Visitor behavior analysis for an ancient Greek technology exhibition

Dimitrios Kosmopoulos and Kali Tzortzi

University of Patras, Rio GR 26500, Greece
{dkosmo,ktzortzi}@upatras.gr

Abstract. The paper reports the findings from research aimed at the analysis of visitor behavior in the Herakleidon Museum in Athens - Greece, which hosts an ancient Greek technology exhibition. Based on behavioral data gathered by direct observation, we aim to implement services to assist museum curators and enhance the visitors' experience. We describe the data collection, analysis and prediction of the visitors' preferences concerning the exhibits of the museum given their past preferences.

Keywords: museum · visitor experience · recommendation systems.

1 Introduction

Museums are among the most important institutions for the preservation and the dissemination of the global cultural heritage. They are places of education and enjoyment, as well as drivers of tourism and economic development. A key objective in the design of museums and exhibitions is to produce a deeper engagement with the displays and create experiences that are meaningful to visitors. A quantification on measuring the success of addressing this objective would be very useful, however the information collected by curators for measuring an exhibition's impact is mostly qualitative.

The prediction of visitors' interest can be such a quantitative measure and can be associated to recommender systems. These are systems that offer "relevant" suggestions to users of electronic services such as video on demand, song playlists, electronic market places etc. They are based on some popular machine learning algorithms that essentially check the similarity of the users (collaborative filtering), or the similarity among the available items (content-based filtering), or combinations of both.

The recommender systems are often employed as part of the guidance experience in museums or other places of cultural interest. Those are learning spaces that typically offer a large amount of information, including the objects on display, as well as the additional interpretative material.

Visitors with limited time or no prior experience visiting a particular museum, or museums in general, may become overwhelmed by the wealth of information, which in turn may offer an experience which is far below their expectations. Furthermore, curators need to measure in a reliable and objective manner the impact of exhibitions. It is important that they have knowledge of the attraction power of exhibits, and if the supplementary material makes them more attractive; this is an essential part of a constant feedback loop that may lead to exhibition optimization [8].

In this paper we present a part of a larger case study, which aims to develop a guide application, which among others includes a tool for the curator to monitor exhibit popularity and an exhibit - recommendation system. Our contribution includes the collection of a museum-specific dataset. Furthermore, we examine the feasibility of predicting the preferences of visitors using some mainstream methods in order to exploit this knowledge to evaluate the impact of the exhibition. This can also be used to implement electronic recommendation services for the visitors. It is to be installed in the Herakleidon Museum (herakleidon.org) in Athens - Greece, which hosts an exhibition of Ancient Greek technology.

The paper is organized as follows. In section 2 we present the related work on museum visitor behavior tracking and museum recommenders; in section 3 we describe the data collection and visitor analysis that we applied; in section 4 we discuss the experimental results on visitor preferences prediction using collaborative filtering; finally section 6 concludes by identifying further research directions for visitor behaviour analysis.

2 Related work

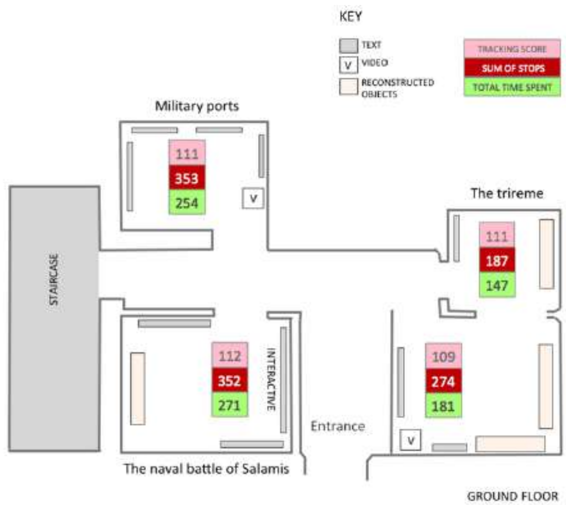
With the goal of understanding how people move, explore and use space and display, we carried out a study of visitor behavior, in particular their patterns of moving and viewing. 'Timing and Tracking' studies, which record traces of visitor movement and activity, were initiated in the early twentieth century, marked by key milestones in the work of Robinson (1928) [13] and Melton (1935) [9]. Since then, they have become a key element in feedback studies on museum performance. These studies commonly aim at improving design (reviewed in [19], [3]), assessing the educational effectiveness of exhibitions (for example, [14]), understanding the characteristics that make exhibits attractive –or unattractive– to visitors [10], identifying styles of moving through gallery space, often described using metaphors, such as ant, butterfly, fish, and grasshopper, as in the case of the study of a natural history museum by Veron and Levasseur (1983) [17], and investigating how far the morphology of movement and encounter in the museum settings is shaped by spatial layout (see, for example, review of syntactic studies of museums in [16]). Timing and tracking and observation studies are also among the standard research methods for understanding interaction patterns and usability issues in the field of HCI [4].

Considering the general recommendation problem, the *filtering methods* can be largely classified into the following: *Demographic filtering - DF* makes the assumption that individuals with certain common personal attributes (sex, age, country, etc.), which are captured during the registration process will also have common interests, e.g., [12]. *Content-based filtering - CBF* makes recommendations based on the semantics of the artifacts. The matching can be based on explicitly asking the users' preferences upon registration to map them to ontology concepts, or by implicit matching via a questionnaire, where representative artifacts have to be rated [18]. It can also be based on previous preferences by the users during the visit. The similarity can be based on (a) the ontology-based semantic attributes (b) the computational analysis of the supplementary text to extract meaning (c) from audio-visual analysis of related digital material. CBF will benefit from the standardization of the metadata for museum artifacts [20]. *Collaborative Filtering - CF* is based on the ratings provided by users of similar profile. The ratings can be acquired explicitly by direct asking or implicitly, acquired (e.g., access, time spent). *Hybrid filtering - HF* uses combinations between the aforementioned filtering techniques to achieve optimal performance to exploit merits of each one of them. For further details on the application of recommender systems in museums the reader can refer to related survey papers such as [8] [1].

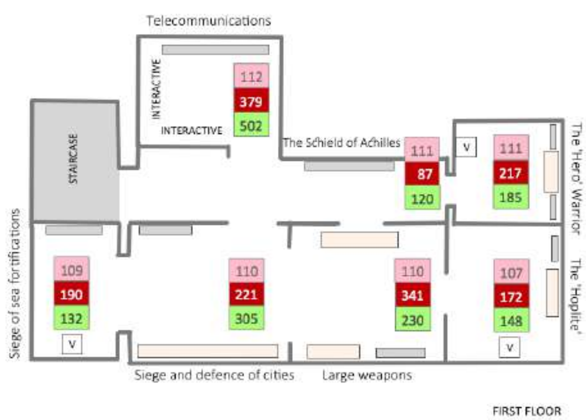
The advent of deep learning has revolutionized several fields, among them the recommendation systems. A taxonomy can be found in [21]. Unfortunately our dataset is currently small, due to manual collection and cannot utilize such models.

Our work is related to [2], where a real-world dataset of visitor pathways was collected at Melbourne Museum (Melbourne, Australia). It suggests that utilising walking and semantic distances between exhibits enables more accurate predictions of a visitor's viewing times of unseen exhibits than using distances derived from observed exhibit viewing times. The time tracking was referring to exhibit areas, while in our approach we measured individual exhibits.

Fig. 1: A view of the "Hero" Warrior of the Mycenaean Period' gallery, showing the combination of reconstructed objects and rich interpretive material.



(a) The ground floor



(b) The first floor

Fig. 2: Basic observation data on the ground and the first floor of the Herakleidon Museum

3 Exhibition features and data collection

Our study focused on the exhibition 'Technology of War in Ancient Greece' (2018-2020), created by the Herakleidon Museum in collaboration with the Association for the Study of Ancient Greek and Byzantine Technology (see Fig. 1). The exhibition was arranged on the two floors of the main building of the Museum (Apostolou Pavlou), and organized in terms of themes, such as 'The trireme', 'The Naval Battle of Salamis' on the ground floor and 'The "Hero" Warrior of the Mycenean Period' and 'Siege of a city" (Figure 1). The display of each room was self-contained and the presentation of objects (mainly reconstructions) was accompanied by rich interpretative material, ranging from informative texts, videos and models to interactive exhibits. Overall, the spatial layout of the display was structured, and choice of route was offered at a localized level.

The field study was carried out in July-August 2019. The visitors were tracked manually during their visit. The approximate time of observation of each of the exhibits was measured as a quantification metric of their interest. The measurements has a resolution if 0.25 minutes.

The visitors were aged between 18 and 65+, and men and women were equally represented in the sample. It should also be noted that the behavioral data gathered by direct observation was combined with questionnaire scores. More details on the demographic data that we collected can be found in [15].

In our study, visitor behavior, was analysed using some established techniques. First, the arrangement of the display was recorded on the building layout as the basis for designing the observation record sheet for mapping visitors' movement and interactions with the displays. Traces of the paths of 112 visitors, who were randomly selected, spreaded across time periods, and had consented to take part in the research, were recorded unobtrusively for their whole visit to the gallery spaces – that is, from the moment they entered the exhibition to the moment of exit. When the visitor stopped to look at a work, read a text, or watch a video, a stopping point was recorded on the plan of the exhibition by the observer. Other symbols were used to clarify where a visitor stopped for longer periods of time. The traces were used to measure a series of space use variables, such as the tracking score of a space, which is the percentage of visitors visiting each space. Furthermore, based on the tracking data, it was possible to obtain a picture of the average rate and distribution of stops made in each space (sum of stops). These are taken to indicate visitors' preference for particular displays and exhibits. The total time they spent in the exhibition (time spent) was also recorded, and we used both to characterize individual visitors and to represent the attraction power of particular displays. The depictions of our visitor measurements are presented in Fig. 2.

The analysis of the behavior data gathered by direct observation enabled us on the one hand to understand whether visitors moved selectively or tended to exhaust all spaces, and to obtain an impression of their preference for particular displays and a range of degrees of interaction with the exhibits; on the other hand, they were used to characterize individual visitors. It was found, for example, that almost a quarter of people observed were individual visitors and half of them were part of a group of two people visiting together. A key feature of the pattern of visitor behavior was the homogeneity of movement: there was little difference in gallery spaces in terms of tracking score, as expected from the relative uniform layout of the museum. In terms of viewing patterns, it was of particular interest that the highest number of stops were found in the last space of the narrative sequence, that is, space 11, on the second level, dedicated to “Telecommunications”. Significantly, timing data confirmed this in that the mean time spent in this gallery is the highest in the sample (twice as high as the average time of stay in each gallery). So stopping patterns are likely to be due to the attraction of exhibits (for instance, in space 11 objects for handling invite the active involvement of the visitor), unaided by their strategic location. Finally, visitors stayed an average of 28 minutes, though many (40%) stayed longer than this (up to a maximum of 80 minutes).

4 Interest prediction

We predicted the interest of the visitors on the basis of their previous observations. To this end used some standard collaborative filtering methods. The methods are briefly described in the following section and the respective results are given in Table 1.

To evaluate the methods we have measured the mean absolute error and the root mean square error of the predictions for the time spent viewing an exhibit. We have done a ten-fold cross validation involving all users and items in each fold. The time was measured in logarithmic scale as some visitors may spend disproportionate amount of time in some exhibits. This is compatible to the approach described in [2].

In our first collaborative filtering approach we used a standard k-NN method for the users (user-based) and the museum exhibits (item-based) for various neighborhood sizes k.

Then we tested factorization methods starting from a basic matrix factorization. It works by decomposing the user-item interaction matrix into the product of two rectangular matrices of lower dimensionality, a low-rank user matrix and a low-rank item matrix. [7]. The amount of factors to use is a method hyperparameter.

We then tested the regularized singular value decomposition of data with missing values, K-means, postprocessing SVD with KNN according to [11]. It extends the set of predictors with the following methods: addition of biases to the regularized SVD, post-processing SVD with kernel ridge regression, using a separate linear model for each item, and using methods similar to the regularized SVD, but with fewer parameters. We finally evaluated the SVD++ method [6]. It merges a latent factor model that captures users and items with a neighborhood model.

As baseline we made predictions using only the items’ scores, i.e., without any personalization. It gave MAE=0.540 and RMSE=0.658. It appears that all

Table 1: Results for recommendation algorithms prediction. MAE stands for mean absolute error, while the RMSE stands for root mean squared error.

Algorithm	**MAE**	**RMSE**
User-KNN, k=1	0.594	0.739
User-KNN, k=5	0.577	0.732
User-KNN, k=10	0.530	0.660
User-KNN, k=20	0.495	0.623
User-KNN, k=30	0.484	0.608
User-KNN, k=40	0.481	0.603
User-KNN, k=50	0.480	0.602
User-KNN, k=60	0.477	0.598
User-KNN, k=70	0.478	0.599
Item-KNN, k=1	0.540	0.680
Item-KNN, k=3	0.500	0.631
Item-KNN, k=5	0.478	0.601
Item-KNN, k=7	0.475	0.597
Item-KNN, k=8	**0.474**	**0.595**
Item-KNN, k=9	0.475	0.595
Item-KNN, k=10	0.479	0.601
Basic matrix factorization, 1 factor	0.488	0.610
Basic matrix factorization, 2 factors	0.501	0.636
Basic matrix factorization, 3 factors	0.507	0.650
Basic matrix factorization, 5 factors	0.559	0.715
Basic matrix factorization, 7 factors	0.587	0.748
Singular Value Decomposition, 1 factor	0.489	0.630
Singular Value Decomposition, 2 factors	0.502	0.640
Singular Value Decomposition, 3 factors	0.527	0.677
Singular Value Decomposition, 5 factors	0.555	0.704
Singular Value Decomposition, 7 factors	0.551	0.690
SVD++, 1 factor	0.503	0.645
SVD++, 2 factors	0.511	0.649
SVD++, 3 factors	0.547	0.696
SVD++, 5 factors	0.580	0.738
SVD++, 7 factors	0.594	0.743
Item - based (baseline)	0.540	0.658

collaborative filtering-based methods can do better, if we select the hyperparameters appropriately.

Surprisingly, the simple item-based k-NN method gave the best results in comparison to the rest and more elaborate of the methods both in terms of mean absolute error and root mean square error. The user-based k-NN follows, but requires significantly more neighbors, which affects efficiency accordingly. Furthermore, regarding the factorization methods, they seem to perform better with a single factor and their performance deteriorates with two or more factors. This is an indication that the personalization factors (factors 2, 3, 4 etc.) are rather weak and lead to overfitting; the available data support models that appear to be closer to the baseline method [5]. We expect this behavior to change significantly in favor of the factorization methods by collecting more data, i.e., more visitors are expected to contribute more knowledge to those latent factors.

5 Conclusions

We have presented our work on the behavior analysis of museum visitors using a custom collected dataset in the Herakleidon Museum, in an Ancient Greek technology exhibition. We have demonstrated the feasibility of predicting the visitors' preferences given past preferences using collaborative filtering. The exhibits' overall popularity may result from summing up the predicted interests of all user groups. We showed that in this application there is some added value compared to the baseline approach that uses only the items' popularity.

In the future we plan to collect more data, manually and automatically using the mobile devices of the visitors. That will enable a better estimate of the exhibition impact and the use of more elaborate deep learning methods for preference predictions. Furthermore, we plan to apply some content-based or hybrid methods given the textual analysis of the descriptions for each exhibit (i.e., via topic modeling).

ACKNOWLEDGMENT This work was supported by the $T1E\Delta K$-00502 MuseLearn project, which is implemented within the framework of "Competitiveness, Entrepreneurship and Innovation" (EPAnEK) Operational Programme 2014-2020, funded by the EU and national funds (www.muselearn.gr).

References

1. Ardissono, L., Kuflik, T., Petrelli, D.: Personalization in cultural heritage: the road travelled and the one ahead. User Modeling and User-Adapted Interaction **22**(1), 73–99 (2012)
2. Bohnert, F., Zukerman, I.: Personalised viewing-time prediction in museums. User Modeling and User-Adapted Interaction **24**(4), 263–314 (2014)
3. Bollo, A.: Analysis of visitor behaviour inside the museum : An empirical study. In: International Conference on Arts and Cultural Management, HEC Montreal (2005)
4. Hornecker, E., C.L.: Human-Computer Interactions in Museums. Morgan Claypool (2019)
5. Jannach, D., Lerche, L., Gedikli, F., Bonnin, G.: What recommenders recommend – an analysis of accuracy, popularity, and sales diversity effects. In: User Modeling, Adaptation, and Personalization, pp. 25–37. Springer Berlin Heidelberg (2013)
6. Koren, Y.: Factorization meets the neighborhood: A multifaceted collaborative filtering model. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p. 426–434. KDD '08, Association for Computing Machinery, New York, NY, USA (2008)
7. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (Aug 2009)

8. Kosmopoulos, D., Styliaras, G.D.: A survey on developing personalized content services in museums. Pervasive Mob. Comput. **47**, 54–77 (2018)
9. Melton, A.: Problems of Installation in Museums of Art, vol. 14. Washington DC (1935)
10. Monti, F., Keene, S.: Museums and Silent Objects: Designing Effective Exhibitions. Ashgate, Farnham (2013)
11. Paterek, A.: Improving regularized singular value decomposition for collaborative filtering (2007)
12. Petrelli, D., Not, E.: User-centred design of flexible hypermedia for a mobile guide: Reflections on the hyperaudio experience. User Modeling and User-Adapted Interaction **15**(3-4), 303–338 (Aug 2005)
13. Robinson, E.S.: The Behavior of the Museum Visitor. American Association of Museums, Washington, DC (1928)
14. Serrell, B.: Paying Attention: Visitors and Museum Exhibitions. American Association of Museums Washington, Washington DC (1998)
15. Styliaras, G.D., Constantinopoulos, C., Panteleris, P., Michel, D., Pantzou, N., Papavasileiou, K., Tzortzi, K., Argyros, A.A., Kosmopoulos, D.: The muselearn platform: Personalized content for museum visitors assisted by vision-based recognition and 3d pose estimation of exhibits. In: Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds.) Artificial Intelligence Applications and Innovations - 16th IFIP WG 12.5 International Conference, AIAI 2020. vol. 583, pp. 439–451. Springer (2020)
16. Tzortzi, K.: Museum Space: Where Architecture Meets Museology. Routledge, London (2015)
17. Veron, E., Levasseur, M.: Ethnographie de l'Exposition. Bibliothque publique d'Information, Centre Georges Pompidou, Paris (1983)
18. Wang, Y., Wang, S., Stash, N., Aroyo, L., Schreiber, G.: Enhancing content-based recommendation with the task model of classification. LNCS, vol. 6317, pp. 431–440. Springer (2010)
19. Yalowitz, S.S., Bronnenkant, K.: Timing and tracking: unlocking visitor behavior **12**(1), 49–64 (2009)
20. Yu, C.H., Hunter, J.: Documenting and sharing comparative analyses of 3d digital museum artifacts through semantic web annotations. J. Comput. Cult. Herit. **6**(4), 18:1–18:20 (Dec 2013)
21. Zhang, S., Yao, L., Sun, A., Tay, Y.: Deep learning based recommender system. ACM Computing Surveys **52**(1), 1–38 (Feb 2019)