. To this end, in the current project, we are focusing on a variety of directions.

We must clarify the principle of facilitators' action selection for alleviating the cognitive loads of human facilitators during web-based discussion. We assume that a facilitator selects an action that maximizes the expected utility corresponding to his/her intention. A previous work [20] described a particular utility function, i.e., the number of non-facilitator utterances in succeeding utterances within a certain period of time, corresponding to a facilitator intention to promote active discussions. The expected utility can be estimated with Random Forest Regression that is trained by a discussion corpus. The experimental results showed that actions selected by the expected utility were consistent with the intention represented by the expected utility. However, the actual actions of the human facilitators were inconsistent with the actions selected by the expected utility. These results indicated that we need to investigate the diverse intentions of facilitators by trying diverse utility functions.

We proposed a method for generating facilitator questions from the extracted opinions of discussion participants in the preceding context [21]. First, the opinions in the preceding context are extracted using clue expressions. A facilitator's question is generated with pattern-matching rules using the case structure of a predicate in the extracted opinion. This method assumes that an appropriate type of question can be selected with a superficial case structure. We evaluated our method through a subjective experiment. The results show that our method has the potential to be utilized for developing autonomous facilitator agents.

4. Case Studies

4.1. Actual Field Social Experiments

Because our idea is to facilitate actual discussions, we believe that evaluation by people in actual fields is the most important aspect for finding valuable insights and new ideas that can contribute to society. We have conducted several social experiments to evaluate our new ideas on a system by real people. Basically, we conducted a mini-size laboratory level experiment to investigate how well our new functions work. If they work well, we can introduce our current system to actual social experiments. If not, we will analyze the reasons for the failure, fix the problems, and apply the new system to another actual field. We review our social experiments as case studies.

We focus on the facilitator effects of online discussion and whether flaming occurred.

4.1. Nagoya Next Generation Total City Planning 2018 [12]

Nagoya in Aichi Prefecture has over three million people. After three months of preparation with city officials, we created an internet-based town meeting on its planning. Nagoya's mayor announced this project in newspapers and on TV as an actual town meeting of the Nagoya Next Generation Total City Planning for 2018. Our experiment ran on the COLLAGREE system during a two-week period from 12:00 on Nov 19, 2013 to 12:00 on Dec 3, 2013 with nine experts from the Facilitators Association of Japan. The participants discussed the following four categories about their perception of an ideal city based on the Nagoya Next Generation Total City Planning 2018: a city where human rights are respected and everyone lives happily; a city that is resistant to disasters where people can live safely; a city with a comfortable urban environment in harmony with nature; and a city with vitality and charm. Over the two weeks, our system gathered 266 registered participants, 1,151 opinions, 3,072 visits, and 18,466 views. The total of 1,151 opinions greatly exceeded the 463 opinions obtained by previous real-world town meetings. On the right in Fig. 4, the results of questionnaires are shown. Both participants and facilitators realized the importance of an online discussion forum to gather opinions for local government.

We did not observe any flaming itself. However, a couple of participants who just posted their own opinions failed to follow the main discussion streams. Even though such actions resemble one kind of flaming, they did not cause any deleterious effects to the other participants or the discussion itself.

4.2. Aichi Design League [11]

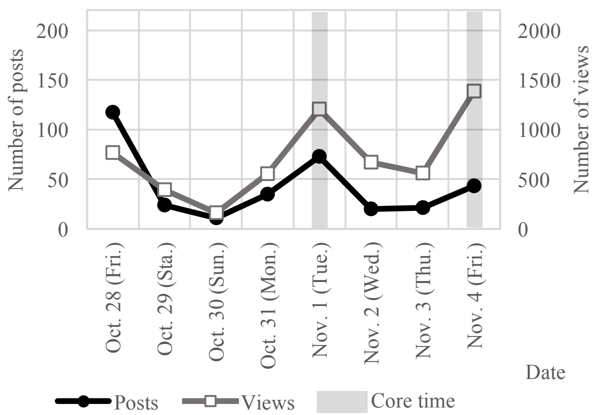


Fig. 7 Core-time effect: we set two core times on Nov. 1 and 4. People were obviously incentivized to act during core times.

We also did a large-scale experiment with local governments in Aichi prefecture. In this experiment, the participants discussed current city planning issues for the towns and cities in Aichi prefecture, which has over seven million people and around 60 local towns and cities. We gathered representative citizens from the local government offices of the towns and cities. On the first day, guest speakers discussed the city planning issues face-to-face, and then the participants continuously discussed them online for another ten days. This ongoing experiment will be summarized soon. So far, we have gathered around 300 opinions from the first two days, and the discussions progressed effectively. Fig. 5 shows the results and the detailed setting of our social experiment. We identified no inflammatory language or flaming.

4.3. Hybrid Discussion Support for Continuous Workshops

The use of city development workshops continues to increase, reflecting the need for citizen participation in city development legislation. City development workshops were carried out continuously over weekly or monthly periods. Even though participants may have opinions or thoughts about discussions during or after the workshops, reflecting on them during workshop discussions is difficult. In our work, we proposed a virtual-world workshop using our developed system and verified our proposed method's effects by social experiments in which continuous workshops were conducted by landowners, residents, and students. Discussions were conducted by a consensus-building support system during and after the workshops. We analyzed the discussion data of both the real- and virtual-world workshops and gave questionnaire surveys to our participants and identified the effects and problems of the proposed method.

Part 1 Lecture by civil servants	
Date	Fri., Oct. 28, 2016
Place	NITech Hall at Nagoya Institute of Technology
Photo	
Contents	Civil servants of nine municipalities in Aichi presented the area and city planning for students.
Part 2 Discussion on the Internet	
Date	Fri., Oct. 28, 2016 – Fri., Nov. 4, 2016
Contents	After part 1, civil servants and students discussed future vision of the area at the internet using our support system.
Participant	Civil servants:21 persons Students:103 persons

Fig. 6 Aichi Design League 2016: we conducted a second large-scale Aichi Prefecture city-design discussion with our system in two parts. In part 1, Aichi prefecture local government officials lectured participants (students) who discussed with our system.

4.4. Aichi Design League 2016

In 2016, we conducted in a large-scale experiment with local governments in Aichi Prefecture [13] (AICHI DESIGN LEAGUE 2016). In this experiment, we verified the core time mechanism that provides the time settings for the facilitator and the participants to gather and discuss. We presented the core time to the participants to encourage them to contribute in the discussion at that timing. Fig. 7 shows an outline of the experiment. 124 people, including 21 civil servants and 103 students, participated on October 28.

The discussion's theme was "town planning in 20 years." This experiment was made in two parts. In "Part 1," nine civil servants and students presented "town planning in 20 years" in the target areas. In "Part 2," the civil servants and students discussed on the internet using COLLAGREE from October 28 to November 4. The "Divergence phase" was conducted until noon on November 1. The "Convergence phase" was conducted until 8:00 p.m. on November 3. The "Evaluation phase" was conducted until midnight on November 5. The core time was set from 10 to 12 o'clock on November 1 and 5 to 7 p.m. on November 4. The core time was set before the discussion phase changed. The core time was announced three days earlier. We explained that the core time is a period during which everyone was encouraged to join the discussion. Participants were not obligated to join the core times.

For verifying how he discussions were influenced by the core time mechanism, we analyzed the number of views and posts per day. Fig. 7 shows the transition of the number of daily posts and views. The number of posts decreased since the discussion's start.

However, on November 1, when the core time was set, there were 78 cases, which is 38 more than the previous 35 cases. In addition, on November 4, when the core time was set, there were 43 cases, 22 more than the previous 21 cases. The number of views decreased since the start of the discussion. However, on November 1, when the core time was set, there were 1203 cases, 649 more than the previous 554 cases. In addition, on November 4, when the core time was set, there were 1385 cases, 826 more than the previous 649 cases. The number of posts and views more than doubled compared with the previous day when the core time mechanism was set up. Therefore, we conclude that the core time mechanism affected the discussion, which was continuously activated until the end. We did not identify any flaming activities in this experiment as well.

4.5. Cyber-Physical Discussion Support

Fig. 9 Cyber-physical experiment: a real-world experiment where upper picture shows the audience, and lower picture shows panelists conferring based on our system's online discussion.

We proposed a hybrid (cyber-physical) environment in which people can simultaneously discuss online and offline. We conducted a large-scale experiment in a panel discussion session at an international conference where participants discussed with our online discussion support system and face-to-face communications as usual. We analyzed the experimental results from the following three metrics: participants' cyber-physical attention, keyword and cyber-physical linkage, and cyber-physical discussion flow. These three analysis results indicated that our methodology can effectively support hybrid large-scale discussions.

We conducted two cyber-physical discussion experiments at the IEEE ICA2017 and AAI2017 conferences and experienced no flaming phenomena in either of them.

5. Lessons Learned: Social Presence of Facilitator

We have conducted more than ten real field social experiments including the experiments presented in the previous session. In these experiments, we did not find any flaming phenomena. Several possible reasons might explain how we avoided flaming in online discussions:

Semi-anonymity: In all of the experiments, participants registered under their real names and e-mail addresses. The system administrator could identify their real names, but the other participants (including facilitators) could not. From the viewpoint of the participants, if they behaved poorly, they might still be identified even without engaging in such activity. But generally, even in such semi-anonymous systems like Twitter, flaming phenomena are very common.

Collaborative discussions: In the experiments so far, the discussions were all collaborative. Thus, participants who behaved anti-socially were ignored and barred. However, even in such collaborative discussions as on Wikipedia articles, sometime such flaming phenomena can be observed in the general internet world.

Social presence: Social presence [22] refers to the feeling of being socially present with another person at a remote location, and this has been largely focused on as a very influential factor in social media in the social psychology field. In our system, in all experiments, we openly informed the participants that facilitators are observing the discussion to facilitate it. Perhaps such a social presence of a facilitator(s) discouraged anti-social behavior by online participants. More work is required on this possibility.

6. Conclusions and Future Work

Large-scale online discussion systems might be an alternative for democratic systems because they enable people to discuss and learn shared problems and could lead to crowd-scale decisions. Toward such a vision, we proposed a facilitator-based online discussion model, implemented intelligent online discussion systems, and conducted real-field social experiments.

The problem of flaming is one critical issue in online discussion systems, including web forums, social networking services, and question-answer systems. Because of flaming, some people avoid online discussions.

As one key idea to attack the flaming problem, we proposed a facilitator-based online discussion system. Actually, our ultimate goal is to create automated software agents that can function as facilitators. In the current stage, we employed human facilitators and provided facilitator support functions. We implemented a large-scale online discussion support system with several functions that can support facilitators and incentivize participants to post opinions. It also has a discussion tree function that enables participants to easily grasp the entire discussion view. We have conducted more than ten social experiments in actual fields with Nagoya, Aichi Prefecture, and international conferences. These experiments are progressing quite well, and most of the participants understand the usability and the possibility of online discussion support systems.

The most critical result is that we have not yet had any flaming phenomena in any of our social experiments. We are currently scrutinizing several reasons that might explain this result. One of our insights suggests that the social presence of facilitators is key.

Future work will investigate the social presence effect in more controlled experiments with social psychologists. If the social presence of a facilitator is effective, then perhaps we can lead discussions in other web forums on the internet. Another direction will address an automated facilitator that is an intelligent software agent that can facilitate discussion among people. We have already had some technological progresses, but we need more investigation.

Acknowledgements

This work was supported by the JST CREST fund (Grant Number: JPMJCR15E1), Japan. I would like to thank all project members and related people for their contribution in the studies and the experiments.

References

[1] Ali Gurkan, Luca Iandoli, Mark Klein, and Giuseppe Zollo, 2010. Mediating debate through on-line large-scale argumentation: Evidence from the field. Information Sciences 180, 19 (2010), 3686-3702.

[2] Joshua Introne, Robert Laubacher, Gary Olson, and Thomas Malone, 2011. The Climate CoLab: Large scale model-based collaborative planning. In Proceedings of International Conference on Collaboration Technologies and Systems (CTS 2011).

[3] Mark Klein, 2007. Achieving Collective Intelligence via Large-Scale On-Line Argumentation. CCI Working Paper 2007-001 (April 2007).

[4] Luca Landoli and Mark Klein, 2007. Can We Exploit Collective Intelligence for Collaborative Deliberation? The Case of the Climate Change Collaboratorium. CCI Working Paper 2008-002 (Dec. 2007).

[5] Thomas W. Malone and Mark Klein, 2007. Harnessing Collective Intelligence to Address Global Climate Change. Innovations 2, 3 (2007), 15-26.

[6] Thomas W. Malone, Robert Laubacher, Josh Introne, Mark Klein, Hal Abelson, John Sterman, and Gary Olson. 2009. The Climate Collaboratorium: Project Overview. CCI Working Paper 2009-003 (Sept. 2009).

[7] Malone, Thomas W., Laubacher, Robert and Dellarocas, Chrysanthos, Harnessing Crowds: Mapping the Genome of Collective Intelligence (February 3, 2009). MIT Sloan Research Paper No. 4732-09.

[8] Dale Hunter, The Art of Facilitation: The Essentials for Leading Great Meetings and Creating Group Synergy, Revised Edition, Wiley, August 2009.

[9] Akihisa Sengoku, et al., Discussion Tree for Managing Large-Scale Internet-based Discussions, Collective Intelligence 2016, June 1-3, 2016.

[10] Kazumasa Takahashi, et al., Incentive mechanism based on quality of opinion for Large-Scale discussion support, Collective Intelligence 2016, 2016.

[11] Takayuki Ito, et al., Incentive Mechanism for Managing Large-Scale Internet-Based Discussions on COLLAGREE, Collective Intelligence 2015, 2015.

[12] Takayuki Ito, Yuma Imi, Takanori Ito, and Eizo Hideshima, "COLLAGREE: A Facilitator-mediated Large-scale Consensus Support System," Collective Intelligence 2014, June 10-12, 2014. MIT Cambridge, USA.

[13] Tomohiro Nishida, et al., "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), 2017.

[14] Katsuhide Fujita, Takayuki Ito, Mark Klein, "Enabling Large Scale Deliberation using Ideation and Negotiation-Support Agents," ICDCS 2017 US-Japan Workshop on Collaborative Global Research on Applying Information Technology, 2017.

[15] Takayuki Ito, et al., "Preliminary Results on A Large-scale Cyber-Physical Hybrid Discussion Support Experiment," The Eleventh 2016 International Conference on Knowledge, Information and Creativity Support Systems (KICSS 2016), 2016.

[16] Akihisa Sengoku, et al., "Towards Intelligent Crowd Decision Support: A Preliminary Result on Large-scale Discussion Support based on the Discussion Tree," International Conference on Crowd Science and Engineering (ICCSE2016), Canada, July 27-30, 2016.

[17] Takayuki Ito, et al., Experimental Results on Large-scale Cyber-Physical Hybrid Discussion Support, International Journal of Crowd Science, Emerald Publishing, ISSN 2398-7294, 2017.

[18] Galen Pickard, et al., 2011. Time-Critical Social Mobilization. Science 334, 6055 (2011), 509-512.

[19] Stephen Robertson and Hugo Zaragoza. 2009. The probabilistic relevance framework: BM25 and beyond. Foundations and Trends in Information Retrieval 3, 4 (2009), 333-389.

[20] Shun Shiramatsu and Yuto Ikeda, An Approach to Discussion Facilitators' Action Selection based on Expected Utility Calculated with Random Forest Regression. Proc. of International Conference on Crowd Science and Engineering (ICCSE2016), pp.1-6, 2016.

[21] Yuto Ikeda and Shun Shiramatsu, Generating Questions Asked by Facilitator Agents using Preceding Context in Web-based Discussion, 2nd IEEE International Conference on Agents (ICA2017), China, 2017.

[22] Tu, C. H. & McIsaac, M. (2002). The relationship of social presence and interaction in online classes. The American Journal of Distance Education, 16(3), 131-150.