IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Service-Oriented Justification of Recommender System Suggestions

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/1798493 since 2023-02-10T15:07:36Z

Publisher:

Springer International Publishing

Published version:

DOI:10.1007/978-3-030-85613-7_23

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

12 October 2024

N. Mauro, Z.F. Hu, L. Ardissono. Service-oriented Justification of Recommender System Suggestions, (Author Copy (PDF)). Human-Computer-Interaction - INTERACT 2021, pp. 321-330, Lecture Notes in Computer Science book series (LNCS, volume 12936). DOI: 10.1007/978-3-030-85613-7_23. ISBN: 978-3-030-85607-6. Springer International Publishing, 2021.

Service-oriented Justification of Recommender System Suggestions

Noemi Mauro[1], Zhongli Filippo Hu[1], and Liliana Ardissono[1]

Computer Science Department, University of Torino, Torino, Italy
{noemi.mauro,zhonglifilippo.hu,liliana.ardissono}@unito.it

Abstract. In the selection of products or services, overviewing the list of options to identify the most promising ones is key to decision-making. However, current models for the justification of recommender systems results poorly support this task because, as they exclusively focus on item properties, they generate detailed justifications that are lengthy to skim. Moreover, they overlook the existence of a complex item fruition process which can impact customer satisfaction as well. For instance, consumer feedback shows that relevant factors in home booking include both the properties of apartments, and previous customers' perceptions of the interaction with the personnel who manages the homes. To address this issue, we propose a visual model that exploits an explicit representation of the service underlying item fruition to generate a high-level, holistic summary of previous consumers' opinions about the suggested items. From this overview, the user can identify the relevant items and retrieve detailed information about them, in a selective way, thus reducing information load. Our model is instantiated on the Airbnb *experiences* domain and uses the Service Blueprints to identify evaluation dimensions for the incremental presentation of data about items. A preliminary user study has shown that our model supports user awareness about items by enabling people to quickly filter out the unsuitable recommendations, so that they can analyze in detail the most relevant options.

Keywords: Summarization of recommendation lists · Service Blueprints · Explainable AI · Sentiment analysis.

1 Introduction

As reported in [6], product comparison is a crucial decision stage that buyers usually perform before they make a choice. Some recommender systems support this type of activity by generating post-hoc justifications of suggestions that highlight the main features of items which might interest the user [17,18,16,6]. However, these systems generate possibly lengthy, item-centric descriptions that only partially represent consumer experience, and might overload the user with too much data. For instance, in the e-commerce domain, it is not only a matter of liking the purchased item. The overall experience also includes the customer care, which might strongly impact the overall satisfaction about products. In service

fruition, the user might have to interact with multiple actors, all of which might positively or negatively influence the final service evaluation.

We claim that, to enhance user awareness about suggestions, recommender systems should present both item-centric data, and perceptions about item fruition, in a compact format that helps the user identify the options to be evaluated in detail. For this purpose, we propose a service-oriented, visual model that supports the summarization of holistic information about items, taking the overall context in which they are experienced into account. The visualization is based on interactive bar graphs that describe customer experience quantitatively. Moreover, our model organizes the aspects emerging from online reviews around coarse-grained and fine-grained evaluation dimensions which support a selective and incremental inspection of consumer feedback. We use the Service Blueprints [3] to define the evaluation dimensions of customer experience for information summarization. Service Blueprints support the design and development of physical and online products and services, by focusing on the actions that the customer carries out, and on the tangibles (s)he encounters, during item fruition. They help organize the feedback about items around the user.

We apply our model to the presentation of Airbnb *experiences*, which consist of informal, leisure activities published by people in the Airbnb platform (`https://www.airbnb.com`). An experience could be a boat tour in a river, or a half day spent with somebody who teaches the preparation of a traditional dish. The *host* of the experience typically participates in it, and an experience can involve multiple *guests*. Thus the user interacts with other people during its enactment. Our service-oriented representation of items, centered around tangibles and human actors, is suitable to model all these stages of fruition. Moreover, it can be applied to other services involving multiple actors, such as home booking.

In the following, Section 2 outlines the related work. Section 3 presents the proposed summarization model, and Section 4 concludes the paper.

2 Background and Related Work

2.1 Service Blueprints

A Service Blueprint [3] is a visual description of the user experience with a service, such as a hotel, or an online retailer. It models the stages a person encounters when (s)he uses the service, from the start point (e.g., enter website or enter shop) to the end one (e.g., customer care), taking both onstage and backstage actions into account. Different from standard workflows, blueprints focus on customer actions. Therefore, they model the actors and the tangibles the user is expected to interact with. A typical Service Blueprint has five components. The *Physical evidence* includes the tangibles that the customer comes in contact with. For instance, in a hotel service, this component represents the hotel website, the parking, the desk, the elevators, the room, and so forth. The *Customer actions* include the actions that the guest performs when (s)he experiences the service; e.g., the reservation and the arrival at the hotel. The *Onstage/visible contact employee actions* are the actions that the contact employees perform while

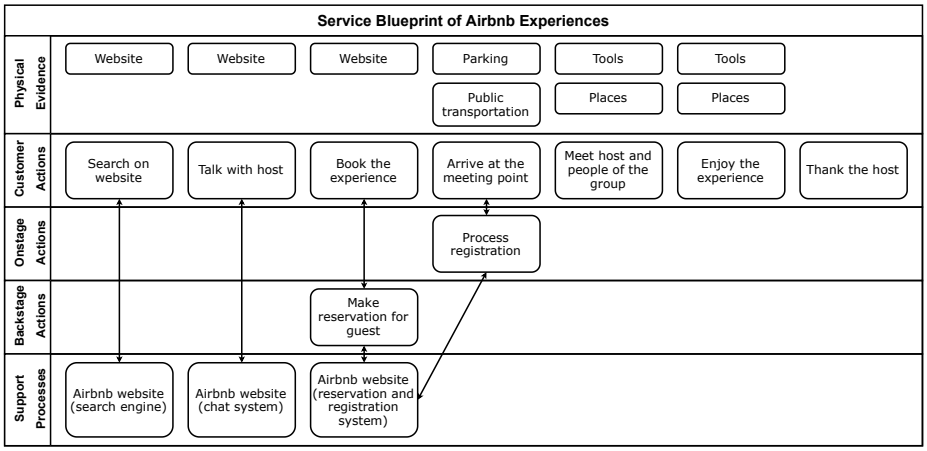

Fig. 1: Service Blueprint of Airbnb experiences.

they interact with the customer; e.g., desk registration. The *Backstage/invisibile contact employee actions* are the actions carried out by the employees that are not visible to the customer but are useful to serve her/him. The *Support processes* include the activities carried out by non-contact employees, such as the reservation and registration systems used by the hotel.

2.2 Explaining and Justifying Recommendations

The research about recommender systems promotes their transparency to enhance user trust and acceptance of results [25,20]. In the systems based on a single item suggestion algorithm, results are typically explained in terms of inference traces [9,19,25]. For instance, some aspect-based recommender systems explain suggestions by highlighting the features of items which match, or mismatch, the target user's preferences [15]. Other ones support feature-based item comparison [5,14], or information exploration based on the visualization of the relevance of items to the keywords of search queries [4,24]. Graph-based recommenders use the connections between users and items as explanations [2,26]. A few systems fuse recommendation and explanation in the same process [15,7,12].

We aim at justifying recommendations, rather than explaining how the system generated them. The reason is that the traces of inference might be too complex for the user [19]. Moreover, if s(he) applies other evaluation criteria than those implemented in the recommender system, the explanations might be hardly convincing. However, different from previous justification approaches [20,6,17,18,16], which directly extract detailed features of items from reviews, we identify the aspects to be visualized by using the underlying service model, represented as a Service Blueprint, to steer the analysis of consumer feedback. This enables us to present items in a high-level, holistic perspective, which covers all the stages of interaction with services that are behind item fruition, and

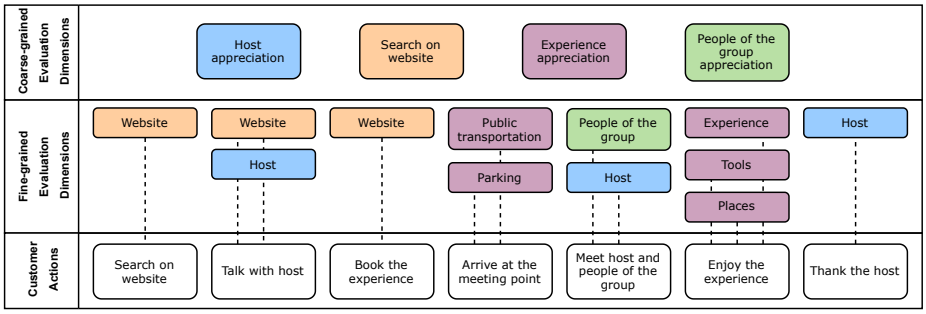

Fig. 2: Evaluation dimensions of Airbnb experiences. Colors represent the associations between fine-grained and coarse-grained dimensions.

provides a synthesis of consumer opinions that facilitates the inspection of large sets of results. The usage of Service Blueprints also distinguishes our work from classical aspect-based justification and explanation [15,20,14], which focus on few peculiarities of items. By applying the blueprints, we extract the sentiment about a small set of evaluation dimensions that describe the overall experience with items.

3 Methodology

During April and May 2019, we scraped the Airbnb web site to collect data about the published experiences. We selected the content written in English. For each experience E, we extracted the reviews about it, and details such as name, host, price, planned activities, and URL of a picture of E. We collected data about 254,253 Airbnb users, 11,086 experiences and 336,288 reviews.

Reviews can cover different evaluation dimensions. For instance, the following example describes multiple aspects (tour, atmosphere) and provides feedback about the host: *"Very nice tour, very different than all the big commercial tours. The guide was really nice and created a friendly atmosphere from the start!"*.

3.1 Service-based Representation of Airbnb Experiences

We introduce a service-oriented representation of items to capture the multiple perspectives about them emerging from online reviews. This representation is obtained in two steps. The first one is the **definition of the blueprint describing the service.** In this step, we represent the customer actions to be carried out during the fruition of the experiences. Moreover, we identify the actors and the tangibles which the guest is exposed to, as these elements contribute to her/his overall satisfaction. The second step is the **identification of evaluation dimensions** measuring consumers' perceptions about items, depending on those actors and tangibles.

We built the blueprint of Airbnb experiences, shown in Fig. 1, by studying both the service, and existing models available for the hotel booking domain [3]. The first tangible that the guest encounters is the Airbnb web site, through which (s)he can interact with the host of the experience and make the reservation. When the experience starts, the guest arrives at the meeting point and (s)he interacts with the host and with the other guests. Then, (s)he participates in the planned activities. Finally, (s)he closes the interaction with the other people.

While the Physical Evidence level of the blueprint describes the objects that the guest is expected to encounter, it does not represent human actors. However, the Customer Actions level provides both types of information. Thus, to obtain a holistic assessment of consumer perceptions, we map Customer Actions to evaluation dimensions associated with their involved tangibles and actors; see Fig. 2. To support both the summarization and a detailed organization of information, we define two levels of dimensions: a fine-grained one, related to the tangibles and actors involved in the Customer Actions, and a coarse-grained one, which describes the perceptions about the overall service at a high level. Fine-grained evaluation dimensions are grouped into coarse-grained ones, as shown by the colors in Fig. 2. For completeness, we introduced the *Experience* dimension, which is not strictly related to tangibles or human actors, because online reviews frequently mention intangible aspects such as ambience and time, which we want to represent to thoroughly model consumer feedback.

3.2 Analysis of Reviews

Given the Service Blueprint and the mappings to evaluation dimensions, we analyze the perceptions about an experience E in the following pipeline:

1. Extraction of the aspects mentioned in the reviews of E. Starting from the lemmatized text of each review R, we use the Spacy dependency parser [8] to build the syntax tree of the sentences of R. For each syntax tree, we apply an extension of the Double Propagation algorithm [22] to identify aspects and corresponding adjectives, as well as the adjectives referring to the host, who is frequently mentioned by name, or using pronouns, in nominal predicates. The output of this step is a list of $< aspect, adjective >$ pairs describing the mentioned aspects, combined with the adjectives used to qualify them. The second term is null when the adjective is missing.

2. Computation of the sentiment about aspects. We compute the polarity of each $< aspect, adjective >$ pair as the mean polarity value returned by the TextBlob [11] and Vader [10] opinion mining libraries. Then, we normalize this value in [1, 5]. We also count the occurrences of the pair in the reviews of E to measure how frequently people express the corresponding opinion.

3. Classification of aspects w.r.t. service evaluation dimensions. For this task, we combine entity recognition, which supports the identification of places and human actors, with the consultation of a dictionary for each fine-grained evaluation dimension. Each dictionary includes the lemmatized terms referring to the related tangibles or actors. The *Places* dictionary includes terms such as park, square, and river. The *People of the group* one includes words

Table 1: Aspects extracted from the reviews of an Airbnb experience. The rows are colored as the coarse-grained evaluation dimensions of Fig. 2.

Aspect	Adjective	Evaluation	Frequency	Fine-grained Evaluation Dimension
experience	great	4.43	39	Experience
tour	great	4.43	19	Experience
boat	big	3.00	17	Tools
time	great	4.43	15	Experience
host	great	4.43	14	Host
host	friendly	3.87	14	Host
experience	amazing	4.19	13	Experience
group	amazing	4.19	10	People of the group
snack	great	4.43	5	Tools

like people, group, and travelers. The *Experience* dictionary includes intangible concepts such as atmosphere, and time. Table 1 shows a subset of the aspects concerning an experience E. Each row shows an aspect and an adjective extracted from the reviews about E, the degree of appreciation (Evaluation) of the $< aspect, adjective >$ pair in [1, 5], and its frequency within these reviews. The last component of the row is the fine-grained evaluation dimension in which the aspect is classified.

To build the dictionaries, we first extracted the terms from a sample set of Airbnb reviews, by applying the techniques described in the step above. Then, three researchers from our University staff collaborated to map terms to the fine-grained evaluation dimensions of Fig. 2. When the researchers disagreed with each other, they discussed the outcome with the authors of the present paper. Notice that Airbnb experiences vary a lot, depending on the proposed activity. Therefore, a complete representation of terms within the dictionaries is practically impossible. Thus, when the aspect extraction returns a term that cannot be classified in any dictionary, we assume by default that it refers to the *Experience* dimension that has a broader scope than the other ones.

3.3 Visual Summary of User Feedback

We aim at enabling the user to overview a recommendation list, looking for the most promising items, and then focus on a few of them for a deeper analysis. This principle is inspired by Ben Shneiderman's zoom and pan information exploration model [1], which puts the user in control of the search process by supporting an incremental access to information. In our model, the classification of aspects in fine-grained and coarse-grained evaluation dimensions, and the computation of the associated sentiment, are the key elements to obtain this type of interaction because they support the presentation of a multi-faceted evaluation of consumer perceptions, at different levels of detail.

Fig. 3 shows the presentation of an Airbnb experience E. Below the picture of E, there are its host name, title, price, and a "Show more" link to retrieve its

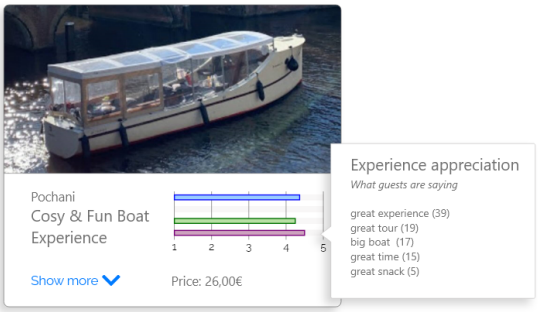

Fig. 3: Visualization of the Airbnb experience of Table 1. A visual legend, not shown, maps the colors of the bars to coarse-grained evaluation dimensions.

details. The bar graph summarizes consumer perceptions about E by showing the values of the coarse-grained evaluation dimensions emerging from its reviews. Each bar represents previous consumers' level of satisfaction about a dimension D, and it takes as value the weighted mean of the evaluations of the aspects classified in D. For each aspect, we use as weight its frequency in the reviews of E to tune its influence based on how many people share the opinion. The *Search on Web Site* bar is missing because the reviews do not refer to it.

By clicking on a bar, a box opens to present a few aspects extracted from the reviews of E that refer to the related coarse-grained dimension. For instance, see the "Experience appreciation" widget in Fig. 3. We aim at selecting diverse attributes to provide the user with rich qualitative data about previous guests' feedback. Thus, within the box of a coarse-grained dimension D, we show the most relevant aspects concerning each of its fine-grained dimensions, with their frequency. For instance, "great experience (39), ..., big boat (17), ..., great snack (5)". We use the frequency of aspects in the reviews of E as a measure of relevance because it reflects the conformity of consumer opinions.

3.4 Preliminary Test with Users in the Home Booking Domain

We tested with users a preliminary version of our model that is applied to the home booking domain and employs the Service Journey Maps [23] to represent items and to analyze online reviews. See [13] for details. This preliminary test makes sense because, similar to Airbnb experiences, the target of this system are young and middle-aged people who select items for leisure and work.

We recruited 11 participants (54.54% women; 45.46% men) representative of the target users. People joined the user study on a voluntary basis, without any compensation, and they gave their informed consent to participate in it.

Participants were between 19 and 57 years old (mean age = 36.36), and their educational level covered middle, high school and university. They declared that they regularly use the Internet. Based on the ResQue questionnaire [21], we asked participants if they tend to trust a person or thing, even though they have

little knowledge about it: 9.09% of people declared that they very probably trust it, 18.18% probably trust it, 54.55% probably do not trust it, and 18.18% very probably do not trust it. We can thus hypothesize that the majority would desire to have a relevant amount of data to evaluate the recommended items.

We defined two tasks to test two versions of the visual model and we investigated their support to decision-making. In `TASK1`, for each presented home, we only showed the bar graph representing the coarse-grained evaluation dimensions we defined for home booking. In `TASK2`, for each presented home, we linked every bar of the graph to a set of aspects that justify its value. The user could view the aspects concerning a specific evaluation dimension by clicking on the associated bar. We asked participants to rate 5 homes for each task, or to opt out if they were not able to make a choice given the presented information.

In `TASK1`, 12.73% of the ratings were "I don't know"; in `TASK2`, participants rated all the homes. This means that, if the bar graph does not provide any qualitative data to justify the values of the evaluation dimensions, it poorly supports decision-making. Notice that, in the post-task questionnaire after completing `TASK1`, 54.54% of people declared that the given information was not enough to rate the homes. Differently, in `TASK2`, the number of people encountering decision-making problems was lower.

These findings suggest that a visual summary of previous experience with items, based on coarse-grained evaluation dimensions, helps the user efficiently filter out unsuitable options. However, users want to explore the details about the promising items before making any selection decisions, a typical attitude towards high-investment products and services. This finding highlights the importance of the method that is applied to select the aspects for summarization. As the Service Journey Maps do not explicitly represent onstage and backstage actions, tangibles and actors, they poorly support the classification of consumers' opinions in fine-grained evaluation dimensions. We thus concluded that we need the power of Service Blueprints, and a two-level representation of evaluation dimensions, to organize the aspects extracted from reviews in an efficacious way.

4 Conclusions and Future Work

We described a service-oriented visual model for the justification of item suggestions that enables the user to overview recommendation lists by means of a holistic analysis of consumer feedback, based on Service Blueprints. Different from previous work, our model summarizes previous consumers' perceptions about the whole item fruition process, rather than only focusing on item properties. In a preliminary user test, we collected positive feedback about the support of our model to user awareness and decision-making. These findings encourage to use the model to enhance information exploration support, and the presentation of recommender systems' results, with a user-centric perspective on items. A systematic user test of our model is part of our future work.

References

1. Ahlberg, C., Shneiderman, B.: Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 313–317. CHI '94, ACM, New York, NY, USA (1994). https://doi.org/10.1145/191666.191775
2. Amal, S., Tsai, C.H., Brusilovsky, P., Kuflik, T., Minkov, E.: Relational social recommendation: application to the academic domain. Expert Systems with Applications **124**, 182 – 195 (2019). https://doi.org/10.1016/j.eswa.2019.01.061
3. Bitner, M.J., Ostrom, A.L., Morgan, F.N.: Service blueprinting: A practical technique for service innovation. California Management Review **50**(3), 66–94 (2008). https://doi.org/10.2307/41166446
4. Chang, J.C., Hahn, N., Perer, A., Kittur, A.: SearchLens: composing and capturing complex user interests for exploratory search. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. pp. 498–509. IUI '19, ACM, New York, NY, USA (2019). https://doi.org/10.1145/3301275.3302321
5. Chen, L., Wang, F.: Explaining recommendations based on feature sentiments in product reviews. In: Proceedings of the 22nd International Conference on Intelligent User Interfaces. p. 17–28. IUI '17, Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3025171.3025173
6. Chen, L., Wang, F., Qi, L., Liang, F.: Experiment on sentiment embedded comparison interface. Knowledge-Based Systems **64**, 44–58 (2014). https://doi.org/https://doi.org/10.1016/j.knosys.2014.03.020
7. Dong, R., Smyth, B.: User-based opinion-based recommendation. In: Proceedings 26th IJCAI. pp. 4821–4825. Melbourne, Australia (2017)
8. Explosion AI: SpaCy - industrial Natural Language Processing in python (2017), `https://spacy.io/`
9. Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work. p. 241–250. CSCW '00, Association for Computing Machinery, New York, NY, USA (2000). https://doi.org/10.1145/358916.358995, `https://doi.org/10.1145/358916.358995`
10. Hutto, C., Eric, G.: VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. pp. 216–225. AAAI, New York, NY, USA (2014), `https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109`
11. Loria, S.: TextBlob: Simplified text processing (2020), https://textblob.readthedocs.io/en/dev/index.html
12. Lu, Y., Dong, R., Smyth, B.: Why i like it: Multi-task learning for recommendation and explanation. In: Proceedings of the 12th ACM Conference on Recommender Systems. p. 4–12. RecSys '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3240323.3240365
13. Mauro, N., Hu, Z.F., Ardissono, L., Izzi, G.: A service-oriented perspective on the summarization of recommendations. In: Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization. ACM, Utrecht, The Netherlands (2021)
14. Millecamp, M., Htun, N.N., Conati, C., Verbert, K.: What's in a user? towards personalising transparency for music recommender interfaces. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization. p. 173–182. UMAP '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3340631.3394844

15. Muhammad, K.I., Lawlor, A., Smyth, B.: A live-user study of opinionated explanations for recommender systems. In: Proceedings of the 21st International Conference on Intelligent User Interfaces. p. 256–260. IUI '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2856767.2856813
16. Musto, C., de Gemmis, M., Lops, P., Semeraro, G.: Generating post hoc review-based natural language justifications for recommender systems. User-Modeling and User-Adapted Interaction **27** (2020). https://doi.org/10.1007/s11257-020-09270-8
17. Musto, C., Narducci, F., Lops, P., de Gemmis, M., Semeraro, G.: Linked open data-based explanations for transparent recommender systems. International Journal of Human-Computer Studies **121**, 93 – 107 (2019). https://doi.org/10.1016/j.ijhcs.2018.03.003
18. Ni, J., Li, J., McAuley, J.: Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 188–197. Association for Computational Linguistics, Hong Kong, China (Nov 2019). https://doi.org/10.18653/v1/D19-1018
19. Nunes, I., Jannach, D.: A systematic review and taxonomy of explanations in decision support and recommender systems. User Modeling and User-Adapted Interaction **27**(3–5), 393–444 (Dec 2017). https://doi.org/10.1007/s11257-017-9195-0
20. Pu, P., Chen, L.: Trust-inspiring explanation interfaces for recommender systems. Knowledge-Based Systems **20**(6), 542 – 556 (2007). https://doi.org/10.1016/j.knosys.2007.04.004
21. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: Proceedings of the Fifth ACM Conference on Recommender Systems. p. 157–164. RecSys '11, Association for Computing Machinery, New York, NY, USA (2011). https://doi.org/10.1145/2043932.2043962
22. Qiu, G., Liu, B., Bu, J., Chen, C.: Opinion word expansion and target extraction through double propagation. Computational Linguistics **37**, 9–27 (03 2011). https://doi.org/10.1162/coli_a_00034
23. Richardson, A.: Using customer journey maps to improve customer experience. Harvard Business Review (2015)
24. Di Sciascio, C., Sabol, V., Veas, E.E.: Rank as you go: User-driven exploration of search results. In: Proceedings of the 21st International Conference on Intelligent User Interfaces. p. 118–129. IUI '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2856767.2856797
25. Tintarev, N., Masthoff, J.: Explaining recommendations: design and evaluation, pp. 353–382. Springer US, Boston, MA (2015). https://doi.org/10.1007/978-1-4899-7637-6_10
26. Wang, H., Zhang, F., Wang, J., Zhao, M., Li, W., Xie, X., Guo, M.: Ripplenet: propagating user preferences on the knowledge graph for recommender systems. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. p. 417–426. CIKM '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3269206.3271739