3H3-E-3-04

Toward machine learning-based facilitation for online discussion in crowd-scale deliberation

Chunsheng Yang*1, Takayuki Ito*2, and Wen Gu *2

*1 National Research Council Canada *2 Nagoya Institute of Technology

The objective of this paper is to develop machine learning-based facilitation agent for facilitating online discussion in collective intelligence, particularly for online discussion in deliberation. The main idea is to model facilitator's human behaviour by using machine learning technique, case-based reasoning paradigm,. After introducing the details of the proposed machine learning-based approach for facilitation of online discussion, the paper presents some preliminary results along with some outline of the on-going research tasks and future work. The results demonstrate that it is feasible and effective to develop machine learning-based agent for smoothing the discussion and achieving a consensus.

1. Introduction

Deliberation is defined as the activity of small group of people who make the best solution for themselves [Ito 2017]. Over centuries, such decision-making process never changed. This deliberation process is controlled by a small group of powerful people who make the policies without incorporation of public opinion from crowd, and excludes the most people's involvement during the decision-making. Such an approach is becoming inadequate because many important ideas are not properly incorporated. Today, democratically, most people or crowds have to be involved in deliberation.

With the rapid development of Internet, the Internet-based online discussion in crowd-scale deliberation [Klein 2011] or in collective intelligence has attracted many efforts from researchers in social science and computer science. Online crowd decision-making support has received an amount of research interests, and some such support systems have been developed. For instance, Climate CoLab at MIT [Introne 2011] was a pioneer project which aims at harnessing the collective intelligence of thousands of people around the world to make arguments on global climate issues. The project developed a web-based crowdsourcing platform to facilitate the online argumentation [Klein 2011, Gurkan 2010, Klein 2007] democratically. Another example is COLLAGREE developed at Nagoya Institute of Technology (NiTech); it is a web-based online discussion platform [Ito, 2014], which provides a facilitator the support for managing online discussion to effectively achieve the consent through various mechanisms, including facilitation, incentives [Ito 2015], discussion-tree [Sengoku 2016], and understanding. The project team has applied the COLLAGREE to political applications such as city planning forum to collect the crowd opinion from public. For example, NiTech and Nagoya City used COLLAGREE for generating the consent for Next Generation Total City Planning. With the help of COLLAGREE, the Nagoya City gathered many opinions from public citizens. On the other hand, the people from city can understand the importance of next generation city plan. Eventually a consent decision can be achieved democratically. Such online argumentation platforms or forums require the facilitators having systematic methodologies to efficiently guild the discussion toward to consensuses by integrating ideas and opinions and avoiding flaming.

Existing online discussion systems or collective intelligence support systems require the human facilitators to conduct facilitation in order to guide/ensure the online discussion towards consensus. However, human facilitators-based online discussion systems remain several challenging issues such as human bias, time/location restriction, and human resources constrains. To address these challenges, relieve some burden of facilitators, and reuse the prior experience and skills of the facilitators, it is desirable that more advanced techniques are available for supporting the automated facilitation to achieve the consensuses efficiently.

Fortunately, the advancement of machine learning and multi-agent systems techniques provides a venue for developing facilitator agent to automate facilitations for large-scale online discussion. One of machine learning techniques available is case-based reasoning (CBR), which provides an effective reasoning paradigm for modeling the human cognition behaviors in solving real-world problems. CBR-based approach has been widely applied to many applications such as fault diagnostics [Yang 2003], recommendation systems, and judge supporting systems [Lopes 2010]. We believe in that machine learning-based facilitation, specifically, CBR-based method should be a good solution to crowd-scale deliberation or online discussion facilitation. Therefore, we propose a CBR approach to facilitating the crowd-scale online discussion in order to achieve a consensus efficiently. The main idea is to develop CBR-based facilitation actions/mechanisms, including better idea generation, smooth discussion, avoiding negative behavior and flaming, and maintaining online discussion, consensus-oriented guidance and navigation, and so on. The paper mainly discuss the basic ideas on developing machine learning-based facilitation agent and some on-going research tasks and future work.

Following this Section, the paper presents the proposed CBR-based approach for facilitating the online discussion in details;

Chunsheng Yang, National Research Council Canada, 1200 Montreal Road, Room370a@M50, Ottawa, ON, Canada, Chunsheng.Yang@nrc.gc.ca, Tel 1-613-993-0262

Section III discusses the on-going research tasks and future work; and the final Section concludes the paper.

2. Machine learning-based facilitation agent

Machine learning techniques have been widely applied to various real-world problems and have been achieved great success in developing machine learning-based modeling technologies. Today, the machine learning-based modeling technology has become a powerful tool for building models to explain, predict, and describe system or human behaviors. The main task is to develop the data-driven models from the historic data or past experience by using machine learning algorithms. The developed models have the given ability to explain, predict, and describe the system or human behaviour. For example, in the prediction applications, the machine learning-based models can forecast the system operating status, including failures or faults. With such predictions the proactive actions can be taken to maintain the system availability. In this work, we contemplate to use a case-based reasoning approach or paradigm to model facilitators' behavior or facilitation by using their experience accumulated in past.

2.1 CBR-based modeling for cognition

CBR is rooted in the works of Roger Schank on dynamic memory and the central role that a reminding of earlier episodes (cases) and scripts (situation patterns) has in problem solving and learning [Schank 1983]. Today, Case-based reasoning is a paradigm for combining problem-solving and machine learning to solve real-world problems. It has become one of the most successful applied intelligences for modeling human cognition. The central tasks in CBR-based methods and systems [Amodt 1994] are: "to identify the current problem situation, find a past case similar to the new one, use that case to suggest a solution to the current problem, evaluate the proposed solution, and update the system by learning from this experience. How this is done, what part of the process that is focused, what type of problems that drives the methods, etc. varies considerably, however". A general CBR-based system or agent can be described by a reasoning cycle composed of the following four steps:

- RETRIEVE the most similar case from existing case bases;
- REUSE the solution in the case to solve the problem such as flaming, wrong post to the issue, distraction post;
- REVISE the proposed solution if necessary;
- RETAIN the parts of this case into a case base for future problem solving.

2.2 Case composition and definition

In general a case documents relationships between problems and its solutions. CBR solves a new problem by adapting similar solutions used for a similar problem in the past. For online discussion facilitation, a case can be defined as three components (as shown as Figure 1): online discussion case description, facilitation action, and case management. Following is the brief description for each components.

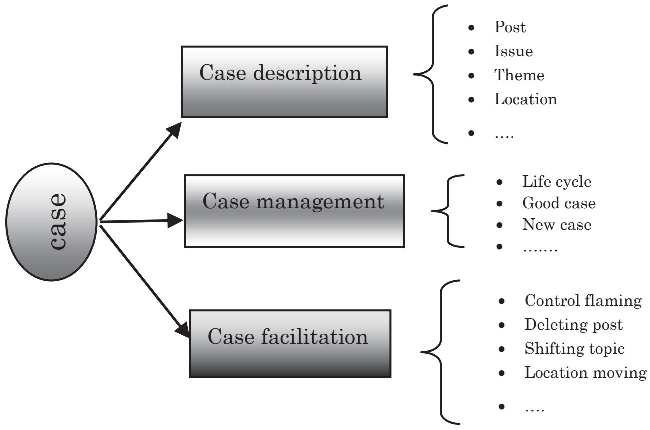

Fig.1 The case composition for online discussion

Case Description: This component contains discussion post, issue related to discussion, topics, theme, and so on. Post could be a free text, or a group of sub posts.

Case Management: This component consists of necessary case management information such as case status, case life cycle, case type, case consistence, and so on.

Case Facilitation: This component records the facilitation actions conducted by human facilitator over the past online discussion. The main facilitation could be flaming control, topic shift, post combination, post deletion, idea promotion, and so on.

From online discussion practice such as Nagoya City Planning, we have collected the data to create cases based on the case definition. It is especially useful to create facilitation cases which documents how a facilitator guided the online discussion; what kind of facilitation was used; how a facilitation action was taken, and so on.

2.3 Similarity computation

Based on the case definition above, a CBR method must provide a similarity algorithm for computing the similarity between two cases. Using computed similarity, the similar cases can be retrieved from a case base. In this work, we provide a global similarity algorithm, which computes the global similarity (sim) using Equation 1.

$$sim = \frac{\sum_{i=1}^{N} \omega_i sim_i(f_i, f_i')}{\sum_{i=1}^{N} \omega_i} \dots\dots\dots\dots\dots (1)$$

where, *sim* is the global similarity of two problems; N is the number of features or attributes that contribute to similarity; ω_i is the weigh coefficient of each feature; sim_i is a local similarity; f_i, f_i' are the i^{th} features in a case and given problem description. It is computed with Nearest Neighbor *(NN)* distance algorithm (*NN* method) for regular types of the features. For a free "text" feature, we use natural language processing techniques to compute local similarity. Particularly, we used IE (Information Extraction) method to compute the text similarity by using the library provided in OpenNPL package [Weber 2001]. We used the Maximum Entropy algorithms implemented in the OpenNLP package to compute the local similarity for two text messages as expressed as Equation 2.

$$sim_i\ (f_i, f_i') = \frac{|f_i \cap f_i'|}{(|f_i| \cup |f_i'|)} \ldots\ldots\ldots\ldots\ldots (2)$$

2.4 CBR-base facilitation agent framework

Once the case base is created, the CBR-based facilitation agent is ready constructed in online discussion system or platform. In general, such a facilitation agent can be implemented following the designed framework showed as in Table 1.

Table 1, the pseudocode for facilitation agent framework
(*Note: CB: case bases, C: case from Post, C': retrieved case from CB, FA: facilitation action from C'*)

```
Input: caseBase (plain text file, CB);
Steps:
  CB =Loadthecasebaseinmomorry (CB text file);
   C= GetOnlineDiscussionCase (post ,issue, theme);
   C'= StartReasonningCycle(C) ;
          ⇨ RrtriveCase(C,SIM());
          ⇨ SolutionAdaptation(C')
          ⇨ ReviseCase( C');
          ⇨ RetainCase( C');
    FA =AdaptatFacilitation(C');
    ExecuteFAcilication (FA);
Stop
```

As described in Table 1, the first step is to load the cases (stored in external files or database) into memory given a case composition and configuration mapping information. Once the case base is loaded into memory, a facilitation agent can execute the CBR reasoning cycles to retrieve the similar case in a case base for obtaining facilitation action give a post, issue and theme. Second step is to adapt the facilitation from the similar case for given case; the third step is to modify the case if necessary; the last step is to retain the revised case and save it back to the case base as a new case. The final step is to execute the facilitation action based on adapted facilitation action to the current post if it requires a facilitation action.

3. Discussion and Future Work

This paper mainly reports our ongoing research project. The objective is to present the ideas for developing machine learning-based agent for online discussion facilitation. Therefore many tasks are ongoing. Since focusing on CBR-based approach for facilitation, we only discuss the CBR related ongoing tasks. The other machine learning-based methods for automated facilitation will be reported in other papers.

3.1 Case structure extension and similarity algorithm

The case defined above is a basic structure. To reflect the various online discussion and complexity of facilitator's behaviour, the case structure may become complicated and complex. The similarity computation algorisms have also to be further investigated and extended from existing simple algorithm. For example, we are exploring a graph-based case structure in order to build case from a group post instead of one individual post [Gu 2018]. On the other hand, we have to investigate new algorithms for computing similarity of graph-based online discussion cases

Cases can be created either from historic data or simulation data. In this work, we conducted an online discussion forum to collect the real data. The forum was set up as a "CBR approach to support facilitation in COLLAGREE". We created the theme for an online discussion in the laboratory. The online discussion was managed and guided by a facilitator who maintains the online discussion in three phases: divergence, convergence, and evaluation. The facilitator used the support vehicle provided to navigate the forum from divergence to convergence to evaluation. Using collected data, we created some cases which reflect the facilitator's facilitation during online discussion. However, to enrich the facilitation more data are required for case creation. One way is to conduct the simulation to generate more facilitation data for creating more cases.

3.2 Machine learning-based case adaptation

In CBR research area, one remaining challenge is case adaptation. It is normal that we can't retrieve a similarity case from a case base to obtain a similar facilitation action for controlling and managing online discussion in practice. Therefore, the CBR-based agent has to adapt a facilitation action. To this end, we have to build the ability for agent to learn a new facilitation action. This motivates us to investigate the machine learning-based case adaptation methods for facilitation agent.

3.3 Case base management

This is a vital research topic for any CBR-based applications. The existing cases are manually created from the forum data collected in COLLAGREE. This is a time-consuming task and requires rich domain knowledge to understand the contents in the post or opinion. With the increasing of the collected data, manual case creation will be a challenge. An automated case creation mechanism is expected and necessary. Therefore two necessary research topics are described as follows:

(1) Automated case generation: As we discussed above, automated case creation is desirable to relieve the burden of manual case creation. From the viewpoint of machine learning, automated case creation is a supervised learning problem. It requests the annotated information to decide the case property or types. To do this sentiment analysis of the post contents is inevitable and vital for determining the case types: positive, natural, and negative. Another challenge is machine translation of language. During the online discussion it may encounter the multiple language. When generating cases from different language the automated machine translation is required.

(2) Case base management: In this work, the case base management still remains a challenge. To manage the case base efficiently, case redundancy and consistence have to be investigated in order to ensure the quality and integrity

of the case bases. Another challenge is case adaptation from the existing case and case updating to existing cases.

3.4 Validation and evaluation

Validation and evaluation for a CBR-based systems is always a challenge issue in developing CBR-based applications. It requires many efforts to design the procedures and methods. In this work, the following tasks will be conducted:

(1) Continue to collect the data from online discussion forum using COLLAGREE and create more cases for evaluation;
(2) Evaluate the performance of CBR-based systems for facilitation support by comparing the results with one from human facilities; and
(3) Validate the scalability of cases crossing different themes, even domains.

4. Conclusions

This paper reported an ongoing research project. The objective is to develop a machine learning-based facilitation agent for online discussion system to perform the automated facilitation in crowd-scale deliberation. After describing the proposed approach, we discussed some on-going research tasks and future work.

ACKNOWLEDGMENT

This work was supported by the JST CREST fund (Grant Number: JPMJCR15E1), Japan. We would like to thank all project members and related people for their contribution in the studies and the experiments. We are also grateful for all participants at Ito-Lab, Nagoya Institute of Technology in online discussion on CBR application to COLLAGREE.

References

[Amodt, 1994] A. Aamodt and E. Plaza, “Case-based reasoning: Foundational issues, methodological variations, and system approaches”, AI Communications, Vol. 7 No. i, 1994.

[Gu, 2018] W. Gu, A. Moustafa, T. Ito, M. Zhang. and C. Yang, “A Case-based Reasoning Approach for Automated Facilitation in Online Discussion Systems”, The Proceedings of The 2018 International Conference on Knowledge, Information and Creativity Support Systems (KICSS 2018), Thailand, Nov. 2018

[Gurkan , 2010] A. Gurkan, L. Iandoli, M. Klein, and G. Zollo, “Mediating debate through on-line latrge-scale atgumentation: Evidence from the field”, Information Science Vol. 180, No. 19, 2010, pp. 3686-3702

[Ito, 2015] T. Ito, Y. Imi, M. Sato, T. K. Ito, and E. Hideshima,”Incentive Mechanism for Managing Large-Scale Internet-Based Discussions on COLLAGREE”, Collective Intelligence 2015, May 31 – June 2, 2015, the Marriott Santa Clara in Santa Clara, CA (poster).

[Ito, 2014] T. Ito, Y. Imi, T. K. Ito, and E. Hideshima, “COLLAGREE: A Faciliator-mediated Large-scale Consensus Support System”, Collective Intelligence 2014, June 10-12, 2014. MIT Cambridge, USA. (poster)

[Ito, 2017] T. Ito, T. Ostuka, S. Kawasa, A. Sengoku, S. Shiramatsu, T.K. Ito, E. Hideshima, T. Matsuo, T. Oishi, and R. Fujita, “Experimental results on large-scale cyber-physical hybird discussion support”, International Jouran of Crowd Science, Vol. 1 No. 1, 2017

[Introne , 2011] J. Introne, R. Laubachar, G. Olson, and T. Malone, “The Climate Colab: Large scale model-based collaborative planning”, Proceedings of International Conference on Collaboration Technologies and Sytems (CTS 2011), 2011

[Klein, 2011] M. Klein, “Toward crowd-scale deliberation”, DOI: 10.13140/RG.2.2.12264.06401 , Massachusetts Institute of Technology, available at https://www.researchgate.net/publication/317613473Towards_Crowd-Scale_Deliberation

[Klein, 2012] M. Klein, “Enabling large-scal deliberation using attention-mediation metrics”, Computer Supported Cooperative Work(CSCW), Vol. 21, No. 4/5, pp.449-473, 2012

[Klein, 2007] M. Klein, “Achieving collective intelligence via large-scale on-line argumentation”, CCI working paper, 2007-001, April, 2007

[Lopes, 2010] E. C. Lopes and U. Schiel, “Integrating Context into a Criminal Case-based Reasoning Model”, the proceedings of 2nd International Conference on Information, Process, and Knowledge management, 2010

[Schank, 1983] R. C. Schank. “Dynamic Memory”, Cambridge Univ. Press, 1983.

[Sengoku, 2016] A. Sengoku, T. Ito, K. Takahashi, S. Shiramatsu, T.K. Ito, E. Hideshima and K. Fujita, “Discussion Tree for Managing Large-Scale Internet-based Discussion”, Collective Intelligence 2016, Stern School of Business New York University , June 1-3, 2016

[Weber, 2001] R. Weber, D. W. Aha, N. Sandhu, H. Mounoz-Avila, “A textual case-based reasoning framework for knowledge management applications”, Proceedings of Knowledge Management by Case-Based Reasioning; Experience and Management as Resue of Knowledge (CWCBR 2001), 2001

[Yang, 2003] C. Yang, R. Orchard, B. Farley, and M. Zaluski, Authoring Cases from Free-Text Maintenance Data, in Proceeding of IAPR International Conference on Machine Learning and Data Mining (MLDM 2003), Leipzig, Germany, July 5-7, 2003, pp.131-140