Analyzing Museum Visitors' Behavior Patterns

Conference Paper · July 2007

DOI: 10.1007/978-3-540-73078-1_27 · Source: DBLP

Massimo Zancanaro
Fondazione Bruno Kessler

Tsvi Kuflik
University of Haifa

Zvi Boger
Ben-Gurion University of the Negev

Dina Goren Bar
Afeka Tel-Aviv Academic College of Engineering

Analyzing Museum Visitors' Behavior Patterns

Massimo Zancanaro[2], Tsvi Kuflik[1], Zvi Boger[3,4], Dina Goren-Bar[1], and Dan Goldwasser[1]

[1] The University of Haifa, Mount Carmel, Haifa, 31905, Israel
[2] ITC-irst, via Sommarive 18, 38050 Povo, Italy
[3] Ben-Gurion University of the Negev, Beer Sheva,84105, Israel
[4] OPTIMAL – Industrial Neural Systems, Be'er Sheva 84243, Israel
zancana@itc.it;tsvikak@is.haifa.ac.il;zboger@bgu.ac.il;
dgb@univ.haifa.ac.il;dgoldwas@cslx.haifa.ac.il

Abstract. Many studies have investigated personalized information presentation in the context of mobile museum guides. In order to provide such a service, information about museum visitors has to be collected and visitors have to be monitored and modelled in a non-intrusive manner. This can be done by using known museum visiting styles to classify the visiting style of visitors as they start their visit. Past research applied ethnographic observations of the behaviour of visitors and qualitative analysis (mainly site studies and interviews with staff) in several museums to define visiting styles. The current work validates past ethnographic research by applying unsupervised learning approaches to visitors classification. By providing quantitative empirical evidence for a qualitative theory we claim that, from the point of view of assessing the suitability of a qualitative theory in a given scenario, this approach is as valid as a manual annotation of museum visiting styles.

1 Introduction

The museum environment is an attractive arena in which to develop and experiment with ambient intelligence in general and personalized information delivery in particular. Many studies have investigated personalized information presentation in the context of mobile museum guides [1]. Regarding the user characteristics that need to be modelled, most approaches focus on history of interaction and user interests. For example, the GUIDE system presented in [2] adapts web-like presentations by adding information about nearby attractions that might be interesting for the visitor of a city. The HIPPIE system proposes personalized tours in a museum by maintaining a model of user interests and knowledge [3]. The REAL system [4] adapts route descriptions according to the actual user position, the limited technical resources of the device, and the cognitive resources of the user. In the context of the PEACH project [5] a spreading activation technique applied on a domain knowledge-base was implemented to predict the interest in concepts related to those for which the system received explicit feedback from the user.

Knowledge-related features are not, however, the only sources of information that are worth considering for modelling a museum visitor. For example, Petrelli and

C. Conati, K. McCoy, and G. Paliouras (Eds.): UM 2007, LNAI 4511, pp. 238–246, 2007.

Not [6] suggest taking into consideration whether the user is visiting the museum alone or with companions, whether she is a first-time or a recurrent visitor, and so on.

Behavioural traits have also been taken into consideration. Sparacino [7] proposed categorization of user types into three main categories: (i) the greedy visitor who wants to know and see as much as possible; (ii) the selective visitor who spends time on artefacts that represent certain concepts only and neglects the others; and (iii) the busy visitor who prefers strolling through the museum in order to get a general idea of the exhibition without spending much time on any exhibits. Her application employs Bayesian networks to model both the user (interest and style) and the appropriateness of the guide's content (length and order).

The same categorization of user types is also used by Hatala and Wakkary [8] together with an ontology-based model of the interests. In both these papers, the validity of such a scheme is justified through qualitative analysis, mainly site studies and interviews with staff at various museums.

In this paper, we will focus on the classification of the visiting style proposed by the ethno methodologists Veron and Levasseur [9]. Starting from ethnographic observations of the behaviour of a number of visitors in several museums, they argued that visitors' movements may be compared to the behaviour of four "typical" animals, and they proposed using this strategy as a way of classifying the "style" of a visitor. Specifically, they suggests that the *ANT* visitor tends to follow a specific path and spends a lot of time observing almost all the exhibits; the *FISH* visitor most of the time moves around in the centre of the room and usually avoids looking at exhibits' details; the *BUTTERFLY* visitor does not follow a specific path but rather is guided by the physical orientation of the exhibits and stops frequently to look for more information; finally, the *GRASSHOPPER* visitor seems to have a specific preference for some preselected exhibits and spends a lot of time observing them while tending to ignore the others. Of course, it might be expected that a given visitor can change her behaviour during a long visit, and it is also possible that the style is affected by the specific interests.

The first attempt to exploit this classification as part of a user model for a mobile guide was in the HIPS project (see mainly [10]) where a Recurrent Artificial Neural Network (ANN) was trained to recognize the visiting style of a visitor given her interaction history. This model was then employed for selecting and tailoring information to the visitor [11]. Although most of the ideas tested experimentally in HIPS underwent user evaluation, the very idea of the existence of visiting styles was taken for granted relying on the qualitative analysis of the original work.

Chittarro and Ieronutti [12] employed Veron and Levasseur's classification in the context of a tool that visualizes users' behaviours in a virtual environment. Their use of the visiting styles was based on qualitative analysis and, again, they did not evaluate the existence of these classes.

In this paper, we are trying to take a step back; we would like to discuss a methodology for validating empirically Veron and Levasseur's model of visiting style. We used log files of 140 visitors exploring a frescoed room with a multimedia museum guide to provide quantitative-based evidence that museum visitors' behavior may effectively be classified according to Veron and Levasseur's model. We used two unsupervised learning techniques (K-means and Auto-Associative ANN) to cluster the visitors' behaviours. The clustering produced by both techniques may be assumed to characterize Veron and Levasseur's four animals. An agreement analysis conducted on the classifications schemes determined by clustering membership revealed a high level of agreement between the two techniques.

This work is intended to complement Veron and Levasseur's ethnographic study by providing empirical evidence for it as well as to provide information in a principled way for further research on user modelling. Our claim, as discussed in the last section, is that this approach may complement—if not replace—reliability analysis of observation schemes derived from qualitative research such as Veron and Levasseur's.

2 Data Collection and Preparation

In the context of a user study of a multimedia mobile guide [13], 143 regular visitors to Torre Aquila[1] in Trento were invited to test the system. Each visitor was requested to visit Torre Aquila with a multimedia guide (although adaptive guides were experimented in Torre Aquila [14], a non-adaptive version was employed for this study). Among the subjects, 61 were males and 82 females. Their age ranged from 20 to 79 years (mean=47, median=50, std.dev=15.9). All were recruited at the entrance of the museum and received a free ticket to visit the castle as a reward for participating in the data collection.

Out of the 143 visit logs, 140 were used for this study; the rest had various errors that prevented their use. The average visit time was 22 minutes, and average time spent in front an exhibit was 4 minutes with standard deviation of 70 seconds. The system automatically logged the visitors' movements in the space (by means of IR sensors) and all their interaction with the museum visitors' guide.

Since we are interested—at this stage—in analyzing the visitors' behavior rather than in predicting the visiting style from the interaction history, we used measures relating to the entire visit rather than temporal-based indices. The measures used for the analysis were the average time spent at each position, the percentage of exhibits visited, a numerical representation between 0 and 1 of the order of the visit (where 1 means that the visitor spent some time on each exhibit and 0 that she did not stop at any exhibit), and a combined description of visitors' behavior, taking into account interaction and whether or not visitors played fully through complete presentations. Further, four cumulative measures were defined considering the percentage of the visit for which the visitor was: (A) interacting with the guide (i.e., asking for more information), but not reaching the conclusion of the presentations; (B) interacting and reaching conclusions; (C) not interacting and not reaching conclusions; and (D) not interacting but reaching conclusions.

Data pre-processing generated 140 7-dimensional vectors including the average time, visit order and completeness, and the percentage of the visit for which the visitor's behavior was according to each of the four types (checking for each and every

[1] Torre Aquila is a tower at the Buonconsiglio Castle in Trento, Italy where a fresco called "The Cycle of the Months," a masterpiece of the gothic period, is to be found. This fresco, painted in the Fifteenth Century, covers all fours walls of a room in the tower and illustrates the activities of aristocrats and peasants throughout the year. The museum guide used to collect visitor data is one of the many prototypes developed in the PEACH project; for more details see [14].

position whether the visitor interacted with the system or not and whether he/she viewed complete presentations or not and then calculating the ratios).

3 Analysis of Museum Visitors' Behavior

The visit logs representation was used as an input to an auto-associative ANN and to a K-means clustering algorithm, both of which clustered the data into four clusters in order to validate the Veron and Levasseur classification and see if their visitors' types might be identified.

3.1 Unsupervised Learning with Auto-Associative ANN

Artificial neural networks are used to form data-driven models. In order to perform unsupervised learning, an auto-associative ANN (AA-ANN), in which the "targets" are identical to the inputs, was used. If the trained AA-ANN succeeds in replicating the inputs as outputs, it means that the hidden neurons are encoding the essential information "distilled" from the inputs features. In most cases the outputs of the hidden neurons are close to either one or zero [15]. Thus all examples that generate the same hidden neurons output pattern are deemed to belong to the same cluster.

Table 1. ANN clustering results

Cluster	# of cases	Av T	A	B	C	D	Order	Completeness
1	53	1.22	<u>0.45</u>	**1.81**	<u>0.14</u>	<u>0.24</u>	1.00	1.02
2	36	0.93	<u>0.24</u>	<u>0.42</u>	<u>0.49</u>	**2.97**	1.03	1.02
3	15	0.74	0.77	<u>0.18</u>	**5.68**	1.07	0.99	0.96
4	36	0.84	**2.69**	0.72	0.84	<u>0.14</u>	0.97	0.97

As explained above, the data consisted of 140 visit summary examples with seven visit attributes named: AvT (Average time), A, B, C, D, Order (of the visit) and Completeness (percentage of frescos visited). The AA-ANN used was a fully-connected, feed-forward ANN of two hidden neurons and seven output neurons, each having the sigmoidal transfer function, which was presented with the dataset with the seven input variables and the identical values as targets. The input data were preprocessed by subtracting the mean value of each attribute column, and dividing by the standard deviation of each column. These values were further re-scaled to the 0.1-0.9 range to serve as the AA-ANN targets. The training was done by the Guterman-Boger set of algorithms that starts with non-random connection weights and employs proprietary algorithms for avoiding entering, and escaping from, local minima encountered during the training [16, 17]. The "binary" pattern of the hidden neurons was used for clustering [18]. The average attribute values of the examples in each cluster were divided by the average of the attribute values of the full dataset. The results are shown in Table 1.

The ratios of the attributes that are higher than 1.5 are marked bold, and those with ratios smaller than 0.5 are underlined. It can be seen that cluster # 1 has a high ratio of the B variable, cluster # 2 has a high ratio of variable D, cluster # 4 has a high ratio of variable A, and cluster # 3 has a high ratio of attribute C. The attributes Av T, Order and Completeness apparently do not contribute much to the clusters' formation, although it may be that cluster 1 may have a somewhat higher mean Av T.

Hence Cluster 1 seems to correspond to an *ANT* type (long, ordered and interactive, and gets complete presentations) and cluster 3 corresponds to a *FISH* type (short visit, without getting complete presentations). Cluster 2 corresponds to a *GRASSHOPPER* (tends to get more complete presentations than *BUTTERFLY*) and cluster 4 to a *BUTTERFLY* (less ordered and does not get complete presentations). After the clustering, we also used an ANOVA [19] with the clusters identified as a factor and the cumulative indexes outlined above as dependent variables. Significant differences were found at $p < 0.001$ along all the variables except Order and Completeness. Table 2 summarizes the results.

Table 2. One-way ANOVA on the ANN clusters

		Sum of Squares	df	Mean Squares	F	sig.
Avg Time	Between Groups	282415.70	3	94138.55	30.36	0.00
	Within Groups	421751.20	136	3101.11		
	Total	704166.90	139			
A	Between Groups	6.13	3	2.04	90.50	0.00
	Within Groups	3.07	136	0.02		
	Total	9.19	139			
B	Between Groups	12.55	3	4.18	137.53	0.00
	Within Groups	4.14	136	0.03		
	Total	16.69	139			
C	Between Groups	2.76	3	0.92	68.26	0.00
	Within Groups	1.83	136	0.01		
	Total	4.59	139			
D	Between Groups	11.87	3	3.96	188.04	0.00
	Within Groups	2.84	136	0.02		
	Total	14.70	139			
Order	Between Groups	0.07	3	0.02	1.52	0.21
	Within Groups	1.96	136	0.01		
	Total	2.02	139			
Completeness	Between Groups	0.06	3	0.02	1.55	0.20
	Within Groups	1.87	136	0.01		
	Total	1.93	139			

A Bonferroni [20] post-hoc analysis validated the analysis of the ANN results above and showed that:

- Visitors in cluster 1 take more time than visitors in the other clusters when visiting the exhibits; they are less "A" than 4; they are more "B" than all the others; they are less "C" than 3 and less "D" than 2 and 3. Therefore visitors in clusters 1 exhibit the traits of the visitors' style defined as *ANT*;
- Visitors in cluster 2 take less time than 1 but more than 3; they are less "A" than 4; they are less "B" than 1 and 4; they are less "C" than 3 and more "D" than 1,3 and 4; therefore visitors in cluster 2 may be ascribed to the visitors' style defined as *GRASSHOPER* (i.e., closer to an ant than to a fish)

- Visitors in cluster 3 take less time than 1 and 2; they are less A than 4; they are less "B" than 1 and 4; they are more "C" than all the others and more "D" than 1 and 4 but less than 2; therefore they belong to *FISH*;
- Finally, visitors in cluster 4 take less time than 1; they are more "A" than all the others; they are less "B" than 1 but more than 2 and 3; they are less "C" than 3 and less "D" than 2 and 3; they appear to belong to the style *BUTTERFLY* (i.e., closer to a fish than to an ant).

3.2 Unsupervised Learning with K-Means

As an alternative way of clustering the cumulative measures, we employed the K-means algorithm [21].

In order to reduce the number of variables, we ran a Factor analysis (Principal Components Analysis with varimax rotation). The results showed that 81% of the variance can be explained by 4 factors while Order and Completeness show very low correlation with any factor. Table 3 shows the contribution of the cumulative variables on the four principal factors.

Table 3. Component matrix extracted by the PCA

	Component			
	1	2	3	4
Average Time	0.824	0.143	-0.045	-0.190
A	-0.346	-0.709	0.580	0.046
B	0.918	-0.213	-0.188	0.031
C	-0.541	-0.060	-0.622	0.354
D	-0.405	0.830	0.114	-0.266
Order	0.197	0.452	0.487	0.265
Completeness	0.209	0.225	0.110	0.838

We classify the visitors in 4 clusters using K-means analysis starting from the factors.

A one-way ANOVA [19], using the cumulative indexes as dependent variables and the clusters determined by K-means as a factor, showed that for all the variables there are statistical differences except for Order and Completeness (see Table 4).

A post-hoc analysis using the Bonferroni [20] test showed that:

- Visitors in cluster 1 take less time than visitors in cluster 2 and more than visitors in cluster 4; they are less A than 2; less "B" than 2 and 3; less "C" than 4; and more "D" than all the others; therefore they may be ascribed to the style of *GRASSHOPPER* (closer to an ant than to a fish);
- Visitors in cluster 2 take more time than all the others; are less A than 3; are more "B" than 1 and 3; less "C" than 4; and less "D" than 1 and 4; therefore they share many of the traits of *ANT*;
- Visitors in cluster 3 take less time than 2; they are less A than all the others; more "B" than 1 and less than 2; less "C" than 4; and less "D" than 1 and 4; therefore they resembles visitors belonging to the style of *BUTTERFLY* (closer to a fish than to an ant);

- Finally, visitors in cluster 4 take less time than 1 and 2; they are less “A” than 3; they are less “B” than 2; more “C” than all the others; and less “D” than 1 but more than 2 and 3; therefore they can be classified as *FISH*.

Table 4. One-way ANOVA on the K-means clusters

		Sum of Squares	df	Mean Squares	F	sig.
Avg Time	Between Groups	321775.10	3	107258.35	38.15	0.00
	Within Groups	382391.80	136	2811.71		
	Total	704166.90	139			
A	Between Groups	6.51	3	2.17	109.87	0.00
	Within Groups	2.69	136	0.02		
	Total	9.19	139			
B	Between Groups	11.83	3	3.96	112.47	0.00
	Within Groups	4.79	136	0.04		
	Total	16.69	139			
C	Between Groups	3.42	3	1.14	132.81	0.00
	Within Groups	1.17	136	0.01		
	Total	4.59	139			
D	Between Groups	11.70	3	3.90	175.50	0.00
	Within Groups	3.00	136	0.02		
	Total	14.70	139			
Order	Between Groups	0.08	3	0.03	1.96	0.12
	Within Groups	1.94	136	0.01		
	Total	2.02	139			
Completen	Between Groups	0.04	3	0.01	0.92	0.43
	Within Groups	1.90	136	0.01		
	Total	1.93	139			

3.3 Comparison of the Two Approaches

In order to assess to what extent the two clustering algorithms agree on classification of the visitors into the different visitors styles, we used the κ statistics [22] which provides a better estimation of the bare percentage agreement since it takes into account the possibility of chance agreement.

Table 5 shows the confusion matrix. The value of the κ statistics in our case is 0.860 with a standard error of 0.035 (p<0.0001; N=140). According to Landis and Koch’s criteria [23], the agreement is very good (κ > 0.8).

Table 5. Confusion matrix for the classifications based on the ANN and K-means clustering

ANN Labels * Kmean Labels Crosstabulation

Count

		Kmean Labels				
		A	B	F	G	Total
ANN Labels	A	50	1	0	2	53
	B	1	33	2	0	36
	F	0	2	12	1	15
	G	2	0	3	31	36
Total		53	36	17	34	140

4 Discussion, Conclusions and Future Work

Qualitative theories from sociology and other disciplines are often used as a starting point for building computational models of human behavior to be exploited in intelligent systems. Usually, a human expert manually labels a number of examples, and a supervised learning approach is employed to predict in a real situation the behavior of users according to the theory, as modelled (or "learned" by the system). In order to test the objectivity of the observation scheme, reliability analysis is often employed: two or more annotators code a number of sequences, and an agreement analysis is performed by computing Cohen's Kappa (or other similar indexes) and by looking at the confusion matrix. In this paper, we tried to provide quantitative empirical evidence for a qualitative theory. We employed two unsupervised learning techniques for clustering museum visitors' behavior patterns and showed how the clusters obtained from them may be explained in the terms of the theory. We claim that from the point of view of assessing the suitability of a qualitative theory in a given scenario, this approach is as valid as a manual annotation with reliability analysis.

Furthermore, the labels automatically produced may then be used by a supervised learning approach to predict the classes to which visitors belong, as they enter the museum. From a pragmatic point of view, this procedure is cheaper—and less error prone—than manual annotation, especially when a large corpus of data has to be annotated.

Of course, we are not proposing that quantitative approaches may simply substitute qualitative approaches in building computational models of human behavior: the two types of approaches have different strengths and, to some extent, different purposes. Rather, we discussed a technique whose aim is to provide a quantitative validation of a particular qualitative theory.

Future research will focus on predicting visitors' behavior type using information collected during the first period of the visit. We intend to evaluate the correlation between the cumulative data representing the whole visit used for clustering in this work, with partial information available at the beginning of the visit of the same visitors. The results may allow us to use the clustering results as labels for prediction visiting style with partial data.

Additional future research will try to correlate the current clustering results with other notions of visitor's types (such as, for example, Sparacino's [7] and others).

References

1. Baus, J., Kray, C.: A Survey of Mobile Guides. Workshop on Mobile Guides. Mobile Human Computer Interaction '03 (2003)
2. Cheverst, K., Davies, N., Mitchell, K., Friday, A., Efstratiou, C.: Developing a Context-aware Electronic Tourist Guide: Some Issues and Experiences. In: The CHI 2000 Conference on Human factors in Computing Systems, pp. 17–24. The Hague, Netherlands (2000)
3. Oppermann, R., Specht, M.: A Context-Sensitive Nomadic Exhibition Guide. In: proceedings of Handheld and Ubiquitous Computing: Second International Symposium, HUC 2000, Bristol, UK, pp. 127–142 (September 2000)
4. Baus, J., Krüger, A., Wahlster, W.: A Resource-Adaptive Mobile Navigation System. In: Proceedings of the 7th International Conference on Intelligent User Interfaces. San Francisco, CA (2002)

5. Kuflik, T., Callaway, Goren-Bar, Rocchi, C., Stock, O., Zancanaro, M.: Non-Intrusive User Modelling for a Multimedia Museum Visitors Guide System. In: Proceedings of UM 2005, Tenth International Conference on User Modelling, Edinburgh (July 2005)
6. 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: The Journal of Personalization Research 15(3-4), 303–338 (2005) http://dx.doi.org/10.1007/s11257-005-8816-1
7. Sparacino, F.: The Museum Wearable: Real-Time Sensor-Driven Understanding of Visitors Interests for Personalized Visually-Augmented Museum Experiences. Museums and the Web, Boston, Massachusetts (2002)
8. Hatala, M., Wakkary, R.: Ontology-Based User Modeling in an Augmented Audio Reality System for Museums. User Modeling and User-Adapted Interaction. 15, 339–380 (2005)
9. Veron, E., Levasseur, M.: Ethnographie de l'exposition, Paris, Bibliothèque Publique d'Information, Centre Georges Pompidou (1983)
10. Marti, P., Rizzo, A., Petroni, L., Tozzi, G., Diligenti, M.: Adapting the Museum: A Non-intrusive User Modeling Approach. In: Proceedings of User Modeling Conference UM 99 (1999)
11. Not, E., Petrelli, D., Sarini, M., Stock, O., Strapparava, C., Zancanaro, M.: Hypernavigation in the Physical Space: Adapting Presentation to the User and to the Situational Context. The New Review of Hypermedia and Multimedia 4, 33–45 (1998)
12. Chittaro, L., Ieronutti, L.: A Visual Tool for Tracing Users' Behavior in Virtual Environments. In: Proceedings of the Working Conference on Advanced Visual Interfaces, Gallipoli, Italy, pp. 40–47 (2004)
13. Goren-Bar, D., Graziola, I., Pianesi, F., Zancanaro, M., Rocchi, C.: Innovative Approaches for Evaluating Adaptive Mobile Museum Guides. In: Stock, O., Zancanaro, M. (eds.) PEACH: Intelligent Interfaces for Museum Visits. Cognitive Technologies, Springer, Heidelberg (2006)
14. Stock, O., Zancanaro, M.: PEACH: Intelligent Interfaces for Museum Visits. In: Cognitive Technologies Series, Springer, Heidelberg (2006)
15. Boger, Z., Guterman, H.: Knowledge Extraction from Artificial Neural Networks Models. In: Proceedings of the IEEE International Conference on Systems Man and Cybernetics, SMC'97, Orlando, Florida, pp. 3030–3035 (October 1997)
16. Guterman, H.: Application of Principal Component Analysis to the Design of Neural Networks. Neural, Parallel and Scientific Computing 2, 43–54 (1994)
17. Boger, Z.: Who is Afraid of the Big Bad ANN? In: Proceedings of the International Joint Conference on Neural Networks IJCNN'02, Honolulu, pp. 2000–2005 (2002)
18. Boger, Z.: Finding Patients Clusters' Attributes by Auto-associative ANN Modeling. In: Proceedings of the International Joint Conference on Neural Networks IJCNN'03, Portland,OR, pp. 2643–2648 (2003)
19. Rutherford, A.: Introducing Anova and Ancova: A GLM Approach, SAGE (2001)
20. Bonferroni, C.E.: Teoria Statistica Delle Classi e Calcolo delle Probabilità, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze vol. 8, pp. 3–62 (1936)
21. MacQueen, J.: Some Methods for Classification and Analysis of Multivariate Observations. In: Proceedings of the 5th Berkeley Symposium, pp. 281–297 (1967)
22. Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20, 37–46 (1960)
23. Landis, J.R., Koch, G.G.: The Measurement of Observer Agreement for Categorical Data. Biometrics 33, 159–174 (1977)