Regular Paper

Context Style Explanation for Recommender Systems

Masahiro Sato[1,a) Koki Nagatani[1,b) Takashi Sonoda[1,c) Qian Zhang[1,d) Tomoko Ohkuma[1,e)

Received: March 4, 2019, Accepted: July 2, 2019

Abstract: Recommender systems support users by helping them choose items, and explanations for the recommendations further enhance such support. Previous explanation styles were based on information about users and items, such as the demographics of users and contents of items. Contexts, such as "usage scenarios" and "accompanying persons," have not been used for explanations, although they influence user's choice of items. In this paper, we propose a context style explanation method, presenting contexts suitable for consuming the recommended items. The expected impacts of context style explanations are as follows: 1) persuasiveness: recognition of a suitable context for usage motivates users to consume items, and 2) usefulness: envisioning a context helps users to make the right choices because the values of items depend on contexts. We evaluate the persuasiveness and usefulness of the context-style explanation by a crowdsourcing-based user study in a restaurant recommendation setting. The context style explanation is compared to the demographic and content style explanations. We also combine the context style and other explanation styles, confirming that hybrid styles improve the persuasiveness and usefulness of the explanation. Further, we investigate the personal preferences for explanation styles and reveal how gender and age relate to such preferences. The contributions of this paper are: the proposal of the novel context style explanation method, the demonstration of the persuasiveness and usefulness of the proposed method by a user study, and the findings of gender- and age-dependence of explanation style preferences.

Keywords: recommender system, explanation, context-awareness, user evaluation

1. Introduction

Recommender systems help users to select items from a large number of candidates. Such systems estimate user's preference for items (such as books, movies, and restaurants) from past histories of user's actions (such as purchases, views, and visits), and then present items that fit the user's tastes. Users can then select their favorite items from the recommended items.

Explaining the reason for recommendations further supports user decision-making, as it helps a user understand why an item is recommended; such an understanding leads to a better decision regarding whether to choose the recommended item. The explanation also invokes the user's interest in the recommended item.

Several explanation styles have been proposed [27], [32], [33], [35]. For example, the neighbor style explanation provides ratings from similar users. The influence style explanation shows items related to those recommended from user's purchase history. The demographic style explanation describes user's age and gender. The content style explanation displays item features, such as keywords for books and user-generated tags for movies. These four styles are based on information related to users or items because these elements influence user's decision processes [11], [28].

Contextual factors such as time, location, companion, and purpose are also essential elements that affect user's decision-making [9]. Context-aware recommender systems [1], [2] have been developed to model user's choices under various contexts, improving the prediction performance for items preferred by users.

However, contexts have never been used for the explanation of recommendations. Contexts are entities different from users and items, and previous explanation methods have used information of either users or items. Considering that user's decisions depend on contexts, explaining recommendations using context will help users.

In this study, we propose a new style of explanation using contexts. The context style explanation presents contexts suitable for the recommended items. For example, *"This restaurant is recommended to you because it is suitable for dates with your girlfriend/boyfriend"*, where "dates with your girlfriend/boyfriend" is the context presented as an explanation. The selected context should also be related to the user; the user might be more interested in the explanation if she or he is familiar with the context. To generate appropriate context style explanations and item recommendations, we select context-item pairs for each user by fulfilling three affinities: a) a user-item match, b) an item-context match, and c) a user-context match.

We expect two impacts of contexts in explanations:

- **Persuasiveness**: the exhibited contexts induce users to picture situations in which they will consume the recommended items in the future, motivating them to make choices.
- **Usefulness**: users select items based on contexts. Therefore, suggested usage contexts should help user's decision-making.

In this study, we investigate these aforementioned impacts of

1 Fuji Xerox Co., Ltd., Yokohama, Kanagawa 220–8668, Japan
a) sato.masahiro@fujixerox.co.jp
b) nagatani.koki@fujixerox.co.jp
c) sonoda.takashi@fujixerox.co.jp
d) qian.zhang@fujixerox.co.jp
e) ohkuma.tomoko@fujixerox.co.jp

Table 1 Overview of the conventional explanation styles and the proposed context style explanation.

Explanation Style	Displayed Information	Example
Neighbor	Ratings or the fact of purchases of similar users	Users similar to you also visit this restaurant
Influence	Items related to recommended ones from users' past consumption	Recommend for those who also visited *Restaurant C*.
Demographic	Gender, age, profession, etc. of users	Recommend for female students in their 20s.
Content	Content of recommended items, represented by extracted keywords or annotated tags	Recommend for those who like hamburgs.
Context (ours)	Context when users would consume recommended items	Recommend for use in a matchmaking party.

context style explanations with the following three aspects: 1) comparison with other explanation styles, 2) hybridization of context style and other styles, and 3) dependence on gender and age. To investigate these aspects, we implement a restaurant recommender system with context style explanations, conducting a user study via crowdsourcing.

The contributions of this paper are summarized as follows.

- We propose a novel context style explanation for recommender systems.
- We verify persuasiveness and usefulness of the proposed explanation methods by conducting a user study.
- We reveal that the personal preference of explanation styles depend on gender and age.

The remainder of this paper is organized as follows. Related work is presented in the next section. Subsequently, the context style explanation method is described. Afterwards, the experimental details are explained, followed by results and discussion. Our conclusion is summarized in the last section.

2. Related Work

This work is mostly related to two research fields: 1) explanation of recommendation and 2) use of context in recommendation.

2.1 Explanation of Recommendation

Explaining recommendation is important for users to understand the reasoning behind them. Explanations have various effects on users [27], [32], [33], [35]. They can help gain users' trust [28] and increase the acceptance of recommendations [14]. They also help users evaluate items accurately [6] and change user's evaluation of items [11].

Various explanation styles have been proposed and evaluated through user studies. Herlocker et al. [14] compared various explanations with different styles and different visualizations, which included histograms of the user's neighbors' ratings (i.e., neighbor style); similarity to other items in the user's profile (i.e., influence style); and the user's favorite actor or actress (i.e., content style). Demographic information has been used for explanations in the tourism domain (i.e., demographic style) [3]. Bilgic et al. [6] demonstrated that explanations using keywords (i.e., content style) or showing items influencing recommendations (i.e., influence style) help users evaluate items effectively. Various kinds of contents and ways to visualize them have been explored for content style explanations. User-annotated tags are used for explanations [34] and are displayed in a tagcloud interface [10]. Musto et al. [24] showed that fusing linked open data and choosing specific properties improves explanations. Organizational explanations show the pros and cons of items extracted from user reviews, according to user's priorities [23]. Chen et al. [8] further elaborated the organizational explanation by grouping similar trade-off items. Recent research has endeavored to generate personalized natural language explanations of items [7], [12], [19], [20], which can be regarded as advanced content style explanations. Chang et al. [7] generated explanations via the collaboration of crowdworkers and intelligent systems. State-of-the-art neural network models are also used for that purpose [12], [19], [20].

Although there is a vast amount of research on explaining recommendations, most rely on the four types of information shown in **Table 1**: neighbor, influence, demographic, and content. In addition to these four types, context significantly influences user's decision-making [9]. Zheng [37] and Papadimitriou [27] alluded to the possibility of using contexts for explanations. However, contexts have not yet been used for explaining recommendations. To the best of our knowledge, this is the first study of context style explanations.

Furthermore, several explanation styles can be hybridized [27], [31]. Symeonidis et al. [31] combined content and influence style explanations. The visualization of complex hybrid explanations has also been investigated in Ref. [18]. In this study, we investigate hybrids of context style with other explanation styles.

The preferences of explanation styles might depend on users, which has not been well explored in the previous research. Recently, McInerney et al. used a bandit algorithm to personalize explanation styles [22]. We also explore the personal preferences of explanation styles. While McInerney et al. focused on model performance, we clarify how the preferences differ by gender and age. In addition, our investigation of the personal preference includes the proposed context style explanation and hybrids of explanation styles, which were not included in the previous study.

2.2 Use of Context in Recommendation

Users evaluate items differently depending on the context, and recommender systems should thus be aware of the influence of contexts [2]. Context-aware item recommendation is a task that involves recommending items suitable for a user in a specific context. The traditional approaches to context-aware recommendations are contextual pre-filtering and post-filtering, in which ratings or items are filtered by relevance to the context either in the initial stage or in the final stage of the recommendation process [1], [26]. Directly modeling user-item-context relation is

Table 2 Comparison of task settings.

Task	Input	Output
Context-aware item recommendation	User + Context	Item
Context recommendation	User (+ Item)	Context
Context style explanation (ours)	User	Context + Item

called contextual modeling and tends to outperform pre-filtering or post-filtering. Since multi-dimensions of user-item-context can be expressed as a tensor, the direct approach involves utilizing a tensor factorization [16]. However, an exact tensor factorization with Tucker decomposition requires a vast amount of computational resources. Approximation for pairwise interactions can achieve comparable or even better performance [5], [29], [30].

Even if a user has already chosen items, there is a room to choose contexts for consuming the items. The notion of recommending contexts to users has recently been investigated [5], [36], [37]. Context recommendation is a task that involves recommending contexts suitable for a user and an item. Baltrunas et al. [5] collected a dataset of best usage context for each music and predicted the context using variants of the nearest neighbor technique. Zheng [36] compared several multi-label classification techniques for the same task to recommend contexts conditioned on users and items. Zheng [37] also recommended contexts to users according to user's preference regarding contexts.

To generate context style explanations, we select context-item pairs for each user. The difference between our task for context style explanations and the conventional tasks are summarized in **Table 2**. Context-aware recommendation generates lists of recommended items for designated users and contexts. Contexts are pre-selected, either explicitly (e.g., users input purpose of travel into a hotel booking site) or implicitly (e.g., current place and activity can be estimated from wearable sensors). There are two kinds of tasks for context recommendation: recommending context for specified user-item pair [5], [36], and recommending context for a user [37]. In the former task, items to be consumed in recommended contexts are fixed. In the latter task, items are irrelevant to context recommendations. In the case of our context style explanation, pairs of items and contexts are provided for each user. Both contexts and items are undetermined. Therefore, the context style explanation is different in terms of the recommendation task setting. We address this difference by a modification of negative sampling in the model training. The primiry focus of this study is to evaluate the impacts of the context style explanation, and the performance improvement of the above task is left for future research.

3. Context Style Explanation

The generation of context style explanations involves two steps: (1) selection of context-item pairs for users, and (2) suggestion of the context of a context-item pair as the item's explanation.

3.1 Selection of Context-item Pairs

Our context style explanation suggests contexts that the user might encounter in the future. This means both the context and the item are unknown in our task, while the context is predetermined in context-aware item recommendation. In this case, the recommender needs to select appropriate pairs of contexts and items for the users. This requires three affinities: a) a user-item match, b) an item-context match, and c) a user-context match. For a restaurant recommendation, the recommended restaurant should match the user's preferences, just as with non-contextualized recommender systems (user-item match). Moreover, the recommended restaurant should match the suggested context (item-context match). If the context of eating with children is suggested in an explanation, then the recommended restaurant should be suitable for that situation. Additionally, the recommended context should be one anticipated by the user (user-context match). If the user does not have children and lacks many opportunities to eat out with children, a suggestion of eating out with children would likely be inappropriate.

These above three affinities can be learned via the latent representation of pairwise interactions among user, item, and context features [5], [15], [29]. We use field-aware factorization machines (FFMs) [15] for their efficiency and performance. The FFM splits features to "fields," and incorporates interaction effects among the features of different fields. The FFM is formulated as,

$$\hat{y} = \sum_{j_1=1}^{n} \sum_{j_2=j_1+1}^{n} (\boldsymbol{w}_{j_1}^{f_2} \cdot \boldsymbol{w}_{j_2}^{f_1}) x_{j_1} x_{j_2}, \tag{1}$$

where $\hat{y}$ is a prediction by the FFM; $\boldsymbol{w}_j^f$ is a latent vector of a feature j that interacts with a field f; and x_j is the value of the feature j.

To model interaction among users, items, and contexts, we prepare a user field, an item field, and a context field. Example features in the user field are gender and age. The latent factors of a female user u in her 30s are expressed as,

$$\boldsymbol{w}_u^{\text{Item}} = \boldsymbol{w}_{female}^{\text{Item}} + \boldsymbol{w}_{30s}^{\text{Item}}, \tag{2}$$

$$\boldsymbol{w}_u^{\text{Context}} = \boldsymbol{w}_{female}^{\text{Context}} + \boldsymbol{w}_{30s}^{\text{Context}}. \tag{3}$$

Features in the item field can be restaurant genres or places. The latent factors of *IndianFood* restaurant i located in *PlaceA* are composed as,

$$\boldsymbol{w}_i^{\text{User}} = \boldsymbol{w}_{IndianFood}^{\text{User}} + \boldsymbol{w}_{PlaceA}^{\text{User}}, \tag{4}$$

$$\boldsymbol{w}_i^{\text{Context}} = \boldsymbol{w}_{IndianFood}^{\text{Context}} + \boldsymbol{w}_{PlaceA}^{\text{Context}}. \tag{5}$$

Similarly for context c of *BusinessEntertaining*,

$$\boldsymbol{w}_c^{\text{User}} = \boldsymbol{w}_{BusinessEntertaining}^{\text{User}}, \tag{6}$$

$$\boldsymbol{w}_c^{\text{Item}} = \boldsymbol{w}_{BusinessEntertaining}^{\text{Item}}. \tag{7}$$

Then, Eq. (1) is expressed as follows,

$$\hat{y} = \boldsymbol{w}_u^{\text{Item}} \cdot \boldsymbol{w}_i^{\text{User}} + \boldsymbol{w}_i^{\text{Context}} \cdot \boldsymbol{w}_c^{\text{Item}} + \boldsymbol{w}_u^{\text{Context}} \cdot \boldsymbol{w}_c^{\text{User}}. \tag{8}$$

Each term in Eq. (8) represents a) a user-item match ($\boldsymbol{w}_u^{\text{Item}} \cdot \boldsymbol{w}_i^{\text{User}}$), b) an item-context match ($\boldsymbol{w}_i^{\text{Context}} \cdot \boldsymbol{w}_c^{\text{Item}}$), and c) a user-context match ($\boldsymbol{w}_u^{\text{Context}} \cdot \boldsymbol{w}_c^{\text{User}}$). Latent factors can be learned using user's past interactions. More specifically, if a user u consumed an item i under a context c, a triplet (u, i, c) is assigned a positive label. Triplets that have not appeared in past consumption logs are assigned negative labels. We train the FFM using these positive and negative samples.

Note that definition of negative samples is different for each task setting in Table 2. In the conventional context-aware item recommendation, negative samples are defined for each pair of user-context existing in the logs. Items not consumed under a user-context pair are negative samples. In our case, negative samples are defined for each user. Item-context pairs that have not appeared in the user's logs are negative samples. While the former negative samples include contexts only experienced by the user, the latter negative samples include contexts not experienced by the user. This enables to learn the affinity between users and contexts.

After the training, the best context-item pairs are selected by the score of Eq. (8).

3.2 Suggestion of a Context as Explanation

Selected context-item pairs are used to produce recommendations and explanations. If there is a context-item pair with context c and item i, then item i is presented to a user as a recommendation and context c is used for an explanation. An explanation is generated using human-crafted templates, for example, *"item i is recommended to you because it is suitable for context c."* In case there are several contexts suitable for the recommended item, then the multiple contexts can be displayed together. In this study, for experimental simplicity, we only use one best context for an explanation.

4. Experiment

First, we collected users' restaurant visit logs with context via crowdsourcing. Second, we trained a context-item pair selector using the acquired logs and prepared recommendations and explanations. Finally, we asked the same users to evaluate explanation styles.

4.1 Collecting Dataset

Restaurant visit logs were collected using a Japanese crowdsourcing platform. There are three entries per restaurant: name of a visited restaurant, the URL to the restaurant within a restaurant information site, and the usage scene of the visit (i.e., context). We asked each crowdworker to input a maximum of 20 restaurants. We recruited crowdworkers who lived in certain urban areas *1 to confine the areas of visit logs. We asked for original contexts of users' visits instead of asking for an evaluation under a provided context, because users behave differently under supposed contexts and real contexts [4], [25]. Usage scenes were selected from 15 options, as described in **Table 3**. We prepared the usage scenes carefully to ensure that they would be familiar and unambiguous to users by consulting the descriptions of usage scenes in several restaurant information sites. If crowdworkers thought that more than one scene can be associated with the visit, then they were advised to select the uppermost scene on the list. The numbers of times each context was chosen by crowdworkers are also shown in Table 3. Most contexts obtained substantial votes and *"none of the above"* received only a small portion of the votes. This supports the validity of the context candidates' design; if the context candidates did not include appropriate contexts for users, the votes for *"none of the above"* would have been large.

Table 3 Candidates of 15 usage scenes (contexts) and counts selected by crowdworkers. The crowdworkers chose one context for each visit. The usage scenes were shown to the crowdworkers in the same order of this list. If crowdworkers thought that more than one scene can be associated with the visit, then they were advised to select the uppermost scene on the list.

Usage scene	Count
Matchmaking party	20
Girls' lunch or night out	184
Business entertaining	39
Banquet or drinking party in a large group	15
With children or grandchildren	163
With parents, sisters, or brothers	212
With a husband or wife	414
Dating with opposite gender	275
With close friends (only eating)	284
With close friends (with drink)	325
With colleagues or acquaintances (only eating)	191
With colleagues or acquaintances (with drink)	108
In solitude	386
Take-away	73
None of the above	16

Table 4 Statistics of collected dataset via crowdsourcing.

data	numbers
total visits	2,884
unique users	155
unique items (restaurants)	2,730
unique contexts (usage scenes)	15
genres of restaurants	210
nearest stations of restaurants	473

We obtained 2,884 visit logs from 155 crowdworkers, after removing logs with improper URLs and crowdworkers who provided improper URLs more than half the time. We tried to remove low-quality crowdworkers by removing careless crowdworkers who input improper URLs. There are 2,730 unique URLs in the remaining visit logs. The statistics of the collected dataset are summarized in **Table 4**. The genders and approximate ages of the crowdworkers were provided from the crowdsourcing platform. Among the 155 crowdworkers, 108 were female and 47 were male. Further, 44 crowdworkers were in their 20s, 39 in their 30s, 26 in their 40s, 13 in their 50s, 2 in their 60s, and the ages of 31 were unknown. The average number of visits per restaurant was 1.056 and the sparsity was 99.32%. We crawled the URLs and collected the restaurant's content information, including genres and nearest stations *2. Note that each restaurant is assigned multiple genres (2.3 genres on average). There are 210 unique genres and 473 unique stations.

4.2 Training Recommender and Preparing Explanations

We trained the context-item selector using the collected dataset. We used the libffm library *3 for the FFM. The features of the user field are user ID, gender, and age. The features of the

*1 The urban areas are Tokyo and Kanagawa, the Japanese capital and a neighboring prefecture of Tokyo, respectively.

*2 Restaurants are located in urban areas where public transportation is well developed.

*3 https://github.com/guestwalk/libffm.

Table 5 Samples of seven explanation styles. Phrases emphasized in italics are tailored to fit users, recommended items, and supposed contexts.

Style	Sample
Non-specific	Recommend based on your visit logs
Demographic	Recommend for *"women in their 30s"*
Content	Recommend for those who often visit *"Italian restaurants"*
Context	Recommend for use *"with husband or wife"*
Demographic + Context	Recommend for use *"in business entertaining"* of *"men in their 50s"*
Content + Context	Recommend for use *"in solitude"* for those who often visit *"noodle shops"*
Demographic + Content + Context	Recommend for use *"with close friends (with drink)"* of *"women in their 20s"* who often visit *"cafes"*

item field are genre and nearest station. Using the demographic features of users and the content features of items alleviated the issue of data sparsity. The features of the context field included context ID, which is assigned to 15 usage scenes.

The training of the recommender proceeded as follows. First, the dataset was randomly split into 80% training and 20% validation data. Hyper-parameters of the FFM were then optimized to maximize the AUC (areas under the curve) of the validation data. The following hyper-parameters were chosen: learning rate 0.05, regularization coefficient 0.0005, and dimensions of factor 100. The obtained AUC with these hyper-parameters was 0.865. Finally, we trained the model using the entire dataset to select context-item pairs.

After training the model, we selected seven best context-item pairs for each user, according to the score of Eq. (8). Then, the order of the selected context-item pairs was shuffled, to ensure that recommendation quality did not correlate to the presentation order. Restaurants visited in the past were removed from the list, whereas contexts experienced in the past were not omitted. The same restaurant was recommended only once per user.

We prepared seven explanation styles as described in **Table 5**. The non-specific explanation did not include any specific information regarding demographics, contents, and contexts. This explanation was the same for all users and all recommended items. For the context style explanation, the context of context-item pair was directly assigned for the explanation. For the demographic style explanation, user age and gender were used for the explanation. Recommended items related to the user age and gender, because the recommender incorporates them as user features. For the content style explanation, we selected a genre common among the recommended restaurants and those that the user visited in the past. Hybrid explanation styles were generated via combinations of the procedures described above.

4.3 Evaluating Explanation Styles

We recruited the 155 crowdworkers who had appropriately submitted restaurant visit logs, and 85 participated in the user study. The participants' demographics are shown in **Table 6**. We presented seven restaurant recommendations with seven different explanation styles to each user. Each recommendation was generated by the same FFM model.

Table 6 Demographics of participants (crowdworkers) in the evaluation of the explanation styles. The participants were recruited from the respondents of the initial data collection; this is necessary for personalizing recommendations and explanations to the participants.

Age	Female	Male	Total
20s	10	9	19
30s	20	4	24
40s	14	2	16
50s	4	2	6
60s	1	0	1
unknown	15	4	19
Total	64	21	85

The order of the explanation styles was randomly shuffled among users in order to cancel any biases related to the display order. We asked the following four evaluation questions using a 7-point Likert scale for each pair of restaurant recommendations and explanations.

- Persuasiveness 1 (P1): The explanation is convincing.
- Persuasiveness 2 (P2): The explanation triggers interest.
- Usefulness 1 (U1): The explanation is useful for choice.
- Usefulness 2 (U2): The explanation is easy to understand.

In addition to these evaluation questions, we asked whether the participants visited the recommended restaurants in the past, and whether they knew of them in advance. There were also free entry fields to express any other comments.

5. Results and Discussion

5.1 Quantitative Analysis

Among the restaurants recommended to the participants, 21% were visited in the past and 20% were known in advance. Note that we recommended restaurants not visited by each user in the collected dataset. This indicates that item recommendation was fairly accurate and that our recommender system works fine.

In this subsection, we investigate the persuasiveness and usefulness of context style explanations with the following three aspects: 1) comparison with other explanation styles, 2) hybridization of context style and other styles, and 3) dependence on gender and age.

5.1.1 Comparison with Other Explanation Styles

We experimented with four single explanation styles: non-specific, demographic, content, and context styles. Responses to the four questions are shown in **Fig. 1**. Responses ranged from strongly disagree (−3) to strongly agree (+3). The average response for the context style explanation was higher than that for the demographic style ($p = 0.008$, 0.10, 0.047, and 0.036 for P1, P2, U1, and U2, respectively, via the Wilcoxon signed rank test). The average response for the context style explanation also tended to be higher than that for the content style, though not statistically significant. The non-specific explanations tended to perform better than other single styles; we discuss the possible reasons in the later section.

5.1.2 Hybridization of Context Style and Other Styles

We combined the context styles with other styles. **Figure 2** shows the comparison between the single and hybrid explanation styles. The combination of the demographic and context styles

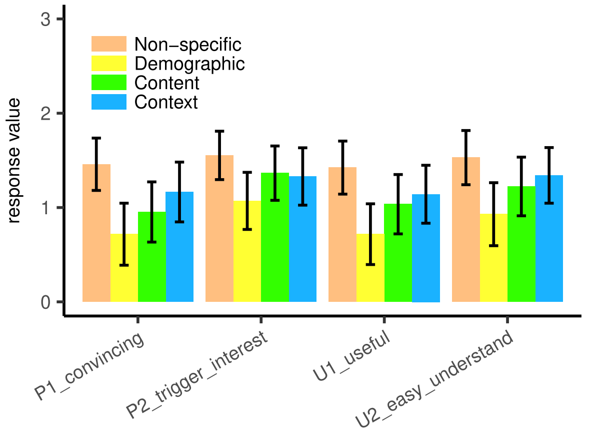


Fig. 1 Responses to four evaluation questions for four single explanation styles. Error bars represent 95% confidence intervals of average response values.

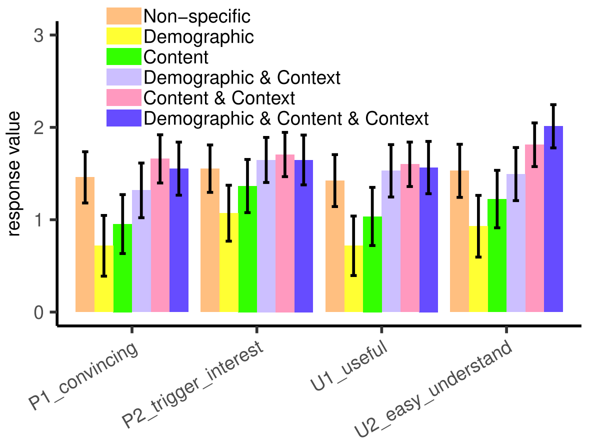
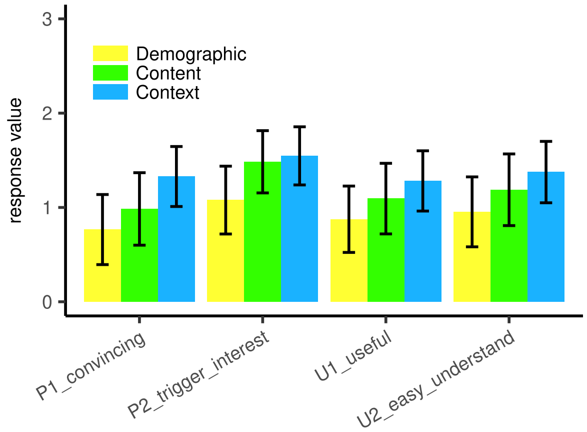
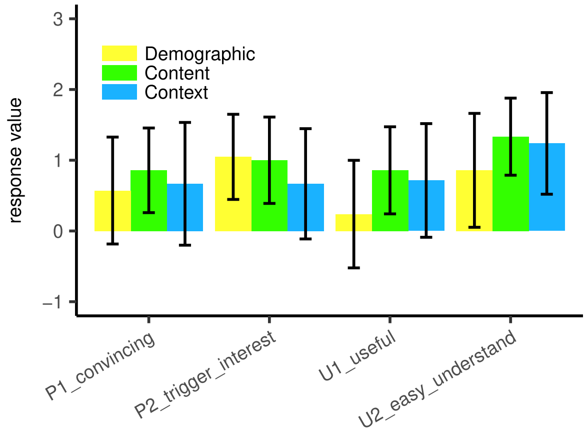


Fig. 2 Comparison between single and hybrid explanation styles. Error bars represent 95% confidence intervals of average response values.



(a) Female users.



(b) Male users.

Fig. 3 Responses for single explanation styles by gender.

outperformed the demographic-only style ($p < 0.01$ for all questions), and the combination of the content and context styles outperformed the content-only style ($p = 0.061$ for P2, and $p < 0.01$ for others). The combination of content and context styles also tended to outperform non-specific style (though not statistically significant). The triple combination of demographic, content, and context styles did show a better performance compared to the dual combination of demographic and context styles.

5.1.3 Dependence on Gender and Age

The preference of explanation styles might depend on users. We investigated the difference in the response for each explanation style in terms of gender and age.

First, we describe the difference by gender. **Figure 3** and **Fig. 4** show the comparisons of single explanation styles and hybrid explanation styles, respectively, for female and male users. While female users tended to prefer the context style over the content style, male users tended to prefer the content style over the context style (Fig. 3). In terms of hybrids (Fig. 4), male users preferred the triple combination more than the dual combinations (not significant for P1 and P2, $p = 0.062$ for U1 and $p = 0.020$ for U2 with comparison of Demographic & Content & Context vs. Content & Context).

Next, we investigated the difference by age. **Figure 5** and **Fig. 6** show the comparisons of single explanation styles and hybrid explanation styles, respectively, for young to elder users. While young users preferred the content style over the context style, middle-age users tended to prefer the context style over the content style (Fig. 5). In terms of hybrids (Fig. 6), there was no clear difference by age.

5.2 Qualitative Analysis

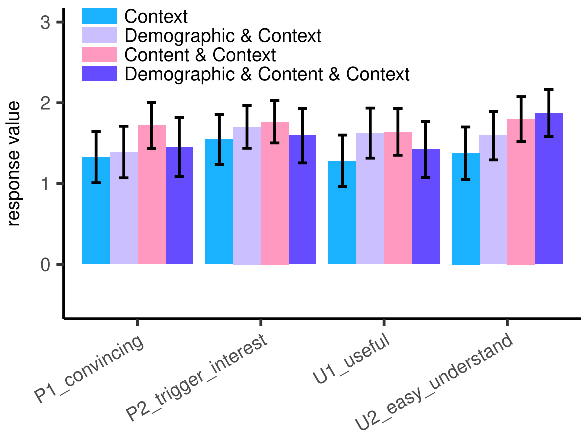
Sixty-four participants input at least one comment, and 293 comments were obtained in total. To further understand the user perception of the context style explanation, we investigated these comments.

User comments indicated two reasons of persuasiveness:

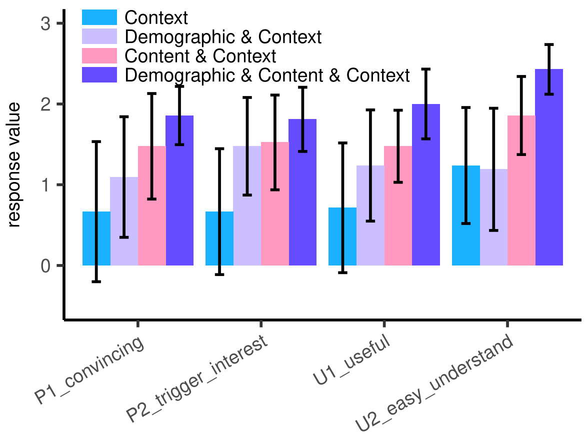
(1) Relevance of the proposed context to users:
- *"Under my current environment, it's a very interesting recommendation, so I became to feel like going."*
- *"I think I want to go because this situation is probable for me."*

(2) Recognition of appropriate context for usage:
- *"I think I am going to use this when I organize a drinking party."*
- *"I have been interested in Japanese rice wine bars, but few of my close friends like it. Visiting here with my colleagues sounds nice."*

Users also mentioned the usefulness of the context for decision-making.



(a) Female users.



(b) Male users.

Fig. 4 Comparison between single and hybrid explanation styles by gender.

- *"I'm afraid of making a wrong choice for a girls' night out, so this explanation is useful."*

These findings from the qualitative analysis support the importance of context for explanation.

On the other hand, there were three kinds of negative responses to the context style explanations.

(1) Mismatch of context and restaurants:

- *"This restaurant is a standing bar, which is not suitable for dating."*
- *"I don't think it's a good idea to use a buffet restaurant for a banquet."*

(2) Context is irrelevant for someone's choice:

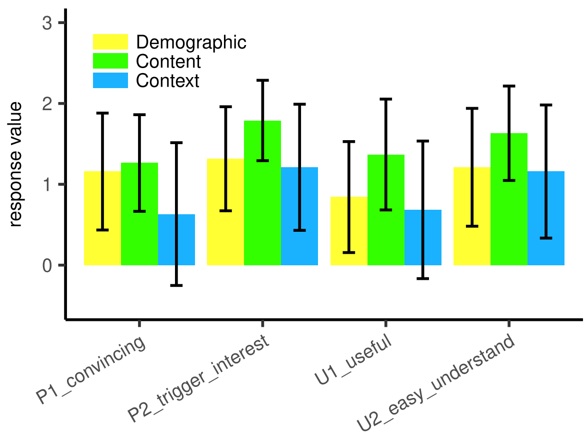
- *"I choose restaurants by whether I like the menu or not."*
- *"I eat out only with my close friends, so information about situation is useless."*

(3) Needs for finer granularity of context:

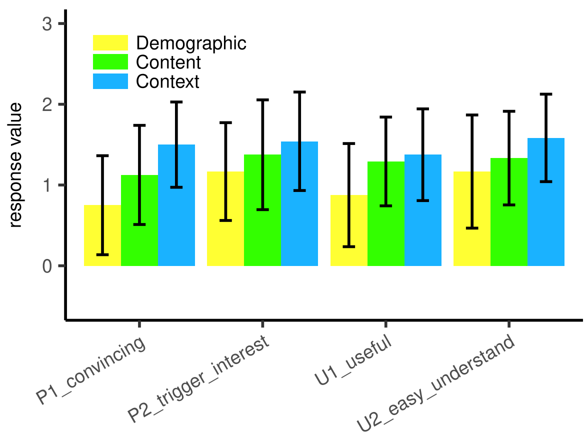
- *"You mention just dating, but is it referring to ordinary dating or anniversary dating?"*

5.3 Discussion

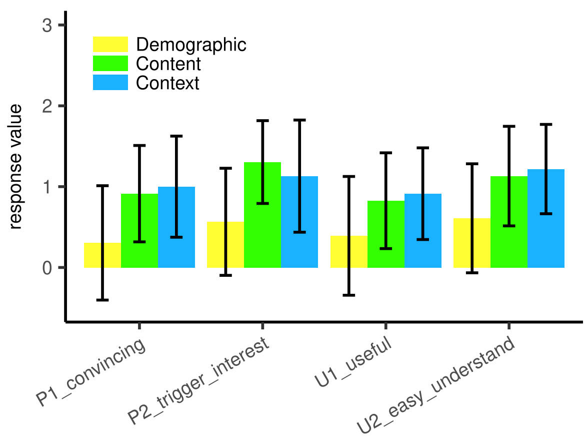
The non-specific explanation tended to perform better than the context style, and the hybrid of content and context styles tended to perform better than the non-specific explanation (Fig. 1 and Fig. 2). Relatively high appraisals of the non-specific explanation might be a result of familiarity with the explanation style.



(a) Young users (20s).



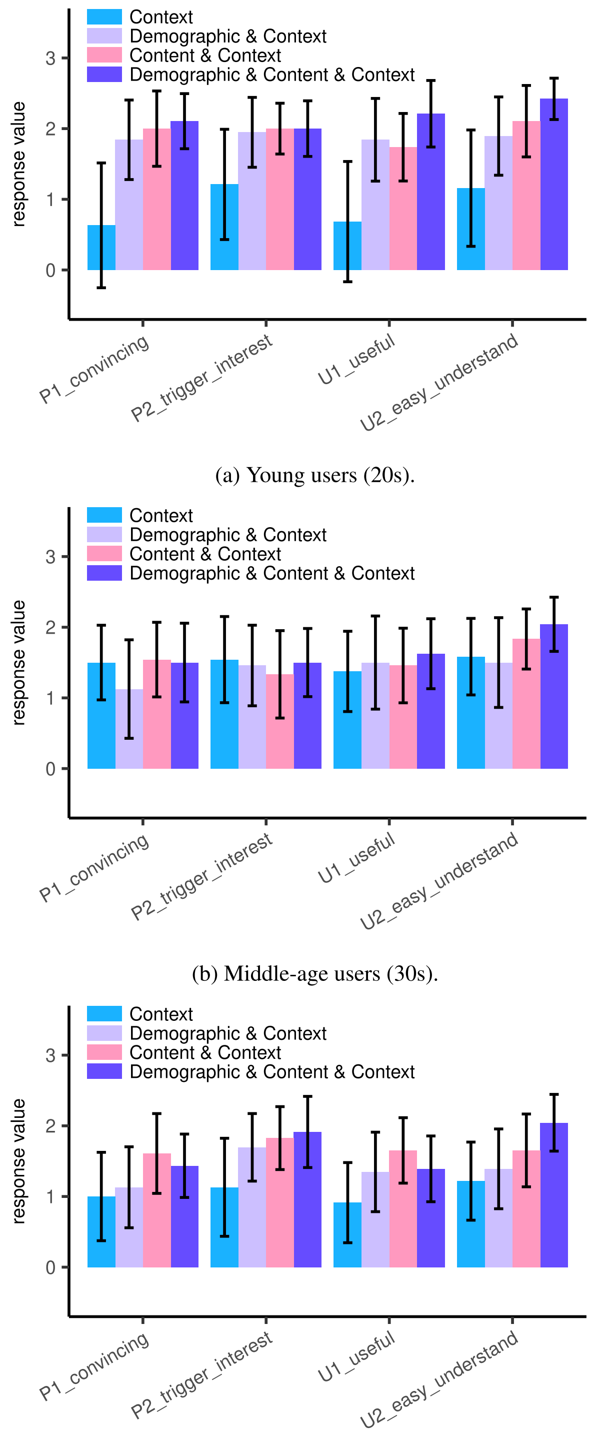
(b) Middle-age users (30s).



(c) Elder users (40s, 50s, 60s).

Fig. 5 Responses for single explanation styles by age.

Some users commented as follows: *"There is a comfort in this type of explanation,"* and *"This writing style suits me the best."* The crowdsourcing platform we used provides task recommendation for users with explanations of this style: *"Recommendation is based on your past task."* Another reason might be the occasional mismatch of the presented context, as seen in the example comments indicating the mismatch shown in the previous section. User evaluation tends to be affected more by negative experiences (i.e., mismatches) than by positive experiences (i.e., good matches). Similar observations were reported in an experiment



(a) Young users (20s).

(b) Middle-age users (30s).

(c) Elder users (40s, 50s, 60s).

Fig. 6 Comparison between single and hybrid explanation styles by age.

of personalizing engaging messages [17]. We plan to improve the accuracy of context-item pair selections in future works. As described in Section 2.2, the selection of context-item pairs is an unexplored new task setting. Thus, we believe there is much room for improvement.

The trio of demographics, contents, and contexts did not produce a significant improvement over the duo of contents and contexts. Users may have felt excessive complexity. Determining an adequate amounts of information for an explanation would be an interesting challenge.

As for gender dependence in the preferences of explanation styles, female users tended to prefer the context style and male users tended to prefer the content style (Fig. 3). This might be a result of difference in personality traits; it is known that women are more tender-minded than men [13]. The used contexts in this study included accompanying persons, which is important information for tender-minded people. Regarding age dependence in the preferences of explanation styles, users in their 20s tended to prefer the content style and users in their 30s tended to prefer the context style (Fig. 5). Tender-mindedness correlates positively with ages [21]; this might be the reason for the age dependence. However, the preference for the context style does not increase for users over the age of 40 years compared to users in their 30s.

This work was conducted in a restaurant recommendation domain. Context is important for recommendations in various domains such as movie, travel, and music [2]. Hence, we believe that context style explanations can be applied to various domains, though relevant contexts should be unique to those different domains. Future research should investigate those other domains. Further, we compared and hybridized context style explanations with demographic and content styles. Future studies should experiment with other conventional explanation styles (e.g., neighbor and influence styles). In addition, we evaluated persuasiveness and usefulness to verify our hypothesis of the effects of the context-style. Evaluation of other factors, such as user trust and decision efficiency, should be conducted in future work.

6. Conclusion

In this paper, we proposed the context style explanation for recommenders. We conducted a crowdsourcing-based user study to measure persuasiveness and usefulness. The context style explanation was better than the demographic style. The context style also tended to perform better than the content style, although the difference was not statistically significant. We further confirmed that the hybrids of the context style and other explanation styles improve persuasiveness and usefulness. Findings from user comments support the importance of contexts for explanations. In addition, we revealed the personal preferences of explanation styles in terms of gender and age. While female or middle-age users tended to prefer the context style over the content style, male or young users tended to prefer the content style.

References

[1] Adomavicius, G., Sankaranarayanan, R., Sen, S. and Tuzhilin, A.: Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach, *ACM Trans. Inf. Syst.*, Vol.23, No.1, pp.103–145 (online), DOI: 10.1145/1055709.1055714 (2005).

[2] Adomavicius, G. and Tuzhilin, A.: *Context-aware recommender systems*, pp.191–226, Springer US (2015).

[3] Ardissono, L., Goy, A., Petrone, G., Segnan, M. and Torasso, P.: Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices, *Applied Artificial Intelligence*, Vol.17, No.8-9, pp.687–714 (2003).

[4] Asoh, H., Motomura, Y. and Ono, C.: An analysis of differences between preferences in real and supposed contexts, *Knowledge Creation Diffusion Utilization* (2010).

[5] Baltrunas, L., Ludwig, B. and Ricci, F.: Matrix Factorization Techniques for Context Aware Recommendation, *Proc. 5th ACM Conference on Recommender Systems, RecSys '11*, pp.301–304, ACM (online), DOI: 10.1145/2043932.2043988 (2011).

[6] Bilgic, M. and Mooney, R.J.: Explaining recommendations: Satis-

faction vs. promotion, *Beyond Personalization Workshop, IUI*, Vol.5, p.153 (2005).
[7] Chang, S., Harper, F.M. and Terveen, L.G.: Crowd-Based Personalized Natural Language Explanations for Recommendations, *Proc. 10th ACM Conference on Recommender Systems*, *RecSys '16*, pp.175–182, ACM (online), DOI: 10.1145/2959100.2959153 (2016).
[8] Chen, L. and Wang, F.: Explaining Recommendations Based on Feature Sentiments in Product Reviews, *Proc. 22Nd International Conference on Intelligent User Interfaces*, *IUI '17*, pp.17–28, ACM (online), DOI: 10.1145/3025171.3025173 (2017).
[9] Chen, W., Hoyle, C. and Wassenaar, H.J.: A choice modeling approach for usage context-based design, *Decision-Based Design*, pp.255–285, Springer (2013).
[10] Chen, W., Hsu, W. and Lee, M.L.: Tagcloud-based Explanation with Feedback for Recommender Systems, *Proc. 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, *SIGIR '13*, pp.945–948, ACM (online), DOI: 10.1145/2484028.2484108 (2013).
[11] Cosley, D., Lam, S.K., Albert, I., Konstan, J.A. and Riedl, J.: Is Seeing Believing?: How Recommender System Interfaces Affect Users' Opinions, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, *CHI '03*, pp.585–592, ACM (online), DOI: 10.1145/642611.642713 (2003).
[12] Dong, L., Huang, S., Wei, F., Lapata, M., Zhou, M. and Xu, K.: Learning to generate product reviews from attributes, *Proc. 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, Vol.1, pp.623–632 (2017).
[13] Feingold, A.: Gender differences in personality: A meta-analysis., *Psychological bulletin*, Vol.116, No.3, p.429 (1994).
[14] Herlocker, J.L., Konstan, J.A. and Riedl, J.: Explaining Collaborative Filtering Recommendations, *Proc. 2000 ACM Conference on Computer Supported Cooperative Work*, *CSCW '00*, pp.241–250, ACM (online), DOI: 10.1145/358916.358995 (2000).
[15] Juan, Y., Zhuang, Y., Chin, W.-S. and Lin, C.-J.: Field-aware Factorization Machines for CTR Prediction, *Proc. 10th ACM Conference on Recommender Systems*, *RecSys '16*, pp.43–50, ACM (online), DOI: 10.1145/2959100.2959134 (2016).
[16] Karatzoglou, A., Amatriain, X., Baltrunas, L. and Oliver, N.: Multiverse Recommendation: N-dimensional Tensor Factorization for Context-aware Collaborative Filtering, *Proc. 4th ACM Conference on Recommender Systems*, *RecSys '10*, pp.79–86, ACM (online), DOI: 10.1145/1864708.1864727 (2010).
[17] Kocielnik, R. and Hsieh, G.: Send Me a Different Message: Utilizing Cognitive Space to Create Engaging Message Triggers, *Proc. 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, *CSCW '17*, pp.2193–2207, ACM (online), DOI: 10.1145/2998181.2998324 (2017).
[18] Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J. and Getoor, L.: User Preferences for Hybrid Explanations, *Proc. Eleventh ACM Conference on Recommender Systems*, *RecSys '17*, pp.84–88, ACM (online), DOI: 10.1145/3109859.3109915 (2017).
[19] Lipton, Z.C., Vikram, S. and McAuley, J.: Generative concatenative nets jointly learn to write and classify reviews, arXiv preprint arXiv:1511.03683 (2015).
[20] Lu, Y., Dong, R. and Smyth, B.: Why I Like It: Multi-task Learning for Recommendation and Explanation, *Proc. 12th ACM Conference on Recommender Systems*, *RecSys '18*, pp.4–12, ACM (online), DOI: 10.1145/3240323.3240365 (2018).
[21] McCrae, R.R., Costa, P.T., de Lima, M.P., Simões, A., Ostendorf, F., Angleitner, A., Marušić, I., Bratko, D., Caprara, G.V., Barbaranelli, C., et al.: Age differences in personality across the adult life span: Parallels in five cultures, *Developmental Psychology*, Vol.35, No.2, p.466 (1999).
[22] McInerney, J., Lacker, B., Hansen, S., Higley, K., Bouchard, H., Gruson, A. and Mehrotra, R.: Explore, Exploit, and Explain: Personalizing Explainable Recommendations with Bandits, *Proc. 12th ACM Conference on Recommender Systems*, *RecSys '18*, pp.31–39, ACM (online), DOI: 10.1145/3240323.3240354 (2018).
[23] Muhammad, K.I., Lawlor, A. and Smyth, B.: A Live-User Study of Opinionated Explanations for Recommender Systems, *Proc. 21st International Conference on Intelligent User Interfaces*, pp.256–260, ACM (online), DOI: 10.1145/2856767.2856813 (2016).
[24] Musto, C., Narducci, F., Lops, P., De Gemmis, M. and Semeraro, G.: ExpLOD: A Framework for Explaining Recommendations Based on the Linked Open Data Cloud, *Proc. 10th ACM Conference on Recommender Systems*, *RecSys '16*, pp.151–154, ACM (online), DOI: 10.1145/2959100.2959173 (2016).
[25] Ono, C., Takishima, Y., Motomura, Y. and Asoh, H.: Context-Aware Preference Model Based on a Study of Difference Between Real and Supposed Situation Data, *Proc. 17th International Conference on User Modeling, Adaptation, and Personalization: Formerly UM and AH*, *UMAP '09*, pp.102–113, Springer-Verlag (online), DOI: 10.1007/978-3-642-02247-0_12 (2009).
[26] Panniello, U., Tuzhilin, A., Gorgoglione, M., Palmisano, C. and Pedone, A.: Experimental Comparison of Pre- vs. Post-filtering Approaches in Context-aware Recommender Systems, *Proc. 3rd ACM Conference on Recommender Systems*, *RecSys '09*, pp.265–268, ACM (online), DOI: 10.1145/1639714.1639764 (2009).
[27] Papadimitriou, A., Symeonidis, P. and Manolopoulos, Y.: A Generalized Taxonomy of Explanations Styles for Traditional and Social Recommender Systems, *Data Min. Knowl. Discov.*, Vol.24, No.3, pp.555–583 (online), DOI: 10.1007/s10618-011-0215-0 (2012).
[28] Pu, P. and Chen, L.: Trust Building with Explanation Interfaces, *Proc. 11th International Conference on Intelligent User Interfaces*, *IUI '06*, pp.93–100, ACM (online), DOI: 10.1145/1111449.1111475 (2006).
[29] Rendle, S., Gantner, Z., Freudenthaler, C. and Schmidt-Thieme, L.: Fast Context-aware Recommendations with Factorization Machines, *Proc. 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, *SIGIR '11*, pp.635–644, ACM (online), DOI: 10.1145/2009916.2010002 (2011).
[30] Rendle, S. and Schmidt-Thieme, L.: Pairwise Interaction Tensor Factorization for Personalized Tag Recommendation, *Proc. 3rd ACM International Conference on Web Search and Data Mining*, *WSDM '10*, pp.81–90, ACM (online), DOI: 10.1145/1718487.1718498 (2010).
[31] Symeonidis, P., Nanopoulos, A. and Manolopoulos, Y.: MoviExplain: A Recommender System with Explanations, *Proc. 3rd ACM Conference on Recommender Systems*, *RecSys '09*, pp.317–320, ACM (online), DOI: 10.1145/1639714.1639777 (2009).
[32] Tintarev, N. and Masthoff, J.: A Survey of Explanations in Recommender Systems, *Proc. 2007 IEEE 23rd International Conference on Data Engineering Workshop*, *ICDEW '07*, pp.801–810, IEEE Computer Society (online), DOI: 10.1109/ICDEW.2007.4401070 (2007).
[33] Tintarev, N. and Masthoff, J.: Explaining recommendations: Design and evaluation, *Recommender Systems Handbook*, pp.353–382, Springer (2015).
[34] Vig, J., Sen, S. and Riedl, J.: Tagsplanations: Explaining Recommendations Using Tags, *Proc. 14th International Conference on Intelligent User Interfaces*, *IUI '09*, pp.47–56, ACM (online), DOI: 10.1145/1502650.1502661 (2009).
[35] Zhang, Y. and Chen, X.: Explainable recommendation: A survey and new perspectives, arXiv preprint arXiv:1804.11192 (2018).
[36] Zheng, Y., Mobasher, B. and Burke, R.: Context Recommendation Using Multi-label Classification, *Proc. 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence* (*WI*) *and Intelligent Agent Technologies* (*IAT*) *- Volume 02*, *WI-IAT '14*, pp.288–295, IEEE Computer Society (online), DOI: 10.1109/WI-IAT.2014.110 (2014).
[37] Zheng, Y., Mobasher, B. and Burke, R.: User-Oriented Context Suggestion, *Proc. 2016 Conference on User Modeling Adaptation and Personalization*, *UMAP '16*, pp.249–258, ACM (online), DOI: 10.1145/2930238.2930252 (2016).

Masahiro Sato was born in Toyama, Japan. He received his M.S. degree from Tokyo University in 2005. He is currently a researcher at Fuji Xerox Co. Ltd. His research interest is in recommender systems, especially those triggering user's actions. He is a member of the ACM.

Koki Nagatani was born in Kagoshima, Japan. He received his B.E. and M.E. degree in from Keio University in 2013 and 2015, respectively. Currently, he is working at Fuji Xerox Co., Ltd.. His research interests are on machine learning and recommendation systems.

Takashi Sonoda was born in Nagasaki, Japan. He received his M.E. and Ph.D. degrees from Nihon University in 1982 and 1985, respectively. In 1985, he joined Fuji Xerox Co., Ltd. His research interests are algorithm of recommender systems and mathematical analysis of neural networks. He is a member of the JSAI.

Qian Zhang was born in China. She received her M.S. degree from Waseda University in 2016. She has been engaged in Research and Technology Group of Fuji Xerox since 2017. Her recent research interest is recommender system and personalization.

Tomoko Ohkuma received her M.S. degree from Keio University in 1996. In 1996, she joined Fuji Xerox. After working on computational linguistics and natural language processing, she received her Ph.D. degree from Keio University in 2009. Currently she is a group manager in Communication Technology Laboratory of Fuji Xerox. She is an editorial committee member of Journal of Natural Language Processing.

(Editor in Charge: *Minoru Sasaki*)