Explaining Recommendations Using Contexts

Masahiro Sato, Budrul Ahsan, Koki Nagatani, Takashi Sonoda, Qian Zhang, Tomoko Ohkuma
Fuji Xerox
Yokohama, Japan
sato.masahiro/budrul.ahsan/nagatani.koki/sonoda.takashi/qian.zhang/ohkuma.tomoko@fujixerox.co.jp

ABSTRACT

Recommender systems support user decision-making, and explanations of recommendations further facilitate their usefulness. Previous explanation styles are based on similar users, similar items, 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 decisions. In this paper, we propose a context style explanation method, presenting contexts suitable for consuming recommended items. The expected impacts of context style explanations are 1) persuasiveness: recognition of suitable context for usage motivates users to consume items, and 2) usefulness: envisioning context helps users to make right choices because the values of items depend on contexts. We evaluate context style persuasiveness and usefulness by a crowdsourcing-based user study in a restaurant recommendation setting. The context style explanation is compared to demographic and content style explanations. We also combine context style and other explanation styles, confirming that hybrid styles improve persuasiveness and usefulness of explanation.

Author Keywords

Recommender System; Explanation; Context-awareness; User Evaluation.

INTRODUCTION

Recommender systems help users to choose items from vast amounts of candidates. Explaining recommendations further supports user decision-making. Several explanation styles have been proposed [21,25,26]. For example, the neighbor style explanation provides ratings from similar users. The influence style explanation shows items related to those recommended from users' purchase histories. The demographic style explanation describes user ages and genders. 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 users' decision processes [11,22]. Contextual factors such as time, location, companion, and purpose are also essential to affect user decision-making [9].

IUI'18, March 7–11, 2018, Tokyo, Japan

ACM ISBN 978-1-4503-4945-1/18/03…$15.00
https://doi.org/10.1145/3172944.3173012

In this study, we propose a new style of explanation using contexts. The context style explanation presents contexts suitable for recommended items. For example, "*This restaurant is recommended to you because the restaurant is suitable for dates with your girlfriend/boyfriend*." We expect two impacts of contexts in explanations:

- **Persuasiveness**: the exhibited context induces 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 decision-making.

In this research, we investigate these aforementioned impacts of context style explanations with the following two aspects: 1) comparison with other explanation styles, and 2) hybridization of context and other styles. To investigate these aspects, we implement a restaurant recommender system with context style explanations, conducting a user study via crowdsourcing.

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

RELATED WORK

Herlocker et al. [12] 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 was 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 how to visualize them have been explored for content style explanations. User-annotated tags are used for explanations [27] and are displayed in a tagcloud interface [10]. Musto et al. [19] showed that fusing linked open data and choosing specific properties improves explanations. Organizational explanations show pros and cons of items extracted from user reviews, according to users' priorities [18]. Chen et al. [8] further elaborated the organizational explanation by grouping similar trade-off items. Recent research has striven to generate personalized natural language explanations of items [7,17,28], which can be regarded as advanced content style explanations. Chang et al. [7] generated explanations via collaboration of crowdworkers and intelligent systems. State-of-the-art neural network models are also used for that purpose [17,28].

Moreover, hybrids of several explanation styles have been investigated [21,24]. Symeonidis et al. [24] combined content and influence style explanations. Visualization of complex hybrid explanations were also investigated in [16].

Whereas there is vast amount of research on explaining recommendations, most rely on the four types of information shown in Table 1: neighbor, influence, demographic, and content. Besides these four, context significantly influences user decision-making [9]. Contexts have also been used in context-aware recommendations [1,2]. Moreover, Zheng [29] and Papadimitriou [21] alluded to the possibility of using contexts for explanation. 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.

Explanation style	Displayed information
Neighbor	Ratings or the fact of purchases of similar users
Influence	Items related to recommended ones from users' past consumption
Demographic	Gender, age, etc. of users
Content	Keywords, tags, pros and cons of recommended items
Context (ours)	Context when users would consume recommended items

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

CONTEXT STYLE EXPLANATION

Generation of context style explanations is composed of two steps: (1) selection of context-item pairs for users, (2) suggestion of the context of a context-item pair as the item's explanation.

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 this case, the recommender needs to select appropriate pairs of contexts and items for the users. This requires 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. Moreover, the recommended restaurant should match the suggested context. If the context of eating with children is suggested in an explanation, the recommended restaurant should be suitable for that situation. Additionally, the recommended context should be one anticipated by the user. 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 properties can be learned via latent representation of pairwise interactions between features [5,13,23]. Among them, we use field-aware factorization machines (FFMs) [13] for their efficiency and performance. FFM splits features to "fields," and incorporates interaction effects among features of different fields. FFM is formulated as,

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

where $\boldsymbol{w}_j^f$ is a latent vector of feature j that interacts with field f; and x_j is the value of feature j. To model interaction among users, items, and contexts, we prepare a user field, an item field, and a context field. If each field takes only one feature, and if the values of features are binary, then Equation (1) is expressed as follows,

$$\boldsymbol{w}_u^{\mathrm{Item}}\cdot\boldsymbol{w}_i^{\mathrm{User}}+\boldsymbol{w}_i^{\mathrm{Context}}\cdot\boldsymbol{w}_c^{\mathrm{Item}}+\boldsymbol{w}_u^{\mathrm{Context}}\cdot\boldsymbol{w}_c^{\mathrm{User}}. \quad (2)$$

Each term in Equation (2) represents a) a user-item match ($\boldsymbol{w}_u^{\mathrm{Item}}\cdot\boldsymbol{w}_i^{\mathrm{User}}$), b) an item-context match ($\boldsymbol{w}_i^{\mathrm{Context}}\cdot\boldsymbol{w}_c^{\mathrm{Item}}$), and c) a user-context match ($\boldsymbol{w}_u^{\mathrm{Context}}\cdot\boldsymbol{w}_c^{\mathrm{User}}$). Context-item pairs are selected by the score of Equation (2).

Suggestion of 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, item i is presented to a user as a recommendation and context c is used for an explanation. Explanation is generated using human-crafted templates, for example, "*item i is recommended to you because it is suitable for context c.*"

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.

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 usage scene of the visit (i.e., contexts). We asked for original contexts of users' visits instead of asking for evaluation under a provided context, because users behave differently under supposed contexts and real contexts [4,20]. Usage scenes were selected from 15 options[1], for example, "with close friends (only eating)," "with colleagues (with drink)," "in solitude," "take-away," and "business entertaining." We asked each crowdworker to input 20 restaurants at maximum. We recruited crowdworkers who lived in certain urban areas in order to confine the areas of visit logs.

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. The genders and approximate ages of the crowdworkers were provided from the crowdsourcing platform. There are 2,730 unique URLs in the remaining visit logs. The average number of visits per restaurant is 1.056 and the sparsity is 99.32%. We crawled the URLs and collected the restaurants' content information, including genres and nearest stations[2]. Note that each restaurant is given many genres (2.3 genres on average). There are 210 unique genres and 473 unique stations.

Training Recommender and Preparing Explanation

We trained the context-item selector using the collected dataset. We use the libffm[3] library for FFM. The features of the user field are user ID, gender, and age. The features of the item field are genre and nearest station. The use of the demographic features of users and the content features of items alleviate the issue of data sparsity. The features of the context field include context ID, which is assigned to 15 usage scenes. Features of the context field include context ID, which is assigned to 15 usage scenes.

Training of the recommender proceeds as follows. First, the dataset is randomly split into 80% training and 20% validation data. Hyper-parameters of FFM are then optimized to maximize the AUC of the validation data. Finally, we train the model using the entire dataset to select context-item pairs.

After training the model, we select seven context-item pairs for each user. Restaurants visited in the past are removed from the list, whereas contexts experienced in the past are not omitted. The same restaurant can be recommended only once per user.

We prepared seven styles of explanations as described in Table 2. Baseline explanation does not include any specific information of demographics, contents, and contexts. This explanation is the same for all users and all recommended items. For the context style explanation, the context of context-item pair is directly assigned. For the demographic style explanation, user age and gender are assigned, because the recommender incorporates gender and age as features, and recommended items should relate to them. For the content style explanation, we select a genre common among the recommended restaurants and those the user visited in the past. Hybrid explanation styles are generated via a combination of the procedures described above.

[1] We consulted descriptions of usage scenes in several restaurant information sites to prepare the options.

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

[3] https://github.com/guestwalk/libffm

Style	Sample
Baseline	Recommend based on your visit logs
Demographic	Recommend for "*women in 30s*"
Content	Recommend for who often visit "*Italian restaurant*"
Context	Recommend for use "*with husband/wife*"
Demographic + Context	Recommend for use "*in business entertaining*" of "*men in 50s*"
Content + Context	Recommend for use "*in solitude*" for who often visit "*noodle*"
Demographic + Content + Context	Recommend for use "*with close friends (with drink)*" of "*women in 20s*" who often visit "*cafe*"

Table 2. Samples of seven explanation styles.

Evaluating Explanation Styles

We recruited the 155 crowdworkers who had appropriately submitted restaurant visit logs, and 85 participated in the user study. We presented seven restaurant recommendations with seven different explanation styles to each user. Each recommendation was generated by the same FFM model. We asked the participants not to consult any other information (e.g., restaurant reviews).

The orders of explanation styles are randomly shuffled among users in order to cancel any biases related to the display order. The orders of context-item pairs are also shuffled, assuring recommendation quality does not correlate to the presentation 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 are also free entry fields to express any other comments.

RESULTS AND DISCUSSION

Among restaurants recommended to participants, 21% were visited in the past and 20% were known in advance. This indicates that item recommendation was fairly accurate and that our recommender system works fine.

Responses to the four questions are shown in Figure 1. Responses ranged from strongly disagree (-3) to strongly agree (+3). The average response of the context style explanation is higher than that of 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 of the context style explanation also tends to be higher than that of content style, though not statistically significant. In terms of the hybrids, the combination of demographic and context styles outperforms demographic-only (p < 0.01 for all questions), and the combination of content and context styles outperforms content-only (p = 0.061 for P2, and p < 0.01 for others).

Sixty-four participants input at least one comment, and 293 comments were obtained in total. User comments indicate two reasons of persuasiveness: 1) relevance of proposed context to users: "*I think I want to go because this situation is probable for me*," and 2) recognition of appropriate context for usage: "*I think I am going to use this when I organize a drinking party*." Users also mentioned usefulness of context for decision making. "*I'm afraid of making a wrong choice for girls' night out, so this explanation is useful*." These findings from qualitative analysis reinforce the importance of context for explanation.

Baseline explanations tend to perform better than the context style, and the hybrid of content and context styles tends to perform better than baseline (both are without statistical significances). Relatively high appraisals of baseline explanation might come from familiarity of the explanation style. Some users commented: "*There is a comfort in this type of explanation*," and "*This writing style fits into me best*." The crowdsourcing platform we used provides task recommendation for users with an explanation of this style: "*Recommendation is based on your past task*." Another reason might be the occasional mismatch of presented context. Evaluation tends to be affected more by negative experiences (i.e., mismatches) than by positive experiences (i.e., good matches). Similar observations are reported in the experiment of personalizing engaging messages [15]. We plan to improve the accuracy of context-item pair selections in future works.

The trio of demographics, contents, and contexts do not produce significant improvement over the duo of contents and contexts. Users may have felt too much complexity. Investigating adequate amounts of information for an explanation would be an interesting challenge.

This work is conducted in a restaurant recommendation domain. We believe that context style explanations can be applied to various domains, though relevant contexts should be unique to those different domains. Those other domains need further study. We compare and hybridize context style explanations with demographic and content styles. Future work will experiment with other conventional explanation styles (e.g., neighbor and influence styles). We evaluated persuasiveness and usefulness to verify our hypothesis of context-style impacts. Evaluation of other factors, such as user trust and decision efficiency, remains for future work.

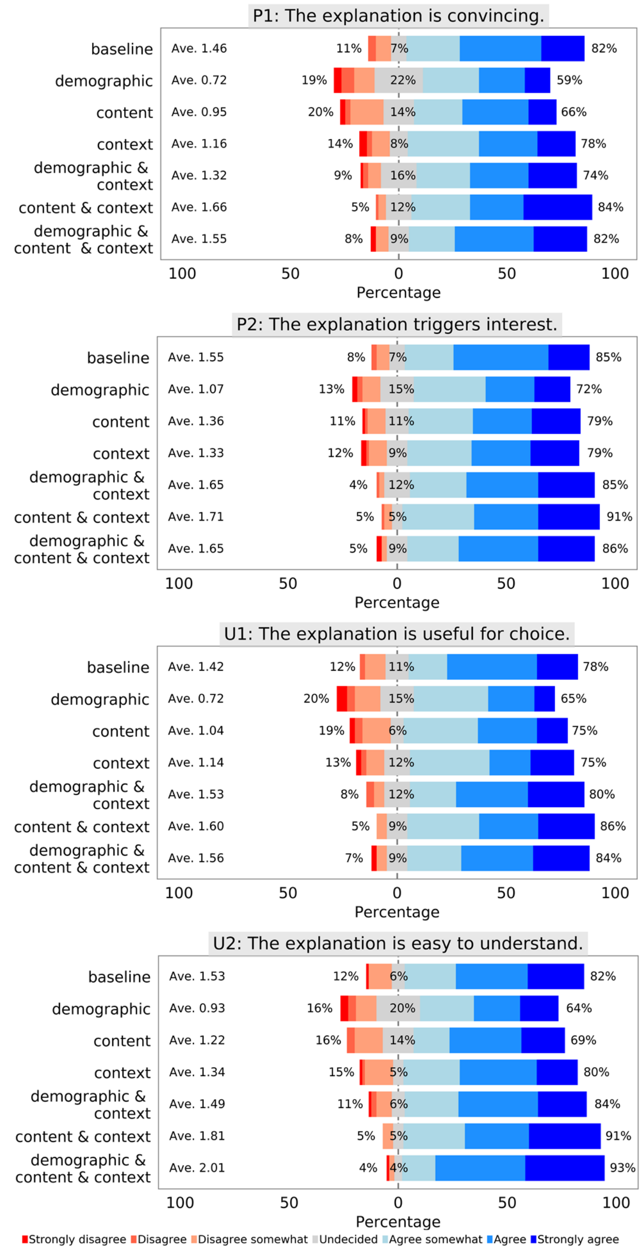

Figure 1. Responses to four questions for each explanation style. Percentages of positive and negative responses are shown in the right and left blanks, respectively. Neutral percentage is shown in the middle. Values of average response are also displayed in the left blanks.

CONCLUSION

In this paper, we proposed the context style explanation for recommenders. We conducted a crowdsourcing-based user study to measure persuasiveness and usefulness. Context style explanations were better than demographic styles. Context styles also tend to perform better than content styles, although not statistically significantly. We further confirmed that hybrids of context style and other explanation styles improve persuasiveness and usefulness. Findings from user comments support these impacts.

REFERENCES

1. Gediminas Adomavicius, Ramesh Sankaranarayanan, Shahana Sen, and Alexander Tuzhilin. 2005. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Trans. Inf. Syst.* 23, 1: 103-145. http://dx.doi.org/10.1145/1055709.1055714
2. Gediminas Adomavicius and Alexander Tuzhilin. 2015. Context-Aware Recommender Systems. In *Recommender Systems Handbook* (2nd. ed.), Springer, US, 191–226.
3. Liliana Ardissono, Anna Goy, Giovanna Petrone, Marino Segnan, and Pietro Torasso. 2003. Intrigue: personalized recommendation of tourist attractions for desktop and handheld devices. *Applied Artificial Intelligence* 17: 687–714.
4. Hideki Asoh, Yoichi Motomura, and Chihiro Ono. 2010. An analysis of differences between preferences in real and supposed contexts. In *Proceedings of the 2nd Workshop on Context Aware Recommender Systems* (CARS '10).
5. Linas Baltrunas, Bernd Ludwig, and Francesco Ricci. 2011. Matrix factorization techniques for context aware recommendation. In *Proceedings of the fifth ACM conference on Recommender systems* (RecSys '11), 301-304. http://dx.doi.org/10.1145/2043932.2043988
6. Mustafa Bilgic and Raymond J Mooney. 2005. Explaining recommendations: satisfaction vs. promotion. In *Proceedings of the Workshop on the Next Stage of Recommender Systems Research*.
7. Shuo Chang, F. Maxwell Harper, and Loren Gilbert Terveen. 2016. Crowd-based personalized natural language explanations for recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (RecSys '16), 175–182. https://doi.org/10.1145/2959100.2959153
8. Li Chen and Feng Wang. 2017. Explaining recommendations based on feature sentiments in product reviews. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (IUI '17), 17–28. https://doi.org/10.1145/3025171.3025173
9. Wei Chen, Christopher Hoyle, and Henk J. Wassenaar. 2013. A Choice Modeling Approach for Usage Context-Based Design. In *Decision-Based Design*. Springer, London, UK, 255-285.
10. Wei Chen, Wynne Hsu, and Mong Li Lee. 2013. Tagcloud-based explanation with feedback for recommender systems. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (SIGIR '13), 945-948. https://doi.org/10.1145/2484028.2484108
11. Dan Cosley, Shyong K. Lam, Istvan Albert, Joseph A. Konstan, and John Riedl. 2003. Is seeing believing?: how recommender system interfaces affect users' opinions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03), 585-592. http://dx.doi.org/10.1145/642611.642713
12. Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (CSCW '00), 241-250. http://dx.doi.org/10.1145/358916.358995
13. Yuchin Juan, Yong Zhuang, Wei-Sheng Chin, and Chih-Jen Lin. 2016. Field-aware factorization machines for CTR prediction. In *Proceedings of the 10th ACM Conference on Recommender Systems* (RecSys '16), 43-50. https://doi.org/10.1145/2959100.2959134
14. Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. 2010. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems* (RecSys '10), 79-86. https://doi.org/10.1145/1864708.1864727
15. Rafal Kocielnik and Gary Hsieh. 2017. Send me a different message: utilizing cognitive space to create engaging message triggers. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (CSCW '17), 2193-2207. https://doi.org/10.1145/2998181.2998324
16. Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2017. User preferences for hybrid explanations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (RecSys '17), 84–88. https://doi.org/10.1145/3109859.3109915
17. Dong Li, Huang Shaohan, Wei Furu, Lapata Maria, Zhou Ming, and Xu Ke. 2017. Learning to generate product reviews from attributes. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* (ECACL '17), 623-632.
18. Khalil Ibrahim Muhammad, Aonghus Lawlor, and Barry Smyth. 2016. A live-user study of opinionated explanations for recommender systems. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (IUI '16), 256–260. https://doi.org/10.1145/2856767.2856813
19. Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. 2016. ExpLOD: a framework for explaining recommendations based on the linked open data cloud. In *Proceedings of the 10th ACM Conference on*

Recommender Systems (RecSys '16), 151–154. https://doi.org/10.1145/2959100.2959173

20. Chihiro Ono, Yasuhiro Takishima, Yoichi Motomura, and Hideki Asoh. 2009. Context-aware preference model based on a study of difference between real and supposed situation data. In *Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization* (UMAP '09), 102–113. https://doi.org/10.1007/978-3-642-02247-0_12
21. Alexis Papadimitriou, Panagiotis Symeonidis, and Yannis Manolopoulos. 2012. A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data Mining and Knowledge Discovery* 24, 3: 555–583.
22. Pearl Pu and Li Chen. 2006. Trust building with explanation interfaces. In *Proceedings of the 11th international conference on Intelligent user interfaces* (IUI '06), 93-100. https://doi.org/10.1145/1111449.1111475
23. Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2011. Fast context-aware recommendations with factorization machines. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (SIGIR '11), 635-644. http://dx.doi.org/10.1145/2009916.2010002
24. Panagiotis Symeonidis, Alexandros Nanopoulos, and Yannis Manolopoulos. 2009. MoviExplain: a recommender system with explanations. In *Proceedings of the third ACM conference on Recommender systems* (RecSys '09), 317-320. https://doi.org/10.1145/1639714.1639777
25. Nava Tintarev and Judith Masthoff. 2007. A survey of explanations in recommender systems. In *Proceedings of 2007 IEEE 23rd International Conference on Data Engineering Workshop*, 801–810. https://doi.org/10.1109/ICDEW.2007.4401070
26. Tintarev, N. and Masthoff, J. 2015. Explaining Recommendations: Design and Evaluation. In *Recommender Systems Handbook* (2nd. ed.), Springer, US, 353–382.
27. Jesse Vig, Shilad Sen, and John Riedl. 2009. Tagsplanations: explaining recommendations using tags. In *Proceedingsc of the 13th international conference on Intelligent user interfaces* (IUI '09), 47-56. https://doi.org/10.1145/1502650.1502661
28. Lipton, Zachary C., Sharad Vikram, and Julian McAuley. 2015. Generative concatenative nets jointly learn to write and classify reviews. arXiv preprint. arXiv:1511.03683
29. Yong Zheng, Bamshad Mobasher, and Robin Burke. 2016. User-oriented context suggestion. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization* (UMAP '16), 249–258. https://doi.org/10.1145/2930238.2930252