A Survey of Techniques for the Evaluation of Explanations in Recommender Systems *

Marta Caro-Martinez, Guillermo Jimenez-Diaz, and Juan A. Recio-Garcia

Department of Software Engineering and Artificial Intelligence
Universidad Complutense de Madrid, Spain
email: martcaro@ucm.es, gjimenez@ucm.es, jareciog@ucm.es

Abstract. Nowadays, explanations are necessary to let users trust the outcomes of recommender systems. However, the evaluation of such explanations is a difficult task. It usually requires ad-hoc solutions focused on the desired effect of the explanation in the target user. Moreover, there are several heterogeneous methodologies in the literature, either online or offline, that could be useful for the evaluation of novel recommender systems. This paper presents a review and classification of 28 evaluations of explanations in recommender systems. The goal is to serve as a reference for researchers facing the task of evaluating explanations in recommender systems.

1 Introduction

Recommender systems are one of the most important tools to make decisions in the task of acquiring new products. The need of explanations in recommenders is obvious when users do not understand why an item (for example, a movie, a book, a song) is recommended to her. This causes mistrust to users, therefore they do not take into account the recommendations provided by these types of systems. Due to this need, research in explanations for recommenders is on the rise currently.

There are many types of explanations, as we reviewed in a previous work [9]. These evaluations are designed in many ways and differ depending on what the goals of the explanation. However, the evaluations have features in common, therefore we can classify them in several types. In this work, we make a review of the different styles of evaluations that we can find in current literature and we discuss the differences among them. We also provide concrete examples.

In Section 2, we review the related work and the previous works regarding the classification of evaluations in recommmender systems. Later, we propose the classification model in Section 3, where we define the two main facets of our proposal. Finally, in Section 4, we provide some examples of the different model types and we give some results from a model validation. Finally, we introduce some conclusions and future work in Section 5.

* Supported by the Complutense University of Madrid (Group 910494) and Spanish Committee of Economy and Competitiveness (TIN2017-87330-R) and the funding provided by Banco Santander in UCM (CT42/18-CT43/18).

2 Background

Nowadays, many research publications address evaluations in recommender systems and explanations. Many of them are surveys or reviews about this subject. In [40], we can observe an overview of evaluations which focuses on the effectiveness of explanations. Thanks to that, we can notice the importance of helping users to make good decisions. The work in [39] identifies some guidelines to evaluate explanations regarding their goals. Furthermore, it considers the benefits of using explanations and how we can measure them in an evaluation. The work in [13] does a review about explanations in recommender systems. It delves into the type of explanation interfaces and the criteria necessary to evaluate explanations. A guideline for making evaluations with users is detailed in [23]. Moreover, we find in [38] some problems produced in the evaluation of effectiveness in explanations and the way to solve them.

3 Evaluation of explanations in recommender systems

After the review of current surveys on this topic and the analysis of the evaluation systems presented in Section 4, we can deduce several common features in the evaluation process. Regarding these features, we can classify evaluations for explanations in recommender systems into two facets: the goal of the evaluation and the type of experimental process. The goal of the evaluation refers to the desired effect the explanation should have in the target user. Depending on that goal, the evaluation must be designed accordingly. The experimental process refers to the type of procedures and resources used to evaluate the explanations. Next, we discuss both facets.

3.1 Goal-directed evaluation

In general, explanations try to provide different properties to a recommendation, that affect the target user. The main goals of an explanation are related to the user and to the effect that we want to cause to this user. In the literature [29, 40], objectives of explanations are reviewed. In our previous work [9], we defined them and proposed two levels of goals: top-level goals, focused on the user; and low-level goals, which try to achieve the top-level goals. As evaluations take into account these goals to measure how good are the explanations designed, we can classify the evaluation of explanations according them.

We classified the top-level goals into three groups [9, 29]: *Improve user retention* (this goal refers to the increment of probabilities of a user returns to the recommendation system), *Improve user experience* (the explanations with this goal try the users enjoy the recommendation activity and help them to make good decisions) and *Justify recommendation* (the goal of these explanations is to help the users to understand the recommendations, explaining why an item was recommended).

The low-level goals can be defined as criteria to design good explanations that help to accomplish the top-level goals. According to this and taking into account the literature, we defined the following low-level goals [9, 29]: *Effectiveness*, (an effective explanation helps the user to get the items that she wants [36]), *Efficiency*, (an efficient explanation helps the user make decisions quickly [36]), *Trust*, (when users understand why the system has recommend to them an item, they use it more, because they rely on the system. Therefore, a good explanation has to provide trust to the user [3, 7, 36]), *Scrutability*, (if an explanation is scrutable, the user can provide feedback about how the recommender system is working [36]), *Persuasiveness*, (an explanation is persuasive when it convinces the user that the recommendation is suitable [7, 36]), *Satisfaction*, (explanations help to increase the user's perceived quality of the recommendations, enhancing user experience, general user enjoyment and pleasure with the system [36]), *Transparency*, (the transparency in an explanation allows to specify how and why a recommendation is made [7, 36]), *Education*, (with explanations, users can learn something that could help them. If users feel that they learn something, the probabilities of a user returning to the recommendation system improves [7]), *Debugging*, (explanations allow users to identify bugs in the recommendation system [7]).

3.2 Experimental Process

Another factor that we can use to classify evaluations in explanations is the process to make this evaluation. In the same way, as in evaluations for recommendation systems, the performance of the evaluation for an explanation depends on the experimental process [18, 23]. We can classify the evaluations according to this:

- *Online evaluation.* It is an evaluation carried out with users. The users make use of the explanation system and decide how good is the explanation shown by the system. This is the most difficult evaluation option, due to the human resources needed, but it is also the most reliable type because we directly assess users' opinions. That is why this is the most used type of evaluation. The performance consists of four important steps to know the users, collect data, and analyze them.
- *Offline evaluation.* This type of evaluation is carried out with an existing dataset. It allows to make an extensive evaluation in a fast way and cheaper. We can even use several datasets and compare many techniques easier. However, this option is much less used than the previous one because is less trustworthy.

3.2.1 Online evaluation We need to carry out online evaluations when we require to collect more information that an offline evaluation can provide to us about the explanation system evaluated. It considers users' behavior while they use the system, and more difficult goals to measure, as the satisfaction or the transparency. The interactions between users and items are real. Therefore, the information extracted from the evaluation is more reliable and accurate [34].

It is important to define what are the goals that we can measure and how is going to be performed the evaluation process. We should detail the questionnaires and the tasks we want the users to carry out. Typically, there are four steps in an online evaluation, that we should prepare with anticipation [18, 23, 34]:

- *Recruiting.* An initial questionnaire is usually asked to complete before the evaluation. This questionnaire tries to collect information about the user's profile, like age, gender, level of studies or personal data about the domain of the system.
- *Tasks description.* Several tasks are asked the user. The user has to do these tasks, while the evaluator observes her behaviour and collect information about.
- *Final questionnaire.* This questionnaire tries to collect the user's perception of how the explanation works for her. It has to be designed to measure the goals defined during the creation of the evaluation process.
- *Analysis of results.* The results are analyzed to check how good are the explanations evaluated. In online evaluations, statistical measures are usually used to get better conclusions from the questionnaire data.

3.2.2 Offline evaluation As we mentioned, offline evaluations use a precollected dataset of users interacting with items, for example, a dataset of ratings. The main goal of this type of evaluation is trying to understand the users' behavior when they use the explanation system that is being evaluated. Although offline evaluations allow measuring the prediction effectiveness of an explanation technique, they are not reliable for measuring other goals like satisfaction or trust. However, they are useful when we can not access to the resources from an online evaluation [34].

The procedure to carry out the evaluation is an iterative process on a dataset. The methodology of these procedures is similar to the methodology used in the evaluation of recommender and information retrieval systems. It consists of four steps:

- *Split the dataset into two partitions.* As a result, there is a training set, to generate the explanations, and an evaluation set, to make the evaluation. The split can be carried out in many ways, for example, by the timestamp.
- *Generation of the explanations.* Explanations are generated with the information included in the training set.
- *Evaluation of results.* It consists of a comparison between the results got in the previous step, and the information included in the evaluation set. If the results are similar, then the explanation system evaluated works effectively. It means that the predicted behavior of the users with the explanations got by the system is the expected behavior.
- *Analysis of results.* The results are analyzed to check if the goals defined for the explanation system are present in the system. Evaluation metrics are used to measure the comparison carried put in the previous step.

4 Evaluation approaches for recommender systems

Now that we have analysed the goals and methodologies for the evaluation of recommender systems, we present a list of such systems analyzed according to those features:

- E_1 [16]. It discusses different ways of explaining recommendations, focusing on the display format. The authors evaluate the explanations in two phases. The first one was an experiment with users interacting with the recommender system. Analysis of usage logs and questionnaires are carried out. The second phase is an interview with users to validate the previous data.
- E_2 [25]. It proposes a system of explanations for Shopr, an Android app for finding interesting clothes. Their explanations are interactive. The evaluation performed is a complete study with users, following the model of a typical online evaluation with a five-point Likert scale questionnaire.
- E_3 [41]. It introduces a novel explanation approach based on tags. The evaluation presents a survey with three parts, where participants answer a five-point Likert scale questionnaire about the proposed explanations.
- E_4 [2]. Authors use graphs and link prediction techniques in order to explain the recommendations. The evaluation is based on measuring the accuracy of predicted links using cross-validation. Moreover, they measure the efficiency comparing those results with a baseline of common neighbors and features.
- E_5 [11]. It introduces an explanation system that uses crowdsourcing to get relevant topics to justify recommendations. In this case, the evaluation is a user study where the proposal is compared with the approach in E_3.
- E_6 [30]. It describes explanations for recommender systems based on social elements and personality features. They carried out an evaluation with three groups of users. Each group receives a specific type of explanation and they analyzed the results got in the subsequent surveys.
- E_7 [37]. This paper introduces an explanation system for movie recommendations. The evaluation consists on a study with users. They observe a movie recommendation with its explanations and, later they answer some questions through a seven-point Likert scale.
- E_8 [43]. It describes a framework to generate knowledge-based explanations to provide transparency in recommendations. In the evaluation, users have to perform some tasks with the tool and they answer a questionnaire based on the TAM model.
- E_9 [32]. Authors propose an approach to explain recommendations by showing the context that satisfies the recommendation and the user's situation. To evaluate the system several recommendations with their explanations are presented to the users. Then, they have to answer some questions with a seven-point Likert scale.
- E_{10} [3]. This paper presents the results of an user study to analyse the influence of several explanation approaches on users. In this case, the evaluation has three phases. In the first phase, they collect demographic data from the users. In the next ones, users have to choose which recommendations or explanations are the best for them.

- E_{11} [31]. The proposal explains recommendations based on matrix factorization to provide transparency. The evaluation in this work is offline. With users that scored similar ratings, they compare the different predictions carried out by the proposal when the ratings change.
- E_{12} [17]. It describes a comparison between the explanations provided by a knowledge-based system and provided by people. The evaluation is online. First, users reply to some personal questions. Later, they watch several videos and explains why they think people in videos are lying or not. Then, they see the KBS explanation and answer questions related to TAM model.
- E_{13} [12]. It proposes an explanation system based on tags. This evaluation is similar to previous online ones. First, users give some demographic data and they rate movies according to the explanation style provided by the system, before and after of watching its trailer. Finally, users rate the explanations taking into account the system goals.
- E_{14} [42]. This publication introduces a probabilistic model based on matrix factorization which makes a sentiment analysis on user reviews to generate explanations. The evaluation consists of an online evaluation where users rate a product before and after of interacting with it, using a five-point Likert scale.
- E_{15} [22]. It proposes reciprocal explanations for recommendations in reciprocal environments, like online dating apps. The evaluation includes several modalities, where users reply to five-point Likert scale questions or the acceptance rate is measured if users interact with the recommended ones.
- E_{16} [6]. This proposal uses LinkedIn API to explain recommendations based on social and semantic knowledge using interactive visualization. In this work, the evaluation is a typical online evaluation: with a step to know the user, a step where users carry out tasks and the last step, where users fill surveys about the system transparency using five-point Likert scale questions.
- E_{17} [5]. It introduces a visual and interactive explanation using hybrid techniques from several social and semantic web resources. This is also an online evaluation with a phase to know the user, a phase to perform tasks with the system, and a last one to users evaluate the system, using a five-point Likert scale.
- E_{18} [21]. It proposes an rule-based explanation approach for user-based collaborative recommendations. There perform an online evaluation and an offline evaluation. In the online evaluation, users have to fill 5 questionnaires. In the offline evaluation, each rule is evaluated through several measures, like the similarity and novelty of the items got with the rule.
- E_{19} [24]. This publication describes an hybrid recommendation system with explanations that personalized the visualization. Again, we find a user study, with the classical steps in an online evaluation. We also find a seven-point Likert scale
- E_{20} [28]. It proposes a framework to generate explanations using natural language and Linked Open Data as a resource to build a graph of descriptive features. They performed an evaluation that consist on collecting data from

users and evaluating explanation with surveys that use a five-point Likert scale.
- E_{21} [20]. This work proposes a case-based explanation method for recommender systems based on matrix factorization. Authors evaluate the explanation system with an offline evaluation, comparing the ratings predicted by the recommender system with the actual ratings of the items retrieved by the explanation system.
- E_{22} [15]. This work uses Formal Concept Analysis to build user profiles employed to explain recommendations. The system is evaluated offline, analyzing several metrics, like height or average children, to measure the efficiency of FCA explaining user profiles.
- E_{23} [10]. We introduce a new explanation method that uses interaction graphs and link prediction techniques to explain black box recommendations. The evaluation is performed in the same way that the E_{21} approach.
- E_{24} [27]. It introduces a case-based explanation system for hotel recommendations that extracts features from user reviews. It is evaluated offline, analysing some properties of the items employed for the explanation.
- E_{25} [1]. Authors use an explainability graph in order to justify recommendations provided by a matrix factorization system. To evaluate the explanations, they propose two measures: explainability precision and explainability recall. Therefore, they performed an offline evaluation.
- E_{26} [35]. This work proposes a content-based recommender system that explains why an item was recommended. This conversational tool uses CBR. The evaluation is offline and measures the distance between the recommended items and the items found in the evaluation set.
- E_{27} [26]. This is a classic work that introduces a CBR conversational system that explains the recommendations provided to improve user acceptance. In this case, the evaluation is offline and it measures the effectiveness through sets of attributes. A recommendation is good when differs from the target query only in a set of attributes.
- E_{28} [19]. This publication describes a recommender system for groups that tries to satisfy all of the members' constraints. It includes explanations for the recommendations showing all of the users' preferences. The evaluation carried out here is an online evaluation, where users have to answer questions about their preferences using the system.

5 Conclusions

Table 1 provides a classification of the studied systems. Regarding the goal facet, we can observe that most of the evaluations have defined their goals. These goals are diverse, but we can conclude that Effectiveness and Satisfaction are the most analyzed ones. We can also find a pattern in the evaluations that not declare what the goals studied specifically are. These evaluations are made in an offline experimental process. We can assume that the main goal of an offline evaluation

Approach	Goal	Experimental Process
E_1	SA / T / EFY / EF / PE	Online
E_2	T / SCR / EFY / SA	Online
E_3	T / EF / EFY	Online
E_4	EF	Offline
E_5	EFY / EF / TR / SA	Online
E_6	PE / EFY / TR / SA	Online
E_7	EF / TR / PE / T / SA	Online
E_8	TR / SA	Online
E_9	PE / EF	Online
E_{10}	TR / T / EF	Online
E_{11}	T	Offline
E_{12}	EF / SA	Online
E_{13}	EF / SA	Online
E_{14}	EF / PE / SA	Online
E_{15}	SA / TR / T / EF	Online
E_{16}	SA / EF	Online
E_{17}	SA / EF	Online
E_{18}	EF	Offline / Online
E_{19}	PE / SA / EF	Online
E_{20}	T / PE / SA / TR / EF	Online
E_{21}	EF	Offline
E_{22}	EF	Offline
E_{23}	EF	Offline
E_{24}	EF	Offline
E_{25}	EF / T	Offline
E_{26}	EF / EFY	Offline
E_{27}	EF	Offline
E_{28}	EF / SA / T / PE	Online

Goals: Transparency (T), Scrutability (SCR), Trust (TR) Persuasiveness (PE), Effectiveness (EF), Efficiency (EFY) and Satisfaction (SA)

Table 1: Classification of evaluations studied according to our model.

is the effectiveness because it is the feature measured by the metrics used in this type of evaluations.

As we mentioned, online evaluation is the most used type of evaluation because it provides more information than offline ones. We can see more details about the opinions of users, and we can measure other more difficult goals to measure, like Satisfaction. For this reason, in our validation, there are more online evaluations than offline ones.

References

1. Behnoush Abdollahi and Olfa Nasraoui. Using explainability for constrained matrix factorization. In *Procs. of the 11th ACM Conf. on Recommender Systems*, pages 79–83. ACM, 2017.
2. Nicola Barbieri, Francesco Bonchi, and Giuseppe Manco. Who to follow and why: link prediction with explanations. In *Procs. of the 20th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 1266–1275. ACM, 2014.
3. Shlomo Berkovsky, Ronnie Taib, and Dan Conway. How to recommend?: User trust factors in movie recommender systems. In *Proceedings of the 22nd Int. Conf. on Intelligent User Interfaces*, pages 287–300. ACM, 2017.
4. Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. Tasteweights: a visual interactive hybrid recommender system. In *Procs. of the 6th ACM Conf. on Recommender systems*, pages 35–42. ACM, 2012.
5. Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. Linkedvis: exploring social and semantic career recommendations. In *Procs. of the 2013 Int. Conf. on Intelligent User Interfaces*, pages 107–116. ACM, 2013.
6. Bruce Buchanan and Edward Shortliffe. *Rule-based Expert System – The MYCIN Experiments of the Stanford Heuristic Programming Project*. 1984.

7. Marta Caro-Martinez, Guillermo Jimenez-Diaz, and Juan A Recio-Garcia. A theoretical model of explanations in recommender systems. In *XCBR: 1st Workshop on Case-based Reasoning for the Explanation of Intelligent Systems*.
8. Marta Caro-Martinez, Juan A. Recio-Garcia, and Guillermo Jimenez-Díaz. An algorithm independent case-based explanation approach for recommender systems using interaction graphs. In *Procs. of the 27th Int. Conf. on Case-Based Reasoning*, page In press. Springer, 2019.
9. Shuo Chang, F Maxwell Harper, and Loren Gilbert Terveen. Crowd-based personalized natural language explanations for recommendations. In *Procs. of the 10th ACM Conf. on Recommender Systems*, pages 175–182. ACM, 2016.
10. Wei Chen, Wynne Hsu, and Mong Li Lee. Tagcloud-based explanation with feedback for recommender systems. In *Procs. of the 36th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 945–948. ACM, 2013.
11. Julie Daher, Armelle Brun, and Anne Boyer. *A Review on Explanations in Recommender Systems*. PhD thesis, LORIA-Université de Lorraine, 2017.
12. Belen Diaz-Agudo, Marta Caro-Martinez, Juan A. Recio-Garcia, Jose Luis Jorro-Aragoneses, and Guillermo Jimenez-Díaz. Explanation of recommender systems using formal concept analyisis. In *Procs. of the 27th Int. Conf. on Case-Based Reasoning*, page In press. Springer, 2019.
13. Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. How should I explain? a comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies*, 72(4):367–382, 2014.
14. Justin Scott Giboney, Susan A Brown, Paul Benjamin Lowry, and Jay F Nunamaker Jr. User acceptance of knowledge-based system recommendations: Explanations, arguments, and fit. *Decision Support Systems*, 72:1–10, 2015.
15. Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5–53, 2004.
16. Anthony Jameson. More than the sum of its members: challenges for group recommender systems. In *Procs. of the Working Conf. on Advanced Visual Interfaces*, pages 48–54. ACM, 2004.
17. Jose Luis Jorro-Aragoneses, Marta Caro-Martinez, Juan A. Recio-Garcia, Belen Diaz-Agudo, and Guillermo Jimenez-Díaz. Personalized case-based explanation of matrix factorization recommendations. In *Procs. of the 27th Int. Conf. on Case-Based Reasoning*, page In press. Springer, 2019.
18. Marius Kaminskas, Fred Durao, and Derek Bridge. Item-based explanations for user-based recommendations. In László Grad-Gyenge and Roy Oberhauser, editors, *Procs. of eKNOW 2017, The Ninth Int. Conf. on Information, Process, and Knowledge Management*, pages 65–70, 2017.
19. Akiva Kleinerman, Ariel Rosenfeld, and Sarit Kraus. Providing explanations for recommendations in reciprocal environments. In *Procs. of the 12th ACM Conf. on Recommender Systems*, pages 22–30. ACM, 2018.
20. Bart P Knijnenburg and Martijn C Willemsen. Evaluating recommender systems with user experiments. In *Recommender Systems Handbook*, pages 309–352. Springer, 2015.
21. Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. Personalized explanations for hybrid recommender systems. In *Procs. of the 24th Int. Conf. on Intelligent User Interfaces*, pages 379–390. ACM, 2019.
22. Béatrice Lamche, Ugur Adıgüzel, and Wolfgang Wörndl. Interactive explanations in mobile shopping recommender systems. In *Joint Workshop on Interfaces and Human Decision Making in Recommender Systems*, page 14, 2014.

23. David McSherry. Similarity and compromise. In *5th Int. Conf. on Case-Based Reasoning*, pages 291–305. Springer, 2003.
24. Khalil Muhammad, Aonghus Lawlor, Rachael Rafter, and Barry Smyth. Great explanations: Opinionated explanations for recommendations. In *23rd Int. Conf. on Case-Based Reasoning*, pages 244–258. Springer, 2015.
25. Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. Linked open data-based explanations for transparent recommender systems. *Int. Journal of Human-Computer Studies*, 121:93–107, 2019.
26. Ingrid Nunes and Dietmar Jannach. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3-5):393–444, 2017.
27. Lara Quijano-Sanchez, Christian Sauer, Juan A Recio-Garcia, and Belen Diaz-Agudo. Make it personal: a social explanation system applied to group recommendations. *Expert Systems with Applications*, 76:36–48, 2017.
28. Bashir Rastegarpanah, Mark Crovella, and Krishna P Gummadi. Exploring explanations for matrix factorization recommender systems. 2017.
29. Masahiro Sato, Budrul Ahsan, Koki Nagatani, Takashi Sonoda, Qian Zhang, and Tomoko Ohkuma. Explaining recommendations using contexts. In *23rd Int. Conf. on Intelligent User Interfaces*, pages 659–664. ACM, 2018.
30. Guy Shani and Asela Gunawardana. Evaluating recommendation systems. In *Recommender Systems Handbook*, pages 257–297. Springer, 2011.
31. Hideo Shimazu. ExpertClerk: A Conversational Case-Based Reasoning Tool forDeveloping Salesclerk Agents in E-Commerce Webshops. *Artificial Intelligence Review*, 18(3-4):223–244, 2002.
32. Nava Tintarev and Judith Masthoff. A survey of explanations in recommender systems. In *2007 IEEE 23rd Int. Conf. on Data Engineering Workshop*, pages 801–810. IEEE, 2007.
33. Nava Tintarev and Judith Masthoff. The effectiveness of personalized movie explanations: An experiment using commercial meta-data. In *Int. Conf. on Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 204–213. Springer, 2008.
34. Nava Tintarev and Judith Masthoff. Evaluating recommender explanations: Problems experienced and lessons learned for evaluation of adaptive systems. In *Workshop on User-Centred Design and Evaluation of Adaptive Systems in association with UMAP'09*. Citeseer, 2009.
35. Nava Tintarev and Judith Masthoff. *Designing and Evaluating Explanations for Recommender Systems*, pages 479–510. Springer US, Boston, MA, 2011.
36. Nava Tintarev and Judith Masthoff. Explaining recommendations: Design and evaluation. In *Recommender systems handbook*, pages 353–382. Springer, 2015.
37. Jesse Vig, Shilad Sen, and John Riedl. Tagsplanations: explaining recommendations using tags. In *Proceedings of the 14th Int. Conf. on Intelligent User Interfaces*, pages 47–56. ACM, 2009.
38. Hanshi Wang, Qiujie Fi, Lizhen Liu, and Wei Song. A probabilistic rating prediction and explanation inference model for recommender systems. *China Communications*, 13(2):79–94, 2016.
39. Markus Zanker and Daniel Ninaus. Knowledgeable explanations for recommender systems. In *2010 IEEE/WIC/ACM Int. Conf. on Web Intelligence and Intelligent Agent Technology*, volume 1, pages 657–660. IEEE, 2010.