CienTec Guide: Application and Online Evaluation of a Context-Based Recommender System in Cultural Heritage

Felipe F. Laskoski[1], **Alfredo Goldman**[1]

[1]Instituto de Matemática e Estatística – Universidade de São Paulo (USP)
05508-090 – São Paulo – SP – Brazil

felipeflaskoski@outlook.com, gold@ime.usp.br

Abstract. *A Recommender System (RS) is best applied in situations where users have to decide to choose among a list of usually many options and visits in cultural heritage sites are an example of that. Visitors may also face problems in finding how to reach their options. This research addresses both problems with a mobile app consisting of a hybrid context-based RS that suggests personalized visiting routes with the goal to maximize user satisfaction and minimize the length of the recommended route. Unlike most published RS papers related to cultural heritage, the system in this research was built for the mobile platform and its effectiveness was evaluated with actual visitors of a museum. The results were consistent in indicating the improved system achieved high user satisfaction, with all the recommender attributes average ratings between 4.3 and 4.7 (in a scale of 1 to 5), and accuracy, with a Mean Average Error (MAE) of 0.69.*

1. Introduction

Situations in which one has to take a decision considering many options, and sometimes with very little information, are not rare in our daily lives. Those situations, defined as Information Overload, are well known to be considered a good context for the application of Recommender Systems (RSs) [Ricci et al. 2015]. One common information overload situation is experienced when a person is visiting a cultural heritage site, such as a museum, since usually there are many attractions to see and a limited time available for the visit. The visitor may also not know much about what is available on display, especially if it is his/her first time in the location. Actually, first-time visitors are very common, often corresponding to the majority of them [Costa et al. 2015].

There are a relevant number of systems presented in the literature with the objective to improve museums visits, using different techniques. However, few of those works propose a RS applied to a mobile device and with use of its user context data generation possibilities, as well as few of them had been tested in an online environment. Most of them were tested with an offline database or an experimental setting with few users.

This paper presents an app with a Context-Based (CB) recommender system that aims to better explore the application of that type of RS for cultural heritages in a mobile platform considering an online experiment. The RS takes user's GPS location, initial preferences, time restrictions and his/her history of ratings given to previous visited items to create a personalized route for the visitor to guide through his most recommended items

This research is part of the INCT of the Future Internet for Smart Cities funded by CNPq proc. 465446/2014-0, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, FAPESP proc. 14/50937-1, and FAPESP proc. 15/24485-9.

in a path that minimizes the total distance needed to visit them and within a time constraint defined by the user. The user's preferences and ratings are processed in a memory-based collaborative filtering (CF) algorithm based on the k-nearest-neighbors (kNN) approach.

With this system we aimed to answer the question: Is it possible to use modern distributed technologies together with recommender systems to offer a satisfactory self-guided experience for visitors in cultural heritage sites? The user satisfaction with the system and its accuracy were both evaluated in an online experiment in the Parque Cien-Tec, an open-air museum in São Paulo, with 76 participants. The app was tested several days with actual museum visitors and after obtaining a ratings database with a reasonable size, it was used in an offline evaluation with several other implementations of the RS so that a better combination of parameters could be found. The implementation with the best results was deployed to the mobile app and tested with 31 new museum visitors.

The rest of the paper is organized as follows. In Section 2, published work related to RSs applied to cultural heritage are presented. Section 3 covers some of the main challenges that RSs have to consider. In Section 4, the proposed RS and its application is explained in details. Then, in Section 5, the experiment and its results are presented.

2. Related Work

In this section we present some of the papers that applied RSs in locations with a relatively large number of items. A RS that deals with the problem of lack of semantic relationship among items was proposed in [Benouaret and Lenne 2015]. A semantic model was built to produce a knowledge base which is consumed by its Content-Based Filtering (CBF) method. The final recommendation rating is composed of the weighted sum of the output values from the CBF recommended and two other, CF and demographic system. Their weights vary according to the number of items evaluated by the active user. Then a post-filtering step combines the result with some contextual information. No evaluation was applied to that system.

The ART recommender [van Hage et al. 2010][Wang et al. 2008] was applied to the Rijksmuseum museum. The RS, based on the CBF method, used a knowledge base with information from the museum and the user preferences, location, a graph representing the museum rooms and the items' locations in order to produce its recommendations.

In [Wang et al. 2008], the work evaluates in an experiment with 48 participants which of the semantic relationships among the items are the most useful to be used in the CBF method. The accuracy levels of the recommendations that were made based on each relationship were measured and compared among one another. The accuracy metric defined was obtained by the ratio of the number of items considered relevant to the user and the total of recommended items the user rated. The relationship among different items which were created by the same author yielded the best result in terms of accuracy.

Table 1 compares the main features of five different RS papers with our work. All of them either conducted a controlled experiment to evaluate their proposed RSs or did not evaluate them in any way. In contrast, our work was evaluated with an online experiment in a production environment with the participation of a higher number of users, implying in more reliable results. Also the RS methods, CF and CB, do not require information about items' attributes or its relationships in order to start working properly, implying in less start-up efforts to apply the system on different locations/items, when compared to

Table 1. Main features in this and other cultural heritage RS papers compared

System	Recommender	Evaluation Type	Num. Users
CienTec Guide	CB/CF kNN	Online	76
CIDOC-CRM based	CF kNN/CBF	None	N/A
ART Recommender	KB/CBF	Controlled Experiment	48
SR for groups [Rossi et al. 2016]	CF MF	Controlled Experiment	29
PGR [Huang et al. 2012]	CF/CBF	Controlled Experiment	72
PLN [Mathias et al. 2014]	CBF	Off-line Simulation	N/A

the CBF or Knowledge-Based (KB) methods used by most of the other papers.

3. Challenges Involved

A RS applied in a cultural heritage site has several other challenges to deal with besides the information overload problem, some of the most important which are presented below.

User's restrictions and needs - Distance/Time: Can limit the number of items a user is willing to experience; Check not recommended items: Even if the RS recommends a list of items to be seen, the user could decide to visit other items not included in the list; Visits in group: Individuals may have very different preferences, what would make it difficult to make recommendations to be followed by a group. But [Rossi et al. 2016] tackles this issue by allowing all group members to add their own ratings.

Location's restrictions and characteristics - Items hours: Some items may not be available all the time. They could be visited at a specific day of the week or with specific starting hours, like presentations and other events; Fixed route: Certain places or a subset of items may have a fixed order of visitation, which reduces the need for a RS since there are not so many choices to be made by the user but to follow the given route; Temporary items: Items may be available just once for a specific period, then any effort to extract their' features will have a more limited value and ratings given may also not be useful anymore; Synergy among items: The user's satisfaction on a particular item may be heavily impacted by items seen previously. There are proposals to deal with that [Pavlidis 2019b][Rossi et al. 2016], although not trivial and with additional effort to gather relationship information.

General RS Challenges - Cold start: It is the difficulty a RS may have to obtain good coverage or accuracy for recommendations of new users/items [Ricci et al. 2010] [Pavlidis 2019a] and is related to the lack of information a RS may have over a set of users/items to provide a recommendation; Large sparse matrices: Caused by the number of possible relationships among users and items, which usually are mostly empty [Ricci et al. 2010]. It may result in high memory or processing consumption and reduced accuracy for items with few user ratings [Fernández 2018]; Processing Performance: It can be unfeasible to develop RSs that run on devices with slower processors, or to retrain the system frequently, like, for instance, every time after receiving new feedback.

4. The Recommender System: *CienTec Guide* App

The proposed RS is a hybrid of context RS with a Collaborative Filtering method. It implements two types of context RSs [Ricci et al. 2015] [Aggarwal 2016]: Context modeling and post-filtering. The CF method was adapted from an implementation of the java

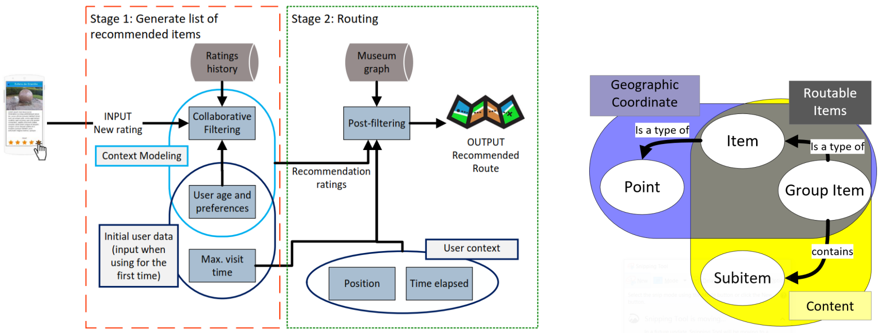

Figure 1. RS's main components (left) and the element types relationships (right)

library[1] [Guo et al. 2015]. It was configured to use a user-based approach, retrieving the 4 most similar users to the current one with the k-nearest-neighbors algorithm. The users were compared with one another through the Pearson's Correlation Coefficient (PCC).

The RS has two main stages (Fig. 1): Firstly, it generates a list of recommended items based on the user's preferences, using CF with contextual modeling. Then it will apply a post-filtering algorithm to define an efficient route for the user to visit the most recommended items given his current position and within an informed time limit.

The first stage starts when a user adds a new rating and ends with the calculated recommended ratings, which indicate the predicted user satisfaction for all the items available in the database. For that, the RS consumes the ratings history together with the new rating just given and information requested when the user starts using the app, which are the context information (*user age* and *user preferences*). Those are transformed in attributes to input into the CF algorithm. The user preferences are represented as an interest level, in a scale of 1 to 5 stars, for several categories related to the place of visit, similarly to the preferences acquisition proposed in [Cao et al. 2018]. Here the preferences categories are represented by some sentences that try to better express the concepts related to the items available at the place. They aim to reduce the cold start problem since the users have to rate those categories according to his/her preference before starting the route.

The recommendation ratings are then used in the second stage, the context post-filtering. It checks the *max. visit time* available (defined by the user) in order to limit the number of items in the recommendation route so that it will be possible to visit all of them within the time restriction. That aims to address the time and distance challenges mentioned in Section 3. The visit route is formed by checking the coordinates of the items available. Since there are commonly exhibitions with numerous items close to one another, four categories were defined to better group the items, as shown on Fig. 1. If a single element is geographically apart from others, it is called an *item*. If there are more than one, the elements are grouped in a *group item*. Each individual element of a group item is a *subitem*. The Content items (items, group items and subitems) have a time variable representing the average time a visitor usually would spend on it. The *point* represents the geographic coordinates used by the routing algorithm in the second stage.

The second stage starts by adding any content items manually added by the user to the list of recommended items (addressing the third challenge mentioned in Section 3). If there is still time available after subtracting the estimated time needed to visit the items

[1] https://www.librec.net/

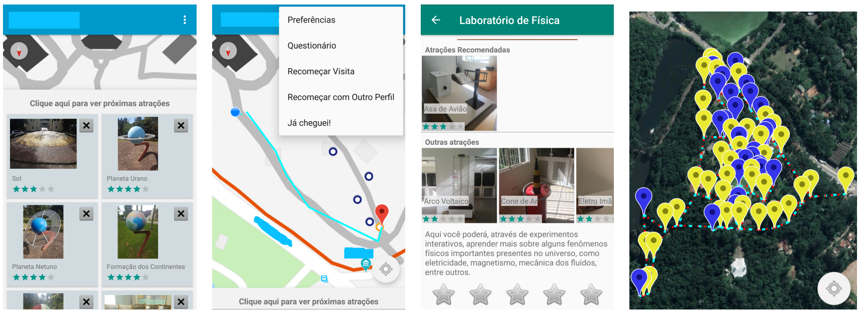

Figure 2. From left to right: app's list of recommended items; recommended route to the next item; group item's details page; coordinates mapped

already added, it verifies how many, and which specifically, of the most recommended content items could be added without surpassing the available time. It will also consider the time needed to move through the route. If it exceeds the time available, the least recommended item is removed and the process is executed again. The application then defines the route using a greedy algorithm considering the user's current location.

The presented RS was developed in an Android app, the *CienTec Guide* [23], with the use of backend features from the Firebase services[4], and is licensed under the MIT Licence. Fig. 2 shows some of the app's main features. It was configured to be used in an open-air museum of science in São Paulo called *Parque CienTec*, that has a wide area and several rooms with fixed expositions with different themes. There were 92 content elements from the museum added to the RS, which are related to 7 themes. So 7 sentences, 1 for each theme, were defined to try to extract the initial preferences of a new user, as for example "What the electricity is and how it works", which refers to one of the museum's main themes (Physics). The rightmost image in Fig. 2 shows all coordinates mapped from the museum. Some of the items had specific working hours, hence those elements are not considered by the recommendation algorithm, but the user can still add it in his route if he wishes so. That addressed the "items hours" challenge cited in Section 3.

5. Results

The app was used at the Parque CienTec museum from 10/Aug/2019 on Saturdays, until 10/Nov/2019. Overall, 76 users data were analyzed, adding up to 1472 ratings. Right after entering the museum, most of the visitors on the days mentioned were invited to use the app, which they installed on their devices. The only explanation given was that they would need to provide some information (the context information) on the app and they would be able to visit the places on their own, that means without the need to follow the fixed order defined by the museum staff. That aimed to address the fixed route problem explained in Sec. 3. The RS was measured in terms of accuracy and user satisfaction. For the accuracy, MAE and Root Mean Square Error (RMSE) metrics were calculated. For the user satisfaction evaluation, four questions were shown to the user, at the end of the visit, asking about his satisfaction level regarding his visit, the museum app, the items recommended to him/her and the recommended route to reach each item.

The experiment was divided in four steps: On the first step, in order to overcome

[2]https://play.google.com/store/apps/details?id=flaskoski.rs.smartmuseum
[3]App's repository: https://github.com/flaskoski/SmartMuseum
[4]https://firebase.google.com/

Table 2. MAE and RMSE results for the proposed RS in the first online evaluation.

Metric	**Groups of Ratings (Chronological Order)**						***All***
	1-20	***31-60***	***51-75***	***61-80***	***81-100***	***101-113***	
MAE	0.447	0.946	0.603	0.632	1.064	1.266	0.799
RMSE	0.922	1.338	0.736	0.864	1.317	1.557	1.132

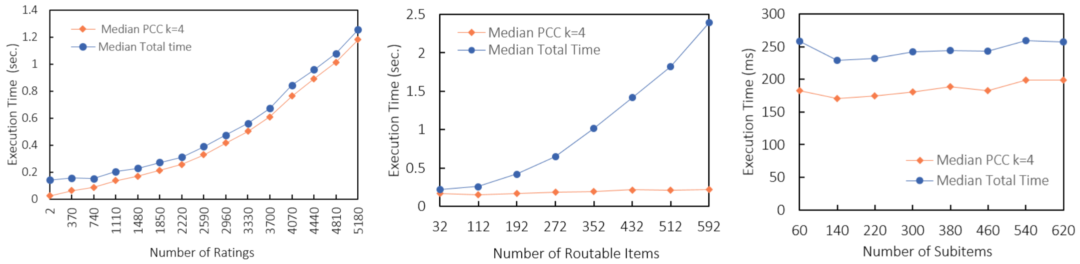

Figure 3. Effect in execution time when increasing the number of routable items (left chart) and subitems (right chart) in the database

the cold start problem related to new items, 14 museum staff members used the app and rated the items they knew, adding up to 592 content items ratings registered with 112 context modeling information ratings (initial user data explained on Sec. 4). The second step consisted on an online evaluation with actual museum visitors, who were recruited as described in the first paragraph. A total of 31 participants used the app and registered a total of 330 ratings (113 for content elements and 217 for context modeling). The RS was configured with a CF kNN user-based method with k=4 and the PCC similarity function. As presented in the Table 2, the system's MAE and RMSE found were 0.799 and 1.132.

Regarding the user satisfaction, the average ratings present a very positive perception of the RS application as whole, especially with the recommendations given to the users and the routes, which obtained an average of 4.4 and 4.2 (up to 5), respectively. The lowest result, with 4.0, was the experience in navigating the app itself.

A performance analysis was also carried out to check execution time when increasing the number of ratings, routable items and subitems. Fig. 3 shows the results. Considering that the application limits by 250 the number of users considered, the max. number of routable items (not considering subitems) as 300 and that the average number of ratings given by the users were equal to 14.5, the execution time remains under 1.5 s.

In step 3 different configurations of the RS could be tested with the users ratings obtained in step 2 by submitting them to an offline evaluation and have their accuracy results compared. The configurations tested were different in three attributes: CF approach, user or item-based, k parameter, 3, 4 or 5 nearest neighbors, and similarity function, PCC or cosine similarity. The first test obtained the MAE and RMSE of each system through 10-fold cross validation using the whole ratings history formed in the previous steps. The second accuracy test simulated the system's usage as how it happened in steps 1 and 2, with the same ratings and order. The configuration CF user-based with k=5 and PCC function achieved the best accuracy on both tests, with MAE/RMSE of 0.753/0.989 on the first test and 0.611/0.802 on the second. Hence that configuration was used on step 4.

On step 4 the app was tested with 31 new visitors, who registered a total of 407 ratings, 160 of which were to content elements. The charts in Fig. 4 show the MAE

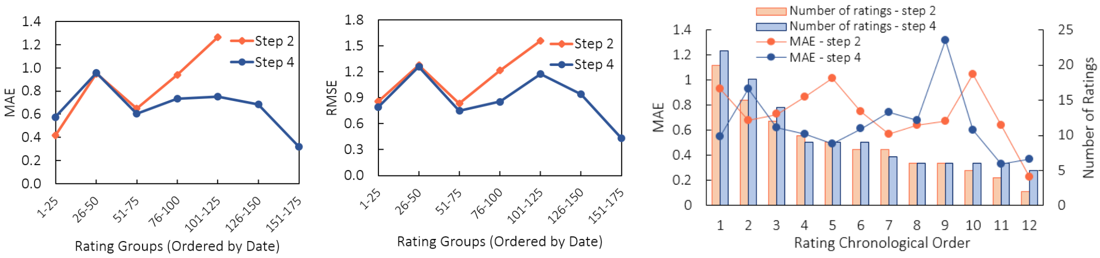

Figure 4. MAE (left) and RMSE (middle) results from step 2 and 4, and MAE of ratings grouped by the position of the rating made by the same user (right).

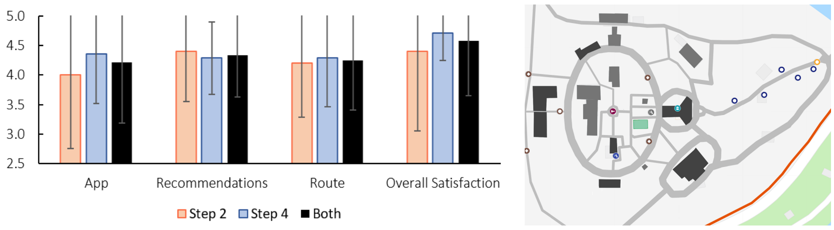

Figure 5. User satisfaction means obtained in step 2, 4 and overall (left) and the new museum map created for the application (right).

and RMSE results obtained, in blue, and compared with those of the step 2. The ratings were sorted chronologically and split in groups of 25. The lines, in general, show an improvement in accuracy. Their global means were of 0.692 and 0.955 on step 4 for MAE and RMSE, respectively, against 0.799 and 1.132 on step 2.

The rightmost chart in Fig. 4 shows the accuracy levels, from step 2 and 4, on chronological order of the ratings. The chart presents, in average, a soft decrease in the error as more user's ratings are added to the database. Moreover, the similarity in accuracy levels of the first content element rating when compared with the others indicate the context modeling information was effective in dealing with the cold start problem for new users, since the error in the user's first ratings was similar to the others.

User satisfaction information was again collected in step 4 and compared with the results of step 2 (Fig. 5). There were larger variations regarding the app and overall satisfaction, with even better average ratings than in step 2, with 4.4 and 4.7, respectively. Besides the improvement in the recommendations, the map of the museum (Google Maps satellite view) on the app used in step 2 was replaced by a new one created for the app (left image on Fig. 5), since the satellite view was not very detailed on the place's paths.

6. Conclusions

There were several different papers proposing alternatives for applying recommender systems in the cultural heritage domain. A subgroup of them applied mobile devices' features in order to get user context information or simply for a better user experience. However, very few of them applied and evaluated them in a production environment.

In this paper a mobile app applying a hybrid RS was developed and evaluated in an online setting to verify if the use of modern distributed technologies with RSs would improve the experience of self-guided visits in cultural heritage sites. Considering the overall user satisfaction averages obtained in the first and second online evaluation (Fig. 5), we concluded that to be true. The research also tested several configurations of the RS and the one with best accuracy was used in the second online evaluation. The results evidenced the context modeling information was effective to remove the cold start for new users, since the MAE for the users' first ratings was very similar to the global MAE.

References

Aggarwal, C. C. (2016). *Recommender systems: the textbook*. Springer, 1st edition.

Benouaret, I. and Lenne, D. (2015). Personalizing the museum experience through context-aware recommendations. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 743–748.

Cao, L., Tao, J., and Bilian, C. (2018). Implementation of personalized scenic spots route recommendation system. pages 1–6.

Costa, A. F., Damico, J. S., Gonçalves, M. d. M., Cazelli, S., and Cruz, S. M. W. d. S. (2015). *Museus de ciência e seus visitantes: pesquisa perfil-opinião 2013*. Fundação Oswaldo Cruz / Casa de Oswaldo Cruz / Museu da Vida, Rio de Janeiro.

Fernández, L. E. M. (2018). Recommendation system for netflix. Technical report, Faculty of Science Business Analytics at Vrije Universiteit Amsterdam.

Guo, G., Zhang, J., Sun, Z., and Yorke-Smith, N. (2015). Librec: a java library for recommender systems. In *UMAP Workshops*.

Huang, Y.-M., Liu, C.-H., and Lee, C.-Y. (2012). Designing a personalized guide recommendation system to mitigate information overload in museum learning. *Educational Technology and Society*, 15:150–166.

Mathias, M., Moussa, A., Zhou, F., Torres-Moreno, J. M., Poli, M.-S., Josselin, D., El-Bèze, M., Linhares, A. C., and Rigat, F. (2014). Optimisation using natural language processing: personalized tour recommendation for museums. In *2014 Federated Conference on Computer Science and Information Systems*, pages 439–446.

Pavlidis, G. (2019a). Recommender systems, cultural heritage applications, and the way forward. *Journal of Cultural Heritage*, 35:183 – 196. Modern and Contemporary Art.

Pavlidis, G. (2019b). Towards a novel user satisfaction modelling for museum visit recommender systems. In Duguleană, M., Carrozzino, M., Gams, M., and Tanea, I., editors, *VR Technologies in Cultural Heritage*, pages 60–75.

Ricci, F., Rokach, L., and Shapira, B. (2015). *Recommender systems handbook*. Springer Publishing Company, Incorporated, 2nd edition.

Ricci, F., Rokach, L., Shapira, B., and Kantor, P. B. (2010). *Recommender systems handbook*. Springer-Verlag, Berlin, Heidelberg, 1st edition.

Rossi, S., Barile, F., Galdi, C., and Russo, L. (2016). Artworks sequences recommendations for groups in museums. In *2016 12th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, pages 455–462.

van Hage, W. R., Stash, N., Wang, Y., and Aroyo, L. (2010). Finding your way through the rijksmuseum with an adaptive mobile museum guide. In *Proceedings of the 7th International Conference on The Semantic Web: Research and Applications*.

Wang, Y., Stash, N., Aroyo, L., Gorgels, P., Rutledge, L., and Schreiber, G. (2008). Recommendations based on semantically enriched museum collections. volume 6, pages 283 – 290. Semantic Web Challenge 2006/2007.