PAPER *Special Section on Intelligent Information Processing Technology to be Integrated into Society*

Spotlight Contents Extraction from Text-Based Online Discussion

Zhizhong WANG[†], **Wen GU**[††a)], **Zhaoxing LI**[†††], **Koichi OTA**[†], ***Nonmembers***, ***and*** **Shinobu HASEGAWA**[†], ***Senior Member***

SUMMARY To understand the development of online discussions and engage effectively, it is a vital issue for both individual participant and facilitator to grasp the contents that the discussion group is focusing, i.e., spotlight contents. However, it becomes extremely challenging to catch up with the spotlight contents in the text-based consensus decision-making online forums (TCDOF) with the increasing of participants and post generation. In this paper, we endeavor to address this challenge through the introduction of a novel framework that leverages topics derived from post contents and inter-post structure to extract spotlight contents from TCDOF. In addition, the extracted spotlight contents are presented in the form of succinct natural language sentences, enhancing accessibility and comprehension. Furthermore, we devise a time-based spotlight contents extraction (TSCE) algorithm to extract spotlight content from a temporal perspective. The effectiveness of the proposed approach is demonstrated with real-world online discussion experiments.

key words: *consensus, decision-making support, spotlight contents extraction, discussion facilitation support*

1. Introduction

Text-based consensus decision-making online forums (TCDOF) have attracted much research interest recently as a promising approach to foster group consensus [1]. These platforms offer participants a low barrier to entry, facilitating seamless engagement in discussions and providing efficient access to diverse viewpoints. A critical challenge in TCDOF pertains to staying abreast of contents that the discussing group is focusing on with the discussion development. Without an awareness of timely spotlight contents of the discussion, participants may encounter obstacles in contributing to discussions, while facilitators may struggle to intervene effectively to promote discussion development. However, due to the asynchronous nature of TCDOF, it is crucial for human to extract spotlight contents, and this task exacerbates when the number of participants increases [2]. To support understanding spotlight contents in online discussions, numerous studies have explored the development of supportive functions, including keywords or hashtags [3], [4]. In spite of these functions, apprehending spotlight contents solely through isolated words poses challenges in the absence of contextual framing. Furthermore, grasping the dynamic evolution of spotlight contents in tandem with discussion progression presents additional complexities.

In pursuit of this objective, we present a content, structure and context based framework (CSC-framework) tailored to the extraction of spotlight contents from TCDOF. Acknowledging the potential variance in spotlight contents across diverse perspectives, CSC-framework initiates with discussion sampling as its foundational step. Subsequently, upon determining the specific perspective, CSC-framework proceeds to extract spotlight topics by integrating both discussion contents and structure into its analytical architecture. After deriving the extracted topics, we employ generative models to facilitate the generation of spotlight contents by incorporating the context of the ongoing discussion. Furthermore, we devise a time-based spotlight contents extraction (TSCE) algorithm to extract spotlight content from a temporal perspective. In summary, this paper makes the following noteworthy contributions to the field.

- We propose a general CSC-framework to extract spotlight contents from TCDOF.
- We develop a TSCE Algorithm that extracts spotlight contents from time-span perspective.
- We conduct real-world online discussion experiments to evaluate the effectiveness of our proposed algorithm.

The rest of this paper is organized as follows. Section 2 introduces the related work of this research. Section 3 provides a detailed explanation of the proposed system. Experiments and experimental results are demonstrated in Sect. 4 and discussed in Sect. 5. Section 6 concludes this paper.

2. Related Work

2.1 Support in Online Consensus Decision-Making

Online consensus decision-making support has attracted much research interest because the difficulties for human facilitators to handle with the increasing of posts. Many researchers focus on supporting from the contents perspective. Shiramatsu et al. developed an agent that supports the web-based discussion by providing relevant information [5]. Ishizuka et al. noticed that participants in a discussion may fall into the spiral of silence if most of them generate similar opinions [6]. They tried to mitigate this situation by

Manuscript received April 8, 2024.
Manuscript revised September 30, 2024.
Manuscript publicized December 20, 2024.
[†]Japan Advanced Institute of Science and Technology, Nomishi, 923–1292 Japan.
[††]Nagoya Institute of Technology, Nagoya-shi, 466–8555 Japan.
[†††]University of Southampton, Southampton, SO17 1BJ, UK.
a) E-mail: wgu@nitech.ac.jp
DOI: 10.1587/transinf.2024IIP0009

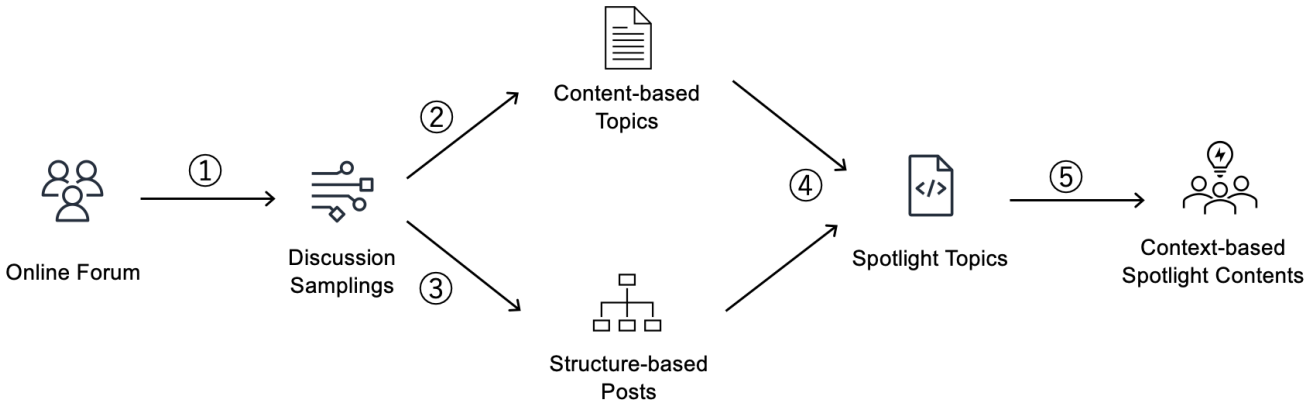

Fig. 1 Content, structure and context based spotlight contents extraction framework

developing an agent that reinforces minority opinions to encourage the sublation and improve the consensus level in the discussion. Meanwhile, another group of researchers paid more attention to the structural perspective of supporting online consensus decision-making. Klein built the TCDOF system called Deliberatorium where posts are directly generated as a tree-structure deliberation map instead of free text [7]. Sengoku et al. proposed the discussion tree to utilize tree structure to represent free text based posts [8]. Gu et al. utilized the post reply relation to identify influential participants [9].

From this line of research we understand that both content feature and structure feature are important to consider in supporting online discussions development. Furthermore, the dynamics of online discussion development is also essential to be considered.

2.2 Spotlight Contents Extraction from Online Discussion

Contents extraction is an vital technique that can be utilized in various applications. Keywords extraction is one famous task that aims to pick up keywords that represent the topic of a document [10]. Latent Dirichlet Allocation (LDA), which considers the contents of multiple documents and assumes that each document is a mixture of topics and that each word in the document is attributable to one of those topics, has been widely utilized in text mining [11]. Extractive approaches are faster while the results can be far different from human expected [12]. Abstractive approaches are considered to be more functional if get equipped with rich representations, which can be resolved by using large language models [13].

To this end, we believe extractive approaches can be utilized in extracting topics and large language model (LLM) can be utilized to support the rich representations. And the combination of these techniques can be utilized to develop novel spotlight contents from online discussions.

3. Content, Structure and Context Based Spotlight Contents Extraction Framework

The proposed content, structure and context based spotlight contents extraction framework (i.e., CSC-framework) is introduced in this section. A time-based spotlight contents extraction algorithm (TSCE) is formally defined according to the CSC-framework.

3.1 CSC-Framework

As depicted in Fig. 1, CSC-framework comprises six main components. The process of extracting spotlight contents from online forum discussion is delineated through a sequence of five procedures. The initial step entails the identification of pertinent feature for discussion sampling. Given the multifaceted nature of spotlight contents, attributes such as temporal aspects, thematic relevance, and user characteristics warrant consideration. Following the delineation of the discussion sample, features germane to spotlight content are extracted through a dual perspective. Firstly, from a content-centric stance, the objective lies in distilling pivotal topics, thereby encapsulating the essence of discourse within the discussion sample. Concurrently, an examination from a structural standpoint seeks to pinpoint seminal posts, thereby elucidating the interrelation among discourse elements. Subsequent to the extraction of salient topics and posts, the determination of spotlight topics ensues. Finally, the formulation of spotlight contents, rendered in natural language, ensues by integrating the identified spotlight topics within the broader context of the sampled discussion.

3.2 Time-Based Spotlight Contents Extraction

In elucidating the operational intricacies of each constituent within the proposed framework, we rigorously define the nuanced attributes of every component, specifically focusing on time-based spotlight content extraction. In alignment with this, we introduce the TSCE algorithm meticulously designed to facilitate the extraction of time-based spotlight content.

3.2.1 Online Forum

In the CSC-framework, a discussion in the online forum is defined as a rooted tree graph $\mathcal{G} = (V, E)$ where V is a set

of vertices and $E \subseteq \{(v, v') \in V^2 \mid v \neq v'\}$ is a set of edges. Each $v \in V$ of $\mathcal{G}$ represents a post that is generated by a participant. Each $e \in E$ of $\mathcal{G}$ represents the existence of the reply relation between two posts. A discussion is defined as a rooted tree graph $\mathcal{G}$ because we assume that each discussion starts with a topic and develops with the increasing directed replies between users. $v^* \in V$ represents a discussion topic, which is the root node of $\mathcal{G}$. Each post v is defined as a tuple $v = (c, t)$, where c denotes the contents of the post and t denotes the time when the post is generated.

3.2.2 Time-Based Sampling Determination

Since spotlight contents in the discussion dynamically change with discussion development, time span is one the typical perspective to sample the discussion in detecting spotlight contents. A time span is defined as a set $T = \{t | t_m \leq t \leq t_n\}$, where t_m denotes the beginning of the time span and t_n denotes the ending of the time span. While the longest time span in a discussion is $\mathcal{T} = \{t | t_b \leq t \leq t_e\}$, where t_b denotes the beginning of the discussion and t_e denotes the ending of the discussion. The set $V_T (V_T \subseteq V)$ represents the posts that are generated during T. Each set becomes one time-based sampling of the discussion.

3.2.3 Content-Based Topic Detection

After determining the target sampling of the discussion, two procedures become necessary. The first one is to distill the pivotal topics generated in the discussion. In the proposed algorithm, Latent Dirichlet Allocation (LDA) [11] is utilized to detect the relevant topics of the discussion. LDA is a statistical model used for topic modeling which considers the contents of multiple documents and assumes that each document is a mixture of topics and that each word in the document is attributable to one of those topics. Specifically, a discussion sample V_T is considered as a corpus and each post $v (v \in V_T)$ is regarded as a document in the CSC-framework. By conducting LDA algorithm, word distribution θ for each topic and topic distribution ϕ for each document are generated.

3.2.4 Structure-Based Post Selection

The second procedure of handling discussion sample is to pinpoint seminal posts based on the structure of discussion development. Since V_T is a subset of V, the graph G which consists of the nodes in V_T can be considered as a subgraph of $\mathcal{G}$. The number of children N_v of each node $v (v \in V_T)$ can be counted. The post (i.e., node) which has a higher number of children means that those contents related to this post are discussed more.

3.2.5 Spotlight Topics Determination

In the proposed TSCE algorithm, the time-span of distilling pivotal topics is set to be $T_t = \{T_t | t_b \leq T_t \leq t_n\}$, which is from the beginning of the discussion to a particular end point. This aims to include all the important topics that have been discussed from the beginning of the discussion. While the time-span of pinpointing seminal posts is set to be $T_s = \{T_s | t_m \leq T_s \leq t_n\}$, which aims to target the recently generated posts. Furthermore, a constant K is set to only target the K nodes that have highest number of children N_v in the structure. The corresponding topics C in these K nodes are determined as the spotlight topics.

Algorithm 1 TSCE Algorithm

Input: $\mathcal{G}, t_b, t_m, t_n, K, \mathcal{P}$
Output: $\mathcal{S}$
1: **for** each post v in $\mathcal{G}$ **do**
2: **if** v is generated during (t_b, t_n) **then**
3: add v to set V_{T_t}
4: **end if**
5: **if** v is generated during (t_m, t_n) **then**
6: add v to set V_{T_s}
7: calculate v's children number N_v
8: **end if**
9: **end for**
10: θ, ϕ = LDA(V_{T_t})
11: V_K = Top K nodes with highest number of N_v
12: $C = V_K$'s topics based on θ, ϕ
13: $\mathcal{S}$ = LLM($\mathcal{P}, C, V_{T_t}$)
14: **return** $\mathcal{S}$

3.2.6 Context-Based Spotlight Contents Generation

In pursuit of generating context-based spotlight contents in natural language, TSCE adopts large language model in accordance with sampled discussion V_{T_t} and extracted spotlight topics C. The prompt $\mathcal{P}$ devised for undertaking this task needs to be formulated according to the feature of utilized large language model. And the output of context and topic based spotlight contents is denoted as $\mathcal{S}$.

3.2.7 TSCE Algorithm

Algorithm 1 encapsulates the procedural framework designed for generating spotlight contents (i.e., output $\mathcal{S}$) based on six inputs that are $\mathcal{G}, t_b, t_m, t_n, K$ and $\mathcal{P}$. $\mathcal{G}$ is an online forum discussion. t_b and t_n represent the beginning time and ending time of distilling pivotal topics, respectively. t_m is the beginning of pinpointing seminal posts. K is the constant number that is used in selecting top K nodes with highest number of children. $\mathcal{P}$ is the prompt that is used in generating spotlight contents with a large language model. Specifically, line 2 through 4 delineate the selection of content-based nodes. Subsequently, line 5 through line 8 detail the process of structure-based posts detection alongside the calculation of each node's children count. Line 10 denotes the initiation of topic generation and distribution, employing the LDA model. Line 11 denotes the posts selection process contingent upon their children count. Following this, in line 12, the determination of spotlight topics ensues, drawing from the selected posts and their associated topics. Finally, line 13 orchestrates the generation of spotlight contents tailored

to the context of variable V_{T_t} and prompt $\mathcal{P}$, utilizing a large language model.

4. Experimentation

In this section, we explain the experiment settings and experimental results of evaluating the proposed approach. Real-world experiments are conducted to answer the research questions that we propose. In general, an online forum platform is developed and utilized for holding online discussions. Spotlight contents are generated based on the proposed framework and the generated contents are evaluated by participants' questionnaires.

4.1 Text-Based Consensus Decision-Making Online Forum

To conduct real-world experiments, we built a text-based consensus decision-making online forum platform based on the open source system *Discourse Meta*[†]. *Discourse Meta* is a flexible, customisable online discussion software, which is suitable for holding online discussions. The detailed server settings are shown in Table 1.

4.2 Experiment Settings

4.2.1 Real-World Discussions

Six real-world online discussions were held by utilizing the online forum introduced in Sect. 4.1 Graduate school students in the Japan Advanced Institute of Science and Technology (JAIST) were recruited to be the participants in the discussions. For each discussion, we asked five participants to discuss the topic for 30 minutes in Japanese. Each discussion topic was selected from the Sustainable Development Goals (SDGs)[††]. The specific topics are translated as follows. And the experiments were approved by the JAIST Life Science Committee (H 05-061).

- Experiment 1. How can we eliminate poverty?
  (貧困をなくすにはどうすればいいですか?)
- Experiment 2. How can we reduce hunger to zero?
  (飢餓をゼロにするにはどうすれば良いですか?)
- Experiment 3. How can we achieve gender equality?
  (ジェンダー平等を実現するにはどうすれば良いですか?)
- Experiment 4. How can we deliver high quality education to everyone?
  (質の高い教育をみんなに届くにはどうすればいいですか?)
- Experiment 5. How can we eliminate inequalities between people and countries?
  (人や国の不平等を無くすにはどうすればいいですか?)
- Experiment 6. What concrete measures are there to tackle climate change?
  (気候変動に具体的な対策はなにかありますか?)

4.2.2 Settings in Topic Detection

To conduct topic detection in the online discussion data, morphological analysis is necessary. In the experiments, MeCab[†††] was selected to conduct morphological analysis to the collected Japanese discussion data [14]. Specifically, "mecab-ipadic-NEologd"[††††] was utilized as the dictionary to handle the words such as new words, slang and internet slang [15]. In addition, gensim library was utilized to conduct LDA algorithm to the topic detection [16]. As a unsupervised learning algorithm, parameters like topic number, passes list, chunksize list, alpha list, update every list and random state in LDA need to be adapted according to the sampled discussion data. The specific settings of these parameters are denoted in Table 2. For each topic extraction, the LDA algorithm was conducted with these parameters. The coherence score for each parameter combination was calculated. We selected the parameter set with the best coherence score and their related topics for use in the experiment.

To test the topic detection with dynamic discussion development, topic detection is conducted every 10 minutes in the 30 minutes discussion. The corresponding time spans for content-based topic detection are $(0, 10]$ minutes, $(10, 20]$ minutes and $(20, 30]$ minutes, respectively. While the corresponding time spans for structure-based post detection are $(0, 10]$ minutes, $(0, 20]$ minutes and $(0, 30]$ minutes, respectively. In addition, three ($K = 3$) nodes are selected in the spotlight topics determination.

4.2.3 Prompt in Contents Generation

To generate spotlight contents in natural language, the large language model GPT-4 [17] is utilized based on the extracted

Table 1 Online forum system settings

OS	Ubuntu18.0
Platform	Discourse meta
SSD	100G
CPU	2 Core
Memory	1 GB
Site	https://disme.hopto.org

Table 2 Parameter settings of LDA

Topic Number	[3, 5, 7]
Passes List	[5, 10, 15, 20]
Chunksize List	[5, 10, 15, 20]
Alpha List	[auto, symmetric]
Update Every List	[2, 1, 0]
Random State	100

[†]https://meta.discourse.org/
[††]https://sdgs.un.org/goals

[†††]mecab-python3, version 1.08
[††††]mecab-ipadic-neologd, version 2020-08-20

Table 3 Statistics of online discussion results

Experiment	1	2	3	4	5	6
Participant Number	4	5	5	5	5	5
Total Posts	27	49	58	28	29	33
New Posts	9	15	21	12	10	11
Reply Posts	18	34	37	16	19	22

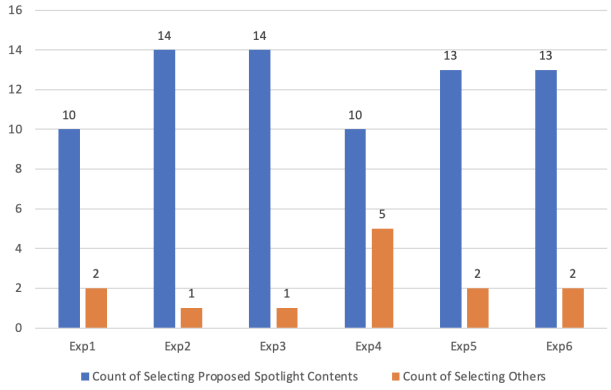

Fig. 2 Questionnaire results of spotlight contents selection

topics, which are represented as a set of words. The designed prompt of generating spotlight contents is elaborated as follows.

- Ignore previous contents.
- You are the assistant in the meetings.
- The contents of the meeting are used to create the LDA model.
- LDA model outputs topics, where topics are chunks of words.
- According to the super-sentence relationship and the topic words, using topic words as much as possible to output a concise one-sentence summary of the topic.
- Here is a good example, "the application of digital technology (AI development) on the issue of lack of educational resources."
- Explanation is not required. Only output the summary of each topic.

4.3 Experimental Results

The statistics of six online discussions are elaborated in Table 3. In the first discussion, there were four participants because one participant could not attend as planned.

After collecting discussion contents, spotlight contents were generated through the proposed TSCE algorithm and questionnaires were collected after the experiments. The questionnaire was built based on each generated spotlight contents as well as an additional option (i.e., "others"). Participants were asked to selected the spotlight contents they agree with. The results of the questionnaire is shown in Fig. 2.

We pick up one of the best results (i.e., experiment 3) and the worst result (i.e., experiment 4) from six experiments and elaborate the detailed results. In each experiment, three time spans are set and optimal parameters of LDA in each time span are calculated by following the settings explained in Sect. 4.2.2. The results of the calculation are shown in Table 4. Furthermore, the detailed questionnaire results of experiment 4 is shown in Table 5, where "◯" represents that the participant selects the spotlight contents generated by the system.

Table 4 Optimal LDA parameters in experiment 3 and experiment 4

	3-1	3-2	3-3	4-1	4-2	4-3
Topic Number	7	3	3	5	7	7
Passes List	15	15	5	15	20	5
Chunksize List	5	20	15	5	5	5
Alpha List	auto	sym	sym	auto	auto	sym
Update Every List	2	2	0	2	2	0
Best Coherence Score	0.576	0.483	0.490	0.477	0.547	0.489

Table 5 Users' individual questionnaire in experiment 4

	4-1	4-2	4-3
Participant 1	◯	others	others
Participant 2	◯	◯	◯
Participant 3	◯	◯	◯
Participant 4	others	others	others
Participant 5	◯	◯	◯

The selected posts with their related topic and generated spotlight contents of Experiment 3-1, 3-2, 3-3, 4-1, 4-2 and 4-3 and elaborated in Tables 6, 7, 8, 9, 10 and 11, respectively. Translation is added to the original data that is written in Japanese.

5. Discussion

Experimental results in Sect. 4.3 demonstrated that the proposed framework can be utilized to extract spotlight contents which is rendered as one simple sentence in natural language. The extracted spotlight contents change according to the dynamic discussion development. In addition, we understand that the extracted spotlight contents are mostly consistent with the participants' thoughts to the spotlight contents. This indicates that extracted information can be helpful for participants as well as facilitators in grasping the timely discussion contents that participants pay attention to. At the same time, experiment 4 received more selections of "others" than other situations. For instance, participant 1 who chose "others" and input "高質な教育をみんなに届くのは可能であるか (Is it possible to make high-quality education reach everyone)." as the spotlight contents in experiment 4-2. In addition, participant 4 who chose "others" and input "ネットによる教育格差の軽減 (Reducing education inequalities through the internet.)" as the spotlight contents in experiment 4-3. This indicates that too abstract spotlight contents may be less acceptable than the spotlight contents with useful details. In addition, participant 4 chose three "others" in a row in the experiment 4. This implies that personal opinions may also vary in the the definition of spotlight contents.

Table 6 Selected posts with their related topic and generated spotlight contents in experiment 3-1

	Post ID: 1213	Post ID: 1226	Post ID: 1215
Related Topic ID	Topic 5	Topic 6	Topic 5
Words in Topic 5 Translation	女性, 男性, 思う, 生理, 抑圧, 弱い, 先天的, 格差, 強い, 社会的 Female, Male, Think, Physiological, Oppression, Weak, Congenital, Disparity, Strong, Social		
Generated Spotlight Contents Translation	男性と女性の生理的な力の格差と社会的抑圧 Physiological power disparity and social oppression between male and female.		
Words in Topic 6 Translation	日本, トイレ, 傾向, 入れる, 感じる, 信頼, 会社, 一時, 労働力, しまう Japan, Toilet, Trend, Put, Feel, Trust, Company, Temporary, Workforce, Put an end to		
Generated Spotlight Contents Translation	日本のジェンダー問題と社会的傾向 Gender issues and social trends in Japan.		

Table 7 Selected posts with their related topic and generated spotlight contents in experiment 3-2

	Post ID: 1238	Post ID: 1239	Post ID: 1240
Related Topic ID	Topic 2	Topic 2	Topic 2
Words in Topic 0 Translation	平等, 女性, GENDER, 思う, 男性, 難しい, くる, 実現, 定義, 感じる Equality, Female, GENDER, Think, Male, Difficult, Come, Realisation, Definition, Feel		
Generated Spotlight Contents Translation	ジェンダー平等の実現とその難しさ Achieving gender equality and the difficulties involved.		

Table 8 Selected posts with their related topic and generated spotlight contents in experiment 3-3

	Post ID: 1253	Post ID: 1255	Post ID: 1261
Related Topic ID	Topic 2	Topic 1	Topic 1
Words in Topic 2 Translation	女性, 思う, 男性, くる, 多い, 平等, 男女, 議論, GENDER, 見る Female, Think, Male, Come, Many, Equality, Gender, Discussion, GENDER, View		
Generated Spotlight Contents Translation	男女間の平等とジェンダー問題の議論 Discussion of equality and gender issues between male and female.		
Words in Topic 1 Translation	平等, 女性, GENDER, 男性, 思う, 実現, 難しい, 社会, 定義, 仕事 Japan, Toilet, Trend, Put, Feel, Trust, Company, Temporary, Workforce, Put an end to		
Generated Spotlight Contents Translation	ジェンダー平等の実現とその難しさ Achieving gender equality and the difficulties involved.		

Table 9 Selected posts with their related topic and generated spotlight contents in experiment 4-1

	Post ID: 1276	Post ID: 1277	Post ID: 1279
Related Topic ID	Topic 4	Topic 0	Topic 4
Words in Topic 4 Translation	教育, 高位, 質, 提供, 考える, 費用, 状況, できる, 必要, 授業料 Education, High ranking, Quality, Provision, Think, Cost, Situation, Can, Need, Tuition		
Generated Spotlight Contents Translation	高質な教育の提供とその費用問題 Provision of high quality education and the issue of its costs.		
Words in Topic 0 Words in Topic 0	成長, 幅, 複雑, 方向, 違う, 趣味, 定義, 考える, 内容, 重視 Growth, Breadth, Complexity, Direction, Different, Hobby, Definition, Think, Content, Emphasis		
Generated Spotlight Contents Generated Spotlight Contents	教育の目標と学生の成長方向 Educational goals and direction of student development.		

Table 10 Selected posts with their related topic and generated spotlight contents in experiment 4-2

	Post ID: 1286	Post ID: 1291	Post ID: 1284
Related Topic ID	Topic 2	Topic 2	Topic 2
Words in Topic 2 Translation	教育, 質, 高位, 受ける, 考える, 側, 途上国, 個性, 提供, 思う Education, Quality, High ranking, Receive, Think, Perspective, Undeveloped Country, Individuality, Provide, Think		
Generated Spotlight Contents Translation	教育の質とその受け手の視点 Quality of education and the perspectives of its recipients.		

Table 11 Selected posts with their related topic and generated spotlight contents in experiment 4-3

	Post ID: 1299	Post ID: 1304	Post ID: 1296
Related Topic ID	Topic 4	Topic 4	Topic 4
Words in Topic 4 Translation	教育, 質, 高位, 考える, できる, 思う, 届く, 子供, 受ける, 以外 Education, Quality, High ranking, Think, Can, Think, Reach, Children, Receive, Other than		
Generated Spotlight Contents Translation	教育の質とその普及の問題 Quality of education and the issue of its dissemination.		

6. Conclusion

In this paper, we proposed a general framework that considers content feature, structure feature and context feature to extract spotlight contents from online discussions. Based on the proposed framework, an algorithm that emphasizes extracting spotlight contents from time perspective is developed. The effectiveness of generating spotlight contents and the quality of generated spotlight contents are evaluated with real-world online discussion experiments. In the future, we plan to improve the efficiency of the proposed approach and explore the details of topic extraction. In addition, investigating the rationale behind determining spotlight contents is essential to enhance the quality of the extracted spotlight contents. Furthermore, additional practical demonstrations with various combinations of discussion samples and user groups should be conducted to ensure more comprehensive validation.

Acknowledgements

This work was supported by JST CREST JPMJCR20D1 and AIP Challenge, and JSPS KAKENHI Grant Number JP24K20900.

References

[1] R. Shortall, A. Itten, M.v.d. Meer, P. Murukannaiah, and C. Jonker, Reason against the machine? future directions for mass online deliberation, Frontiers in Political Science, vol.4, 2022.
[2] W. Gu, A. Moustafa, T. Ito, M. Zhang, and C. Yang, "A Case-based Reasoning Approach for Supporting Facilitation in Online Discussions," Group Decis Negot, vol.30, pp.719–742, 2021.
[3] T. Ito, Y. Imi, T.K. Ito, and E. Hideshima, Collagree: A faciliator-mediated large-scale consensus support system, Proc. 2nd ACM Collective Intelligence Conference, 2014.
[4] P. Raman, T. Avery, C. Brett, and J. Hewitt, "Exploring the Use of #Hashtags as an Easy Entry Solution to Enhance Online Discussions," IJEDE, vol.35, no.1, 2020.
[5] S. Shiramatsu, K. Kitagawa, S. Naito, H. Koura, and C. Cai, Four approaches to developing autonomous facilitator agent for online and face-to-face public debate, In: A. Nolte, C. Alvarez, R. Hishiyama, I.-A. Chounta, M.J. Rodríguez-Triana, and T. Inoue (eds.), Collaboration Technologies and Social Computing, pp.191–200, Springer, Cham, 2020.
[6] H. Ishizuka, S. Shiramatsu, and K. Ono, Prototyping agents for resolving opinion biases toward facilitating sublation of conflict in web-based discussions, 2022 IEEE International Conference on Agents (ICA), pp.18–23, 2022.
[7] M. Klein, How to harvest collective wisdom on complex problems: An introduction to the mit deliberatorium, CCI working paper, 2011.
[8] A. Sengoku, T. Ito, K. Takahashi, S. Shiramatsu, T. Ito, E. Hideshima, and K. Fujita, Discussion tree for managing large-scale internet-based discussions, Proc. 4th ACM Collective Intelligence Conference, 2016.
[9] W. Gu, S. Kato, F. Ren, G. Su, T. Ito, and S. Hasegawa, "Influence propagation based influencer detection in online forum," IEICE Trans. Inf. & Syst., vol.106, no.4, pp.433–442, 2023.
[10] N. Firoozeh, A. Nazarenko, F. Alizon, and B. Daille, "Keyword Extraction: Issues and Methods," Natural Language Engineering, vol.26, no.3, pp.259–291, 2020.
[11] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent dirichlet allocation," Journal of machine Learning research, pp.993–1022, Jan. 2003.
[12] W.S. El-Kassas, C.R. Salama, A.A. Rafea, and H.K. Mohamed, "Automatic text summarization: A comprehensive survey," Expert Systems with Applications, vol.165, 113679, 2021.
[13] H. Jin, Y. Zhang, D. Meng, J. Wang, and J. Tan, A comprehensive survey on process-oriented automatic text summarization with exploration of llm-based methods, arXiv preprint arXiv:2403.02901, 2024.
[14] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying conditional random fields to Japanese morphological analysis," Proc. EMNLP, pp.230–237, 2004.
[15] T. Sato, T. Hashimoto, and M. Okumura, "Implementation of a word segmentation dictionary called mecab-ipadic-NEologd and study on how to use it effectively for information retrieval," Proc. 23rd Annual Meeting of the Association for NLP, pp.NLP2017–B6–1, 2017 [in Japanese].
[16] R. Rehurek and P. Sojka, Gensim–python framework for vector space modelling, NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic, vol.3, no.2, 2011.
[17] OpenAI, Gpt-4 technical report, arXiv preprint, 2023.

Zhizhong Wang received his M.S. from Japan Advanced Institute of Science and Technology in 2024. Currently working at Subaru Inc. While a student, he was engaged in research and development of online discussion support system.

Wen Gu is an associate professor at Nagoya Institute of Technology. He received his Ph.D. from Nagoya Institute of Technology and University of Wollongong in 2022 as the first runner of the Joint Degree Doctoral Program in Informatics. His current research interests lie in consensus informatics, agent-based systems, social network analysis and learning support. Wen has published his research work in the journals such as Group Decision and Negotiation, Applied Intelligence and international conferences such as IJCAI, PRICAI, IEEE ICA and KICSS.

Zhaoxing Li is a Research Fellow at the University of Southampton. He received his PhD from the Durham University in 2024. His current research interests lie in Deep Reinforcement Learning, Large Language Models (LLMs) and designing AI systems for diverse end users. Zhaoxing has published his research work in conferences, such as Human Factors in Computing Systems (CHI), IEEE VR, ACM Conference on Intelligent User Interfaces (IUI), the Designing Interactive Systems Conference (DIS), the Artificial Intelligence in Education Conference (AIED) and the Journal of Neural Computing and Applications (NCA). Key projects include the EPSRC-funded AGENCY: Assuring Citizen Agency in a World with Complex Online Harms' (EP/W032481/1) and Citizen-Centric Artificial Intelligence Systems (EPSRC). Zhaoxing actively participated in academic circles as a peer reviewer and committee member in several international conferences and journals, like the AAAI, CHI, IEEE VR, AAMAS, AIED, ITS and NCA.

Koichi Ota received his B.S., M.S., and Ph.D. degrees in Information and Communication Engineering from The university of Electro-Communications in 2005,2007, and 2011 respectively. He is now an assistant professor of Research Center for Advanced Computing Infrastructure in Japan Advanced Institute of Science and Technology. Since 2004, he has been engaged in research on web based self-directed learning support. His main research interests lie in designing of learning models, meta-cognition in hyperspace, and scaffolding with cognitive tools.

Shinobu Hasegawa is currently a professor at Center for Innovative Distance Education and Research, JAIST. He received his B.S., M.S., and Ph.D. degrees in systems science from Osaka University in 1998, 2000, and 2002, respectively. The primary goal of his research is to facilitate "Human Learning and Computer-mediated Interaction" in a distributed environment. His research field is mainly learning technology which includes support for Web-based learning, game-based learning, cognitive skill learning, affective learning, distance learning system, and community-based learning.