Data Mining Applications in Tourism: A Keyword Analysis

Mirjana Pejić Bach  
Faculty of Economics & Business  
University of Zagreb  
Trg J.F. Kennedyja 6, 10000 Zagreb, Croatia  
mpjic@efzg.hr

Markus Schatten  
Faculty of Organization and Informatics  
University of Zagreb  
Pavlinska 2, 42000 Varaždin, Croatia  
markus.schatten@foi.hr

Zrinka Marušić  
Institute for Tourism  
Vrhovec 5, 10000 Zagreb, Croatia  
zrinka.marusic@iztzg.hr

Abstract. The paper reviews applications of data mining in academic papers dealing with tourism industry, both from demand and supply side. Web of science, SCOPUS, and in particular key tourism journals published in 1995 – 2013 period have been searched with the usage of appropriate keywords. Literature searches revealed 88 papers that present applications of data mining in tourism. Keyword and conceptual network analysis were conducted with the usage of Wordle and LaNet-vi tools. Papers from tourism related journals and ICT related journals were analysed separately. In order to detect historic trends, analysis was conducted separately for the two periods before 2005, and since 2006. The conclusion of the paper is that tourism steps on a path to evolve to being both people-driven and data-driven, thus utilizing data mining approach as leverage towards increased competitiveness and profitability.

Keywords. data mining, tourism, keywords, conceptual network, review

1 Introduction

Data mining emerged in the 80’s from the fields of machine learning, statistics, and databases, and has eventually taken place as one of the most important tools to get additional value from information gathered in organizational databases [1]. Its application is large and it usually results in specific knowledge that is a basis for action. Examples are segmentation, lifetime value, and credit scoring and churn prediction. Data mining in most cases explores data generated transactions, such as sales transactions, web logs, and process generated transactions by the use of machine learning and statistical techniques [3].

Modern technology has had a great impact to tourism since the 90s, when Internet technologies became the most dominant communication channel [2]. Changes occurred both on the demand and the supply side. Tourists’ expectations are related not only to tourism services, but also to technology. Tourists expect that tourism companies actively implement modern technologies as the part of the value chain. Modern trends and technologies generated a whole set of new tools, such as recommendation systems [5]. In addition, new technologies help tourism companies to establish one-to-one connection with tourists, thus increasing the customer loyalty [8].

The number of reviewed papers has investigated the usage of modern technologies in tourism, but to our knowledge, attempts to review data mining and tourism are rare. The goal of this paper is therefore to review data mining applications in tourism based on the keywords analysis.

2 Methodology

The following steps were performed in order to collect literature that investigates applications of data mining in tourism. First, Scopus and Web of Science databases were searched with the use of following keywords: data mining, knowledge discovery in databases, tourism, tourist, destination, travel and hotel. Since words tourism and tourist, have many derivatives, a lemmatization feature of Web of Science was turned-on in order to include similar words. We set the time frame to the period from 1995 to 2013. In addition, tourism journals included in Web of Science and Scopus were also searched with the usage of the above keywords. We limited the search to peer-reviewed journals only. Just few articles that contain the above keywords were found that actually do not tackle the issue of the paper, and were therefore excluded from the analysis. Based on such approach, we found 88 papers in journals indexed in Web of Science and/or Scopus that deal with data mining applications in tourism.

In order to conduct keyword and conceptual network
analysis, we used Wordle\(^1\) and LaNet-vi\(^2\) tools. The three goals of the research were set: (i) to detect main concepts in tourism data mining applications, (ii) to detect historic trends, (iii) to detect differences between ICT and tourism related journals. Keywords were first analysed with the usage of conceptual network visualization using LaNet-vi. Then, keyword analysis was conducted for the two periods (before 2005, and since 2006) and for the two groups of journals (ICT and tourism related).

3 Results

The obtained results are summarized in the following two subsections.

3.1 Conceptual network visualization

In order to gain insight into the mutual interconnection between the various keywords in papers that investigate data mining applications in tourism, we constructed a conceptual network based on co-affiliation; e. g. two keywords are connected if they are used in the same article which is similar to the approaches given in [4,6,7]. The conceptual network has been visualized using LaNet-vi and is shown on Figure 1.

The size of the node (shown on the left node size scale on the image) visualizes the number of connections a node has, whilst the colour of the node depicts its interconnection with other nodes in the same core (depicted in the right scale of node colours in the image).

As can be seen on the image there are 6 cores representing each a certain area of mainstream research as well as an outer sparsely connected residue representing non-mainstream areas of research. The six cores, from inside to outside, can be recognized as: (i) data mining and forecasting; (ii) machine learning and personalization; (iii) tourism management; (iv) tourism systems (recommendation systems, geographic information systems, mobile systems etc.); (v) segmentation, and (vi) advanced techniques (support vector regression, multi agent systems, particle swarm optimization etc.).

3.2 Keyword analysis

In the following part of the paper we present four keyword visualizations in order to gather better insight into the matter. With the use of Wordle as a visualization tool four visualizations were created: (i) keywords visualization of papers published prior to and including 2005; (ii) keywords visualization of papers published since 2006; (iii) keywords visualization of papers published in ICT related publications, and (iv) keywords visualization of papers published in tourism related publications.

The first visualization is depicted in Figure 2. From the image one can conclude that the most important keywords in papers published prior to 2005 were: tourism, segmentation, neural, information, mining, marketing, networks and forecasting.

Figure 2: Visualization of keywords until 2005

Figure 3 depicts the visualization of keywords in papers published since 2006. The most important keywords here were: tourism, segmentation, analysis, systems, data, mining, fuzzy, forecasting, travel, and recommender.

Figure 3: Visualization of keywords since 2006

Figure 4: Visualization of keywords in ICT related publications

These two depictions allow us to conclude that there was a shift in research in the observed intervals regarding to technology. Whilst in the former neural networks seem to be the important mainstream approach, in the latter approaches and recommender systems seem to be a surface. In both intervals tourism, data/information mining, forecasting and segmentation play an important role.
Figure 4 depicts the visualization of keywords in ICT related publications. Most important keywords include: systems, tourism, recommender, data, and fuzzy.

The visualization of keywords in tourism related publications on the other hand is quite different (Figure 5). Here the most important keywords include: segmentation, analysis, mining, tourism, and data.

When compared it becomes obvious that ICT related publications are especially focused on various types of (software) systems and their implementation while tourism related publications are more interested in analysis of actual data.

4 Discussion and conclusion

In order to get insights in applications of data mining in tourism we have conducted a thorough literature search with the selected keywords related to both research areas. After careful investigation, 88 peer-reviewed articles were selected and analysed applying the method of conceptual networks and keyword analysis. The following results emerged. First, six core areas of data mining applications in tourism were detected by the use of conceptual networks. These fields cover applications of data mining in forecasting, personalization, tourism management, tourism systems (such as recommendation systems), and machine learning techniques such as support vector regression, multi agent systems, particle swarm optimization etc. Second, keyword analysis comparing the two time periods (1995 to 2005 vs. 2006 to 2013) was conducted. The first period revealed keywords oriented towards specific applications, such as segmentation, forecasting, and marketing. The second one revealed keywords oriented towards more advanced systems and technologies, such as recommender and fuzzy, although segmentation and forecasting still remained highly pondered. As expected, papers from tourism journals were more prac-
tically oriented, while papers from ICT journals were more technically oriented.

Our research opens an interesting area of data mining applications in tourism with keywords analysis based on the investigation of Web of Science and Scopus journal articles. However, this also generates the limitations of our study. First, more elaborate analysis with the in-depth analysis of the articles should be conducted in order to get more insight into what decisions in tourism are being driven by the use of data mining, and what methodology, sample and time are being used in the process. In addition, usage of additional sources such as case study reports and papers from highly respected, practically oriented conferences, should also be considered as candidates for future studies.

References


Appendix – List of reviewed papers


[33] Hsu, C-I; Shih, M-L; Huang, B-W; Lin, B-Y; Lin, C-N. Predicting tourism loyalty using an integrated Bayesian network mechanism. Expert Systems with Applications, 36(9):11760-11763, 2009.


[44] Law, R; Rong, R; Vu, HQ; Li, G; Lee, HA. Identifying changes and trends in Hong Kong outbound tourism. Tourism Management, 32(5):1106-1114, 2011.


[48] Li, G; Law, R; Vu, HQ; Rong, J. Discovering the hotel selection preferences of Hong Kong inbound travelers using the Choquet Integral, Tourism management, 36:321-330, 2013.


[85] Xia, JC; Evans, FH; Spilsbury, K; Ciesielski, V; Arrowsmith, C; Wright, G. Market segments based on the dominant movement patterns of tourists. Tourism Management, 31(4):464-469, 2010.


[87] Ye, Q; Law, R; Li, S; Li, Y. Feature extraction of travel destinations from online Chinese-language customer reviews, International Journal of Services Technology and Management, 15(1/2):106–118, 2011