A Literature Overview on Explanations in Recommender System and Possible Criteria for Designing Brief Explanation

Yifei Zhan, Sarah Theres Völkel

Abstract— This paper provides a literature overview of explanations in recommender systems. It introduced two standards/criteria on how to design and evaluate explanations in recommender systems. In general, the type of explanations depends on the goal of the system. For example, in some scenarios, like in a driving context, pages of explanations might not be suitable. In this case, drivers may prefer brief explanations. In order to find possible criteria for it, we extracted some principles from two recent studies and what we have found is that efficiency, persuasiveness and why question-type-based explanations might be suitable for brief explanations. However, more studies need to be done in order to explore other possible criteria for it.

Index Terms—Explanation, Recommender system, Personalisation

1 Introduction

Personalization is no more a novelty. The first emergence of this phenomenon dated back to antiquity when experienced merchants provided different customers with different products or services [1]. However, the real interest of personalization did not arise until the late 1990s with the advancement of the Web technologies [11]. From then on, more studies have been done in this field, and people try to explore the essence of personalization.

In the world of personalization, one thing we might often hear about is recommender systems (RS), which intelligently give users personalized suggestions. The recommendations can vary from algorithm to algorithm. The most common forms are the content-based recommendation, and the collaborative filtering recommendation. The first one, content-based recommendation, observes utility of previous items and then use this attribute to predict the utility of other items [9]. An example can be "User A has bought a novel. And based on this bought history, a magazine is more likely to be recommended to him than a pair of shoes, because novels and magazines are more in common than novels and shoes ". The second type, collaborative filtering, observes utility for user-item pairs which is then used to predict utility for unobserved user-item pairs [9]. One example can be "User A(1,2) has bought items 1 and 2, user B(1,2,3) has purchased items 1,2 and 3 and user C(2,4) has bought items 2 and 4. As a recommendation for user A, item 3 is more likely to be suggested to user A than item 4, since user A and user B are more similar than user A and user C".

A good recommendation algorithm is no doubt indispensable for a recommender system, and the way of explaining its complex algorithm to users plays also an important role. With the development of the Internet, the privacy issue becomes more and more important. The newly EU regulation even requires explanations for complex algorithms [14]. People are willing to know which information is "consumed" by the system and they may lose their trust in one system if it behaves differently as they expected. In this case, a proper explanation of a recommender system can help to inspire users' trust and loyalty, increase satisfaction, make it quicker and easier for users to find what they want [25]. In order to answer the question, "What makes a good explanation," we check through different studies

and try to summarize a set criteria in this paper. Besides, we particularly focus on brief explanations.

The rest of the paper is divided into four parts. The first section provides an overview of different studies on explanations in recommender systems. The second part focuses on two frameworks, which are often used by other researchers to evaluate and design explanations of recommender systems. The third section extracts some principles from the two frameworks, which can be used to design "brief explanations" in an automotive scenario. And in the last part, we summarize our contribution and propose several questions for future studies.

2 Literature Review

In this section, we try to give readers an overview of state of the art. We introduce in this part some relevant studies which focus on the recommendation and its explanations.

User experience research is increasingly attracting researchers' attention. Some researchers, like Li Chen [21], and Bart P.Knijnenburg [16], put emphasis on recommender system and try to figure out what constitutes an effective and satisfying recommender system. Other researchers focus on its explanations. They have proved that a good explanation can boost users' trust towards the system and help them in decision making process [25] [27] [20].

One possible way to evaluate explanations of a recommender system is to look at the soundness and completeness [18]. The soundness means the extent to which an explanation describes all of the underlying systems. Meanwhile, the completeness means how truthful each element in an explanation is, in the perspective of a user, in the underlying system. The author has found out that LSHC(which means low soundness and high completeness) provides participants with the best-perceived cost/benefit ratio of attending to the explanations.

Another possible method to evaluate the explanations is to categorize explanations in different types of questions. In Lim's work, Toolkit Support Intelligibility in Context-Aware Applications [19], they divided the explanation of a context-aware system into eight questions, which are Why, Why Not, How To, What, What If, Inputs, Outputs, and Certainty. In the findings, they emphasized on streamlining explanations while maintaining access to the rich explanation capabilities, which makes explanations usable and quickly consumable.

Finally, a significant breakthrough is made by Nava Tintarev and Judith Masthoff. They have proposed what they called "seven possible advantages of an explanation facility [25]." These seven advantages are often taken as a standard criteria in evaluating and designing explanations for a recommender system.

- Yifei Zhan is studying Media Informatics at the University of Munich, Germany, E-mail: yifei.zhan@campus.lmu.de

3 Explanation Factors

In this part, we take a look at explanation factors from two studies and try to find requirements for successful explanations. The first research suggests a standard way to evaluate explanations while the second study aims at designing explanations.

3.1 The first framework - Seven Explanatory criteria

Like what we have already mentioned in the second part, literature review, which the "seven explanatory criterion" [25](see table 1) from Nava Tintarev and Judith Masthoff is the most popular standard used to evaluate the explanation. It is comprehensive since it summarized all advantages which explanation can bring to the system.

Aim	Definition
Transparency	Explain how the system works
Scrutability	Allow users to tell the system it is wrong
Trust	Increase users' confidence in the system
Effectiveness	Help users make good decisions
Persuasiveness	Convince users to try
Efficiency	Help users make decisions faster
Satisfaction	Increase the ease of use or enjoyment

Table 1: Explanatory criteria and their definitions

Although they may have different names in different studies ("explanation attributes" [7], "seven possible aims for explanations" [26] or "quality factors" [13]). The basic idea is the same, that is to make a recommender system more understandable by users.

Transparency: Transparency in a recommender system is related to the capability of a system to expose the reasoning behind a recommendation to its users [15]. A recommender system without explanation works like a "black box", which may lose users' trust. Thus, transparency is considered as an important factor to build user's trust in a recommender system [24].

Scrutability: Scrutability means, in short, allow users to tell the system it is wrong. It can also be seen as a kind of User Control, which allows users to correct reasoning from system [10].

Trust: Trust is sometimes linked with transparency. Studies have already shown that transparency and the possibility of interaction with recommender systems increase user trust [12] [23].

Effectiveness: An effective explanation can help users to make a better decision. If an item suggested by the system is the one the user likes, such an explanation can be considered as effective.

Persuasiveness: Persuasiveness, sometimes referred as promotion, is strongly related to effectiveness and can be defined as the ability of an explanation type to convince the user to accept or disregard certain items [13].

Efficiency: An explanation is usually considered as efficient when it helps the user to decide more quickly or when it helps to reduce the cognitive effort required in the decision process [13]. The most common approach to measure it is to look at the interaction time between users and the recommender system.

Satisfaction: The user's overall satisfaction with a recommender system is assumed to be strongly related to the perceived quality of its recommendations and explanations [24].

What we need to bear in mind, is that it is hard to design an explanation which obeys all of these seven criteria. In most of the time, we need to have a trade-off. For instance, an explanation that offers good transparency may impede efficiency as the user may spend more time in reading the explanations [25]. In reality, the choice of criteria depends on the goal of the system.

3.2 The second framework Question-type-based Explanation

Another possible way to evaluate and design explanations in context-aware applications is to focus on different question types. Lim et al. have proposed eight question-type-based explanations [19].

These eight question types are:

1. Input: Explanations inform users which information sources are used to generate the recommendation.

2. Output: Explanations inform users which output/result is generated by the system.

3. What: System explains the current state to users (what it is now).

4. What If: Which results will be generated, if users alter his/her input values(change the behavior).

5. Why: Explanations inform users why the application derived its output value from the current (or previous) input values[19].

6. Why Not: Explanations inform users why another output/result is not chose.

7. How to: Explanations tell users what they can do so that they can get another output/result.

8. Certainty: Explanations tell users, how certain/uncertain the output/result is.

Although the original use case of eight questions is a context-aware application, we can find explanations from other recommender systems which also apply these principles. For example, an e-commercial website which tells you that your recommendations are based on your purchase record. This kind of explanation can be seen as a Why-type-question. The next section will show a more detailed example.

3.3 Example – Explanations in Google Advertisement Recommendation Service

In this section, we apply both of the frameworks introduced above with a real-world use case. We try to illustrate how these explanation principles work with a personalised online advertisement service.

Users see nowadays more and more frequently online advertisement when they browser on the Internet. For example, Google Search results page may show ads similar to a search keyword the user just typed. It is a kind of personalised service provided by Google since every user may get their customised results, in other words, see different advertisements [6].

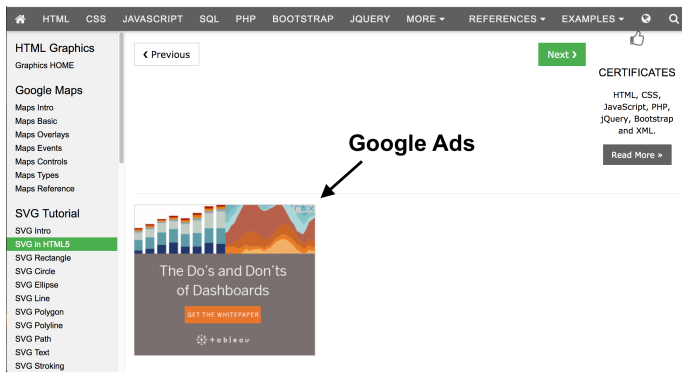Figure 1 shows an example of Google Ads.

Figure 1: An example of Google Advertisement[5]

Behind the advertisement, users can find pages of explanations provided by Google, which tells its users how the recommendation works. We can see here in figure 2. It is the first level of explanations, and Google Ads mainly explains two things here. "how does it work" and "which information is used". From the perspective of the first framework - Seven Explanatory Criteria, "how does it work" improves the transparency and increases users' trust, since it exposes the reasoning behind the recommendation to its customers. From the perspective of another framework - Question Type Based Explanation, "which information is used" tells users which information Google Ads uses in producing Advertisements. In this way, it can be seen as a Input type question because it explains users which information sources are used to generate the recommendation.

Figure 2: Explanations from Google Advertisement[2]

Figure 3 shows the second level of explanations in Google Ads. It again improves the transparency and increases users' trust in terms of the first framework . Meanwhile, from the perspective of Question-Type-Based explanation, information like, "You may see ads on Google Search...your location and the time of day", can be seen a Why and Why Not question types since they explain the reason why users see certain advertisements.

Figure 3: Explanations from Google Advertisement[3]

Figure 4: Explanations from Google Advertisement[4]

In the next level, Google even allows its users to control their Ads personalization. In figure 4, explanations are divided into users' authority("What you can do"), users' controls("Your controls"), users' ability to turn on/off personalization("Ads Personalization") and how to alter the personalization("Where your Ads settings apply"). From the perspective of Seven Explanatory Criteria again, such accessibility to control the personalisation can be seen as Scrutability since it allows users to avoid the Ads that they are not willing to see. And concerning Question-Type-Based Explanation, it provides users with What If and How to explanations because it teaches users how to alter their personalisation.

Users who care a lot about privacy issues may find explanations of Google Ads helpful because they are transparent and comprehensive. However, it can not match all cases. In some scenarios, like in a driving-context, where brief explanation may be preferred. The next section extracts part of two frameworks mentioned above and tries to figure out which explanation principles fit in with brief explanations.

4 Brief Explanation

In this section, we will dive into brief explanation and introduce how it works in a driving-context and which criteria or factors are helpful when we try to design brief explanation.

4.1 Brief Explanation based on Seven Explanatory Criterion

Roland Bader and his colleagues used part of Seven Explanatory Criteria and described an automobile assistant system that can reduce driver's distraction while he/she search for some information during driving [8]. The main function of the system is to provide a driver with proactively delivered recommendations. For example, this system can recommend some gas stations to a driver when it detects that the gas level is low. Such proactively delivered recommendations may not be accepted by the driver without a suitable explanation. Thus their goal is to enhance transparency of proactively delivered recommendations by means of explanations and tried to figure out the best way to explain.

In the study, they built the explanations based on efficiency and persuasiveness extracted from the first framework - Seven Explanatory Criteria [8]. The reason why they chose these two principles lies in his application environment. In a driving context, brief explanation may be preferred, and an efficient explanation (efficiency) guarantees short interaction time for decision making, while persuasiveness makes the recommendation convincible (In other words, it persuades the driver that the recommendation right now is relevant to your current state). Roland Bader's and his colleagues then set up a desktop prototype to test users' opinion on their explanations. With each recommendation (like "The system recommends xxx gas station to you"), explanations are distributed over several levels of detail. The lowest level (first phase), which contains maximal two arguments, like: "this gas station is on the route or the price of this gas station is low", provided automatically by the system. Then gradually more and more information can be accessed when the user requests. In the paper, Roland Bader takes the explanations in the lowest level as short explanations [8]. The first criterion persuasiveness is measured by asking the respondents whether they are satisfied with the explanation (brief explanations) provided in the first phase. And the second criterion efficiency was measured by looking at how often the respondents need to switch to deeper phases with more information.

The result is as expected. Almost all the experimental respondents were satisfied with the first level explanation, i.e., brief explanation. Besides, they found the information at this stage was adequate for decision-making in normal scenarios, which means that a brief and convincing explanation (persuasiveness) meets users' need. The information in the second level was selected only when particular details were needed, which means that users also prefer an efficient brief explanation (efficiency).

4.2 Brief Explanation based on Question-Type-Based-Explanation

The second framework, Question-Type-Based-Explanation is also a good starting point for designing brief explanations. Based on this framework, Martin Steinert and Larry John Leifer shows how brief explanations of a car's autonomous action affect the driver's attitude and safety performance [17].

In their study, two types of information (brief explanation) were provided to explain the auto-braking behavior (car's autonomous action) to the driver during the simulated driving. The first one, "how" information, which tells the driver the current state of the car (like: "the car is braking"), equals to "'what' question type of Question-Type-Based-Explanation in Lim's Toolkit [19]. The second one is "why" information, which tells the driver the reason why it behaves in certain ways (like: "Obstacle ahead.").

In the user test, they employed a two-by-two between-participants experimental design (see table 2). In table 2, they defined the "Why" type explanation ("Obstacle ahead") as external information while the "What" type explanation is defined as internal information which describes a car's activity ("The car is braking"). The autonomous action(automatically braking) took place during driving and each time, one of the following four different types of explanations would be presented:

1. No explanation.

2. "what" explanation ("The car is braking") short before the automatically braking.

3. "why" explanation ("Obstacle ahead") short after the automatically braking.

4. A combination of "what" and "why" explanation("The car is braking due to obstacle ahead").

	Why message	without Why message
what message	Both internal and external information referring to automation	Internal (car activity) information
Without what message	External (situation) information	No information (control condition)

Table 2: Structure of study[17]

Drivers' attitudes and safety performance were measured in order to learn how different types of information affect the driver.

The result shows something different than the original hypothesis. A combination of explanation ("why" + "how") was supposed to improve driving behaviors. However, the final result indicated that it affected driver attitude negatively. When people were told both how and why the car was about to act on their behalf, they felt anxious. One possible reason for it might be that the driver must process two types of information (the machine's status and the situational status) at the same time, which created large cognitive load. Although it was perceived negatively, the combination of explanation ("why" + "how") contributed to safer driving by minimizing off-lane excursions.

With "how" explanation alone, drivers performed the worst: they drifted out of their lane. It seems that drivers felt like

they took a passive role in driving if they only received "how" explanation. One possible reason is that human beings expect polite machine behavior [22]. Thus, only explaining the car's behavior, "Car is braking," without explaining the reason might be taken as "impolite."

And "Why" explanation alone created the least anxiety and highest trust. This kind of behavior from the system can be seen as transparent, which can, in turn, gain more trust from users. Compared with the combination of explanation ("why" + "how"), the amount of time that it takes to hear and cognitively process the why-only message is shorter, which allowed a quicker response/reaction from the driver. Besides, drivers might find the information valuable and credible when the content satisfies their cognitive need (explain the system's behavior). For this reason, the "why" explanation is more important to drivers and safer because drivers can anticipate upcoming events [17].

5 Discussion, Future Work and Conclusion

With the development of the automobile and mobile industry, more and more recommender systems need to work in a context-aware environment and compared with the desktop environment, users can not fully concentrate on the device because their environment requires more attention at the same time. In such cases, an efficient and brief explanation which informs the system's behaviors is especially important.

However, unlike typical explanations in a recommender system (explanations for desktop recommender systems), which has already some "standard" methods to design them. There is no "standard" approach to design and evaluate brief explanations.

In the first study, mentioned in section 4.1, Roland Bader extracted two criteria from the first framework - Seven Explanatory Criteria for brief explanations. He has successfully proved that these two principles (efficiency and persuasiveness) from the first framework are suitable for brief explanations. However, he did not mention other principles. Thus, it might be helpful if further studies can examine and evaluate other criteria (transparency, scrutability, effectiveness, trust and satisfaction). In the second study, Martin Steinert only evaluated two types of explanations (what and why) extracted from the second framework, Question-Based-Explanation. Therefore, some further studies can focus on other question-type-based explanations (why not, how to, what if, inputs, outputs and certainty).

In conclusion, in this paper, we firstly provided a literature overview of explanations in personalization and recommender systems. Secondly, we introduced two frameworks (Seven Explanatory Criteria and Question-Type-Based Explanations) which are used to design and evaluate explanations. Moreover, we illustrated how these two frameworks are applied in Google Advertisement recommender system. Thirdly, we introduced two studies which extracted some principles (efficiency, persuasiveness and why question-type-based explanations) from the two frameworks mentioned above, in order to find a proper approach to design brief explanations for a assistant recommender system in a car. And finally, we summarized our contributions and proposed some possible directions for future studies.

References

[1] Gediminas Adomavicius, Zan Huang, and Alexander Tuzhilin. "Personalization and recommender systems". In: State-of-the-Art Decision-Making Tools in the Information-Intensive Age. INFORMS, 2008, pp. 55–107.

[2] Google Advertisement. Explanations of personalization from Google Advertisement, the first level. A snapshot, [Online; accessed 1-Dec-2017]. 2017. URL: https://support.google.com/ads/answer/1634057?hl=en-GB&ref_topic=7049263.

[3] Google Advertisement. Explanations of personalization from Google Advertisement, the second level. A snapshot, [Online; accessed 2-Dec-2017]. 2017. URL: https://support.google.com/ads/answer/7029660.

[4] Google Advertisement. Explanations of personalization from Google Advertisement, the third level. A snapshot, [Online; accessed 2-Dec-2017]. 2017. URL: https://support.google.com/ads/answer/2662856.

[5] Google Advertisement. Google Advertisement in w3cschool website. a snapshot of google advertisement, [Online; accessed 1-Dec-2017]. 2017. URL: https://www.w3schools.com/graphics/svg_inhtml.asp.

[6] Google Advertisement. Introduction of Google Advertisement. [Online; accessed 1-Dec-2017]. 2017. URL: https://support.google.com/ads/answer/1634057?hl=en-GB&ref_topic=7049263.

[7] MOHAMMED Z Al-Taie. "Explanations in recommender systems: overview and research approaches". In: Proceedings of the 14th International Arab Conference on Information Technology, Khartoum, Sudan, ACIT. Vol. 13. 2013.

[8] Roland Bader12, Andreas Karitnig, Wolfgang Woerndl, and Gerhard Leitner. "Explanations in proactive recommender systems in automotive scenarios". In: First International Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems (DEMRA 2011). 2011, p. 11.

[9] Marko Balabanović and Yoav Shoham. "Fab: content-based, collaborative recommendation". In: Communications of the ACM 40.3 (1997), pp. 66–72.

[10] Marek Czarkowski and Judy Kay. "A scrutable adaptive hypertext". In: International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems. Springer. 2002, pp. 384–387.

[11] Peppers Don and Rogers Martha. Managing customer relationships: a strategic framework. 2011.

[12] Alexander Felfernig, Erich Teppan, and Bartosz Gula. "Knowledge-based recommender technologies for marketing and sales". In: International Journal of Pattern Recognition and Artificial Intelligence 21.02 (2007), pp. 333–354.

[13] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. "How should I explain? A comparison of different explanation types for recommender systems". In: International Journal of Human-Computer Studies 72.4 (2014), pp. 367–382.

[14] Bryce Goodman and Seth Flaxman. "European Union regulations on algorithmic decision-making and a" right to explanation"". In: arXiv preprint arXiv:1606.08813 (2016).

[15] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. "Explaining collaborative filtering recommendations". In: Proceedings of the 2000 ACM conference on Computer supported cooperative work. ACM. 2000, pp. 241–250.

[16] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. "Explaining the user experience of recommender systems". In: User Modeling and User-Adapted Interaction 22.4-5 (2012), pp. 441–504.

[17] Jeamin Koo, Jungsuk Kwac, Wendy Ju, Martin Steinert, Larry Leifer, and Clifford Nass. "Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance". In: International Journal on Interactive Design and Manufacturing (IJIDeM) 9.4 (2015), pp. 269–275.

[18] Todd Kulesza, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng-Keen Wong. "Too much, too little, or just right? Ways explanations impact end users' mental models". In: Visual Languages and Human-Centric Computing (VL/HCC), 2013 IEEE Symposium on. IEEE. 2013, pp. 3–10.

[19] Brian Y Lim and Anind K Dey. "Toolkit to support intelligibility in context-aware applications". In: Proceedings of the 12th ACM international conference on Ubiquitous computing. ACM. 2010, pp. 13–22.

[20] Pearl Pu and Li Chen. "Trust-inspiring explanation interfaces for recommender systems". In: Knowledge-Based Systems 20.6 (2007), pp. 542–556.

[21] Pearl Pu, Li Chen, and Rong Hu. "A user-centric evaluation framework for recommender systems". In: Proceedings of the fifth ACM conference on Recommender systems. ACM. 2011, pp. 157–164.

[22] Byron Reeves and Clifford Nass. How people treat computers, television, and new media like real people and places. CSLI Publications and Cambridge university press, 1996.

[23] Rashmi Sinha and Kirsten Swearingen. "The role of transparency in recommender systems". In: CHI'02 extended abstracts on Human factors in computing systems. ACM. 2002, pp. 830–831.

[24] Kirsten Swearingen and Rashmi Sinha. "Interaction design for recommender systems". In: Designing Interactive Systems. Vol. 6. 12. 2002, pp. 312–334.

[25] Nava Tintarev and Judith Masthoff. "A survey of explanations in recommender systems". In: Data Engineering Workshop, 2007 IEEE 23rd International Conference on. IEEE. 2007, pp. 801–810.

[26] Nava Tintarev and Judith Masthoff. "Evaluating the effectiveness of explanations for recommender systems". In: User Modeling and User-Adapted Interaction 22.4 (2012), pp. 399–439.

[27] Jeroen Van Barneveld and Mark Van Setten. "Designing usable interfaces for TV recommender systems". In: Personalized Digital Television: Targeting Programs to Individual Viewers, Kluwer Academic Publishers, Dordrecht, The Netherlands (2004), pp. 259–285.