

# Chapter 14

## Exploring User Behavior in Destination Websites: An Application of Web Mining Techniques



Maria Francesca Cracolici and Furio Urso

**Abstract** The development of the Internet has so strongly affected the way a tourist destination can be promoted that portals now play an important role in defining the marketing strategies of tourist destinations. Hence, investigating website surfing behavior has become crucial to understanding the needs of potential tourists. Using Web Mining techniques, our study explores the information-seeking behavior of those who log on to a website promoting the island of Sicily, a well known tourist destination in the South of Italy. The study explores whether a tourist's country of origin affects their information needs and surfing behavior. Actually, our results do show differences in behavior between users of the site from Western countries and those from China.

**Keywords** Tourist information search · Online browsing behavior · Web usage mining · Destination promotion · Big data

### 14.1 Introduction

In the network economy, the rapid development of the Internet is having an important role in the promotion and marketing of tourist destinations. For an information-intensive industry such as tourism, websites have become a huge information platform.

In recent decades there has been an increase in tourism promotion sites owned by Destination Management Organizations (DMOs) and the like. Online portals have rapidly become crucial for the marketing strategies of these organizations (Wang and Fesenmaier 2006) and are designed to act as the interface between potential visitors and the destination promoted, and to allow users to compare holiday packages (Kim et al. 2003).

---

M. F. Cracolici · F. Urso (✉)

Department of Economics, Business and Statistics, University of Palermo, Palermo, Italy  
e-mail: [mariafrancesca.cracolici@unipa.it](mailto:mariafrancesca.cracolici@unipa.it); [furio.urso@unipa.it](mailto:furio.urso@unipa.it)

© Springer Nature Singapore Pte Ltd. 2021

S. Suzuki, R. Patuelli (eds.), *A Broad View of Regional Science*, New Frontiers in Regional Science: Asian Perspectives 47,  
[https://doi.org/10.1007/978-981-33-4098-5\\_14](https://doi.org/10.1007/978-981-33-4098-5_14)

259

Websites promoting destination tourism (both official and unofficial) have to provide a systematic and overall picture of a destination, bearing in mind that a touristic service is complex and varied, covering various consumer goods such as transport, accommodation, recreational activities, touristic hubs, and other services. A tourism destination promotion portal should provide structured information that allows potential visitors to plan the holiday best suited to their needs, providing a single access point to the tourist destination rather than requiring visits to a number of different websites (Rita 2000).

Promotion by websites is different from the more traditional channels because the potential visitors are not known, but it is still possible to gather information about the characteristics of web users by extracting it from the server. Knowledge about the users of a website by country, information needs and surfing behavior inside a website is crucial to understanding the needs of potential tourists and consequently to creating effective online marketing strategies (Biswas and Krishan 2004; Dias and Vermunt 2007).

Despite the obvious importance of these websites to DMOs, analysis of user behavior has received little attention in tourism literature (Gretzel et al. 2006). Indeed, previous studies about the analysis of online behavior and information preferences have mainly dealt with the topic of e-commerce, rather than e-tourism and promotion sites and they mostly focused on comparative analysis (Rebón et al. 2015).

To analyze web users' information-seeking behavior we need to handle the huge amount of data which is generated as they explore the website. These big data can be managed by means of Web Usage Mining methods (WUM). These statistical techniques allow the most requested information to be identified, as well as how users reached it and how much time they spent on each page. Using this knowledge, site owners can enhance the most popular features with additional information, create new hypertext links where necessary and make the site structure easier to navigate.

In the light of these considerations, our study explores whether potential tourists differ as to their information preferences. The empirical analysis has been performed by taking Sicily, an internationally known Italian tourist destination, as a case in point. We use data from the website of PalermoTravel<sup>1</sup> which is an unofficial destination website promoting Sicily as a tourist destination. Using Web Mining techniques, we explore the behavior of its website users searching for information on immaterial tourism services (i.e. local attractions, experiences, seasonal events) and/or material tourist services (i.e. accommodation, transport, etc.).

Additionally, to explore regional differences in the needs of potential tourists, our analysis also takes into account the origin of users; hence we have discriminated between Italian, American, and European users. The browsing behavior of Chinese users was also analyzed in order to understand whether there are substantial differences between potential tourists from western countries and those from Asia.

---

<sup>1</sup>A pseudonym has been used, instead of using the real name of the company.

The chapter has been structured as follows: in Sect. 14.1 we will present how the web mining analysis of e-tourism portals can be used to obtain important information about the needs of potential tourists. Specifically, WUM techniques such as association analysis and Markov chains will be illustrated, as will a presentation of the dataset relating to the visits to the PalermoTravel website. In the second Section, we discuss the main results we obtained. Finally, in the last part some concluding remarks about user behavior will be made.

## 14.2 Web Usage Mining and Data

Web Usage Mining is the application process of Data Mining techniques used to discover hidden browsing patterns from Web Data. The goal is to determine profiles of users who are exploring the contents of a given website.

The Web Usage Mining process can be divided into three separate phases: the data collection and pre-processing phase, the pattern identification phase, and the pattern analysis phase. In the pre-processing phase the dataset containing the list of web objects requested by the users is cleaned up and the user sessions are identified. A session is the sequence of web pages that have been viewed by the same IP address in a defined short period of time and represents a user's activity on the website (Liu and Keselj 2007).

We can see in Table 14.1 a section of a Path Matrix, each row being a user session.

In the pattern identification phase, the data is processed in order to obtain the hidden patterns that reflect the behavior of the users, and indices are calculated that are representative of users, sessions, and site components. In the final phase, the patterns and statistics are further processed and aggregated to be used by Data Mining algorithms such as Association rules and the Markov chain (Cooley et al. 2000).

**Table 14.1** Path matrix example

	V1	V2	V3	V4	V5	V6
1.0.153.1.0	Page 1					
1.1.176.1.0	Page 1	Page 5				
1.129.97.1.0	Page 1	Page 2	Page 3	Page 2	Page 4	
1.129.97.1.1	Page 1	Page 2	Page 6	Page 4		
1.132.105.1.0	Page 1					
1.132.110.1.0	Page 1	Page 6	Page 5	Page 4		
1.136.106.1.0	Page 1					

### 14.2.1 Association Rules

Association rules analysis is one of the most important Data Mining techniques as it enables us to find co-occurrence relationships among data items (Agrawal et al. 1993).

Let  $I = \{i_1, i_2, \dots, i_m\}$  be a set of items and  $T = \{t_1, t_2, \dots, t_n\}$  a set of transactions where  $t_i$  is a set of items such that  $t_i \subseteq I$ . An association rule is an implication of the form:

$$X \rightarrow Y, \text{ where } X \subset I, Y \subset I \text{ and } X \cap Y = \emptyset \quad (14.1)$$

It is possible to measure the effectiveness of the rule by identifying whether its *Support* and *Confidence* values are greater than or equal to the user-specified minimum. The *Support* of a rule is the percentage of transactions in the set  $T$  that contains  $X \cup Y$ , and can be seen as an estimate of the probability  $\Pr(X \cup Y)$ . So, it is a measure of the rule reliability in the transaction set  $T$ . The *Confidence* of a rule is the percentage of transactions containing  $X$  that also contain  $Y$ . It can be seen as an estimate of the conditional probability  $\Pr(Y|X)$  and determines the predictability of the rule. If the *Confidence* value of a rule is low, it is not possible to predict  $Y$  from  $X$  with sufficient reliability.

Another interesting index is the *Lift* which provides an estimate of the improvement in the predictive capacity of the rule. The *Lift* is calculated as the ratio between the conditional probability of an event given a sequence (the rule *Confidence*) and the probability of the event occurring in the absence of the sequence.

$$\text{Lift}(X \rightarrow Y) = \frac{\text{confidence}(X \rightarrow Y)}{P(Y)} \quad (14.2)$$

If the *Lift* value is above the unit, then visiting pages from other thematic areas increases the probability that the goal page will be displayed, whereas *Lift* values lower than one indicate that the display of the other thematic areas decreases this probability.

Association rules analysis in the context of Web Usage Mining enables us to identify which groups of objects or pages are purchased or visited at the same time and how often, providing a greater understanding of user preferences. This enables website managers to modify and organize content more efficiently and provide suggestions which take into account pages already selected by the user.

### 14.2.2 Markov Chain

A sequential patterns analysis is a useful tool to investigate the navigation path of users. Knowledge of such “routes” allows site owners to make visualization/non-

visualization predictions and to improve website structure. Among the methods employed to analyze user behavior is the Markov model.

A time homogeneous Markov chain of order  $k$  is a stochastic process  $X^n$  that, at each time  $n$  takes state  $s_n$  from a finite set of state  $S$ , with probability that is independent of time  $n$  and that depends only on the states attained in the previous  $k$  times. This process can be described by transition matrices  $P^k$ , where the generic element  $P^k_{i,j}$  is the probability of transitioning from state  $i$  at time  $n - k$  to state  $j$  at time  $n$ .

The Ching et al. (2013) variation of Raftery’s model supposes that the weighted sum of the  $k$  previous transition probabilities is an approximation of state probabilities distribution  $X$ .

$$X^{n+k+1} = \sum_{i=1}^k \lambda_i Q_i X^{n+k+1-i} \tag{14.3}$$

$$Q_i = [P^i_{j,h}]_{m \times m} = \begin{bmatrix} P^i_{1,1} & \dots & P^i_{1,m} \\ \vdots & \ddots & \vdots \\ P^i_{1,m} & \dots & P^i_{m,m} \end{bmatrix} \sum_{i=1}^k \lambda_i = 1, \lambda_i \geq 0 \forall i$$

with  $Q_i$  a  $m \times m$  lag-specific transition probability matrix and  $\lambda_i$  the weight for each lag  $i$  in the model.

The Markov model, considering a sequence of possible events, allows the state of a given event to be predicted given the states of previous events. In the context of WUM, a state is a user’s request for a site object (viewing a page) and the transition probabilities are the probabilities that the user requests that object knowing what objects he previously requested (Moe 2003).

### 14.2.3 Data

The above web mining techniques have been used to explore the behavior of users who log on to a destination website. The extracted data is related to visits to the website of a Sicilian tourism promotion company called PalermoTravel. It provides different hospitality services, such as the booking of luxury apartments, the booking of transfers, information on experiential tourism, and so on. The website is structured into thematic areas which can be divided into two macro areas.

The first macro area is “immaterial” tourism services which consists of three subsections: *Attraction* providing information on touristic hubs or general information on Sicily; *Experience* where additional activities can be booked; *Event* providing an overview of seasonal concerts, festivals or theatrical performances.

The second macro area is “material” tourism services which has two subsections: *Accommodation*, where users can view pages of bookable apartments and carry out thematic research by city, period, number of guests etc.; *Service*, which offers auxiliary services (e.g. taxi transfer service, bike rental, baby equipment rental).

In addition, the site provides general information on the company and its staff, on partners and access to bloggers and user reviews (i.e. the *Info* subsection).

Data were collected in 2017 for the months of September, October, November, and December. The dataset consists of 2,487,802 lines divided into 17 log fields. The web server log data is in *IIS-W3Cex Extended* format.

### 14.3 Results

The data were cleaned up, eliminating all the lines that did not concern the page views, and pre-processed. The dataset obtained consists of 95,201 lines by IP address and chronological order. Then we were able to extract 43,182 User Sessions, which allows us to follow the path of users within the site in the form of a sequence of pages viewed. We also chose not to consider the actual names of the pages viewed but only their thematic areas.

Table 14.2 reports descriptive statistics listing user accesses from Italy, the EU (except Italy; in the following we refer to it as EU users), the USA, and China.

The reported average values suggest that users accessing from Italy, EU, and the USA behave in a similar way, while those from China view a greater number of pages on average per session, spend less time on each page and also have shorter sessions.

Association rules analysis was applied to the “Area” variable referring to the thematic category of the displayed pages. We used the Paths Matrix of 43,182 user sessions. The analysis was carried out using the R “arules” package. The association rules have been obtained through an *apriori* function that allows the same algorithm to be implemented by entering in input a matrix of transactions (the sessions) and *Support* and *Confidence* threshold values (Agrawal and Srikant 1994). The algorithm is then implemented by selecting a 5% *Support* threshold level and a 30% *Confidence* threshold level.

Tables 14.3 and 14.4 show association rules obtained from users who have logged on from Italy, EU, the USA, and China.

**Table 14.2** Italy, EU, USA, China access

	Average page number	Average page duration (s)	Average session duration (s)
Italy	6.056	65.497	198.247
EU	5.570	64.529	175.436
USA	4.519	76.155	181.142
China	15.074	16.936	83.289

**Table 14.3** Association rules for Italy access data

Rule	Support	Confidence	Lift
{Service} => {Accommodation}	0.05089323	0.5326087	1.0223199
{Info} => {Attraction}	0.06086415	0.4599686	0.8649566
{Info} => {Accommodation}	0.05068550	0.3830455	0.7352397
{Experience} => {Attraction}	0.08475280	0.4880383	0.9177407
{Experience} => {Accommodation}	0.11466556	0.6602871	1.2673931
{Attraction} => {Accommodation}	0.22642293	0.4257812	0.8172691
{Accommodation} => {Attraction}	0.22642293	0.4346093	0.8172691
{Homepage} => {Attraction}	0.29372663	0.5039202	0.9476061
{Homepage} => {Accommodation}	0.30930619	0.5306486	1.0185576
{Attraction, Experience} => {Accommodation}	0.06481097	0.7647059	1.4678206
{Accommodation, Experience} => {Attraction}	0.06481097	0.5652174	1.0628736
{Experience,Homepage} => {Attraction}	0.05670960	0.5083799	0.9559925
{Experience,Homepage} => {Accommodation}	0.08039053	0.7206704	1.3832964
{Attraction,Homepage} => {Accommodation}	0.14831741	0.5049505	0.9692311
{Accommodation,Homepage} => {Attraction}	0.14831741	0.4795165	0.9017157

**Table 14.4** Association rules for EU access data

Rule	Support	Confidence	Lift
{Service} => {Attraction}	0.06324405	0.6390977	1.0565158
{Event} => {Accommodation}	0.07068452	0.5475504	1.3527717
{Event} => {Attraction}	0.05319940	0.4121037	0.6812638
{Experience} => {Accommodation}	0.06584821	0.5315315	1.3131955
{Experience} => {Attraction}	0.07068452	0.5705706	0.9432311
{Accommodation} => {Attraction}	0.20796131	0.5137868	0.8493597
{Attraction} => {Accommodation}	0.20796131	0.3437884	0.8493597
{Homepage} => {Accommodation}	0.22656250	0.3677536	0.9085678
{Homepage} => {Attraction}	0.38244048	0.6207729	1.0262224
{Accommodation, Homepage} => {Attraction}	0.13913690	0.6141215	1.0152267
{Attraction, Homepage} => {Accommodation}	0.13913690	0.3638132	0.8988327

On the left-hand side of the tables we can see the rules divided into *antecedent* and *consequent*. We notice there are just two *consequent* categories, *Attraction* and *Accommodation*, which are the most visited areas of the website. In the case of access from Italy, the EU, and the USA, we immediately notice in Table 3.2 how few rules have satisfied the thresholds established for *Support* and *Confidence*. The rules referring to Italian access in Table 3.2 include many *Lift* values above 1 when *Accommodation* is *consequent* of the rule, for example, we have a *Lift* value of 1.467 and a *Confidence* value of 0.764 when the user has viewed *Attractions* and *Experiences*. It should be remembered that a high *Confidence* value means that assuming that *Attractions* and *Experiences* are visualized, in 76% of cases *Accommodation* was also displayed. With *Attraction* as the *consequent*, the only *Lift* value greater than 1 is related to the visualization of both the *Accommodation* and

**Table 14.5** Association rules for USA access data

Rule	Support	Confidence	Lift
{Event} => {Accommodation}	0.05428954	0.4576271	1.1853814
{Event} => {Attraction}	0.07372654	0.6214689	0.9375446
{Info} => {Attraction}	0.09048257	0.6585366	0.9934647
{Service} => {Accommodation}	0.05093834	0.3877551	1.0043934
{Service} => {Attraction}	0.08981233	0.6836735	1.0313861
{Experience} => {Accommodation}	0.07640751	0.3562500	0.9227865
{Experience} => {Attraction}	0.13739946	0.6406250	0.9664434
{Homepage} => {Accommodation}	0.09517426	0.3776596	0.9782432
{Homepage} => {Attraction}	0.13873995	0.5505319	0.8305294
{Accommodation} => {Attraction}	0.20978552	0.5434028	0.8197745
{Attraction} => {Accommodation}	0.20978552	0.3164813	0.8197745
{Accommodation, Experience} => {Attraction}	0.05630027	0.7368421	1.1115960
{Attraction, Experience} => {Accommodation}	0.05630027	0.4097561	1.0613821
{Accommodation, Homepage} => {Attraction}	0.05227882	0.5492958	0.8286646
{Attraction, Homepage} => {Accommodation}	0.05227882	0.3768116	0.9760467

*Experiences* pages. In the case of access from another EU member country, we have two significant *Lift* values of 1.35 and 1.31 when *Accommodation* is the *consequent*, meaning that there is an increase in the probability of viewing *Accommodation* given that *Event* or *Experiences* has been viewed, and three *Lift* values greater than 1 with *Attraction* as the *consequent*, when the *antecedent* was either *Service* or *Homepage* or *Homepage* and *Accommodation*. As far as access from the USA is concerned we can see three *Lift* values greater than 1 with *Accommodation* as *consequent* when *Events*, *Services* and a combination of *Attraction* and *Experiences* have been viewed, and two *Lift* values greater than 1 with *Attraction* as *consequent* when *Service* and a combination of *Accommodation* and *Experiences* have been viewed. We chose not to display the rules related to Chinese accesses because, differently from previous cases, many rules (441 in total) satisfied the thresholds; and also all page combinations show *Lift* values greater than 1 (Table 14.5).

In general, it is noticeable that in the case of access from Italy, the EU, and the USA, visiting other areas just slightly increases the probability of visualizing the housing and touristic attractions, while in all three cases the visualization of the *Accommodation* pages decreases the probability of displaying the *Attraction* pages and vice versa.

Users from China tend to view any and all of the website areas, and the joint visualization greatly increases the probability of displaying a page in the *Accommodation* or *Attraction* category.

If we look at visits from western countries, it should be noted, however, that despite the minimum *Support* threshold of 5%, which may seem too generous, very few rules have met the requirements and rules related to users who have viewed more than three page categories are not listed because the percentages are too low.



Using Markov chains in this context enables us to follow the user movements in probabilistic terms. This analysis was conducted using the *FitMarkovChain* function of the “Clickstream” package.

Looking at the transition matrices for Italian accesses shown in Fig. 14.1, it is clear that users tend to remain in the same subject area.

For example, let us consider first order transition probabilities. A user viewing a page in the *Accommodation* category has a 78% chance of moving to another page in the same category. Similarly, a user on a page in the *Attraction* category has a 56% chance of remaining in the same category.

Other thematic areas have greater transition probabilities in the case of a passage within the same category, but to a lesser extent compared to the two previous categories (values between 27% and 52%).

When we increase the order of transition matrices, the situation remains almost the same but with slight decreases in probability in the main diagonal. At the sixth click, we note that, with the exception of the *Accommodation* and *Experiences* categories, the transition probabilities in the same category fall below 36% and users in the *Attraction* category also have a 36% probability of moving to an *Event* page, close to the probability of remaining in the same category (42.6%).

Moving to EU and USA accesses (Figs. 14.2 and 14.3), we can see similar structures with high first order probabilities of remaining in the same category if it is *Accommodation*, *Attraction* or *Info* (probabilities greater than 50%). There is a 42% transition probability that a USA user will move from an *Event* to an *Attraction*. After six clicks we notice an increase in the probability of a move to the *Attraction* area, especially in the case of USA accesses where the transition probability values are between 23% and 57%.

As regards Chinese accesses, we notice in Fig. 14.4 high first order transition probabilities of remaining in the *Homepage* area (58%) and *Attraction* area (73%) and of moving from an event page to an attraction page (51%). The Six order transition probability matrix shows that there is a 66% probability of still being in the *Attraction* area when moving from a page in that area, and that after six clicks there is a 61% or higher probability that users will be in the *Attraction* area having come from either *Service*, *Event* or *Experiences*. Differently from other users, it seems that the Chinese do not stay in *Accommodation* very long but move to another area after one click.

Finally, we can see in Table 14.6 the session exit probabilities. Users from a western country are more likely to end their sessions either in an *Attraction* page or in an *Accommodation* page, with Americans preferring the former and Italians the latter. Chinese users tend to end their sessions in an *Attraction* area.

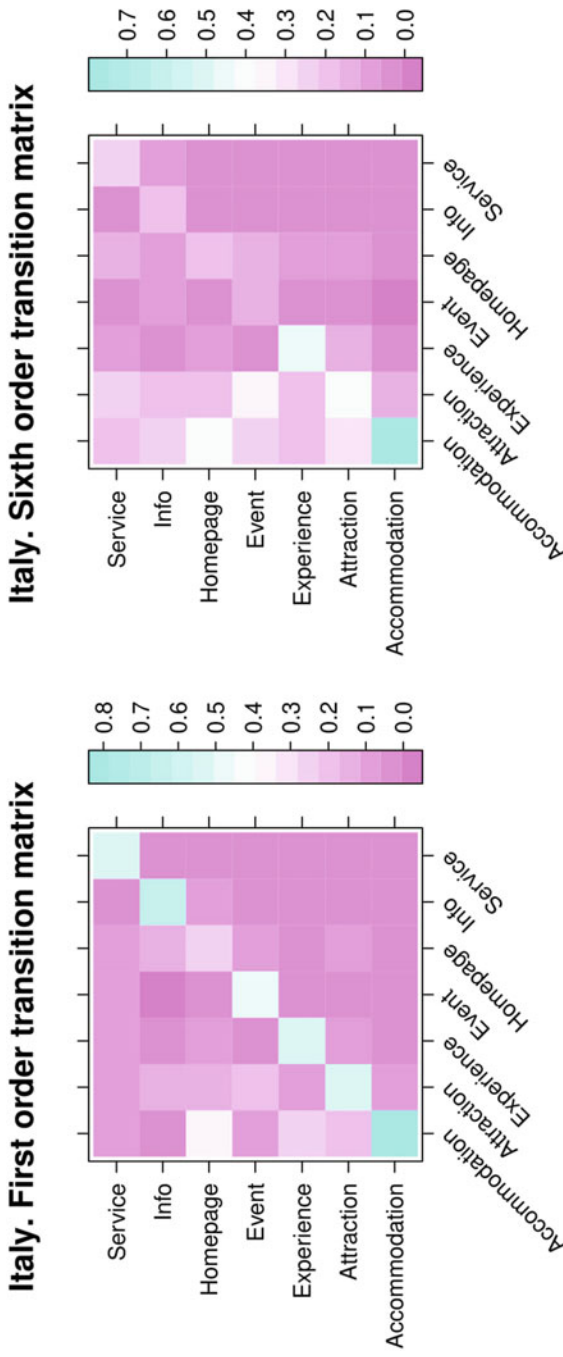


Fig. 14.1 Transition probability matrices, Italian accesses: order 1 and 6

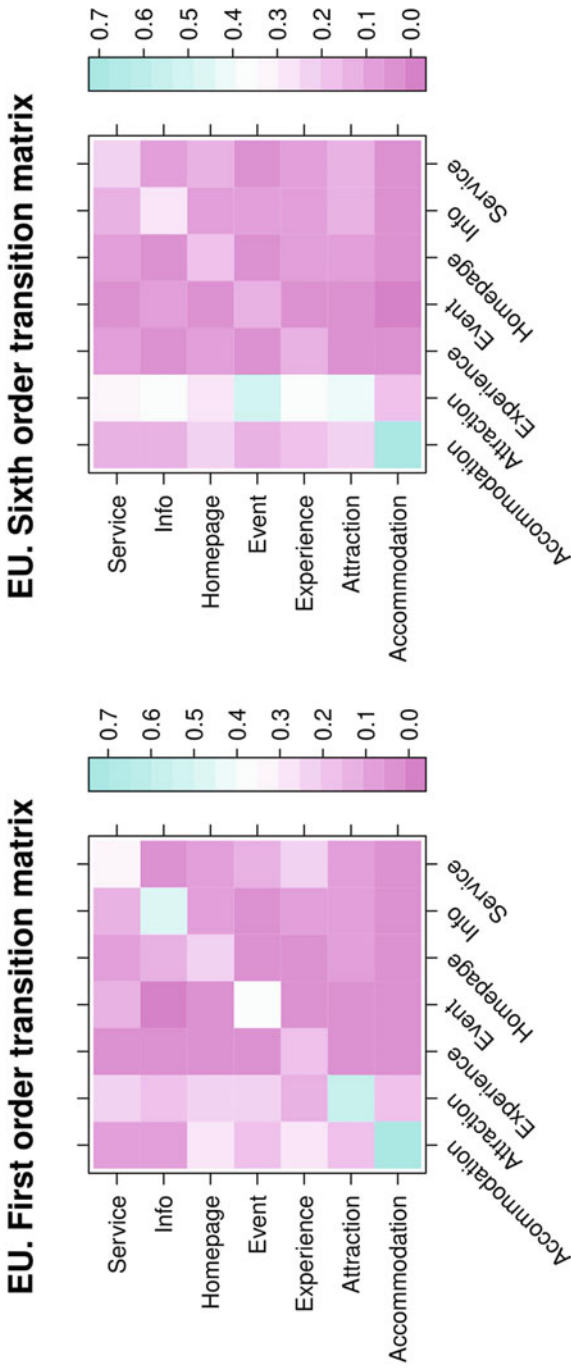


Fig. 14.2 Transition probability matrices, EU accesses: order 1 and 6

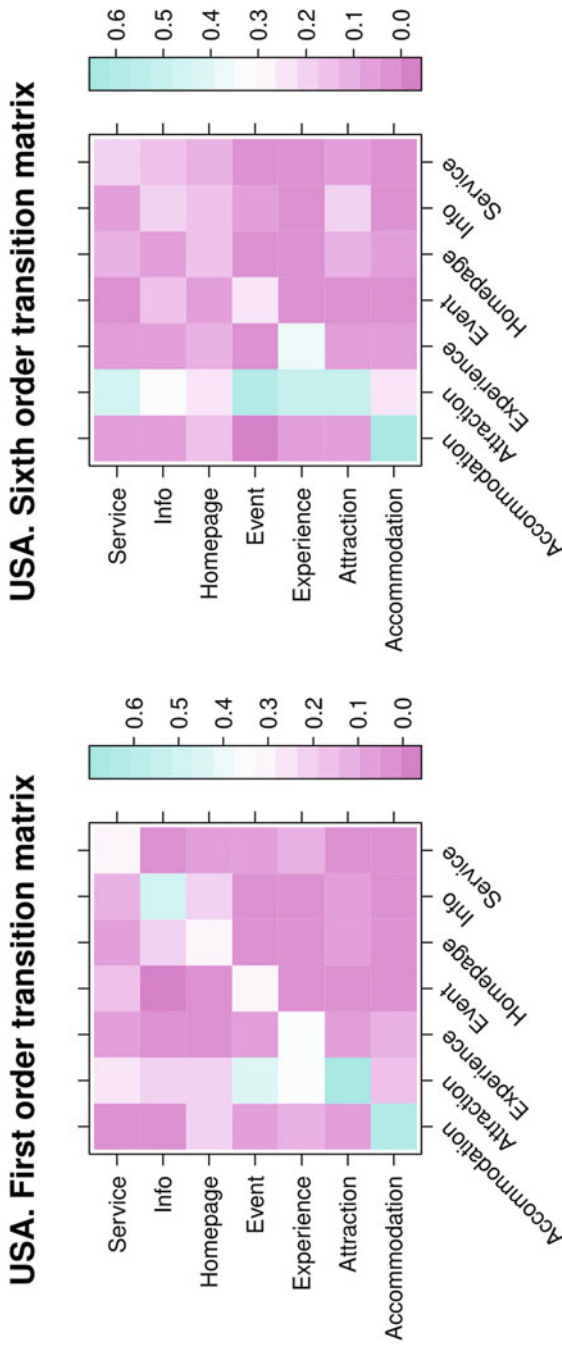


Fig. 14.3 Transition probability matrices, USA accesses: order 1 and 6

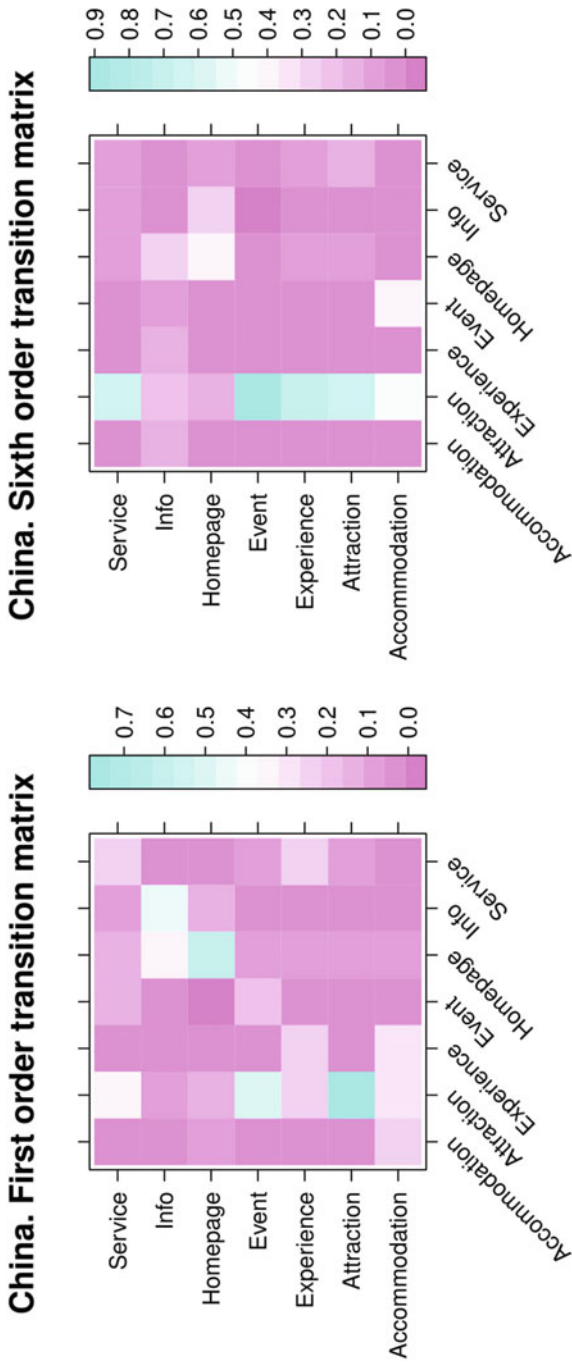


Fig. 14.4 Transition probability matrices, China accesses: order 1 and 6

**Table 14.6** Exit probabilities for Italian, EU, USA, and China users

	Accommodation	Attraction	Experience	Event	Homepage	Info	Service
Italy	0.406	0.273	0.100	0.032	0.061	0.071	0.057
EU	0.327	0.367	0.039	0.059	0.076	0.076	0.056
USA	0.223	0.440	0.118	0.062	0.047	0.066	0.043
China	0.050	0.355	0.033	0.033	0.317	0.039	0.172

## 14.4 Concluding Remarks

The analysis we carried out has shown how WUM techniques can be used to identify the navigation patterns of users of a Sicilian tourist portal and to understand their information needs. The results obtained highlight profound differences in behavior between users who log on to the site from Western countries and those from China. The latter tend, on average, to view more pages per session, but both the time spent on each page and the total length of the session is considerably less than the time spent by other users.

Association rules analysis shows that western users tend to visit no more than three different thematic areas, and that the *Attraction* and *Accommodation* areas seem to be mutually exclusive, dividing users into those who are most interested in information about apartments from those looking for information about Sicilian cultural attractions. In contrast, Chinese users explore the site in its entirety by visiting each thematic area.

Using Markov chains we have observed that Italian users view pages in the same thematic area in sequence, showing a greater interest in the area relating to apartments, which is generally also the last area they visit, indicating that they are not interested in searching for any other kind of information. American users begin by viewing pages from either the *Attraction* or *Accommodation* areas, and after some movements they either remain in the *Accommodation* area or move from other areas to the *Attraction* area, in most cases concluding the session there. European users seem to be a mix of the two. Chinese users, on the other hand, show little interest in the *Accommodation* area and quickly move to the *Attraction* area where they conclude their session.

These results indicate important directions for further research. For instance, additional information on user behavior inside each area, like accommodation or attractions, should allow us to verify the assumption on the lack of homogeneity in search behavior in the set of the countries considered, and also to define potential visitor segments, which might also be better determined by merging information on user search and booking behavior.

In conclusion, when planning their holidays, users from different countries displayed different information needs. To be specific, the analysis reveals a sort of dependency between cultural distance and information preferences. Presumably, since Italians are already aware of the information on the region and its cultural heritage, they are more likely to explore pages on accommodation whereas users from other European countries view pages both on attractions and seasonal events

and on accommodation. Americans, on the other hand, focus much more on pages of cultural interest, because Sicily is most likely unknown to many of them.

It is noteworthy that although the Chinese seem really interested in cultural heritage and naturalistic sites these results should be viewed with caution because they actually visit each area of the site at least once, as highlighted by the analysis of the association rules. In addition, the short sessions and the little time spent per page seem to indicate an erratic movement within the site, perhaps due to linguistic problems and therefore to a poor understanding of the information presented.

## References

- Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proceedings of international conference on very large data bases (VLDB)
- Agrawal R, Imieliski T, Swami A (1993) Mining association rules between sets of items in large databases. In: Proceedings of ACM SIGMOD international conference on management of data (SIGMOD-1993)
- Biswas A, Krishan R (2004) The Internet's impact on marketing: introduction to the JBR special issue on "marketing on the web—behavioral, strategy and practices and public policy". *J Bus Res* 57(7):681–684
- Ching WK, Huang X, Ng M, Siu TK (2013) *Markov chains: models, algorithms and applications*, 2nd edn. Springer, Boston
- Cooley R, Mobasher B, Srivastava J (2000) *Web mining: information and pattern discovery on the world wide web*, SIGKDD exploration, vol 1. Department of Computer Science and Engineering, University of Minnesota, Minneapolis
- Dias JG, Vermunt JK (2007) Latent class modeling of website users' search patterns: implications for online market segmentation. *J Retail Consum Serv* 14:359–368
- Gretzel U, Fesenmaier DR, Formica S, O'Leary JT (2006) Searching for the future: challenges faced by destination marketing organizations. *J Travel Res* 45(2):116–126
- Kim S, Shaw T, Schneider H (2003) Web site design benchmarking within industry groups. *Internet Res* 13(1):17–26
- Liu H, Keselj V (2007) Combined mining of web server logs and web contents for classifying user navigation patterns and predicting users' future requests. *Data Knowl Eng* 61(2):304–330
- Moe W (2003) Buying, searching, or browsing: differentiating between online shoppers using in-store navigational clickstream. *J Consum Psychol* 13(1–2):29–39
- Rebón F, Ocariz G, Argandoña J, Gerrikagoitia JK, Alzua-Sorzabal A (2015) Behaviour of virtual visitor based on e-shop and DMO websites: a comparative study by means of data mining techniques. In: Tussyadiah I, Inversini A (eds) *Information and communication technologies in tourism*. Springer, Cham, pp 201–210
- Rita P (2000) Internet marketing destinations in the global tourism marketplace. In: Proceedings INET 2000. <http://www.isoc.org/inet2000/cdproceedings.htm>
- Wang Y, Fesenmaier DR (2006) Identifying the success factors of web-based marketing strategy: an investigation of convention and visitors bureaus in the United States. *J Travel Res* 44 (3):239–249