How Does Netflix "Understand" Me?: Exploring End-user Needs to Design Human-centered Explanations

Yoonseo Choi
School of Computing, KAIST
yoonseo.choi@kaist.ac.kr

Eun Jeong Kang
School of Computing, KAIST
eunxkang@kaist.ac.kr

Juho Kim
School of Computing, KAIST
juhokim@kaist.ac.kr

Abstract

In this work, we explore the end-user needs of the explanations when using recommender systems and potential actions that the users could do with explanations. We conducted an online survey (N=14) and a think-aloud study (N=3) to investigate user needs for explanations in entertainment-domain recommender systems. The study revealed that users want to get 'self-referenced' explanations on how the system understands them. Furthermore, these insights confirmed that the users want to directly manipulate the algorithms either immediately or long-term, even they do not have technical understandings or abilities.

1 Introduction & Background

AI-based recommender systems have become increasingly popular in recent years and are utilized in various areas, including movies, music, news, search queries, social tags in general [1]. These recommender systems use user data (e.g., users' history, interaction log, behavior, etc) as inputs to train and improve the model to get personalized recommendation results. Comparatively, explanations about how these systems produce recommendations are rarely or vaguely provided to users. For example, Spotify [1] automatically generates a playlist based on the users' activities. However, it does not explain how the playlist was created. Netflix [2] shows the similarity between recommended items as a percentage or changes the arrangement of items in a row based on an algorithm while the results do not contain explanations on why the recommendations and adjustments happened. Since current recommender systems do not explicitly explain recommendation results to lay users, it is challenging for these users who are not experts to understand the algorithmic process happening in systems. To be specific, recipients of the algorithm's output have difficulty understanding how or why the inputs lead to a particular outcome [2].

An explanation is actually an answer to a question [3]. In human-AI interaction, a user would want to ask questions to understand incomprehensible results given by the AI. When people could not ask, sometimes they tend to imagine algorithm's operation behind the recommendation and use it in their own way of interaction in the system, based on folk theory [4] — a person's informal, intuitive understanding about a system that guides one's thoughts, beliefs, and actions with that system [5]. Even if explanations are provided, it is unclear whether the explanations are fully understandable for users. To represent user needs for explainability, Liao et al. developed an algorithm-informed

[1]https://www.spotify.com/
[2]https://www.netflix.com/

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

explainable AI (XAI) question bank filled with prototypical questions driven by interviewing UX and design practitioners to uncover user needs [6]. Yet, they did not evaluate whether existing XAI methods answered the identified questions by end-users who are usually non-experts. Ribera et al. pointed out that explanations should be provided differently to each targeted stakeholder group such as developers, AI researchers, domain experts, and lay users [7]. Therefore, more research should cover what kind of explanations non-expert users need and how these can be supported in daily recommender systems.

In this workshop, to focus on designing explanations from the lay user's perspective, we explore the user needs regarding explanations and what kind of potential actions the users could do to understand and utilize the system. We used a mixed-method approach: survey and think-aloud study. The domain of recommender system we focused on is 'entertainment'. In this domains, users make lightweight and small decisions by exploring and navigating recommendations. Based on our observations, we suggest a potential form of explanation that represents how the system understands the users. Furthermore, we investigated the user needs for controllability based on the provided explanation.

2 Survey

To understand what kind of explanations are needed by lay users when interacting with recommender systems, we conducted an online survey with 14 respondents (6 undergraduates, 8 graduate students) at our institution who had prior experiences with recommender systems for entertainment content (e.g., movie, music, visual).

Inspired by Liao et al.'s work [6], we tried to investigate what kind of questions the end-user could ask about recommendation results. The survey comprised open-ended questions about the recommender systems and focused on collecting users' questions and feedback about the systems. We provided four examples of well-known entertainment platforms: Netflix, Spotify, Pinterest [3], and Medium [4]. All of these platforms (1) leverage user data to recommend content and (2) are related to entertainment domain. Respondents could choose one platform among examples and proceed with the survey. To focus on situations in which the end-user required explanations while using the systems, we designed survey questions to explore three points: (1) unclear points on recommendation given, (2) satisfaction about recommended results, and (3) what kind of explanations do users want and need. More than half of the respondents (N=8) chose Netflix for the survey. Three respondents chose Spotify, two chose Medium, and two chose Pinterest. We have a respondent who submitted responses for both Netflix and Pinterest. Using content analysis [8], we found three notable insights.

2.1 Findings

Users want to know how the system works in more detail. Respondents had many questions on the unclear points about the recommender systems and wanted to learn about them. Their questions could be mapped to the conceptual process of general machine learning (Input - Model - Output) [9]. The respondents wanted to know about **the input data** used for training ("Why don't you consider my whole listening catalog when providing recommendations?" - R2), **model/algorithm** to predict the potential need ("How much weighting each song had, [and] which songs were considered [to recommend this song?]." - R2) and **output** ("Is customization based on what I saw, or does everyone get the same result?" - R1). Beyond the specific recommendation process, respondents wanted to know more about the **high-level mechanism** of recommender systems ("The 'Why and How' are not explained" - R8) and **user interface** ("What are the criteria for sorting the shows?" - R10, "Does each row mean something?" - R5). In summary, we found that lay users ask specific questions related to the recommendation process and the UI representation of recommendations.

Granular explanations are requested by users to better understand the system. Even if the platforms provided explanations, our respondents wanted to get further details to understand the brief explanations. In some platforms, such as Netflix, items that are similar to others are displayed with similarity percentages. However, some respondents noted that there was no explanation for why the items were similar. For example, ("Can you show the percentage of matches for each criterion respectively?" - R13) ("How were the matching criteria defined?" - R10). The granular explanations

[3]https://www.pinterest.co.kr
[4]https://medium.com

they needed were different from person to person. Some respondents wondered about the meaning of vague terms used on the interface. For example, R4 and R10 asked, “what is the meaning of ‘content you’d like’?”. Several participants wanted to know more deeply about terms such as ”How were that the content you’d like calculated?”(R6, R13). In addition, most of the respondents answered that they had unclear points while using recommender system, however, a few respondents were satisfied with the current explanations.

End-users wonder how the system knows about them. Respondents are interested in ‘defined identity’ from the systems’ perspective. Since the platforms recommend content by utilizing personal history and user’s input, respondents mentioned that they want to know about how the system perceives them. For instance, respondents expressed curiosity about their identified profile (“How does the system profile me?” - R8). Also, they assumed that there are some keywords that explain each user’s identity and profile; thus they were curious about what kind of keywords are used to provide similar recommendations to themselves (“[Which] keyword [of] this recommended image [is] similar to my past record?” - R11). Moreover, R5 expressed desire to edit the identified profile if it helps to get better recommendations (“What are the movie genres or other components that are identified by the system [as my favorite]? Maybe I have to change some, because maybe the system misunderstood me...” - R5), which could be linked to need for controllability.

3 Think-aloud Study

Based on the findings from the survey, we developed research questions aimed at understanding how users’ interact with real recommendation interfaces. We wanted to answer two questions with a think-aloud study: (1) When do users want to get an explanation about how the system understands them? and (2) With that explanation, what do they want to do more with the system? Again, we limited the participants to non-experts who do not have high technical knowledge about recommender systems. To answer those questions, we recruited three participants who identified themselves as heavy users of recommender platforms. We asked participants to choose a recommendation-based service or app that they most frequently use in their daily life; P1 chose Spotify, and P2 and P3 chose Netflix.

To understand participants’ context when using the recommender system, we first ran a short interview about the frequency of recommender system use, technical knowledge level about recommendation systems, and perceptions about recommendation and explanations. Then we provided a scenario to begin the experiment within a specific context, which was inspired by the interview guide in Ngo et al.’s work [10]: *“Suppose you have sudden free time in the evening. Please freely explore [the app] as you would generally use. You will choose a [content] that you didn’t know previously. After that, please imagine that you realized that [the chosen content (media)] was not as good as expected. Let’s try to improve your [app] experience for the next time.”* Participants were instructed to freely use the app and think-aloud with this scenario. They were asked to look back at the process they had proceeded with and whether they needed explanations and controls for the system and think aloud about their exploration behaviors and questions in their minds. Sometimes, we asked provoking questions to lead the sessions. All participants were heavy users of more than one of recommender systems, but each person had different usage scenarios. P1, who used Spotify, barely uses the interface. They usually control music with button features in their headphones so that they were unfamiliar with the Spotify interface during the study even though they have been using the app for over six months. P2 and P3 chose Netflix, which is a movie and TV show streaming platform, so they were more familiar with the interface itself as they had to interact with it to consume the content. To analyze the qualitative data from the think-aloud and interview sessions, we first transcribed all the audio recordings and two researchers conducted an open coding and thematic analysis [11] on it.

3.1 Findings

Everything is hidden from the users. The thematic analysis revealed that there are specific reasons for why participants are not satisfied with the systems’ recommendations. For example, P3, a heavy user of Netflix, was unhappy because they had already watched all the recommended content, which is not frequently updated. On the other hand, P2 watches movies heavily during subscription period but does not always subscribe to Netflix. Thus, they assumed that sometimes they do not get enough personalized recommendations due to a lack of user data. Participants felt that the algorithm behind the recommended results they got is too complex. They assumed that there were so many features that could be related to their own recommendation, e.g., genre, actors, storyline, etc. Furthermore, they could not guess how these features actually influenced personal recommendation algorithm and results. Even if there was a gap between the original purpose of the recommendation algorithm and the user expectation toward it, participants said that they could not understand what is happening behind them.

"How does the recommender system understand me?" Interestingly, similar to the survey response, one of participant (P2) was curious about how the recommender system understands them. Similarly, prior work observed that users tend to identify the image of themselves based on the output of recommendation results [10]. P2 mentioned that "I want to know what Netflix thinks of me." They believed that some unsatisfactory recommendation might be because the system understood them incorrectly. If there was no content that they liked, then they might think, "What does Netflix think of me to make such a bad recommendation?". Thus, they want to get information on how Netflix understands the user.

"What if I could control the given explanation?" Participants expressed the need for correcting the system's understanding about themselves. Such need for controllability is derived from receiving an unsatisfactory recommendation. "I was sometimes unsatisfied with what Netflix recommended, then I thought 'What do you (Netflix) think of me to recommend this?"' P2 stated that they would like to amend the features directly in the reasoning process if Netflix explain how it provide the recommendation. They pointed that the current system (Netflix) seemed to provide explanation focusing on genres they would like, but P2 think the explanations should provide a process of how they recommend the movies, rather than what the movie is about. Also, it limits users to only provide binary feedback such as thumbs up and down, which is hard to be seen as they control their recommendations based on the given explanation. P3 also showed a desire for control over the system, but instead, they wanted to control the input side of the algorithm, which differs from P2. P3 wanted to check the movies they've seen and remove them directly from their history data that the algorithm considers. In YouTube or Facebook, a similar feature exists but does not explain the recommendations or affect recommendations at a granular level.

4 Design Implications

In this workshop, we want to discuss design implications derived from our study, which contributes to the literature on human-centered explanation design.

4.1 "Tell me about myself"

Survey respondents and study participants want to get explanations on how the recommendations are related to them, and they tend to centralize themselves, regarding themselves as the center, when interacting with the recommender system. In literature, when users describe how Netflix works, some focus on how the system uses their information and provides the recommendation results, while considering themselves as central components [10]. Furthermore, users tend to identify an image of themselves based on the recommendation results [12]. In a recommender system, it may be natural for users to think of themselves as the center/standard because most of these systems construct models based on individual interaction logs, browsing history, and ratings. From psychology, people tend to remember information better when the information has been linked to themselves by the "Self-reference effect" [13]. Likewise, linking and reflecting oneself into the process of the algorithm could help the users understand the algorithm quickly and utilize the system more actively. Rather than providing vague and algorithm-centered explanations (e.g., 'N% Match', 'Daily Mixes'), presenting explanations from the perspective of each user while considering their circumstances might enhance the user satisfaction, trust, and understanding so that they could gain more agency and actively explore the system.

4.2 "One step closer to controllability"

Interestingly, users not only want to get self-referenced explanations but also desire to control the recommendations so that they could change the model both immediately and long-term. Even though we did not ask specific questions regarding controllability in the survey, some survey respondents mentioned that they wanted to control and manipulate either the recommendation algorithm or the recommended items. R2 wondered, "Why doesn't [the system] consider my entire listening catalogue [as input] when providing recommendations?", as they felt Spotify does not equally consider all songs they have played. On the other hand, R11 wanted to make a change on the recommended outputs by categorizing them with some keywords. R5 wanted to control 'preference elicitation' made by the system because they worried that the system might misunderstand them. R5 mentioned that "What are the movie genres or other components that are identified by the system [as my favorite]? Maybe I have to change some, because maybe the system misunderstood me...". In the findings that participants require controllability as well as explanations while using a recommender system, we confirmed Smith-Renner et al.'s study [14] that explainability can be supported by user's feedback through the study. Especially in recommender systems that use user's data and preferences, their desire to give feedback is for benefiting themselves. For example, when systems provide how they analyze user's data logged in the system, users could choose to discard some of their information in the process. Such controllability would be able to give users a higher sense of agency in using the system.

5 Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-01347, Video Interaction Technologies Using Object-Oriented Video Modeling).

References

[1] Prem Melville and Vikas Sindhwani. Recommender systems. *Encyclopedia of machine learning*, 1:829–838, 2010.

[2] Jenna Burrell. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1):2053951715622512, 2016.

[3] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019.

[4] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. First i "like" it, then i hide it: Folk theories of social feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 2371–2382, New York, NY, USA, 2016. Association for Computing Machinery.

[5] Megan French and Jeff Hancock. What's the folk theory? reasoning about cyber-social systems. *Reasoning About Cyber-Social Systems (February 2, 2017)*, 2017.

[6] Q Vera Liao, Daniel Gruen, and Sarah Miller. Questioning the ai: informing design practices for explainable ai user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2020.

[7] Mireia Ribera and Agata Lapedriza. Can we do better explanations? a proposal of user-centered explainable ai. In *IUI Workshops*, volume 2327, page 38, 2019.

[8] Philipp Mayring. *Qualitative Content Analysis: Theoretical Background and Procedures*, pages 365–380. Springer Netherlands, Dordrecht, 2015.

[9] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. `http://www.deeplearningbook.org`.

[10] Thao Ngo, Johannes Kunkel, and Jürgen Ziegler. Exploring mental models for transparent and controllable recommender systems: A qualitative study. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '20, page 183–191, New York, NY, USA, 2020. Association for Computing Machinery.

[11] Virginia Braun, Victoria Clarke, Nikki Hayfield, and Gareth Terry. Thematic analysis. In Pranee Liamputtong, editor, *Handbook of Research Methods in Health Social Sciences*, pages 843–860. Springer Singapore, 2019.

[12] Nedim Karakayali, Burc Kostem, and Idil Galip. Recommendation systems as technologies of the self: Algorithmic control and the formation of music taste. *Theory, Culture & Society*, 35(2):3–24, 2018.

[13] Timothy B Rogers, Nicholas A Kuiper, and William S Kirker. Self-reference and the encoding of personal information. *Journal of personality and social psychology*, 35(9):677, 1977.

[14] Alison Smith-Renner, Ron Fan, Melissa Birchfield, Tongshuang Wu, Jordan Boyd-Graber, Daniel S Weld, and Leah Findlater. No explainability without accountability: An empirical study of explanations and feedback in interactive ml. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.