Service-based Presentation of Multimodal Information for the Justification of Recommender Systems Results

Service-based Presentation of Multimodal Information for the Justification of Recommender Systems Results

Zhongli Filippo Hu
University of Torino
Torino, Italy
zhonglifilippo.hu@unito.it

Noemi Mauro
University of Torino
Torino, Italy
noemi.mauro@unito.it

Giovanna Petrone
University of Torino
Torino, Italy
giovanna.petrone@unito.it

Liliana Ardissono
University of Torino
Torino, Italy
liliana.ardissono@unito.it

ABSTRACT

The current models for the explanation and justification of recommender systems results focus on qualitative and quantitative data about items, overlooking the power of images to describe the different aspects of experience that the consumer should expect from their selection to post-sales. In the present paper, we extend previous justification models by exploiting object recognition on images to support a service-oriented presentation of multimodal (textual, quantitative, and images) information about items. As a testbed for our model, we chose the home-booking domain. In a user study, we found that item comparison can be enhanced by empowering the user to filter multimodal data based on a set of evaluation dimensions describing the experience with items. These results encourage the introduction of service-based filters for multimodal information retrieval in product and service catalogs.

CCS CONCEPTS

• **Information systems** → *Web searching and information discovery*; *Recommender systems*; • **Human-centered computing** → *Interaction techniques.*

KEYWORDS

Justification of Recommender Systems Results, Images, Service Models

ACM Reference Format:
Zhongli Filippo Hu, Noemi Mauro, Giovanna Petrone, and Liliana Ardissono. 2023. Service-based Presentation of Multimodal Information for the Justification of Recommender Systems Results. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (UMAP '23), June 26–29, 2023, Limassol, Cyprus.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3565472.3592962

UMAP '23, June 26–29, 2023, Limassol, Cyprus

ACM ISBN 978-1-4503-9932-6/23/06…$15.00
https://doi.org/10.1145/3565472.3592962

1 INTRODUCTION

Transparency has become a core property of recommender systems and several models have been developed to explain, or justify, the suggestions [11]: the explanation models describe how a recommender system generates the results [18, 32, 41, 42]; the justification models generate post-hoc descriptions of the suggested items [29, 31]. In [27], we introduced a service-based justification model of recommender system results, having recognized that items are complex entities that might impact the consumer experience by involving people in stages of interaction with multiple services and actors [39]. For instance, when renting a hotel room, not only the location but also the interaction with the clerks might influence the customer's overall experience, determining multiple evaluation dimensions of experience to be taken into account. However, similar to the other explanation and justification models, we did not exploit item images as vectors of specific impressions about the system's suggestions.

We now extend our previous work to manage *multimodal information.* We propose a justification model that allows filtering qualitative and quantitative data and images of the system's suggestions by fine-grained evaluation dimensions describing the expected consumer experience during the stages of interaction with items. By exploiting object recognition on images [37], and a Service Blueprint [3] representation of the service underlying item fruition, we classify multimodal data in the service stages it is about, and by keywords. This supports the exploration and comparison of results, which is a crucial decision stage that buyers usually perform before making a choice [8].

As we test our model in the home-booking domain, we consider service stages such as the interaction with the host renting the home and the in-apartment experience. Moreover, we index images and textual data by keywords like the kitchen of a home (that belongs to the in-apartment experience stage). We are interested in investigating whether this information filtering support is useful in users' selection decisions. Thus, we pose the following research questions:

- RQ1: *Does a service-based presentation of images and textual data about items enhance the comparison of recommendations w.r.t. a flat presentation that does not take service stages into account?*

- RQ2: *Does the possibility to filter images and textual data about items by keywords help the comparison of recommendations?*

To answer these questions, we developed a test application that guides the user in the exploration of a list of homes to choose the preferred one. The application manages the following visualization models:

- FILTER-WITH-IMG enables the user to filter multimodal information by fine-grained evaluation dimensions describing the expected experience during the stages of interaction with items; e.g., the in-apartment experience, or the host appreciation. Moreover, the model offers a keyword-based filtering function to further project data on details, such as the kitchen of the homes.
- FULL-DATA shows the same data as FILTER-WITH-IMG but, similar to standard catalogs, it does not support service-oriented information filtering.
- FILTER-WITHOUT-IMG supports service-oriented information filtering but omits images.

In a user study involving 50 participants, we found that they perceived the FILTER-WITH-IMG model as useful to item comparison. However, they appreciated and used the filter by fine-grained evaluation dimension more than the keyword-based one. This suggests introducing service-oriented information filtering in product and service catalogs through a simpler user interface.

In the following, Section 2 presents the related work. Section 3 describes the data and its pre-processing used by the visualization models, which are presented in Section 5. Section 6 details the methodology applied in the user study and Section 7 discusses its results. Section 8 describes the limitations of our work and concludes the paper.

2 RELATED WORK

Several researchers found that enriching recommender systems with explanation functions increases the acceptance of results [9, 34, 42], even though different types of explanations (e.g., visual, textual or hybrid) might be suitable depending on the user profile [23, 40] and, in some circumstances, it is best not to provide an explanation at all [5, 43].

Most existing explanation [14, 33, 41] and justification models [29, 31] overlook the role of images in the preview of the expected user experience with products and services. Some systems show a single photo for each item but they focus on textual and quantitative data for its presentation [7, 10, 13, 28]. Review-based recommender systems [6, 17], and the models supporting item comparison, such as [8], show item images but they fail to link them to the stages of interaction with items because they are service-agnostic.

Images are key to previewing the aspects of items in product catalogs (e.g., Amazon [2] and Zalando [45]) and service ones (e.g., Airbnb [1] and Booking [4]). However, these systems propose lists of photos showing diverse types of information, such as the indoor and outdoor scenes of a hotel, which the user has to navigate to retrieve the relevant data. Differently, we aim at supporting multimodal information filtering to enhance item comparison. We do this by extending service-based recommender systems [26, 27] with a justification model that combines images with qualitative and quantitative data.

Several works analyze images to retrieve extra information about items, such as the extraction of clothes characteristics in MMFashion [25] and the recognition of ingredients for recipe retrieval [20]. Some recommender systems employ image analysis to build user profiles [21], or to identify sets of similar items, e.g., clothes that look like those selected by the user in fashion recommender systems [10]. Some systems suggest images that the user may like by exploiting their metadata [22]. In comparison, we analyze images to recognize the presence of key elements to compare items based on the expected experience during the stages of interaction with them. For instance, by combining object recognition [37] with service modeling [3], we can distinguish the indoor and outdoor photos of a home, its rooms (kitchen, bedroom, etc.), and the presence of specific elements such as a table or a TV.

Our visualization model builds on the literature about faceted user interfaces [16] and on Shneiderman's mantra "Overview first, zoom and filter, and details on demand" [38]: we use keywords and fine-grained evaluation dimensions of experience as filters to enable the user to view the multimodal information (s)he is most interested in.

3 DATA

For this work, we used a dataset Δ of Airbnb reviews about homes located in London city, which we downloaded from http://insideairbnb.com/get-the-data.html in January 2021, and we filtered on the reviews written in English. For each home h, Δ contains the title of h, the link to its first image, the list of amenities it offers (TV, balcony, etc.), the link to its host's Airbnb page, and the list of reviews that h received from previous guests. To enrich the basic data provided by Δ, we used the results of [27], where we extracted the aspects of the homes emerging from their reviews to retrieve the sentiment of previous guests [36]. That work applied a standard NLP approach, based on lemmatization and dependency parsing, to extract the aspects, and the aspect-adjectives pairs occurring in the reviews.

For our experiments, we needed more than one image per home. However, we planned to use 15 homes in total; see Section 7. Thus, we decided to do a limited scraping of the Airbnb website to retrieve that information. We worked as follows, in September 2022: first, we sorted the homes of Δ by their number of reviews, in decreasing order. Then, we manually browsed the sorted list to select the first 15 homes that were still available on the Airbnb website and had at least 15 photos. For each of the selected homes, we downloaded the full list of images available on the Airbnb website, obtaining 321 photos. Then, we analyzed such images through object recognition to identify the entities appearing in them, e.g., a bed, a TV, and so forth, and we annotated them accordingly.

To perform the object recognition, we trained a YOLOv5 model [19] with transfer learning using the Scene Understanding Database (SUN2012) [44]. SUN2012 includes 16,873 images and specifies, for each one, the types of objects appearing in it and the coordinates of their boundaries. It has a large set of "classes" denoting different types of objects, among which those relevant to our work. With this newly trained model, we performed object recognition on the 15 images used in our user study. The result of this analysis is a vector representation of each image, storing the list of classes that have been identified (we overlook coordinates). Unfortunately, the

Table 1: Coarse-grained and fine-grained evaluation dimensions with references to physical evidence and keywords.

Coarse-grained dimensions	Fine-grained dimensions	Physical Evidence	Dictionary
Host appreciation	Host	-	advice, communication, host, tip, …
Check-in/Check-out	Check-in	Check-in tangibles	arrival, access, check-in, wait, key, …
	Check-out	Check-out tangibles	check-out, departure, goodbye, …
In-apartment experience	Ambiance	Ambiance	air conditioning, atmosphere, smell, …
	Bathroom	Bathroom amenities	towel, shower, soap, hair-dryer, …
	Kitchen	Kitchen amenities	kitchen, fridge, microwave, oven, …
	Laundry	Laundry	dryer, ironing board, washer, …
	Relax	Relax amenities	balcony, wi-fi, tv, swimming pool, …
	Bedroom	Bedroom amenities	bed, pillow, wardrobe, blanket, …
Surroundings	Surroundings	Surroundings	attraction, gym, lake, street, sunset, …
	Services	Services	transportation, atm, bus, grocery, …

trained model has low precision because some classes of SUN2012 are poorly represented; e.g., "bathtub" and "parking". For this reason, the model did not correctly detect some objects included in the images of the Δ dataset. Thus, for the purpose of our user study, we manually checked the extracted labels to adjust the misclassified or unclassified objects. To improve the precision of the automatic annotations, in our future work we plan to look for other state-of-the-art object detection algorithms.

4 SERVICE-BASED CLASSIFICATION OF MULTIMODAL INFORMATION ABOUT ITEMS

To support a service-oriented justification of recommender system results, we have to model (i) the stages of interaction with the tangibles and actors involved in the item fruition process, and (ii) the evaluation dimensions of experience concerning such stages [27, 39]. Moreover, the information presented by the system has to be classified with respect to the evaluation dimensions of experience, so that the user can filter data depending on her/his interests. In the following, we describe the building blocks of this knowledge representation approach; see [26, 27] for details.

Coarse-grained and fine-grained evaluation dimensions of experience. To support the presentation of multimodal information according to the expected consumer experience with items, we use a subset of the evaluation dimensions of experience described in [27]. In that work, we defined a Service Blueprint [3] to describe the stages of interaction with homes in the home-booking domain (see Figure 1 of [26]) and we derived a set of coarse-grained and fine-grained evaluation dimensions of experience. Table 1 shows those relevant to the present work:

- The *coarse-grained dimensions* represent high-level evaluation dimensions to assess the consumer's experience during the main stages of the service; e.g., the interaction with the host of the home ("host appreciation"), and experience while being there ("in-apartment experience").
- The *fine-grained dimensions* represent a more specific assessment of the consumer experience deriving from the actions performed during the service stages. For instance, while being in a home, the user is expected to use the kitchen and bathroom, and these actions induce experiences with the associated tangibles and actors, contributing to her/his overall experience. The tangible and actors are defined in the Physical Evidence layer of the service blueprint and reported in the third column of Table 1.

Keywords. The tangibles and actors, such as "Bathroom amenities", are generic and cannot be directly applied to classify the data about the homes; therefore, we use a set of dictionaries defined in [27] to specify the terms relevant to each of them. The fourth column of Table 1 shows some sample keywords from such dictionaries.

Information classification. As the dictionaries are coupled with the fine-grained evaluation dimensions of experience, they support the classification of both the aspects of items extracted from their reviews and their images. The latter are classified using the vector representation described in Section 3, having mapped the classes defined in the SUN2012 dataset to the keywords included in the dictionaries.

5 JUSTIFICATION MODELS

5.1 FILTER-WITH-IMG

In all the justification models we propose, the test application we developed presents five homes to choose from. Figure 1 shows a portion of the user interface of the **FILTER-WITH-IMG model**. For each home h, the application shows the offered amenities, a bar graph that summarizes previous guests' experience with h (see [27] for details), the images of h, its reviews, and the "Select home" button to set h as the preferred home of the list. The names of the homes are hidden to prevent the user from searching for them on the Airbnb website.

The top of the page includes the information filtering menus. Each of them represents a coarse-grained evaluation dimension (e.g., "Surroundings" in Figure 1) and the user can open it to focus on a fine-grained dimension d ("surroundings"). When the user clicks on a filter from the menus, the application focuses the presentation on

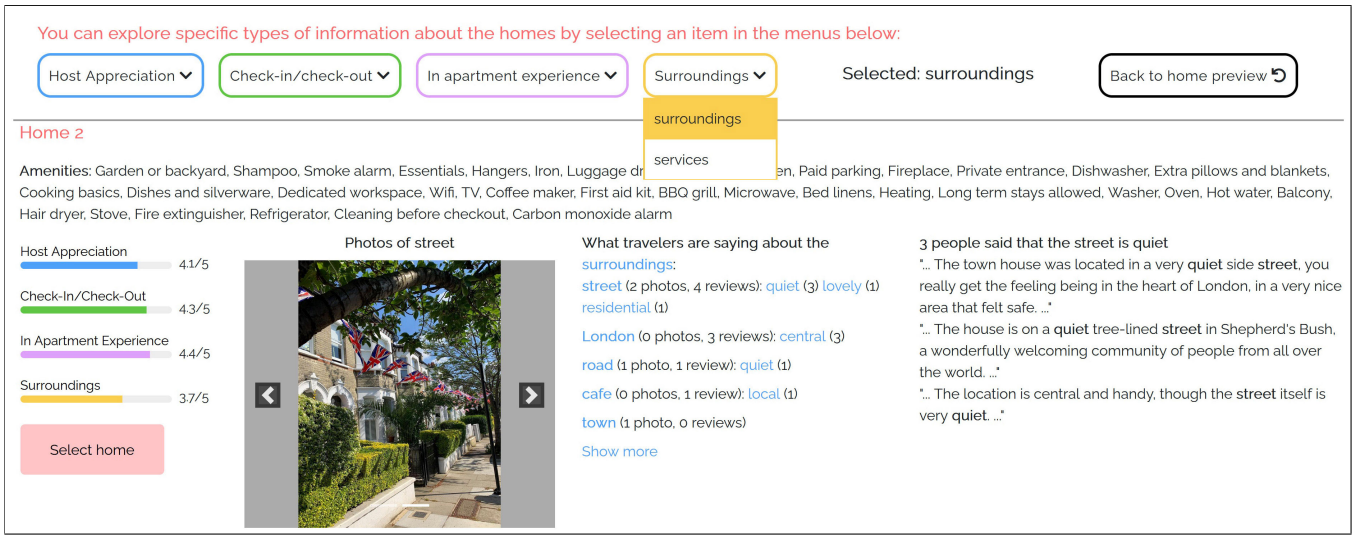

Figure 1: Portion of the user interface of the FILTER-WITH-IMG justification model.

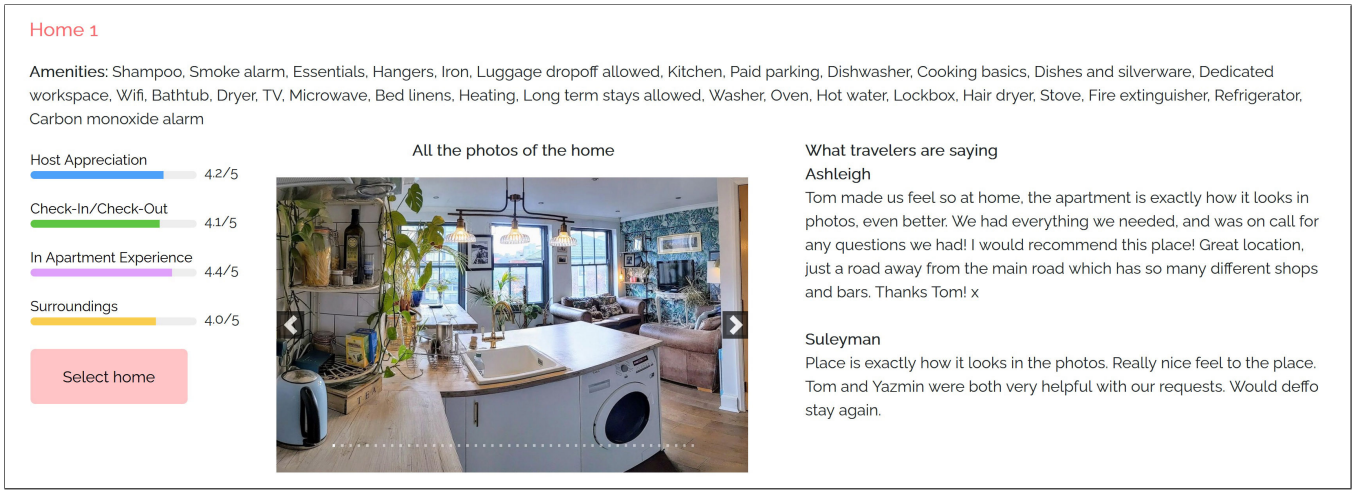

Figure 2: Portion of the user interface of the FULL-DATA justification model.

the selected fine-grained evaluation dimension.[1] For each home h, a clickable carousel displays the images of h classified in d. Moreover, the application shows the reviews of h classified in d (denoted as R_{hd}), and two scrollable widgets:

- The left one provides keyword filters to focus the presentation on the aspects (nouns) and adjectives extracted from the reviews of R_{hd}.
- The right one presents such reviews. If the user clicks on a keyword to further restrict the focus, the application shows the sentences of R_{hd} that mention the selected term. In the figure, the user has clicked on the "street" keyword to restrict the information about the "surroundings" of h.

[1]In case of empty results (images and reviews), the application shows a default "no information" image.

The filters also impact the images: having clicked on "street", the carousel is restricted to "Photos of the street".

5.2 Baselines

Figure 2 shows a portion of the user interface of the **FULL-DATA model**. Similar to FILTER-WITH-IMG, it shows the offered amenities and the bar graph to summarize the consumer experience with the home. However, it shows all the available photos within a clickable carousel and all the reviews in a scrollable widget. Moreover, it does not provide any widgets to filter data by fine-grained evaluation dimension or by keyword.

The **FILTER-WITHOUT-IMG model**, not shown for brevity, is derived from FILTER-WITH-IMG by omitting the carousel of the images. It enables the user to filter the reviews of the homes by

fine-grained evaluation dimension of experience and to further restrict data by keyword.

6 VALIDATION

We conducted a user study to evaluate the users' experience with the three justification models.[2] For this purpose, we developed the test application shown in Figures 1 and 2. The application guided the participants through the experiment without our supervision, logging the click and scroll actions they performed to support the analysis of their behavior. To guarantee users' privacy, the application did not collect their names or any other identifying data and used numerical identifiers to tag the anonymous data it acquired during the interaction sessions.

In the study, we used a within-subjects approach. We managed each treatment condition (FILTER-WITH-IMG, FULL-DATA, FILTER-WITHOUT-IMG) as an independent variable, and each participant received all the treatments. The application presented a different order of tasks to the users to reduce the effect of fatigue and practice, and the result biases. It did not impose any time limits for the completion of tasks.

(1) Initially, the application showed the informed consent (see https://bit.ly/3X3Myg4), and asked participants to give their explicit agreement. Moreover, it asked them to declare that they were 18 years old or over.
(2) Next, it asked users some questions about demographic information, cultural background, and familiarity with booking and e-commerce platforms.
(3) Then, it asked users to interact with the three justification models, in counterbalanced order. For each model, the participant explored five homes (the same for all the users, to support the comparative analysis of their behavior), and the application asked her/him to choose which one (s)he would have liked to book. Immediately after having completed the interaction with an individual model, the application administered a post-task questionnaire to evaluate the user's level of agreement with the statements of Table 2. These statements are taken from [12, 24, 35] and are based on the ResQue recommender system questionnaire. They measure the experience and perceptions of the user interface. Participants answered on a {Strongly disagree, …, Strongly agree} scale, mapped to [1, 5].

The questionnaires include some attention checks to verify that people worked with care during the user study.

We recruited people using social networks and public mailing lists, specifying in the invitation message that we searched for adult people. The invitation message included the link to the URL of the test application. All the participants joined the experiment voluntarily, without any compensation.

7 RESULTS

We conducted the user study from November 1st to December 20th, 2022. The entire user study took on average 19.89 minutes per participant, with a Standard Deviation = 10.36.

[2]Our experiment has been approved by the Ethics Committee of the University of Torino (Protocol Number: 0421424).

We recruited 59 people but we excluded 9 of them because they did not pass the attention checks. Thus, the sample size of the user test is N = 50. The subjects were 21 females, 29 males, 0 not-binary, and 0 not declared, with the following age distribution: ≤ 20 (13 people), 21-30 (33), 31-40 (3), and 41-50 (1). Education level: high school (15), university (30), and Ph.D. (5). Background: technical (19), scientific (20), humanities and languages (5), economics (2), and another background (4). 23 participants classified themselves as advanced computer users, 24 as average ones, and 3 as beginners. We asked people to evaluate their familiarity with online booking or e-commerce platforms and 13 people declared that they used those platforms a few times a week, 26 a few times a month, and 11 a few times a year.

7.1 Analysis of participants' experience with the visualization models

Table 2 shows the results of the post-task questionnaire. We conducted a *post-hoc* comparison using a Mann-Whitney test which showed a limited statistical significance of the difference between FILTER-WITH-IMG and FULL-DATA. However, the test clearly differentiated the perception of the models that show the images of the homes from the FILTER-WITHOUT-IMG one, which omits them ($p < 0.01$).

The participants perceived FILTER-WITH-IMG as the model that best supports the understanding of why homes are good or bad (Q1), and which helps compare homes in the most effective way (Q2, statistically different from both FILTER-WITH-IMG, $p < 0.01$ and FILTER-WITHOUT-IMG, $p = 0.08$). People also perceived FILTER-WITH-IMG as the most informative model (Q3) and they declared that the provided data about homes is sufficient to select a home (Q5). Overall, we can say that this model achieves a good evaluation as far as the support in item selection and comparison is concerned. We explain this finding with the information filtering support it provides (a function that FULL-DATA does not offer), combined with the provision of the images of the homes. Differently, FILTER-WITHOUT-IMG, which omits that information, is badly evaluated in all these aspects.

Regarding the perception of the user interface, the participants declared that, compared to the other models, FULL-DATA is less cluttered and confusing (Q4); moreover, the information about the homes is easier to interpret and understand (Q6). The preference for FULL-DATA can be explained by the fact that this justification model offers the same interactive functions as well-known platforms like Airbnb and Booking. Those platforms show all the reviews of the homes without providing any data filtering tools. Even though filtering reviews by fine-grained dimension and possibly also by keyword might have challenged the users, they were aware of the value of this function and declared that they found the information about homes more quickly (Q7) when using FILTER-WITH-IMG.

We investigated participants' overall satisfaction, with a focus on item comparison, through statements Q8 and Q9. The models reporting both images and textual data scored comparably. Participants declared that they preferred to frequently use the FILTER-WITH-IMG model in the comparison of homes (Q8), and secondly the FULL-DATA one. Moreover, they felt equally confident in using both models for this task (Q9). We explain these results as follows:

Table 2: Post-task questionnaire results. We report the mean value of users' replies with Standard Deviation. The best values for each statement are in boldface (minimum for Q4, maximum for the other statements). Stars denote the statistical significance of the difference between the best-performing model and the other ones. Significance levels: (**)$p < 0.01$, (*)$p = 0.08$.

	FILTER-WITH-IMG	FULL-DATA	FILTER-WITHOUT-IMG
Q1: It was easy to understand why some homes were good and others not.	**3.52(0.81)**	3.34(1.02)	2.78(1.13)**
Q2: The system helped me to compare the homes.	**3.62(0.85)**	3.18(1.08)*	3.00(1.11)**
Q3: The system was sufficiently informative.	**3.76(0.72)**	3.72(0.90)	2.84(1.22)**
Q4: The system was cluttered or confusing.	2.50(0.95)	**2.32(1.17)**	2.90(1.15)**
Q5: The information about the homes was sufficient for me to select a home.	**3.88(0.77)**	3.86(0.86)	2.80(1.21)**
Q6: The information about the homes was easy to interpret and understand.	3.70(0.89)	**3.78(0.86)**	3.04(1.03)**
Q7: I found the information about homes quickly.	**3.62(0.78)**	3.52(0.91)	3.00(1.01)**
Q8: I think that I would like to frequently use this system to compare homes.	**3.30(0.95)**	3.16(1.13)	2.38(1.03)**
Q9: I felt very confident using this system to compare homes.	**3.36(0.85)**	**3.36(0.90)**	2.72(0.97)**

Table 3: Log analysis. Time is measured in seconds, # denotes the mean number of events per user.

	FILTER-WITH-IMG	FULL-DATA	FILTER-WITHOUT-IMG
Mean time spent to explore 5 homes	170.06	178.26	169.2
# scrolling on homes	28.84	20.92	29.18
# scrolling on reviews	45.56	33.06	54.74
# visualized reviews	32.30	17.00	32.49
# clicks on fine-grained dimensions	8.90	-	5.65
# clicks on keywords	1.62	-	2.1
# clicks on photos	15.56	48.50	-

FILTER-WITH-IMG supports item comparison in a more effective way than FULL-DATA; however, as these systems provide similar information about items, they make the user equally confident in the selection decisions.

7.2 Log analysis

During the interaction with the participants of the user study, our test application logged the clicks and the scrolls on the components of the user interface. There are two types of scrolls: the former is aimed to visualize the hidden portion of the list of homes presented to the user. The latter enables the user to view hidden reviews. Table 3 shows the most relevant data we collected. It reports the mean values per user, during the interaction with a specific justification model:

- "Mean time spent to explore 5 homes" is the average time that participants spent exploring the visualized homes, and selecting the preferred one.
- "# scrolling on homes" is the mean number of times a specific home became visible on the screen for more than 2 seconds. It represents the mean amount of scrolling performed by users to explore a list of homes. We did not consider the visibility for less than 2 seconds because it is too short to represent a reading event; e.g., it could be an accidental visualization while the user browses the list to reach the homes placed at its ends.
- "# scrolling on reviews" is the mean number of times a review became visible on the screen for more than 2 seconds (a minimum time to capture the reading of very short reviews such as "Amazing view!" and overlook quick scrolls). It measures the mean amount of scrolling activity on the reviews of the homes.
- "# visualized reviews" is the mean number of distinct reviews visualized on the screen for more than 2 seconds.
- "# clicks on fine-grained dimensions" is the mean number of times participants filtered the information about the homes by clicking on some fine-grained dimensions. This type of event is only available in FILTER-WITH-IMG and FILTER-WITHOUT-IMG.
- "# clicks on keywords" is the mean number of times users filtered the information about the homes by clicking on aspects (for instance, a noun such as "kitchen") or adjectives ("beautiful"), only available in FILTER-WITH-IMG and FILTER-WITHOUT-IMG.
- "# click on photos" is the mean number of times participants clicked on the carousels of the homes to change the visualized images, available in FILTER-WITH-IMG and FULL-DATA.

The mean time spent by users on the three user interfaces is rather similar, with a slightly higher value in FULL-DATA, which does not provide the information filtering functions offered by the other two models.

Interestingly, "# scrolling on homes" shows that participants moved in the user interface to view the homes much more frequently in FILTER-WITH-IMG and FILTER-WITHOUT-IMG than in FULL-DATA. The two models supporting information filtering obtained almost 38% more scrolls than FULL-DATA. This means that people visualized the homes in the list a larger number of times, denoting higher comparison activity. Similarly, # scrolling on reviews" shows that, with FILTER-WITH-IMG and FILTER-WITHOUT-IMG, users scrolled the review list more frequently than with FULL-DATA, and "# visualized reviews" shows that they read twice as many reviews as in FULL-DATA. However, FILTER-WITHOUT-IMG received the larger amount of scrolling, probably because users needed to check the reviews in depth, lacking the support of images to evaluate the homes.

A direct comparison of the models providing the information filters shows that the participants clicked on average 8.9 times on the fine-grained evaluation dimensions of experience when using FILTER-WITH-IMG (i.e., about 2 clicks per home) and 5.65 times when using FILTER-WITHOUT-IMG. As the main difference between the two models is that FILTER-WITH-IMG shows the images while FILTER-WITHOUT-IMG does not, the filtering activity is probably aimed at selecting relevant ones to view. Differently, users used the filters by keyword in a more or less equivalent way on the two models but the usage of these filters is very limited and thus does not provide much information.

The last comparison concerns the two models that show the images of the homes. The number of clicks on the carousels shows that, when participants interacted with the user interface of FULL-DATA, they browsed the photos 3 times as much as with FILTER-WITH-IMG and visualized about 15 and 5 images per home, respectively. We explain this observation with the fact that, in FILTER-WITH-IMG, they could focus on the photos describing the most relevant scenes, such as the indoor environment of the homes. Differently, in FULL-DATA, they had to browse the image lists searching for relevant photos in an indiscriminate way.

7.3 Discussion

The participants of the user test perceived the FILTER-WITH-IMG and FULL-DATA justification models, which show the photos of the homes, as better than FILTER-WITHOUT-IMG. In other words, the combination of textual and pictorial information is useful for item comparison.

The most interesting results concern the trade-off between the power of information filtering on the exploration of multimodal data about items, and the complexity brought by the service-based filters we propose. On the one hand, the FILTER-WITH-IMG model combines the helpfulness of visualizing the images of homes with the conciseness of a service-aware presentation of information and empowers users to quickly find what they need, reducing the effort in the analysis of the reviews. However, it was perceived as more cluttered than the baselines. Anyway, the log analysis confirms with objective data about user behavior that the information filters based on the fine-grained evaluation dimensions of experience (offered by FILTER-WITH-IMG) clearly favored the comparison activity compared to FULL-DATA. Moreover, the log analysis reveals that the filter of images is useful to explore the data about the suggested items. We can thus positively answer research question RQ1.

The results of the user study (concerning both user experience and log analysis) also show that the participants were not interested in filtering images and reviews by keyword to focus on detailed aspects of items. We interpret this finding as evidence that having filtered information by fine-grained evaluation dimension of experience, the user finds the retrieved data as relevant and does not need to reduce it further. This provides a negative answer to research question RQ2. Finally, the observation that some people perceived the user interface of FILTER-WITH-IMG as moderately cluttered or confusing confirms the suggestion to simplify this user interface by reducing the filters based on keywords and aspects.

8 CONCLUSIONS

We presented a service-based model for the justification of recommendation results that makes it possible to filter multimodal information about items by keywords and by fine-grained evaluation dimensions describing consumer experience during the stages of interaction with items. For the filtering of images, we employ object recognition techniques. As a testbed for our model, we chose the home-booking domain and we carried out a user study involving 50 participants. The results show that empowering the user to filter data based on fine-grained evaluation dimensions of experience enhances the comparison activity during item selection. These results encourage the introduction of service-based multimodal information filtering in product and service catalogs and suggest exploiting filters based on such dimensions to empower the user to steer the presentation of data.

The main limitation is the small sample of users involved in the experiment. We plan a larger one to retrieve more information about users' perceptions of the service-based justification models. Moreover, as the user study has revealed the need to simplify the information filtering functions offered by the FILTER-WITH-IMG justification model, we plan to personalize the suggestion of filters to the user's interests. Finally, we plan to test our models on other domains such as the e-commerce one to assess the applicability of our approach to heterogeneous types of items. In this respect, it is worth mentioning that the service modeling research has produced specifications that can be adapted to the selected domain; e.g., service blueprints for the online retailer platform [15], and for food and beverage service systems [30].

ACKNOWLEDGMENTS

This work has been funded by the University of Torino.

REFERENCES

[1] Airbnb. 2022. Airbnb. https://airbnb.com.
[2] Amazon.com. 2022. Amazon.com: online shopping for electronics, apparel, clothing, etc. http://www.amazon.com.
[3] Mary Jo Bitner, Amy L. Ostrom, and Felicia N. Morgan. 2008. Service Blueprinting: A Practical Technique for Service Innovation. *California Management Review* 50, 3 (2008), 66–94. https://doi.org/10.2307/41166446
[4] Booking.com. 2022. Booking.com. https://www.booking.com.
[5] Owen Chambers, Robin Cohen, Maura R. Grossman, and Queenie Chen. 2022. Creating a User Model to Support User-Specific Explanations of AI Systems. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (Barcelona, Spain) *(UMAP '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 163–166. https://doi.org/10.1145/3511047.3537678

[6] Li Chen, Guanliang Chen, and Feng Wang. 2015. Recommender systems based on user reviews: the state of the art. *User Modeling and User-Adapted Interaction* 25, 2 (2015), 99–154. https://doi.org/10.1007/s11257-015-9155-5

[7] Li Chen and Feng Wang. 2017. Explaining recommendations based on feature sentiments in product reviews. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) *(IUI '17)*. Association for Computing Machinery, New York, NY, USA, 17–28. https://doi.org/10.1145/3025171.3025173

[8] Li Chen, Feng Wang, Luole Qi, and Fengfeng Liang. 2014. Experiment on sentiment embedded comparison interface. *Knowledge-Based Systems* 64 (2014), 44–58. https://doi.org/10.1016/j.knosys.2014.03.020

[9] Henriette S. M. Cramer, Vanessa Evers, Satyan Ramlal, Maarten van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob J. Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction* 18, 5 (2008), 455–496. https://doi.org/10.1007/s11257-008-9051-3

[10] Yashar Deldjoo, Fatemeh Nazary, Arnau Ramisa, Julian Mcauley, Giovanni Pellegrini, Alejandro Bellogin, and Tommaso Di Noia. 2022. A Review of Modern Fashion Recommender Systems. https://doi.org/10.48550/ARXIV.2202.02757

[11] Tommaso Di Noia, Nava Tintarev, Panagiota Fatourou, and Markus Schedl. 2022. Recommender Systems under European AI Regulations. *Commun. ACM* 65, 4 (mar 2022), 69–73. https://doi.org/10.1145/3512728

[12] Cecilia Di Sciascio, Peter Brusilovsky, Christoph Trattner, and Eduardo Veas. 2019. A Roadmap to User-Controllable Social Exploratory Search. *ACM Transaction on Interactive Intelligent Systems* 10, 1, Article 8 (aug 2019), 38 pages. https://doi.org/10.1145/3241382

[13] Ayoub El Majjodi, Alain D. Starke, and Christoph Trattner. 2022. Nudging Towards Health? Examining the Merits of Nutrition Labels and Personalization in a Recipe Recommender System. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (Barcelona, Spain) *(UMAP '22)*. Association for Computing Machinery, New York, NY, USA, 48–56. https://doi.org/10.1145/3503252.3531312

[14] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. 2014. How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies* 72, 4 (2014), 367–382. https://doi.org/10.1016/j.ijhcs.2013.12.007

[15] Sarah Gibbons. 2017. Service Blueprints: Definition. https://www.nngroup.com/articles/service-blueprints-definition/.

[16] Marti A. Hearst. 2006. Design recommendations for hierarchical faceted search interfaces. In *Proceedings of SIGIR 2006, Workshop on Faceted Search*. 26–30.

[17] María Hernández-Rubio, Iván Cantador, and Alejandro Bellogín. 2019. A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews. *User Modeling and User-Adapted Interaction* 29, 2 (2019), 381–441. https://doi.org/10.1007/s11257-018-9214-9

[18] Dietmar Jannach, Michael Jugovac, and Ingrid Nunes. 2019. Explanations and User Control in Recommender Systems. In *Proceedings of the 23rd International Workshop on Personalization and Recommendation on the Web and Beyond* (Hof, Germany) *(ABIS '19)*. Association for Computing Machinery, New York, NY, USA, 31. https://doi.org/10.1145/3345002.3349293

[19] Glenn Jocher. 2022. YOLOv5. https://github.com/ultralytics/yolov5.

[20] Yoshiyuki Kawano, Takanori Sato, Takuma Maruyama, and Keiji Yanai. 2013. [Demo paper] mirurecipe: A mobile cooking recipe recommendation system with food ingredient recognition. In *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. 1–2. https://doi.org/10.1109/ICMEW.2013.6618222

[21] Risa Kitamura and Takayuki Itoh. 2018. Tourist Spot Recommendation Applying Generic Object Recognition with Travel Photos. In *2018 22nd International Conference Information Visualisation (IV)*. 1–5. https://doi.org/10.1109/iV.2018.00011

[22] Kirill Kobyshev, Nikita Voinov, and Igor Nikiforov. 2021. Hybrid image recommendation algorithm combining content and collaborative filtering approaches. *Procedia Computer Science* 193 (2021), 200–209. https://doi.org/10.1016/j.procs.2021.10.020 10th International Young Scientists Conference in Computational Science, YSC2021, 28 June – 2 July, 2021.

[23] Sébastien Lallé, Cristina Conati, and Giuseppe Carenini. 2017. Impact of Individual Differences on User Experience with a Real-World Visualization Interface for Public Engagement. In *Proceedings of the 25th Conf. on User Modeling, Adaptation and Personalization* (Bratislava, Slovakia) *(UMAP '17)*. ACM, New York, NY, USA, 369–370.

[24] James R. Lewis and Jeff Sauro. 2009. The Factor Structure of the System Usability Scale. In *Human Centered Design*, Masaaki Kurosu (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 94–103.

[25] Xin Liu, Jiancheng Li, Jiaqi Wang, and Ziwei Liu. 2021. MMFashion: An Open-Source Toolbox for Visual Fashion Analysis. In *Proceedings of the 29th ACM International Conference on Multimedia* (Virtual Event, China) *(MM '21)*. Association for Computing Machinery, New York, NY, USA, 3755–3758. https://doi.org/10.1145/3474085.3478327

[26] Noemi Mauro, Zhongli Filippo Hu, and Liliana Ardissono. 2022. Service-Aware Personalized Item Recommendation. *IEEE Access* 10 (2022), 26715–26729. https://doi.org/10.1109/ACCESS.2022.3157442

[27] Noemi Mauro, Hu Zhongli Filippo, and Liliana Ardissono. 2022. Justification of recommender systems results: a service-based approach. *User Modeling and User-Adapted Interaction* (2022). https://doi.org/10.1007/s11257-022-09345-8

[28] Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. 2020. 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* (Genoa, Italy) *(UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 173–182. https://doi.org/10.1145/3340631.3394844

[29] Cataldo Musto, Marco de Gemmis, Pasquale Lops, and Giovanni Semeraro. 2021. Generating post hoc review-based natural language justifications for recommender systems. *User-Modeling and User-Adapted Interaction* 31 (2021), 629–673. https://doi.org/10.1007/s11257-020-09270-8

[30] Ki Woong Nam, Bo Young Kim, and Bruce W Carnie. 2018. Service Open Innovation; Design Elements for the Food and Beverage Service Business. *Journal of Open Innovation: Technology, Market, and Complexity* 4, 4 (2018). https://doi.org/10.3390/joitmc4040053

[31] Jianmo Ni, Jiacheng Li, and Julian McAuley. 2019. 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)*. Association for Computational Linguistics, Hong Kong, China, 188–197. https://doi.org/10.18653/v1/D19-1018

[32] Ingrid Nunes and Dietmar Jannach. 2017. A Systematic Review and Taxonomy of Explanations in Decision Support and Recommender Systems. *User Modeling and User-Adapted Interaction* 27, 3–5 (Dec. 2017), 393–444. https://doi.org/10.1007/s11257-017-9195-0

[33] Alexis Papadimitriou, Panagiotis Symeonidis, and Yannis Manolopoulos. 2012. A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data mining and knowledge discovery* 24, 3 (2012), 555–583. https://doi.org/10.1007/s10618-011-0215-0

[34] Pearl Pu and Li Chen. 2007. Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems* 20, 6 (2007), 542 – 556. https://doi.org/10.1016/j.knosys.2007.04.004

[35] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, Illinois, USA) *(RecSys '11)*. Association for Computing Machinery, New York, NY, USA, 157–164. https://doi.org/10.1145/2043932.2043962

[36] Jiayin Qi, Zhenping Zhang, Seongmin Jeon, and Yanquan Zhou. 2016. Mining customer requirements from online reviews: A product improvement perspective. *Information & Management* 53, 8 (2016), 951 – 963. https://doi.org/10.1016/j.im.2016.06.002

[37] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 779–788. https://doi.org/10.1109/CVPR.2016.91

[38] Ben Shneiderman. 1992. Tree Visualization with Tree-Maps: 2-d Space-Filling Approach. *ACM Transactions on Graphics* 11, 1 (Jan. 1992), 92–99. https://doi.org/10.1145/102377.115768

[39] Marc Stickdorn, Jakob Schneider, and Kate Andrews. 2011. *This is service design thinking: Basics, tools, cases.* Wiley.

[40] Maxwell Szymanski, Martijn Millecamp, and Katrien Verbert. 2021. Visual, Textual or Hybrid: The Effect of User Expertise on Different Explanations. In *26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) *(IUI '21)*. Association for Computing Machinery, New York, NY, USA, 109–119. https://doi.org/10.1145/3397481.3450662

[41] Nava Tintarev and Judith Masthoff. 2012. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 399–439.

[42] Nava Tintarev and Judith Masthoff. 2022. *Beyond Explaining Single Item Recommendations.* Springer US, New York, NY, 711–756. https://doi.org/10.1007/978-1-0716-2197-4_19

[43] Xinru Wang and Ming Yin. 2021. Are Explanations Helpful? A Comparative Study of the Effects of Explanations in AI-Assisted Decision-Making. In *26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) *(IUI '21)*. Association for Computing Machinery, New York, NY, USA, 318–328. https://doi.org/10.1145/3397481.3450650

[44] Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. 2010. SUN database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3485–3492. https://doi.org/10.1109/CVPR.2010.5539970

[45] Zalando. 2022. Zalando. https://www.zalando.com.