Leveraging Newly Available Big Data for Urban Architectural Heritage: Designing a Recommendation System for Heritage Sites through the Lens of Social Media

S.Sezi Karayazi, Gamze Dane, Bauke de Vries

(MSc, S.Sezi Karayazi, Eindhoven University of Technology, P.O. Box 513, 5600MB Eindhoven, s.s.karayazi@tue.nl)
(Dr. Gamze Dane, Eindhoven University of Technology, P.O. Box 513, 5600MB Eindhoven, g.z.dane@tue.nl)
(Prof. Dr. Bauke de Vries, Eindhoven University of Technology, P.O. Box 513, 5600MB Eindhoven, b.d.vries@tue.nl)

1 ABSTRACT

Urban heritage sites are essential part of the cities, because they reflect the historical background of societies and create attraction for tourism industry. However, as tourism industry focuses exclusively on economic growth, usually historical urban cores are under the pressure of mass tourism and urbanization, causing negative experiences for residents and visitors such as overcrowd, nuisance and waste. Therefore, there is a need to understand what attributes attract visitors to certain heritage sites and which heritage sites are overrepresented in space and time, so that recommendations can be given to the visitors and local government in order to reduce the negative impacts of mass tourism. On the other hand, the rise of Internet usage has fundamentally changed the perception for built environment. People are able to reflect own ideas or opinions leaving behind their digital footprints within urban areas. Such digital footprints can be collected as datasets that reflect people's behaviors and opinion in time and space. In this respect, the aim of this paper is to define a common framework for extracting information on the attractiveness and representation of heritage sites by using spatial big data. This paper reports a conceptual framework in order to investigate the motivations of visitors to visit the heritage sites and the influences of their visitations to the heritage sites by exploiting spatial big data and analytics. Moreover, a bibliometric network among the keywords related to existing state-of-the-art is revealed in the literature review section by using VOSviewer. The paper will conclude with discussions on how the results of the proposed framework can contribute to designing positive tourist experiences in overly touristic historical cities. Furthermore, Destination Management Organizations (DMO's) can benefit from the results of this proposed framework since they can develop urban facilities in more peripheral areas instead of heavily touristified zones.

Keywords: urban heritage sites, overtourism, crowdsourcing, big data, context-aware recommendation

2 INTRODUCTION

Urban fabrics consist of different elements such as landscape, built environment, infrastructure, and open space. Historical urban core is assessed within the built environment and it is under the pressure of mass tourism and urbanization. Historic city centres are the essence of European cultural heritage and these are protected considering the each country's rule (García-Hernández et al., 2017). These places are considered as magnet for visitors because of their relevance regarding history and they attract many visitors.

The attractiveness is an important component for visitors to many historical cities (Kourtit et al., 2019) and the attractiveness of a heritage is dependent on the characteristics of heritage (i.e. typology, uniqueness), characteristics of built environment where the heritage is located (i.e. distance to other attractions, facilities, transport network) and visitors' characteristics (i.e. interests, activity-schedules, opinions). In order to understand contribution of heritage attributes and people's motivations behind their visitation, heritage attributes can be analysed under the stated choice analysis. In this research, heritage attributes will be considered in two groups; visitor (tourist) oriented and heritage oriented (Vong & Ung, 2012).

Nowadays, the usage of smartphones and cameras not only have influence on heritage interaction, but also effect on people's experiences and interpretations of heritage sites (King et al., 2016). The growth of social media has been influential on people's choices of destination. The fast development of Information and Communication Technology (ICT) is contributed to the flow of a large amount of data for urban studies and it is becoming a prominent component of urban research (Long & Liu, 2016). Big data provides various data sets such as mobile phone activities, geotagged photographs, travel trajectories, and recommendation on web platforms. These type of data can be utilized to track people's behaviour at very fine scale, since human trace is an element to understand interaction between people and urban areas, in that sense spatial distribution with time stamp is an important source for analysis. The excessive spread of tourism in urban neighbourhoods in particular heritage sites led to an overtourism, since a large amount of visitors stuck in certain locations. For instance, around 17.5 million people visited Amsterdam in 2016 and it is projected to double by 2030

(Boucher, 2019). In order to combat with overtourism in Amsterdam's hotspot such as Dam Square, Vondelpark and Zeedijk, less visited areas are offered by Amsterdam Marketing which is an organization that is funded by local government. A VR (Virtual Reality) experience was installed in Amsterdam Centraal Station in 2018 to push tourists away from the city center; therefore, visitors were encouraged to visit other nearby areas. The VR introduces to tourists other less known places around the Amsterdam such as Haarlem, Volendam, Zaanse Schans and Zandvoort was promoted as a beach resort which is located west coast of North Holland and it renamed the "Amsterdam Beach" to draw tourists' attention. Some popular places are almost exceed carrying capacity and these places are not capable of coping with such amount of visitors. Innovative solutions could be helpful to manage with visitor influx, and less-visited places including heritage sites can be recommended in order to scatter people evenly by highlighting the hidden treasures of the cities.

In order to understand the overtourism phenomenon from the perspective of heritage sites and their visitors, and to solve this problem in the cities, we propose a methodology in four steps (fig. 1). Big datasets that have location annotation can explain visitor behavior and opinions spatio-temporally. Current literature already emphasizes the importance of utilizing newly available big data sources for volunteered based and data-driven management of historical cities (Ginzarly et al., 2018; Koutras et al., 2019). However, there is no study to utilize and fuse newly available big datasets for better understanding the relation of heritage and visitors by taking into consideration heritage attributes, and heritage tourism problems related to overtourism. The conceptual framework is based on the four phase mode of research including qualitative and quantitative analysis. First, the attributes that have an influence on heritage attractiveness to understand people's motivations behind their visitations are determined (Falk & Dierking, 1992; Vong & Ung, 2012). The evaluation on the importance of each attribute is done by stated choice experiments. This method provides weights per attributes; it is based on observations of responses that made by participants in controlled hypothetical situations. Participants select the attractiveness of the levels of each attribute that contribute to the evaluation on rating scale (Kemperman, 2000). Second, big data sources such as Flickr, Twitter, TripAdvisor, and Google Places are investigated in order to align with the result of first step. Most useful dataset alternatives (based on the availability of possible attributes in the datasets) are brought together and utilized to explain the attractiveness of heritage sites. Next step describes the leveraging of urban big data for scattering people within historical urban core by giving recommendations on alternative less popular heritage sites considering the weights per attributes. Fourth and the last step is to propose a recommendation/replicable system that can be applied to cities that have the similar scenario regarding overtourism. Finally, this paper will propose a conceptual framework for big data-based recommendation system and it can be concluded that the proposed method will provide knowledge for future practice in the relevance of newly available big data and heritage studies.

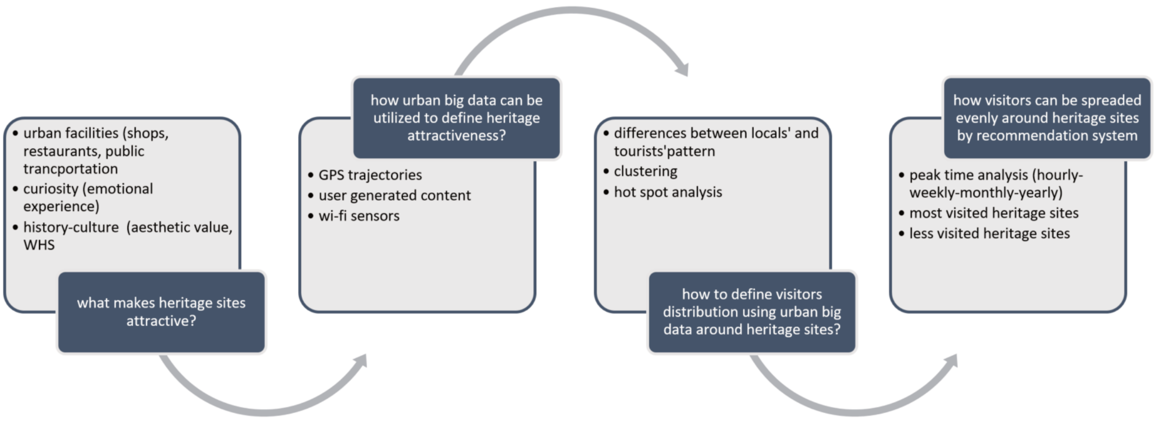

Fig. 1: Conceptual framework of proposed recommendation system

Next section shows the related works. In order to define the knowledge gap in the literature, text mining technique is applied using VOSviewer, therefore; the relation among the selected keywords are visualized. After, the methodology section explaines intented methods, research questions and expected results. Moreover, the relevance and the relation of each research questions are described in this section. Final section refers to the results, discussion and future works.

3 RELATED WORKS

The purpose of this research is four-fold; the current state-of-the-art is explained per the four steps of conceptual framework.

3.1 Heritage attractiveness and people's motivations

Heritage sites consist of different levels such as existing built environment, intangible heritage, cultural diversity, socio-economic factor and environmental factor (Centre, 2019). These levels complete each other, for instance; perception of existing built environment depends on people's background and they are linked with intangible heritage e.g., customs, traditions. According to the United Nations Educational, Scientific and Cultural Organization (UNESCO) approach, historic urban landscape covers not only preservation of the built environment, but also it focuses on the whole human environment with tangible and intangible factors. Heritage sites are one of the key elements from past to present and they can be assessed to understand interaction between people's motivations and heritage attractiveness. Falk & Dierking (1992), explain the attractiveness of heritage sites with a complex spectrum of associative components such as personal (motivation and interest), social (experience and behaviour), and physical settings (the atmosphere of heritage sites). It can be suggested that both internal and external factors are worth-stressing components of heritage meaning.

Heritage attributes can fall into two categories; human oriented is related to visitor experience, and heritage oriented is connected to component of cultural heritages (Vong & Ung, 2012). A. Morozov & M. Morozov (2018) highlight the factors that contribute to attractiveness of the destination with 9 parameters; the presence of cultural and historical tourist resources, transport availability, socio-economic development in destination, general infrastructure, urban facilities, natural factors, information security, site security and attitude of local residents. Karunanithy (2013) explaines the heritage attributes by the tourism perspective. The heritage attributes are evaluated with five indicators; tourist package, style of historical building, cultural village and entertainment. Gaffar et al., (2011) highlight that characteristics on cultural heritage sites are evaluated with six aspects; attractions, activities, accessibilities, amenities, anchillary services and available services. The combination of attributes and levels are shown in table 1.

Tourist (visitor) oriented		Heritage oriented	
Urban facilities and services	Heritage attractiveness	History and culture	Heritage interpretation
Proximity -shopping -museum -public transportation	Curiosity -emotional experience -positive expectation	Historical value -aesthetic -social -spiritual	Ample relevant information -old-fashioned -audio-visual -digital
	Overcrowding -local -tourist	Site uniqueness -WHS -non-WHS	

Table 1: Heritage attributes and levels (Vong & Ung, 2012)

3.2 Newly available big data and distribution of people

The concept of big data face rapid growth in recent years. The amount of information has been growing continuously, because it is produced automatically by different form of sensors. Big data consists of a wide range of information and it presents data-driven evidence on the basis of numbers instead of anecdotes, stories or experiences (Song & Liu, 2017). This type of data contains three key concepts "3V's"; Volume represents a large amount of quantity, Velocity is described as the measurement of how fast data arrives from sources and Variety is the range of data types (Laney, 2001). After, this concept is updated by adding Veracity (Laney & Beyer, 2012), which represents the accuracy and applicability of data, and Value (Mao et al., 2014) is the potential of big data; it can be transform into desired information. Such data influx is coming with various information that is why newly available big data have been becoming essential tool for urban studies and planning practices.

Urban big data is divided into five categories; sensor systems, user generated content (UGC), administrative data, private sector data and hybrid data. Volunteered geographic information is placed under the UGC and it supplies real-time analytics to researchers (Ginzarly et al., 2018; van Zanten et al., 2016). People can be considered as sensor, because they contribute to generate the data. UGC is at individual level and collected at fine levels of spatio-temporal scale; therefore, it allows understanding and modelling human behaviour in urban scale. Social media services such as Twitter, TripAdvisor, Foursquare, which contains volunteered geographic information, have a wide range of digital footprints and these location based services provide data on the urban services, and such data can allow monitoring of events, emotions and preference of users (Thakuriah et al., 2017). Planners can utilize the newly available big data sources with the observations and surveys, hence big data can be used as a supportive way to access information related to human and urbanscape interaction (Frias-Martinez et al., 2012).

The newly available big data based urban heritage studies have different dimensions such as destination management (van der Zee et al., 2018), tourist activity in historical cities (Kádár, 2014), mapping historical values (Ginzarly et al., 2018), and investigating historical places (Koutras et al., 2019). As far as heritage attributes are concerned, UGC big data sets can be utilized to investigate relation between heritage sites and people. For instance, “facilities-services” are placed under the visitor-oriented heritage attributes, and these are influential factors for the attractiveness of heritage locations (Vong & Ung, 2012). Therefore, if tourist movements are traced around the shopping and eating locations, the relation between facilities and tourists’ behaviour can be described by statistical methods (as shown at Dane et al., 2019), and this relational behaviour can also be visualized by mapping and simulation techniques. Another example is “overcrowding” which is associated with heritage attractiveness and in order to investigate whether overcrowding has influence on heritage locations or not. In that sense UGC big data is a valuable source for understanding overcrowding in space and time, because the rising of social media results in increasing trend of leaving digital traces (Paldino et al, 2015).

Overtourism and ever-growing tourism influx have negative impacts on cities throughout the world. It can be described as too many visitors in a particular destination. The identification of highly-visited areas and reasons behind the over visitation can be helpful to reduce the pressure of overcrowds in historical places. Historic city centers consist of tangible and intangible heritages, monuments and cultural landscapes (van der Zee et al., 2018). The service providers such as, hotel, restaurant, and tour guides are shaping functional places for tourists (Ashworth & Page, 2011). The combination of historical places and services can create an attractive historic district.

People distribution in the cities can be investigated as local and tourist, since they are followed different patterns. While tourists are clustered in the city center, local movements are extended such as parks and recreational areas (García-Palomares et al., 2015). It results in unbalanced dispersion, because heavily concentrated areas are placed in the city centre and urban core is the main attraction for tourists and tourism products (Ioannides et al., 2018; Kádár, 2014; van der Zee et al., 2018; Zhang et al., 2017).

In order to define tourist and local, researchers have used different threshold; Girardin et al. (2009), Garcia-Palomares et al. (2015), and Koutras et al. (2019) use 30-days, Kádár (2014) and Huang (2016) use 5-days limitation to separate tourist and local. If user upload more than one photograph within assigned threshold, it can be named as tourist. Otherwise, it can be accepted as local residents. The issues related to overcrowding result in degrading of the environments for local people; therefore, they are seeking more mature urban heritage destinations (Ganzaroli et al., 2017; van der Borg et al., 1996; van der Zee et al., 2018). It is highlighted by case study in Venice, overtourism has destructive effect on urban heritage areas (Ganzaroli et al., 2017).

It can be concluded that historical urban core draws tourists’ attention and they should be distributed evenly within the city in order to avoid detrimental factors of overcrowding.

3.3 Recommendation system

Tourists often need to help effective travel planning when they visit to city. Recommender system can be beneficial tool for users to identify their need from a vast amount of data. The observation of interaction between users and objects are the base of the recommendation system. It is able to combine different characteristics; user preferences and past behaviours, preferences and behaviour of the user community, items’ features and how they can match user preferences, user feedbacks, context information and how recommendations can change together with the context (Amanto et al., 2016). The experience from previous users in similar context can be valuable information to current users who would like to visit certain destinations.

Recommendation techniques are classified into three groups; collaborative filtering (recommendations based on previous user with similar preferences), content-based recommendation (provide a user based on her/his formerly preferred) and hybrid approaches (Huang, 2016). Generating recommendation based on predicting users' interest can enhance the tourists' experience, because the system can suggest locations in which are derived from UGC. In that sense location recommendation using GPS trajectories or aggregating geotagged social media data has a valuable potential to create identify locations when people visiting the heritages.

3.4 Distribution of big data based heritage studies considering tourism and recommendations

Table 2 classifies the distribution of literature example regarding subject, data collection, and method and research question. It consists of examples that are used to propose recommendation system. It starts with what makes heritage sites attractive, and follows how urban big data is utilized to define hotspots & POI(s), and how recommendation system can be developed using UGC.

Author (s)	Subject	Data collection	Method	Research question
(Kempiak et al., 2017)	Heritage tourism & attraction and visitor experience	Self-administrated questionnaire	Univariate-bivariate analysis, exploratory factor analysis	What are they key factors influencing the visitor experience at heritage attractions?
(Vong & Ung, 2012)	Heritage attributes and heritage tourism	Survey	Principal component factor analysis, The Kaiser-Meyer-Oklin test, The Bartlett test	What are the critical factors that are essential to enhance tourist experience when visiting Macau's heritage sites?
(Trinh & Ryan, 2017)	Heritage visitors and analysis of cultural tourism site	Questionnaire	Textual analysis by Leximancer and CatPac	Is there an articulation of differences arising from different national groups when visiting a site representing a culture different from their own?
(Ganzaroli et al., 2017)	Heritage tourism	TripAdvisor and the number of hotel arrivals	Correlation between reviews and concentration ratio	Does TripAdvisor contribute to strengthening the popularity of already known restaurants in spite of their ranking?
(García-Palomares et al., 2015)	Tourism	Panoramio	Density map and correlation relations	How can the popular attractions be identified using photo sharing services?
(Girardin et al., 2009)	Tourism	Flickr and network data (AT&T)	Density map and spatio-temporal distribution	How do locals and visitors share the space?
(Koutras et al., 2019)	Tourism	Flickr	Density-based algorithm	How can GIS analysis be employed to identify tourist behavior in the Athens?
(Huang, 2016)	Recommendation	Flickr Weather Underground API	Clustering method (DBSCAN) and collaborative filtering	Does context-aware methods provide location recommendations matching a tourist's travel interests and visiting context based on geotagged photos?

Table 2: Distibution of existing studies

In order to analyze attractiveness of heritage sites, questionnaires are applied to visitors, and analysis are carried out by the Kaiser-Meyer-Oklin Measure of Sampling Adequacy (KMO) value. While Kempiak et al. (2017) reveal that heritage settings (atmosphere), special events related to heritages, availability of well-informed staff and the conservation are the important factors for heritage experience, Vong & Ung (2012) emphasize that respondents have high opinions of heritages' historical value and preserving the local heritage in a good condition. In addition, history and culture, facilities and services at heritage sites, heritage interpretation and heritage attractiveness are distinctive factors for heritage tourists in Macau. Trinh & Ryan (2017) analysed the motivations of heritage visitors from different nationalities and their interpretations of heritage sites in New Zealand. It is found that culture is an essential determinant how tourists perceive a place and their experiences of visit.

ISBN 978-3-9504173-8-8 (CD), 978-3-9504173-9-5 (print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI, Clemens BEYER, Judith RYSER, Christa REICHER, Canan ÇELIK

Big data based studies are processed with different data sets and methods. Ganzaroli et al., (2017) analyse the efficiency of TripAdvisor on the quality of a restaurant as part of the cultural heritage of Venice and it is concluded that ranking of restaurants is strongly related to visitors' expected quality in Venice. Garcia-Palomares et al., (2015) focus on identification of tourists' hot spot based on social networks, and they reveal that uploaded photos are concentrated around monuments, tourist attractions, and museums. Tourists' photographs are clustered in the city center; however, locals' movements are extended such as parks and recreational areas. Girardin et al., (2009) carried out quantifying urban attractiveness using digital footprints and they are revealed that waterfront attractiveness is shown positive growth over the summer. Koutras et al., (2019) focus on tourist behavior using social network data in Athens and it is possible to define temporal tourist concentration in every POI, weekly, monthly and yearly time intervals by using spatio-temporal characteristic of Flickr data set.

Huang (2016) proposes context-aware location recommendation using geotagged photos and research suggests that experiences from past users in similar context can be helpful to choose where to visit. The experiment results in aggregating other tourists' travel histories matching in current users travel preferences and the context of the visit.

From this point o forward, it can be said that the visitation of heritage sites can be motivated by different attributes and they can be investigated for better understanding for what makes heritages attractive. It is possible to reveal people movement with high resolution spatio-temporal data by means of GPS and WiFi enabled devices and social media networks. Such data has location, time and user characteristic and they enable to researchers to investigate human behaviour in urban scale. It can be stated that overcrowding has a negative impact on heritages (van der Zee et al., 2018) and historical places are exposed to tourist pressure. UGC and WiFi sensors can be utilized to find solution and they can be employed to propose recommendation system to disperse people within historical urban core.

In order to define the network within the existing literature, VOS clustering technique is applied to articles which are downloaded from Scopus database. Keywords are selected considering the conceptual framework. VOSviewer is a software tool to construct and visualize bibliometric network and it offers text mining to construct the occurrence of important terms extract from a body of literature (VOS, 2019). Total number of 2149 articles are uploaded to VOSviewer. The association strength method is selected to normalize strength of the links between keywords, and visualization is done by occurrences. Figure 2 shows the connection between keywords and the relevance between keywords are emphasized by the distant of each frame.

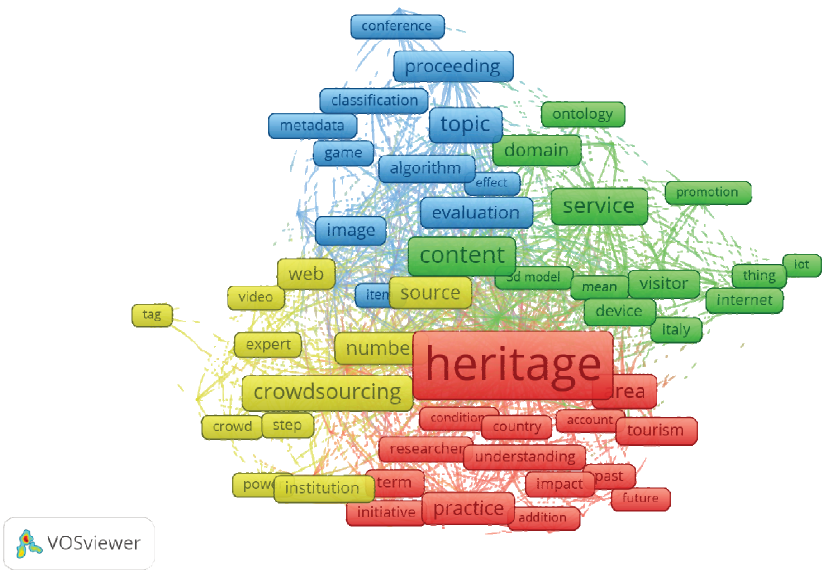

Figure 2: The network visualization of the most occured keywords

As it can be seen in the figure 2, the closest keywords for heritage related studies generally associated with crowdsourcing, content and device. On the other hand, there is a lack of interrelation between understanding, tag, classification, crowd, service, evaluation, tourism and visitor. The blue cluster and yellow cluster has the same number of keywords and they show the relation between "heritage AND context-aware" and "heritage

AND crowdsourcing" respectively. The second cluster (green) depicts the connection between "heritage AND big data", and the red cluster represents topic related to "heritage AND tourism".

Current literature already emphasizes the importance of utilizing newly available big data sources for volunteered based and data-driven management of historical cities. However, there is no study to utilize newly available big datasets for better understanding relation of heritage and visitors by taking into consideration heritage attributes, and heritage tourism problems related to overtourism.

4 METHODOLOGY

The intended research adopts different methodologies throughout the process to correspond to the objectives of each step. Table 3 shows the the planned methods; the combination of qualitative and quantitative analysis can be used to develop research.

Step	Research Question	Relevance	Methodology	Expected results
Step 1	What attributes contribute to attractiveness of heritage sites and what are the people's motivations behind the visitation of heritage sites?	Initial point	Literature review Stated choice experiment	The relevant attributes will be derived from state-of-the-art. A stated choice experiment will be designed considering these attributes. It is vital to understand people's interpretation and understanding to the heritage sites.
Step 2	What is the urban big data and how can big data be utilized within heritage sites?	Utilize the output of the stated choice analysis and to provide detailed understanding about urban big data and heritage sites	Qualitative (tag mining) and quantitative (regression analysis, DBSCAN algorithm, machine learning algorithm, neural networks)	The identification of hotspots and attractive heritage sites, and their relations between urban facilities will be investigated. People's flow between landmarks will be investigated by GPS and geotags. People's opinions and experiences will be derived from tags and images. There is not enough implementation of big data analysis with each attribute of heritage sites.
Step 3	How can urban big data (UBD) be leveraged for identifying the attractiveness of heritage sites in a dynamic way, and how can urban big data be used for distributing people evenly within heritage sites?	Respond the findings of the 2nd step, to provide detailed information about people' distribution within heritage sites and it explains what makes heritage sites attractive using big data	Case study in Amsterdam using big data sets (online textual/photo) Semi-structured interview around heritage sites	Developing different scenarios derived from spatio-temporal analysis and people's understanding are the main input of the recommendation system
Step 4	What are the visitor recommendation systems and how can new systems be developed using big data?	Final output (can be adapted as an application or website)	Context-aware location recommendation with collaborative filtering	Proposed system can contribute to reduce visitor pressure in heavily touristified areas and can be combined with real-time datasets

Table 3: Methods and expected results

The research question of step 1 will be responded by literature review and stated choice experiment. The literature review can be contributed to develop heritage attributes, since relevant publications can provide evidence based results. The experiment enables researchers to control certain factors (attributes), and it can be conducted with binary choices (two alternatives) or multinomial choices (more than two alternatives) (Johnson et al., 2007). The main aim of experiment is to have an understanding on people's preferences considering different heritage attributes. The descriptive statistics of users can be useful to analyse socio-demographic characteristics and representation of population. The model that emerged can have six attributes (Table 1); accessibility, curiosity, overcrowding, historical value, site uniqueness and ample relevant information. Intended duration for experiment can be a month with 750-800 respondents through the

ISBN 978-3-9504173-8-8 (CD), 978-3-9504173-9-5 (print)
Editors: Manfred SCHRENK, Vasily V. POPOVICH, Peter ZEILE, Pietro ELISEI, Clemens BEYER, Judith RYSER, Christa REICHER, Canan ÇELIK

web-based questionnaire. Results can be evaluated by multinomial logit (MNL), mixed logit and latent class logit models. As a result, coefficients of attributes' levels can be evaluated to interpret which choices are the best for respondents, so the importance of weight for each attribute can be identified to use in the recommendation system. Current studies are based on surveys or on-site questionnaires; however, they have some limitations. They can be applied a limited number of visitor, because the collection of traditional data takes more time than the collection of big data. Moreover, questions are asked on site; only visitors of certain heritage sites are able to respond questionnaire. Online survey can be applicable to large number of people, and it can be collected and analysed more efficiently compare to traditional survey.

Second phase of research focuses on the utilizing big data considering weights per attributes of heritage sites. Step 1 can result in extensive data set, and it can be important to analyse the results from questionnaire one by one, because each attributes can contribute to understand degree of influences each other. The step 2 focuses on newly available big data experiment to combine with the results of the step 1. Peoples' evaluations of heritage attributes can improve to develop better insight for understanding, since only big data cannot sufficient to interpret peoples' viewpoint to the heritage sites. Data collection can be done by coding such as HTTP-GET and GO. The Application Programmers Interface (API) of each data sources e.g., Flickr, Facebook, Instagram, Twitter contains metadata and it can be downloadable using parameters. The dataset from social media mainly contains time, location (lat.&lon.), tags, reviews, and photographs. The time stamps and locations of each photographs/reviews are valuable sources for urban research. Combination of heritage data and the spatio-temporal distribution of people can provide essential insight of their movement and preferences. In order to analyse relation among the urban facilities, the attributes of heritage sites, and people's behaviour, statistical methods such as geographical regression analysis can be done (Ioannides et al., 2018). Sentiment analysis can be carried out to investigate curiosity, site uniqueness and ample relevant information, because the tags of photographs/reviews contain textual description about visitors' experience. Moreover, tags can be utilized to visualize heritage values using Tag Clouds which represent frequency of tags (Ginzarly et al., 2018). The degrees of overcrowd can be assessed clustering methods such as DBSCAN, K-means and hierarchical clustering. Results can reveals the hotspots/POI(s), and they can be accepted as attractive points for people. The motivations of visitors to visit the heritage sites, and the influences of their visitations to the heritage sites by exploiting spatial big data and stated choice experiment can provide better understanding to the reasons of heritage sites attractiveness/popularity; therefore, results can contribute to develop recommendations in order to reduce visitors' pressure in heritage sites.

As suggested in the literature, overcrowd is an impotrant issue in historical cities and many destinations are facing the problem of overtourism (Seraphin et al., 2019). The step 3 attempts to explain how people can be distributed evenly in historical cities. The step 2 is the main input of this phase, as peoples' opinion about the heritage sites and big data analysis can provide novel approach to spread people around the heritage sites. After investigate the most attractive heritages and their attributes, unvisited/less visited heritage sites can be subtracted by means of GIS software such as ArcGIS, QGIS. The step 1 also can be evaluated to identify the reasons of less visitation; for example audio-visual guidance is available for some heritages and the weight of heritage interpretation can be analyzed whether it is effective to use such an audio-visual tool to increase the visitation or not. It is possible to apply spatial statistics within the GIS environments; therefore, it could improve to analyze relation between urban facilities and unvisited/less visited heritage sites visually. Spatio-temporal distribution of people in Amsterdam considering heritage sites can be analyzed, and it can be used as an input of step 4. The case study brings to better understanding for step 3 by involving tourist and local participation and semi-structured interviews could be designated for both target groups. Reducing concentration of tourists in hot spots by offering less crowded heritage sites can be a solution to minimize overcrowds in historical core. Big data-based analysis can provide space-time relation and it can be utilized to spread visitor flows' throughout the city even county level. Furthermore, it can be possible to identify areas where might be exposed to under tourism browsing previous space-time tags and less-known and forgotten heritage sites can be promoted as new hotspots.

The step 4 focuses on developing recommendation system for the visitors of heritage sites. It can be useful tool for visitors to find what they need from retrieving a wide range of data. Although many websites and applications are available to access relevant information before the trip, personalized recommendations which can combine several aspects such as purpose of trip (leisure), urban facility (museum), weather (summer), time (evening) enhance the heritage experience. It can help to visitor in retrieving information that affiliate with own preferences by recommending locations from a wide range of choice. Suggested heritage sites can be offered by selecting less visited heritage sites in which are exposed to under tourism. The results of stated choice experiment (step 1), big data analysis (step 2), and semi-structured questionnaire (step 3) can be input to find under touristified areas, because local residents have tendency to visit less touristic places and they can contribute to develop new recommendations.

Proposed system can be based on different datasets and knowledge, and it can allow to manage all the different kinds of information. The 4 steps can complete each other and results from previous step can be the input of the next step. The aim of this paper is to elaborate the approach for investigating heritage attractiveness, and to utilize big data sets to understand interactions between people and heritage through the lens of social media. These methods can provide location recommendation to visitors where to visit using similar context other people often visit. However, the aim is to reduce visitor pressure in heritage sites; therefore, suggested locations should be chosen considering less visited heritage sites. Timeframe of geotagged information can provide real-time data and, it can be developed as an application or web-sites which serve as a tailored (customized) guide for the visitors of heritage sites. It can provide personalized recommendation matching visitor's preferences including type of attraction (museum, park, art gallery), experience (positive), proximity to public transport (train, tram), historical value of heritage (aesthetic, social, spiritual). The crowdsourcing can support the maintenance of the system's knowledge base enriching visitor activities and recommendations can be supported by real-time data.

5 RESULTS AND FUTURE WORKS

Proposed recommendation system which consists of state choice experiment, urban big data analysis with user generated content and context-aware recommendation can result in the creative an effective way to distribute people in heritage sites. The first part of the research can contribute to understand the relevance of heritage attributes as heritage oriented and visitor oriented. After the stated choice experiment, results can be evaluated to analyse the impact of attributes over the heritage tourism. The most/least attractive points can show the people perception toward the heritage sites by considering weight of attributes. The big data based analysis can contribute to better understanding in fine spatio-temporal scale. The methodology that will be applied in this research can reveals the hotspots and it can be used as an input to distribute people evenly in the historical core. People's attitude to the heritage sites will be analysed with tags and it can enhance the knowledge about emotional expectations/experiences. Lastly, the case study in Amsterdam can be supportive to make an observation in a real conditions. The research not only contribute to cover a knowledge gap in literature, but also provide a holistic viewpoint relation between heritage and people.

The final output will contains deeper analysis about the reason of heritage visitation and attractiveness. Spatio-temporal pattern of people can provide the information about the usage of space in time. These type of information can be analysed by municipalities and companies to take measure in overcrowd areas such as it is possible that increase the frequency of public transportation in rush hours, and to provide special route for shuttles to promote less visited heritage sites. Destination Management Organizations can benefit from the result of the research, they can develop urban facilities in less touristic areas instead of heavily touristified zones. As a visitor perspective, final output can be helpful while making a decision where/when to visit, because real-time data flow can be embedded into proposed recommendation system.

This recommendation system can be tested using two different methods with a focus group around 30-50 people, who have experiences about overtourism. If the proposed system will be tested as a conceptual framework, participants can be informed about projected crowds in advance so that they can make an observation in the overtouristified heritage sites (i.e Amsterdam). On the other hand, if it will be tested as an application or website, notifications about expected crowds can be posted to the focus group and they can evaluate on site whether the proposed recommendation system work or not. Therefore, the system can be improved by considering focus group's experience.

6 REFERENCES

A. Morozov, M., & M. Morozov, M. (2018). The Influence of Cultural Heritage on the Attractiveness of the Tourist Destination. Proceedings of the 4th International Scientific Conference - SITCON 2018, 69–75. https://doi.org/10.15308/Sitcon-2018-69-75

Amanto, F., Moscato, V., Picariello, A., & Sperlı, G. (2016). Recommender Systems and Social Networks: An application in Cultural Heritage. 8.

Ashworth, G., & Page, S. J. (2011). Urban tourism research: Recent progress and current paradoxes. Tourism Management, 32(1), 1–15. https://doi.org/10.1016/j.tourman.2010.02.002

Boucher, P. (2019, May 20). Europe's Vacation Hot Spots Have a Message for Tourists: Sorry, We're Full. Retrieved October 9, 2019, from https://fortune.com/2019/05/19/europe-vacation-overtourism/

Centre, U.N.E.S.C.O.W.H. (2019). Historic Urban Landscape approach explained. Retrieved October 21, 2019, from https://whc.unesco.org/en/news/1026

Dane, G., Borgers, A., & Feng, T. (2019). Subjective immediate experiences during large-scale cultural events in cities: a geotagging experiment. Sustainability, 11(20), [5698]. https://doi.org/10.3390/su11205698

Falk, J.H., and L. D. Dierking. 1992. The Museum Experience. Washington, DC: Whalesback.

Frias-Martinez, V., Soto, V., Hohwald, H., & Frias-Martinez, E. (2012). Characterizing Urban Landscapes Using Geolocated Tweets. 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, 239–248. https://doi.org/10.1109/SocialCom-PASSAT.2012.19

Gaffar, Vanessa, & Wetprasit, Prateep. (2011). Comparative Study of Tourist Characteristics on Cultural Heritage Tourism Sites: Survey on Tourist in Indonesia and Thailand Heritage Sites. 3(3), 16.

Ganzaroli, A., De Noni, I., & van Baalen, P. (2017). Vicious advice: Analyzing the impact of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice. Tourism Management, 61, 501–510. https://doi.org/10.1016/j.tourman.2017.03.019

García-Hernández, M., de la Calle-Vaquero, M., & Yubero, C. (2017). Cultural Heritage and Urban Tourism: Historic City Centres under Pressure. Sustainability, 9(8), 1346. https://doi.org/10.3390/su9081346

García-Palomares, J. C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. Applied Geography, 63, 408–417. https://doi.org/10.1016/j.apgeog.2015.08.002

Ginzarly, M., Pereira Roders, A., & Teller, J. (2018). Mapping historic urban landscape values through social media. Journal of Cultural Heritage. https://doi.org/10.1016/j.culher.2018.10.002

Girardin, F., Vaccari, A., Gerber, A., & Biderman, A. (2009). Quantifying urban attractiveness from the distribution and density of digital footprints. International Journal of Spatial Data Infrastructures Research, 4, 26.

Huang, H. (2016). Context-Aware Location Recommendation Using Geotagged Photos in Social Media. ISPRS International Journal of Geo-Information, 5(11), 195. https://doi.org/10.3390/ijgi5110195

Ioannides, D., Röslmaier, M., & van der Zee, E. (2018). Airbnb as an instigator of 'tourism bubble' expansion in Utrecht's Lombok neighbourhood. Tourism Geographies, 1–19. https://doi.org/10.1080/14616688.2018.1454505

Johnson, F. R., Kanninen, B., Bingham, M., & Özdemir, S. (2007). Experimental Design for Stated-Choice Studies. In B. J. Kanninen (Ed.), Valuing Environmental Amenities Using Stated Choice Studies (Vol. 8, pp. 159–202). https://doi.org/10.1007/1-4020-5313-4_7

Karunanithy, M. (2013). The Impact of Heritage Attributes on the Satisfaction of Tourism in Sri Lanka. European Journal of Business and Management, 7.

Kádár, B. (2014). Measuring tourist activities in cities using geotagged photography. Tourism Geographies, 16(1), 88–104. https://doi.org/10.1080/14616688.2013.868029

Kemperman, A. D. A. M. (2000). Temporal aspects of theme park choice behavior: modeling variety seeking, seasonality and diversification to support theme park planning. Eindhoven: Technische Universiteit Eindhoven. https://doi.org/10.6100/IR542240

Kempiak, J., Hollywood, L., Bolan, P., & McMahon-Beattie, U. (2017). The heritage tourist: An understanding of the visitor experience at heritage attractions. International Journal of Heritage Studies, 23(4), 375–392. https://doi.org/10.1080/13527258.2016.1277776

King, L., Stark, J. F., & Cooke, P. (2016). Experiencing the Digital World: The Cultural Value of Digital Engagement with Heritage. Heritage & Society, 9(1), 76–101. https://doi.org/10.1080/2159032X.2016.1246156

Kourtit, Nijkamp, & Romão. (2019). Cultural Heritage Appraisal by Visitors to Global Cities: The Use of Social Media and Urban Analytics in Urban Buzz Research. Sustainability, 11(12), 3470. https://doi.org/10.3390/su11123470

Koutras, A., Nikas, I. A., & Panagopoulos, A. (2019). Towards Developing Smart Cities: Evidence from GIS Analysis on Tourists' Behavior Using Social Network Data in the City of Athens. In V. Katsoni & M. Segarra-Oña (Eds.), Smart Tourism as a Driver for Culture and Sustainability (pp. 407–418). https://doi.org/10.1007/978-3-030-03910-3_28

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META Group Research Note, 6, 70.

Laney, D. Beyer, M. (2012). The Importance of Big Data: A Definition. Retrieved April 01, 2015, from: https://www.gartner.com/doc/2057415

Long, Y., & Liu, L. (2016). Transformations of urban studies and planning in the big/open data era: A review. International Journal of Image and Data Fusion, 7(4), 295–308. https://doi.org/10.1080/19479832.2016.1215355

Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., & González, M. C. (2015). Urban magnetism through the lens of geo-tagged photography. EPJ Data Science, 4(1), 5. https://doi.org/10.1140/epjds/s13688-015-0043-3

Seraphin, H., Gowreesunkar, V., Zaman, M. and Lorey, T. (2019), "Limitations of Trexit (tourism exit) as a solution to overtourism", Worldwide Hospitality and Tourism Themes, Vol. 11 No. 5, pp. 566-581. https://doi.org/10.1108/WHATT-06-2019-0037

Song, H., & Liu, H. (2017). Predicting Tourist Demand Using Big Data. In Z. Xiang & D. R. Fesenmaier (Eds.), Analytics in Smart Tourism Design (pp. 13–29). https://doi.org/10.1007/978-3-319-44263-1_2

Thakuriah, P., Tilahun, N. Y., & Zellner, M. (2017). Big Data and Urban Informatics: Innovations and Challenges to Urban Planning and Knowledge Discovery. In P. Thakuriah, N. Tilahun, & M. Zellner (Eds.), Seeing Cities through Big Data (pp. 11–45). https://doi.org/10.1007/978-3-319-40902-3_2

Trinh, T. T., & Ryan, C. (2017). Visitors to Heritage Sites: Motives and Involvement—A Model and Textual Analysis. Journal of Travel Research, 56(1), 67–80. https://doi.org/10.1177/0047287515626305

UNESCO (n.d.). Retrieved November 5, 2018, from http://www.unesco.org/new/en/culture/themes/illicit-trafficking-of-culturalproperty/unesco-database-of-national-cultural-heritage-laws/frequently-askedquestions/definition-of-the-cultural-heritage/

van der Borg, J., Costa, P., & Gotti, G. (1996). Tourism in European heritage cities. Annals of Tourism Research, 23(2), 306–321. https://doi.org/10.1016/0160-7383(95)00065-8

van der Zee, E., Bertocchi, D., & Vanneste, D. (2018). Distribution of tourists within urban heritage destinations: A hot spot/cold spot analysis of TripAdvisor data as support for destination management. Current Issues in Tourism, 1–22. https://doi.org/10.1080/13683500.2018.1491955

van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., Smith, J. W., Tieskens, K. F., & Verburg, P. H. (2016). Continental-scale quantification of landscape values using social media data. Proceedings of the National Academy of Sciences, 113(46), 12974–12979. https://doi.org/10.1073/pnas.1614158113

Vong, L. T.-N., & Ung, A. (2012). Exploring Critical Factors of Macau's Heritage Tourism: What Heritage Tourists are Looking for when Visiting the City's Iconic Heritage Sites. Asia Pacific Journal of Tourism Research, 17(3), 231–245. https://doi.org/10.1080/10941665.2011.625431

VOS, Visualizing scientific landscapes. (2019). Retrieved November 13, 2019, from https://www.vosviewer.com

Zhang, J., Zhang, J., Huo, X., Zheng, W., Zheng, X., & Zhang, M. (2017). Research on the Positioning of Protection and Utilization of Historic Districts under Big Data Analysis. Isprs - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W5, 731–735. https://doi.org/10.5194/isprs-archives-XLII-2-W5-731-2017