Implementation and Evaluation of a Collaborative Conversational Recommender in a 3D Virtual World

David Contreras, Maria Salamó, Inmaculada Rodríguez, and Anna Puig

***Abstract*—A 3D Virtual World facilitates users' interaction as they feel immersed and engaged in a shared virtual space. This type of interface may be specially useful when consumers employ home electronics for accessing online personalized services. In a previous research we focused on a Collaborative Conversational Recommender framework, where a synchronous and online 3D interface for multiple consumers integrates with a recommender. In this paper we go further and define a state-based model of user-recommender interaction that allows users move from different states of interaction (i.e., individual and collaborative) among users. Then, it is evaluated with users and compared with an individual approach. Our results demonstrate that the collaborative capacities proposed in the framework improves user experience and significantly increases the performance of the recommendation process, i.e users take less time in achieving the desired service.**

***Index Terms*—Application/Implementation $<$ Entertainment & Services Technology, Human-Computer Interface $<$ Human-Device Interaction, Interactive Technology $<$ Human-Device Interaction.**

I. Introduction

Nowadays, users need to personalize services when they use electronic equipment in their everyday. For example, to personalize what content to watch in a smart TV, what music to listen in an audio device, or what video game to play in a videogame console [1]. Either from entertainment or from home-office devices, users are more than ever searching for products to consume among a large volume of content. The wide range of products and their specific and varied characteristics make it difficult for a user to search for an item. Recommender Systems [2] assist users in this search and provide them with suggestions of items to consume or to buy, taking into account their requirements. Recommender systems based on collaborative filtering [3] techniques have been used to suggest highly rated items for a target user based on the products that similar users have experienced in the past.

However, for high-risk product domains[1], where users are likely to search and buy products for the first time, the recommender cannot establish a meaningful profile for many of its recommendation seekers [5]. To overcome such cold start problems, Conversational Recommenders Systems [6] have been broadly recognized as an effective preference-based search and recommender technology. Conversational Recommender Systems use product's features to help users to navigate through a product space, making alternatively product suggestions and eliciting user feedback [7].

Nevertheless, most of these recommender systems lack of online collaboration to enable the user to be aware of and interact with other users that are simultaneously searching for a product. To address this issue, in a previous research we proposed a Collaborative Conversational Recommender (CCR) that integrates a conversational recommendation process in a 3D collaborative environment [8]. In this framework, users have the possibility of engaging in a joined search of a desired product. This framework consists of two main layers in charge of providing the User Interface and the Recommender Systems, and a third layer which is responsible for communicating both of them [9], [10], [11].

In this paper, based on aforementioned research work and encouraged by the results obtained there with a simulator and with a preliminary evaluation with users, we propose a model that defines users' states and the transitions during user-recommender interaction. This state-based model distinguishes individual and collaborative states that allows a more detailed analysis of the efficiency and efficacy[2] of the CCR algorithm with real users in a 3D Virtual World implementation. Accordingly, we compare an Individual Critiquing [13] algorithm (IC), which has no transitions as the user does not collaborate with anyone, with a collaborative approach (where users' states change over time) that enables interaction among users. The evaluation shows good results concerning usability as well as a significant improvement on efficiency and efficacy of our framework with respect to a non-collaborative one.

II. Related Work

In this section, we analyze the main approaches that integrate recommenders within virtual environments according different points of view: application domains, recommenders visualization and interaction platforms, recommendation methods, and recommender collaboration capabilities.

Regarding the application domain, most of the previous studies have been focused on implementing shopping assistants. For example, a solution of a virtual shopping mall on the Internet [14] or the recommendation of virtual objects inside a virtual reality interface [15]. Alternatively, others have focused on recommending locations inside the virtual world [16]. In the cultural domain, a recommender has been used to help users in navigating in 3D spaces [17] (i.e., both museums and galleries). Although the application example of our CCR framework focuses on an e-commerce domain, it is applicable to other domains.

[1] In high-risk product domains (e.g., domains where the products are very expensive), the task of locating a desired choice among a large set of options is indeed becoming intimidating for the average customer [4].

[2] In this work, the efficiency is measured through the number of recommendation cycles to reach a desired product and the efficacy is measured through the Decision Accuracy measure using the same methodology described in [12].

Related to recommenders visualisation and interaction platforms, relatively little studies have focused on exploiting 3D/Virtual Reality interfaces [18]. Bonis et al. in [17] propose a 3D desktop platform based on Second Life[3] protocol. This platform has been widely used by different researchers [16], [19]. In [20] it was proposed the use of the OpenSimulator[4], an open-source 3D virtual world platform that follows Second Life protocols. Other authors used more specific libraries to develop 3D virtual environments, such as Java3D with Virtual Reality Markup Language (VRML) [14], [15]. Our CCR framework uses OpenSimulator platform which is an out-of-the-box solution.

With respect to the recommendation method, the majority of previous studies have used a Collaborative-Filtering (CF) method [16], [14], [20] for generating user recommendations in a 3D virtual environment. CF [3] is based on historical data and does not necessarily imply a direct online interaction among users. Alternatively, the use of a hybrid recommendation method based on both CF [3] and Content-Based[5] [21] filtering has been proposed in [15]. However, this approach does not allow online user collaboration either. Our CCR framework uses a conversational recommender based on eliciting feedback with *critiquing*[6] and allows online collaboration among users. By online collaboration we mean that users may interact together during the search of a desired content.

Finally, regarding collaboration among users, most of the studies proposed thus far have been limited to implement a CF method based on off-line historical data, i.e., the use of aggregated values of other users profiles. Our proposal differs from those approaches because we allow the online collaboration among users in a 3D virtual space. Specifically, our approach is primarily based on collaboration since we aim to tap human nature, in which cooperation and collaboration between users have their roots [22]. Moreover, older social psychological studies applied to marketing and more recent ones in e-commerce conclude that consumers need some kind of decision aid, such as word of mouth or others opinions, to reach their desired product [23]. Therefore, the main advantage of our proposal is that users can interact with each other, as well as collaborate and discuss during the recommendation process.

III. Overview of the CCR Framework

The CCR framework integrates a conversational recommender system within a collaborative 3D virtual environment. The conversational recommender is based on *critiquing*. For example, given a recommended product, the user may perform critiques such as “a cheaper camera”, or “a different manufacturer”. Then, the system provides him with a new product that complies with these critiques. This process can be repeated until the user either finds a product or abandon.

[3]Second Life is a massively online 3D virtual world that permits users to construct, interact, and inhabit their own 3D world.

[4]www.opensimulator.org

[5]Content-based recommendation systems try to recommend items similar to those given user has liked in the past.

[6]Critiquing is a preference-based feedback mechanism where users provide feedback by constraining a feature’s value at the feature-level.

Figure 1 shows the three-layer architecture of our framework: 3D Collaborative Space Client, 3D Collaborative Space Server, and the Collaborative Conversational Recommender.

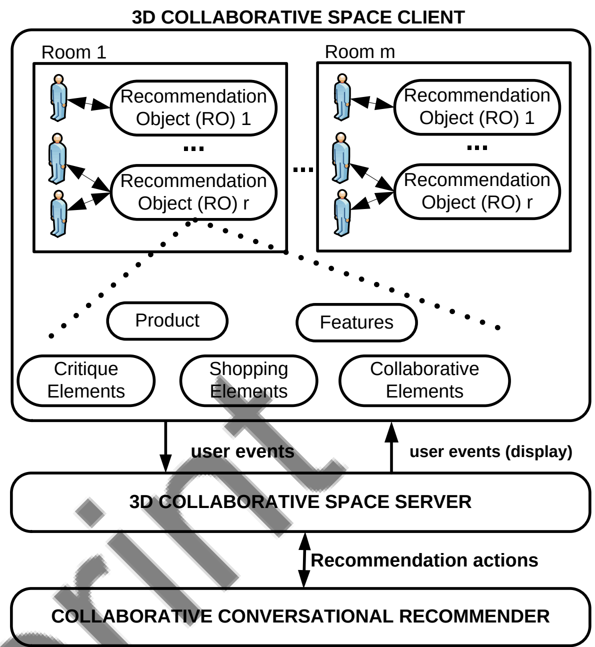


Fig. 1: Conceptual Architecture of Collaborative Conversational Recommendations Framework.

The *3D Collaborative Space Client* is an immersive environment that enables interaction and fosters collaboration between multiple users. This 3D virtual space contains several scenes called *rooms*, each of them representing a store of a particular product. Inside the virtual space each user is represented by an avatar. When users log in the space they can navigate, look for and choose a free Recommendation Object (RO) to start a recommendation session. A RO can be represented by a 3D panel that displays the product and its features, and facilitates users’ interaction, with Critique Elements, Collaborative Elements, and Shopping Elements. Critique Elements allow users to perform critiques, Collaborative Elements are devoted to facilitating collaboration (e.g., start or finish collaboration action), and Shopping Elements let users either start or finish (activation, buying or abandoning) a recommendation process.

By the own nature of the multi-user 3D interfaces users are aware of others’ presence and, as happen in real shops users can see what products are being recommended to others. If users are interested in others’ products, they can collaborate together. In that case, users pass from an individual to a collaborative state.

The middle and the bottom of the Figure 1 show the *3D Collaborative Space Server* (Server layer) and the *Collaborative Conversational Recommender* (Recommender layer), respectively. When a user interacts with an element, it is activated an event in the space client, for example through a Critique Element displayed on a 3D RO, the Server layer receives the critique and sends it to the Recommender layer. Then, the CCR algorithm in the Recommender layer, updates the user preferences with the critique and it selects the next recommendation from the full set of products available in this layer. Next, the Recommender sends to the Server layer the new recommendation to be displayed on the RO of the user (space client layer) that performed the last critique. Moreover, the Server layer stores all users' information connected to the virtual environment. Finally, this layer uses a Standard Virtual Module to manage all functionalities of the 3D Virtual World.

In particular, the 3D Collaborative Space Server is implemented using C#, Open and Linden Scripting Language (OSL and LSL, respectively). It uses as engine the OpenSimulator virtual world platform, which stores an inventory of objects for the correct information display of both product and features in the recommendation panel. Moreover, OpenSimulator controls communication between the *3D Collaborative Space Client* and the *Collaborative Conversational Recommender* using a communication protocol based on messages and script modules. The recommender layer is implemented using Java. It includes the CCR algorithm, which extends the Incremental Critiquing (IC) algorithm to incorporate collaborative capacities allowing two user feedbacks: the traditional critiquing feedback and the collaboration among users. The IC is one of the most well-known conversational recommenders that use *critiquing* [13]. More details for each modules and functionalities of the server and the recommender layers of the CCR can be found in [8], [11].

IV. User States in the Recommendation Process

Users change between different states during the collaborative conversational recommendation process. These states and their interactions can be defined in a finite-state-machine (see Figure 2).

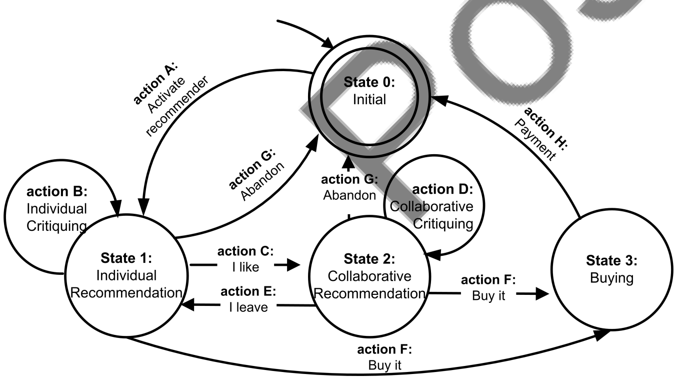


Fig. 2: Model of users' states.

The set of states and their transactions are detailed below.

State 0: Initial. The user is connected to the virtual space and is able to interact with objects and other users. When the user starts a recommendation session (i.e., the *activation* action), by means of a Shopping Element, in an available RO **(action A)**, the recommender creates an individual user model[7], for the user and she changes to **State 1** (Individual Recommendation state).

[7]The user model is the set of critiques that the user has performed during the recommendation process.

State 1: Individual Recommendation. In this state, the recommender returns a product to be shown in the RO. The user can perform four actions at this state.

1) *Individual critiquing* **(action B)**, consists in making unit critiques through the Critique Elements in the RO. For each critique, the recommender provides a new product recommendation and the user model is updated to include the new critique. This action maintains the user in the same state (State 1).
2) *I like* **(action C)**, which is activated by means of a Collaborative Element in a RO of another user (see the top of Figure 1). This action occurs when a user wants to collaborate with another (i.e., when the user likes the recommended product shown in another user's RO). The user who starts the collaboration is called *guest user* and the other user is called *host user*. After this action, both users change to **State 2** (Collaborative Recommendation state).
3) *Buy it* **(action F)** occurs when the user has found a suitable product, and it is activated by means of a Shopping Element in a RO, as depicted in Figure 1. After this action, the user changes to **State 3** (Buying state).
4) *Abandon* **(action G)**, which is activated by means of a Shopping Element in a RO. This action happens when the user explicitly finishes the recommendation process because she does not find a suitable product and she returns to **State 0** (Initial state), where she can start a new recommendation.

State 2: Collaborative Recommendation. A user changes to this state in two situations: when she starts the collaboration (*guest user*) or when someone wants to collaborate with her (*host user*). In a collaborative state the host user is the only one that interacts with the RO. We could say that the *host user* acts as a leader of the whole set of collaborators. Reaching a consensus may be done using communication tools (text or voice chat) in the 3D virtual environment. If guests do not agree, they can decide to leave the collaborative state using the *I leave* Collaborative Element whenever they want. Four actions can be performed at this state:

1) *Collaborative Critiquing* **(action D)** where the host user is in charge of making the agreed critique. As a result of this action host and guest users update their user models and receive a new product recommendation, which is shown to them in the corresponding RO.
2) *I Leave* **(action E)**. At any time, a guest user is free to continue alone the process of searching for a suitable product. To this end, users activate the *I leave* Collaborative Element in the RO of the host user and returns to **State 1** (Individual Recommendation state). Notice that the host user also returns to Individual Recommendation state when there are not guest users.
3) *Buy it* **(action F)**, this action is activated when collaborators have found a suitable product to finish the recommendation process and are moved to **State 3** (Buying state). When host user performs this action, all guest users are returned back to the **State 1** and they

continue with an individual recommendation process.

4) *Abandon* **(action G)**, which is activated by means of a Shopping Element in a RO. This action happens when the host user explicitly finishes the recommendation process because she did not find a suitable product. Then, she returns to **State 0** (Initial state), where she can start a new recommendation. When host user performs this action, all guest users are returned back to the **State 1** and they continue with an individual recommendation process.

State 3: Buying. In this state the user pay for the selected product. After this action (**Action H**), the user changes to **State 0** where at any moment may start a new recommendation process.

V. CCR Implementation

This section first describes a particular implementation of the CCR framework in a 3D Virtual World (VW). Next, it depicts an example walk-through in the 3D VW.

A. Implementation in a 3D Virtual World

The Recommendation Object (RO) in a 3D Collaborative Space Client can be represented by several methods as done by [24]. Our implementation represents the RO through a 3D panel, which consists of several visual and interactive elements as shown in Figure 3:

- (a) is a graphical representation of the current recommended product. Despite the 3D nature of the interface but due to the lack of a database of 3D scanned images, our prototype shows the image of the recommended product. The visualization of 3D real sized products would improve the experience [1].
- (b) are visual affordances representing the features of the current recommended product, with the value of the feature on the right of it,
- (c) displays one (<>, different than) or two (+,−) icons that users touch for critiquing product's features (i.e., the Critique Elements in the Conceptual Architecture). Concretely, button <> is used to change nominal features like *manufacturer* and + or − buttons for critiquing numerical ones, i.e., "I want a cheaper (−) camera or more expensive (+) one",
- (d) are the names of guest users, that are collaborating with the host user,
- (e) are buttons for the collaborative actions (i.e., the *I like* and *I leave* Collaborative Elements described in Section III), and
- (f) are buttons for the shopping actions (i.e., the *Activate*, *Buy it* and *Abandon* Shopping Elements described in Section III).

B. Interaction Walk-through

In this section we present a simple example of collaboration between 3 users and show the users' states they pass through. Users log in the 3D Virtual World (State 0 in Figure 2) and can navigate, look for and choose for a free RO to start a recommendation session by means of the *Activate* button in the panel (see letter (f) in Figure 3).

Fig. 3: Screenshot of the recommendation panel.

Figure 4 summarizes some hypothetical interaction walk-through featuring 3 users (Tester 1, Tester 2 and Tester 3) in a timeline. Figure 5 depicts this interaction using three recommendation panels.

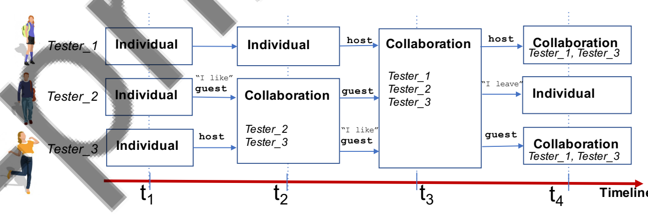


Fig. 4: Timeline of the walk-through featuring 3 users. In the different times, timeline details the type of collaboration for each user.

Figure 5a shows users working individually (see State 1 in Figure 2). Note that users have different products in their own corresponding panels. At this state, users can continue to perform critiques individually and find a desired product, or they can search for another user to initiate a collaboration.

Next, Figure 5b shows that *Tester 2* likes the product shown in *Tester 3*'s panel and decides to start a collaboration by pressing the *I like* button on *Tester 3*'s display panel (see letter e) in Figure 3).

Afterwards, both of them change to Collaborative Recommendation (State 2), in which *Tester 3* is called the host user, u_h, and *Tester 2* is designated the guest user, u_g. Note that the name of the guest user, *Tester 2*, is shown above the collaborative buttons on the panel of *Tester 3*. Additionally, the panel of *Tester 2* is unavailable to other users. Moreover, it depicts the product recommendation that is currently given to *Tester 3*. This Figure 5b shows users are currently using the chat to decide the next critique to perform (see bottom left corner of Figure 5b). Once they agree that they want a cheaper Smartphone, u_h performs the critique and the new recommended product is shown on both panels.

In Figure 5c *Tester 3* notifies *Tester 2* that wants to start a collaboration with another tester, *Tester 1*. In such a situation, *Tester 2* could press the *I leave* button to finish the collabo-

(a) Three users are currently working individually, each one with its own panel (t_1 in Figure 4).

(b) Tester 2 likes the product on Tester 3's panel and presses "I like" button. Tester 3 and Tester 2 are collaborating and they use chat messages for deciding a critique (t_2 in Figure 4).

(c) Tester 3 decides to collaborate with Tester 1, Tester 2 is also guest. All are collaborating and they use chat messages for deciding a new critique (t_3 in Figure 4).

(d) Tester 2 decides to press "I leave" button (t_4 in Figure 4).

Fig. 5: Scenes of the 3D virtual world from individual to collaborative interactions.

ration with Tester 3. But, in this case, she does not press it, meaning she also wants to collaborates with *Tester 1*.

As a consequence when *Tester 3* presses the *I like* button of *Tester 1*, *Tester 2* (so far, the guest of *Tester 3*) automatically becomes the guest of *Tester 1* and is teleported by default to *Tester 1*'s panel, as shown in Figure 5c. At this point the host is *Tester 1*, and *Tester 2* and *Tester 3* are the guests. Additionally, Figure 5c shows the three users talking to each other in order to reach a consensus over the next critique to be applied. All testers have the same product in their panels and every critique performed by the host will be updated in the user model of every guest user. Note that the only remaining active panel for performing critiques or leaving the collaboration is the one that belongs to *Tester 1*.

Next, let's suppose that before performing any critique, *Tester 2*, presses the *I leave* button in the panel of *Tester 1*. Figure 5d depicts the situation produced after performing this action, in which *Tester 2* is back in their panel that is activated for performing actions and the product displayed is the last one that was shown before exiting *Tester 1*'s panel. Also, *Tester 1*'s panel shows that she has only one guest: *Tester 3*.

Finally, *Tester 1* and *Tester 3* continue collaborating and performing critiques, while *Tester 2* continues in an individual critiquing state.

VI. Live-User Evaluation of the CCR Framework

We assess and compare the efficiency of our collaborative framework (CCR) with a non-collaborative one (Incremental Critiquing, IC). Additionally, we aim to evaluate user decision accuracy (efficacy) and gather user opinions on the overall experience (satisfaction).

A. Setup and Methodology

We use a summative test[8], which is adequate for prototypes that already incorporate the major part of the required functionality and focuses on gathering both qualitative and quantitative data [25]. We recruited 20 participants, diverse in features such as age (11 users in the range age of 20-29, 5 users in the range of 30-39, and 4 in the range of 40-49), gender (15 men and 5 women), computer skills and experience in 3D virtual environments (11 users have a medium-high level, 5 users have a medium-low level, and 4 users have a low level).

The test was performed using a SMARTPHONE data set that contains products with two types of features: non-numeric (e.g., manufacturer: Sony, Samsung, etc.) and numeric (e.g., price). The test was conducted by a moderator and an observer. The equipment consisted of 3 computers, the Virtual World (VW) server and two VW clients. Users were requested to perform 5 tasks in pairs, three with a predefined target product and two without a target product.

Task 1, an individual recommendation task (IC algorithm) with a target product; **Task 2**, a collaborative recommendation task (CCR algorithm), where the moderator told both users they could freely decide either to collaborate or not collaborate, and users had similar target products; **Task 3**, a collaborative recommendation task where users did not have similar target products. We designed **Task 2** and **Task 3** with the aim of confirming that collaboration enhances the buying experience even when users have dissimilar targets, corroborating the results obtained with the simulator in [8]. It is important to remark that in **Task 2** and **Task 3** participants did not know how the targets were (similar or dissimilar to the remaining participants). Participants perceived the similarity or dissimilarity just by looking at others' recommendation panels.

Task 4 is a collaborative recommendation task, where again the moderator told both users they could freely decide either to collaborate or not collaborate to buy a desired product; and **Task 5** is an individual recommendation task. We designed **Task 4** and **Task 5** to measure decision accuracy and confirm that it increases when users collaborate. In both, **Task 4** and

[8]Summative usability testing is a Quality Assurance type of test usually performed in medium-near to end stages of development.

Task 5, there was not a target product. That is, users searched for their desired target, which may not exist in the product base. We apply the same methodology described in [12]. Concretely, the decision accuracy was quantitatively measured by the fraction of participants that switched to a different, better option when, once finished the recommendation session, they browsed all alternatives in the data set. A lower switching fraction means that the algorithm allows a higher decision accuracy since most of users are able to find their target choice with it. Since SMARTPHONE data set is too large (1721 products) we selected a subset of 90 products. Specifically, this subset was chosen taking into account the distribution of the $manufacturer$ feature, as it is one of the most representative features in a product. Our selection process is made randomly taking a proportional amount of products for each manufacturer. For example, the Samsung manufacturer (413 products) is close to 24% of original data (1722 products). Thus, we selected randomly 21 Samsung smartphones for this test.

The experiment was designed as a within-subject test. That is, the same group of participants was used for the five tasks. These tasks were rotated in order to avoid carryover effects. We divided participants in two groups of 10 participants and each group performed the tasks (without target 1, 2, 3 and with target 4,5) in a different order. Specifically, first group did the tasks as follows: Task 1, Task 2, Task 3, Task 4, and Task 5. For the second group, the order of the tasks was changed to Task 3, Task 2, Task 1, Task 5, and Task 4. After the test, users did a post-test questionnaire. The next section describes the analysis.

B. Analysis of results

In this section, we show results related to efficiency, efficacy (decision accuracy), and user satisfaction.

1) Recommendation efficiency: Recommendation efficiency is measured through the Average Session Length (ASL)[9] [13]. The bar charts in Figure 6 show recommendation efficiency in terms of both individual (in diagonal lines fill) and collaborative (in dot fill) iteration cycles for each task. The bar chart in Figure 6a shows all tasks with target (Task 1, Task 2, and Task 3) and by the contrary, the bar chart in Figure 6b shows the ones where no target was defined (Task 4 and Task 5).

For tasks with a target (see Figure 6a) ASL decreases when users collaborate (Task 2 and Task 3). A collaborative recommendation with similar targets products (Task 2) obtains the lowest ASL value (16.74 cycles), whereas with dissimilar targets products (Task 3) the ASL value is 21.05 cycles. In contrast, when users work individually their cycles increase up to 29.33. Note that, in Task 2 and Task 3, the collaboration (ASL Collaborative Cycles) is reduced from 11.70 cycles in Task 2 to 6.63 cycles for Task 3. Probably due to users realized that targets were dissimilar and so preferred to continue more time interacting individually.

We apply the ANOVA statistical method to analyze whether the difference between the results of the CCR (Task 2 and Task 3) with respect to the IC (Task 1) are statistically significant. Concretely, we apply the ANOVA in our three algorithms, $k = 3$, (IC, CCR with similar target, and CCR with dissimilar target) with $k - 1 = 2$ degrees of freedom. The ANOVA results show that the differences are significant among the algorithms, we obtained a $p\text{-}value$ of 3.236e^{-6}, that is lower than critical value, $\alpha = 0.05$. Additionally, we apply the multiple significance Bonferroni test [26] which denoted that the efficiencies of CCR in Task 2 and Task 3 are significantly better than IC in Task 1, separately. We obtained a $p\text{-}value$ of 2.4e^{-6} between IC (Task 1) and CCR in Task 2 and 8.1e^{-4} between IC (Task 1) and CCR in Task 3.

In Figure 6b we depict the results of Task 4 and Task 5. In these tasks ASL also decreases from 32.31 cycles for users that work individually (IC algorithm in Task 5) to 19.45 cycles when they collaborate (Task 4). Note that the number of collaborative cycles and the number of individual cycles are more balanced in Task 4 than in Task 2 and Task 3, this is probably due to the fact that likely users started individually with an idea more or less clear of their desired products but some time later they realized this idea was not clear enough and then decided to collaborate, as it is shown in the video recording and in the observer's notes taken during the test. In Figure 6b we only compare two algorithms and therefore we apply the T-TEST measure to evaluate statistically the difference between IC and CCR. T-TEST results show that differences are significant among them, we obtained a $p\text{-}value$ of 2.5e^{-5} between CCR in Task 4 and IC in Task 5, which is lower than critical value, $\alpha = 0.05$.

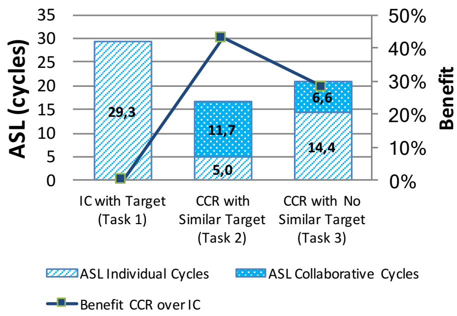
Furthermore, note that charts in Figure 6 depicts a blue line, which shows the benefit[10] in percentage of CCR algorithm over IC in both scenarios, with and without target product. In Figure 6a, the collaborative recommendation achieves a benefit over IC of 42.9% when users have similar target products. This benefit is 28.2% for users with dissimilar target products. In Figure 6b, the collaborative recommendation (Task 4) also achieves a benefit over IC (Task 5) of 39.8%. These results are aligned with the ones obtained using the simulator in the previous work [8].

2) Recommendation efficacy: To evaluate the recommendation efficacy we use the decision accuracy measure. Users were asked to perform two tasks Task 4 and Task 5. Recall that the decision accuracy is measured by the fraction of participants that switched to different better option when, once they finished the recommendation session, they browsed all alternatives in the data set.

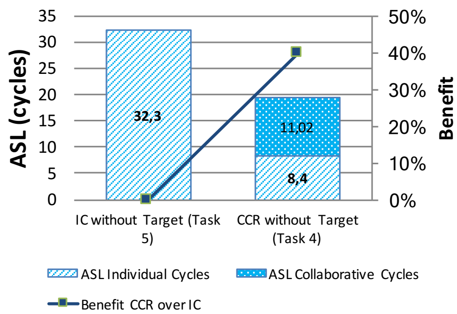
In our experiments we obtained a Decision Accuracy relatively higher in CCR than IC, a 70% and 50% respectively as shown in Figure 7a. This measurement means that in the case of CCR 70% of users maintain the same final recommendation product when they browse the full set of products. The remaining 30% of users switched to a different, better choice when they had the opportunity to browse all products. Furthermore, with the aim of corroborating the decision effectiveness of our collaborative framework we analyzed the users' behavior during the recommendation process in Task

[9] Average Session Length is the number of recommendation cycles to reach a desired product.

[10] We computed the percentage of benefit as $Benefit(x, y) = (1 - \frac{y}{x}) \cdot 100$, where y and x stand for the ASL of the compared algorithm and the base line, respectively.

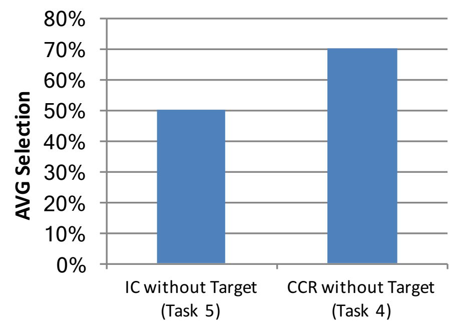


(a) Tasks with Target.

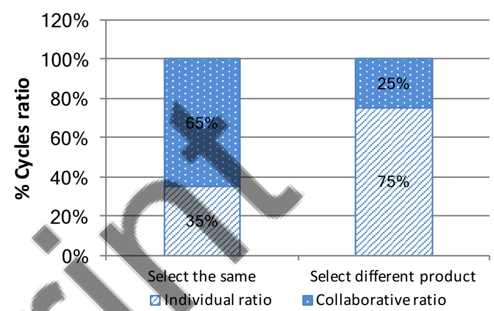


(b) Tasks without Target.

Fig. 6: Recommendation efficiency results in the live-user evaluation.



(a) Decision Accuracy.



(b) Decision Analysis.

Fig. 7: Decision accuracy and decision analysis in the user evaluation.

4. Thus, in Figure 7b we show the ratio of individual (in diagonal lines fill) and collaborative cycles (in dot fill) for both: those users that selected the same product and those that switched to a different product. Specifically, when users bought the same product (i.e., a higher decision accuracy) the majority of times they were collaborating (i.e., a 65% of the cycles) and only a 35% of the cycles users worked individually. By contrast, when users bought a different product, they were more time interacting individually (i.e., a 75% of cycles) than collaborating with other users (i.e., a 25% of cycles). These results show the effectiveness of our approach, which aids users in finding their desired products.

3) User satisfaction: To collect user satisfaction measurements we use a post-test questionnaire that follows the methodology proposed in [27]. Concretely, we focus on analysing the following dimensions (see Figure 8): Learnability, Collaboration, Perceived Accuracy, Perceived Usefulness, and User Satisfaction. First, the *Learnability* dimension analyzes how easy is for the user to learn the use of the recommender. Second, the *Collaboration* shows the perceived user experience when users collaborate with others. Third, the users' *Perceived Accuracy* is the degree to which users feel the recommendations match their interest and preferences. Fourth, users' *Perceived Usefulness* or utility of the recommender is perceived by comparing it with an environment that lacks a recommender. Finally, the *User Satisfaction* refers to the global impression users have about their experience in the framework.

Participants perceive a high level of recommender accuracy

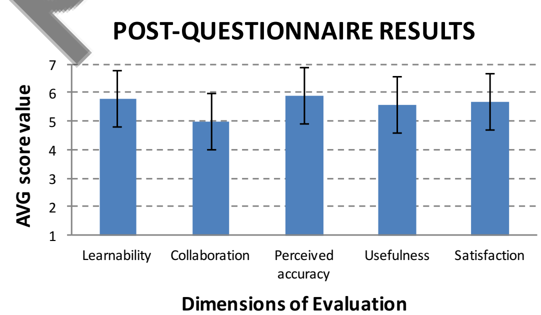


Fig. 8: Average dimensions score value and standard deviation for the post-test questionnaire in a 7-point likert scale.

with an average of 5.9 (see Perceived Accuracy dimension in the graph). It denotes that users perceive the recommender assists them during the recommendation process. Overall, participants found easy to learn how to interact with the recommender and with other users (see the average of 5.78 in the Learnability dimension). Moreover, the User Satisfaction of participants is high with an average satisfaction of 5.68. The good results on these three dimensions encourage us to continue developing our 3D Virtual World.

In general, participants found useful the recommender to find a desired product, with an average of 5.58 in the usefulness dimension in the graph. In our opinion, a bit lower value in the dimension of collaboration (4.98) may reflect that users think that collaboration helps but not fully because they interacted individually during some part of the recommendation process.

VII. Conclusions

This paper, based on a previously proposed Collaborative Conversational Recommender (CCR), defines a state-based model of user-recommender interaction and presents a specific implementation of the CCR framework in a 3D virtual world. Note that a 3D interface may have advantages respect to a 2D one when consumers employ home electronics for accessing online personalised services. It is a shared virtual space that provide consumers with an immersive experience and be aware of how, where and whom to interact with.

We further evaluate the approach with users and compare the collaborative approach with an individual one. Results demonstrate that there is a statistically significant difference between efficiency and efficacy of individual recommendation and the proposed collaborative conversational recommendation. Concretely, the benefit in efficiency of CCR is 39.8% in comparison to a non-collaborative recommender (IC). Moreover, CCR that enables collaborative and individual interaction achieves an efficacy in terms of decision accuracy of 70% whereas the non-collaborative recommender reaches 50%. Additionally, several usability dimensions measured by users' satisfaction questionnaire are also encouraging and promising, being on average near to 6 in a 7-point likert scale.

VIII. About the Authors

David Contreras (dcontrag@maia.ub.es) received his MTech CSE from the University Federico Santa María (Chile). He is currently a PhD student in the University of Barcelona (Spain) and he is professor in the University Arturo Prat (Chile). His research interests include Recommender Systems, Virtual Environments, and User Modeling.

Maria Salamó (maria@maia.ub.es) received both her B.S. in Computer Science (1999) and her Ph.D. (2004) degrees from the Universitat Ramon LLull (Spain). She is associated professor in the University of Barcelona and member of UBICS research institute. Her research covers a broad range of topics within AI including Recommender Systems, User Modeling, CBR, and Natural Language Processsing.

Inmaculada Rodríguez (inma@maia.ub.es) obtained her Ph.D in the University of Alcala (Spain) in 2004. She is associated professor in the University of Barcelona where she is also member of UBICS research institute. Her major research focus lies on the integration of both artificial intelligence and Virtual Worlds technologies, HCI, Serious Games and Gamification.

Anna Puig (anna@maia.ub.es) received her BSc. in Computer Science (1991) and her PhD in Computer Science (1998) from the Polytechnic University of Catalonia. Currently, she is associated professor in the University of Barcelona (Spain). She researches in medical volume visualization as well as the usage of Artificial Intelligence methods in Virtual Worlds, Serious Games, and Gamification.

IX. Acknowledgements

This work has been supported by projects: SGR-623-2014, TIN2015-66863-C2-1-R, and TIN2015-71147-C2-2. We also thanks to UBICS research institute for their support to our research.

References

[1] G. Shapiro, "Consumer electronics association's five technology trends to watch: Exploring new tech that will impact our lives," *IEEE Consumer Electronics Magazine*, vol. 2, no. 1, pp. 32–35, Jan 2013.

[2] J. Konstan and J. Riedl, "Recommender systems: from algorithms to user experience," *UMUAI*, vol. 22, no. 1-2, pp. 101–123, 2012.

[3] Y. Koren and R. Bell, "Advances in collaborative filtering," in *Recommender Systems Handbook*. Springer, 2011, pp. 145–186.

[4] L. Chen and P. Pu, "Critiquing-based recommenders: survey and emerging trends," *UMUAI*, vol. 22, no. 1-2, pp. 125–150, 2012.

[5] F. Cacheda, V. Carneiro, D. Fernández, and V. Formoso, "Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems," *ACM Trans. Web*, vol. 5, no. 1, pp. 2:1–2:33, Feb. 2011.

[6] D. Aha, L. Breslow, and H. Muñoz Avila, "Conversational Case-Based Reasoning," *Applied Intelligence*, vol. 14, pp. 9–32, 2001.

[7] D. Bridge, M. H. Göker, L. McGinty, and B. Smyth, "Case-based recommender systems," *The Knowledge Engineering Review*, vol. 20, no. 3, pp. 315–320, 2005.

[8] D. Contreras, M. Salamó, I. Rodríguez, and A. Puig, "Integrating a Conversational Recommender System within a 3D Collaborative Environment," in *Proc- of the XV Conf. of Spanish Association for Artificial Intelligence*, 2013, pp. 89–98.

[9] ——, "An approach to improve user experience with conversational recommenders through a 3d virtual environment," in *Proc. of the XV Inter. Conf. on Human Computer Interaction*. ACM, 2014, p. 66.

[10] ——, "Supporting users experience in a 3d ecommerce environment," in *Proc. of the XVI Inter. Conf. on Human Computer Interaction*. ACM, 2015, pp. 45:1–45:4.

[11] ——, "A 3d visual interface for critiquing-based recommenders: Architecture and interaction," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 3, no. 3, pp. 7–15, 2015.

[12] L. Chen and P. Pu, "Evaluating critiquing-based recommender agents." in *Proc. of the Nat. Conf. on Artificial Intelligence*, 2006, pp. 157–162.

[13] J. Reilly, K. McCarthy, L. McGinty, and B. Smyth, "Incremental critiquing," *Know.-Based Systems*, vol. 18, no. 4-5, pp. 143–151, 2005.

[14] B. Xu and Y. Yu, "A Personalized Assistant in 3D Virtual Shopping Environment," *Second Int. Conf. on Intelligent Human-Machine Systems and Cybernetics*, pp. 266–269, 2010.

[15] X. Hu and K. Wang, "Personalized Recommendation for Virtual Reality," *2010 Int. Conf. on Multimedia Technology*, pp. 1–5, Oct. 2010.

[16] F. Shah, P. Bell, and G. Sukthankar, "A destination recommendation system for virtual worlds." in *The Florida Artificial Intelligence Research Society Conference*. AAAI Press, 2010.

[17] B. Bonis, J. Stamos, S. Vosinakis, I. Andreou, and T. Panayiotopoulos, "A platform for virtual museums with personalized content," *Multimedia tools and applications*, pp. 139–159, 2009.

[18] Y. Yuan, "Changing the world with virtual/augmented reality technologies," *IEEE Consumer Electronics Magazine*, vol. 6, no. 1, pp. 40–41, Jan 2017.

[19] E. Lacic, D. Kowald, D. Parra, M. Kahr, and C. Trattner, "Towards a scalable social recommender engine for online marketplaces: The case of apache solr," in *Proc. of Int. Conf. on WWW*, 2014, pp. 817–822.

[20] G. Guo and M. Elgendi, "A New Recommender System for 3D E-Commerce: An EEG Based Approach," *Journal of Advanced Management Science*, vol. 1, no. 1, 2013.

[21] P. Lops, M. Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender Systems Handbook*. Springer, 2011, pp. 73–105.

[22] R. Sussman and C. Cloninger, *Origins of Altruism and Cooperation.* Springer, 2011.

[23] F. Vachon, "Can online decision aids support non-cognitive web shopping approaches?" *International Journal of Business and Management*, pp. 16–27, 2011.

[24] F. Pittarello, "The time-pillar world," in *Personalized Digital Television*, ser. HCI Series. Springer, 2004, vol. 6, pp. 287–320.

[25] D. Bowman, J. Gabbard, and D. Hix, "A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods," *Presence: Teleoperators and Virtual Env.*, vol. 11, no. 4, pp. 404–424, 2002.

[26] J. M. Bland and D. G. Altman, "Multiple significance tests: the bonferroni method," *Bmj*, vol. 310, no. 6973, p. 170, 1995.

[27] P. Pu, L. Chen, and R. Hu, "A user-centric evaluation framework for recommender systems," in *Proceedings of the Fifth ACM Conference on Recommender Systems*, 2011, pp. 157–164.