A Location Analytics Method for the Utilisation of Geotagged Photos in Travel Marketing Decision-Making

Citation: SJ Miah, HQ Vu, JG Gammack (2018). A Location Analytics Method for the Utilization of Geo-tagged Photos in Travel Marketing Decision-Making, Journal of Information and Knowledge Management (in press)

A Location Analytics Method for the Utilization of Geo-tagged Photos in Travel Marketing Decision-Making

Shah Jahan Miah
College of Business, Victoria University
Footscray Park Campus, Melbourne, Australia
Email: shah.miah@vu.edu.au

Huy Quan Vu
School of Engineering and Technology,
Central Queensland University,
Melbourne, VIC 3000
Email: h.vu@cqu.edu.au

John G. Gammack
College of Technological Innovation
Zayed University, Dubai, UAE
Email: John.Gammack@zu.ac.ae

Abstract

Location analytics offers statistical analysis of any geo- or spatial data concerning user location. Such analytics can produce useful insights into the attractions of interest to travelers or visitation patterns of a demographic group. Based on these insights, strategic decision-making by travel marketing agents, such as travel package design, may be improved. In this paper, we develop and evaluate an original method of location analytics to analyze travelers' social media data for improving managerial decision support. The method proposes an architectural framework that combines emerging pattern data mining techniques with image processing to identify and process appropriate data content. The design artifact is evaluated through a focus group and a detailed case study of Australian outbound travelers. The proposed method is generic, and can be applied to other specific locations or demographics to provide analytical outcomes useful for strategic decision support.

Keywords: travel agency, travel marketing, social media, geo-tagged data, location analytics

1. INTRODUCTION

Travel management agencies generally offer travel packages and promotional materials for their potential customers such as pleasure tourists or business travelers. Managers of travel agencies must design attractive travel packages/products to best suit various travelers' needs. It is important to closely understand travelers' interests and trends at target locations globally (Lin et al., 2015). A travel agency works behind the scenes to design effective solutions for their future consumers, which are guided by highly specific and up to date information. Along with other methods, data-driven marketing within the tourism and travel industry has become essential (Park et al., 2016).

The availability of social media sites has offered common platforms, where travelers express and share their travel related information such as photos, comments, tips, location, and other content. Such data are embedded with useful information about tourists' interests and activity at destinations. The study of traveler-generated online content for travel marketing and tourism management is an emerging research trend. Studies have been conducted that support travelers by providing travel related information (Xiang and Gretzel, 2010; Fotis et al., 2012; Ayres et al., 2002) and developing travel recommendation systems (Zhuang et al., 2016; Majid et al., 2012). Other studies aim to aid the decision making of managers in tourism destination management, or the forecasting by domestic (i.e. within country) or outbound tour organizers offering packages to demographic groups, by providing further insights into tourist behavior and travel patterns (Chua et al., 2016; Wong et al., 2006; Shen et al., 2011). These methods may identify flows between locations, yield for specific demographics or information seeking and choice behaviors. Such insights, which can result from data-mining, complement the quantitative and aggregated data available from analysis of departure card samples and surveys. There remains however, a lack of an appropriate data-driven approach for capturing and analyzing strategically the rapidly growing location-oriented data becoming available.

There has been a growing realization that organizations can dramatically enhance their insights by adding geographic location to business data (ArcNews, 2012). This allows identification of patterns specific to localities and regional areas at arbitrary levels of granularity, where maps and spatial analysis provide a new analysis context not achievable with traditional tables and charts. Location oriented insights have shown potential in strategic decision support such as for building facility management (Warmerdam and Pandharipande, 2016), for greater understanding of spatial configuration of street and roads networking (Li et al., 2016), and for detecting preferred tourism destinations (Zhow et al., 2015). However, these methods do not offer specific user analysis and support aids relevant to strategic travel management. They were neither designed to support automated searching and analysis nor for discovering the complex behavior of outbound travelers at different locations. Organizational decision-making and product development can benefit by combining highly specific analytics with traditional and independently gained insights.

Aiming to address the abovementioned shortcoming, this paper introduces the design of a location analytics solution that addresses travel agency managers' strategic information needs, described formally as a *method*[1], a type of information systems artifact in design science research. We conducted a design research study by adopting the principles of Hevner et al. (2004) to address a data-driven marketing strategy (Provost and Fawcett, 2013). We propose a generic solution method, architectural framework and functional tools oriented towards informing decisions, specifically for tourism destination marketing. The solution method is capable of collecting, processing, integrating, and representing location-oriented insights from a location-oriented social media site, such as Flickr - one of the most popular photo sharing sites that allow users to post photos and other content.

Tourist photos nowadays are frequently taken by smart photo capturing devices with built-in Global Position Systems (GPS), such as smartphones, smart cameras and tablets. The photos are annotated with information, including title, tags, description, time, and especially geographical location. Geo-tagged photos identify the geographical location of each photo using GPS coordinates, which uniquely specify every location on Earth by a set of numbers, *latitude* and *longitude*. The latitude represents the angular distance of a place north or south of the earth's equator, while the longitude is the angular distance of a place east or west of the prime meridian located in Greenwich, London. The

[1] Method as a solution artifact defines processes and provides guidance on how to solve problems. The methods can range from "formal, mathematical algorithms that explicitly define the search process to informal, textual descriptions of best practice approaches, or some combination" (Hevner et al., 2004, pp. 79).

latitude and longitude values of the geotagged photos are expressed in decimal degrees format in the ranges between ±90° and ±180° respectively.

We evaluate our location analytics approach through a case study demonstration of Australian outbound travelers' activities, one of the fastest growing outbound tourism sectors (TRA, 2016). Photos that are freely available for public access can provide valuable data sources for location oriented insights, and nowadays most people, across all demographics, use smartphones to take and upload images. A large and increasing number of Australians travel overseas every year to various global destinations (TRA, 2016), and are no exception to this pattern. Analysis of images from their diverse travel behavior therefore provides a representative sector with which to demonstrate the advantages of our proposed approach. The outcomes will also be of particular interest to package and tour providers in the identified countries, since it provides insights into what Australian demographics seek in specific destinations.

The remainder of the paper is structured as follows. The next section provides details of background literature relevant to the design of the solution artifact. Details of the research methodology follow and, in the third section, we specify details of the design artifact, demonstrating our theoretical contribution in the context of previous, relevant work. After that we demonstrate the value of the artifact, according to utility and usability criteria, within the case study context. The final section contains an overall discussion, including details of study limitations and identification of avenues for further research.

2. RELATED WORKS

This section first provides an overview of existing work in the areas of location analytics and previous approaches for analyzing geotagged photos, which is then followed by discussion on the need for location analysis in travel marketing.

2.1 Overview of Location Analytics

Location analytics is a segment of business analytics that is used for the purpose of producing geographic intelligence. ArcNews (2012) defined location analytics as focused on "thematic mapping and spatial analysis for the world of business analytics". Various new analytics approaches have been specified under the umbrella of location data analytics paradigm, with applications based on either non-observational or observational data collection approaches (Chua et al., 2016). The former relies on recall diaries or self-administered diaries that require direct contact with participants. The later relies on sensors to passively track movement and behavior through social media or virtual communities. It has been suggested that non-observational approaches face several shortcomings since the quality of the data gathered depends on the subjects' efforts and collaboration (Thornton et al., 1997). Chua et al. (2016) also indicated that the spatial-temporal precision of the gathered data tends to be comparatively lower (in non-observational approaches) than those obtained with observational methods due to the analogue data collection procedure.

Recent studies suggest two main classes of location analytics solutions – location *predictive* analytics and location *extraction* analytics. The former class offers data-driven solutions for forecasting future behavior or trends. For instance, a location predictive analytics solution in terms of a recommender system, called GeoSRS, was developed that utilizes geolocation data to predict and recommend potential locations that the user may be interested to visit next (Capdevila et al., 2016). To suggest routes at a chosen location, work by Okuyama and Yanai (2013) proposed a "trip model" system based on sequences of geotagged photos, overcoming limitations of earlier work which had relied on expensive GPS devices, but does identify demographic or comparative destination popularity data, rather options among possible routes through attractions at a preselected location. Later work (Oku et al., 2015) aimed to improve recommender accuracy by associating geotagged tweets with contemporaneous photos but focused more on algorithmically improving granularity than on strategic application. Social media big-data in terms of collection of geotagged photos generated by visitors have also been studied to discover sequences of visited locations to build travel histories of users, and different methods were proposed to find popular locations or representative travel sequences to address traveling-related queries. For example Majid et al., 2012 aimed to personalize suggestions better by identifying an individual user's interests and relating those to similar attractions in another (Chinese) city. At large scales, such information is of interest to city planners and destination managers.

Big-data from a firm's own web forum has also been used to enhance managers' understanding of stakeholders' concerns. In (Jiang et al., 2014) for example, in-house forum participants were grouped using a probabilistic clustering technique taking similar topics of interest and other clues from written content. At different stages of an event, relatively important stakeholder groups (e.g. "investors") were identified and associated with improvements in prediction of stock market performance.

While this research draws on an approach used in marketing for segmentation, other techniques come from other fields. For example, in such fields as architecture and environmental planning, *space syntax* references the idea that space can be conceptualized as configurable components, which can be mapped with relevance to social usage patterns. Li et al. (2016) developed a location analytics solution using a space-syntax technique to analyze tourist movement pattern based on cell phone locations data. Their aim was to provide understanding of the relationship between location features, such as streets and road networking, and tourist preferences. Such knowledge is important for city planners and designers to develop a rich understanding of city inherent morphological structure, and its impact on, for example, wayfinding. Li et al.'s analytics are not used to analyze and acquire intelligence from geo-tagged big data, but used data related to tourists' activities drawn from sensors, which arguably provides more limited data sets than from social media, due to relative media richness. Related approaches come from building facilities management: e.g. Warmerdam and Pandharipande (2016) used wireless sensors in the lighting system to detect signal strengths transmitted from users' mobile devices, which allows zone position data to be analysed for crowd detection, occupation density and similar applications.

Table 1 illustrates key examples of location analytics solution for data extraction. None of these studies, however, go beyond to explore ways in which users engage and perceive geographical areas with a view towards identifying attitudes and personal interest in these geographical areas (Naaman 2011; Majid et al., 2012).

Table 1: Relevant studies related to location analytics solutions for data extraction.

Model artifact	**Key techniques**	**Purpose of application**

Travel analytics (Huang et al., 2017)	Spatial and statistical analyses	Understanding the destination choice to predict the demand side for businesses in a particular location
Mobile application for supporting travelers (Zhuang et al., 2016)	Simulation technique	Location-centric social networking game that offers recommendations from other people or statistics
Travel recommendation systems (Majid et al., 2012)	Semantic annotation of location	Providing summary of opinions to help individuals for trip planning
Location analytics (Warmerdam and Pandharipande, 2016)	Received Signal Strength Indicator (RSSI)-Based Zone Positioning	Providing occupancy density in locations
Space syntax analytics (Li et al., 2016)	Space Syntax analysis	Understanding of locations of visitation pattern in China
Cloud based analytics (Zhow et al., 2015)	Image processing and spatial analysis	Detecting and ranking places of interest at various locations.

2.2 Issues for Travel Marketing

Tourism and travel industries play an important role in the economic growth and sustainability of many countries. Numerous attempts have been made to study behavior of travelers to support strategic planning and decision making in the tourism industry. For instance, Lim (2004) focused on identifying preferences and factors affecting destination choice of Korean travelers to Australia, such as seasonality and recreation vs. business travel. Other factors influencing travel demand have also received considerable research interest, such as wealth effects (Kim et al., 2012), exchange rates (Yap, 2013), income level and relative difference in the cost of a stay between local and international destinations (Lin et al., 2015), and travelers' demographic profiles (Prayag et al., 2015). Other works have focused on outbound tourism forecasting, which include tourists' own analysis (Furmanov et al., 2012), destination prediction (Wong et al., 2006), seasonal demand analysis (Shen et al., 2011) and demand elasticities modeling (Seetaram, 2012). However, solution design studies to aid decision making that particularly target travel marketing aspects have been limited.

The emerging literature on location analytics can support effective solution design. "Location analytics adds geographic, demographic, economic, and similar types of information to the financial and other data that companies already collect" (Garber, 2013, p. 15). It implies that location oriented data processing may offer promise for identifying demand that more finely supports strategic activities, for example, the design of tour packages and seasonal airline promotions.

This paper attempts to address the issues in travel marketing by proposing a location analytic method for geotagged social media data, whose details are given in the next section.

Our proposed approach differs from the previous approaches in several ways. First, previous works are limited to a more specific focus on recommending popular places, locations and general visitation patterns, or on techniques to improve these. Our proposed method has a strategic focus and is comprehensive in extracting insights into activities and interests by incorporating both textual meta-data and content of the actual photos. Second, prior approaches usually focused on one or only a few specific destinations. Our work detects attractions for different tourism destinations as well as the difference between them using emerging pattern mining techniques, and allows for analysis anywhere in the world, and at arbitrary levels of granularity. Third, our work aims to solve a real-world decision-making problem within the travel industry through applying a systematic design methodology (i.e.. design science research), whereas most of the previous methods describe a technology development and do not necessarily consider a more generalizable contribution to knowledge. Finally, the effectiveness of our method was evaluated with industry practitioners, which was not demonstrated in most of the prior works.

3. METHODOLOGY

In recent years Design Science Research (DSR) has gained a lot of attention among solution development researchers both for better embodying explicit consideration of practical relevance, and promoting theoretical rigor. DSR "seeks to create innovations that define the ideas, practices, technical capabilities, and products through which the analysis, design, implementation, management, and use of information systems can be effectively and efficiently accomplished" (Simon, 1996, p. 76). Through explicit attention to theory, DSR is increasingly viewed as a research-appropriate form of IT artefact development, and one which provides a contribution to knowledge beyond specific solution development for any defined problem solving purposes. DSR approaches provide guidelines and prescriptive methodological support for development, implementation, evaluation, and adaptation of artefacts for problem solving. DSR not only provides solutions for identified organizational problems but also provides a new dimension in designing solutions for many complex problems (Baskerville et al., 2010). Since locational analytics solutions are composed of adaptable GPS platforms, software, and human interfaces and analytic methods we base development of our solution on DSR as we discuss next.

3.1 Approach based on Design Science Research

"Science analyses the existing world to create new knowledge, design uses existing knowledge to create a new world" (Verkerke et al, 2013, p.195). Paraphrasing this, Baskerville et al. (2015, p. 542) describe a design-science study as broadly representing (1) a design-science research project, (2) an artifact "build and evaluate" project that such a research project may entail, (3) the production of new knowledge from design-and-development, and (4) the creation of reports or articles describing this design-science research project. Our study follows this model to produce new understanding from an artifact design. DSR produces five distinct types of artifacts: constructs, models, methods, instantiations and design theories (Gregor and Hevner, 2013). In our study, we design a method, one of the artifact types, and ensure that a construction process can be replicated and applied beyond a simple IT development project. The problem space to which the method applies is scoped, and problems for which solutions were previously difficult or impossible are addressed. In this case, handling unstructured and volatile large datasets for strategic decision support in the travel industry is the goal of the approach. While we focus on decision making aspects relevant to travel decision makers such needs are likely to be in common with other marketing practices, or other spatial applications such as urban designs and crowd control.

Our adopted methodology is based on three broad activities, we designate as preparation, processing and delivery (Figure 1). These firstly *identify business problems and relevant artifact types;* followed by *artifact creation and evaluation*; and finally *outcomes and their communication of the result*. These phases are constructed upon Hevner et al.'s (2004) seven guidelines for design research, (Table 2): a DSR framework selected because it proposes specific activities towards designing IT tools that enhance organizational capabilities. In our case this means understanding travelers better, and the mapping of the design research activities to our specific case is shown in Table 2.

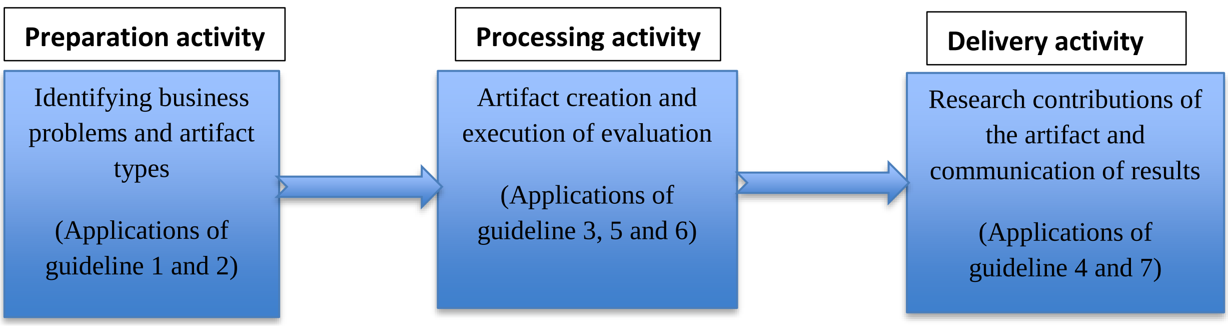


Figure 1: Three activity oriented methods adopted in this study

Table 2: DSR study phases and guidelines (Hevner et al., 2004, p. 83)

Hevner's Guidelines	*Descriptions associated to the artifact design*
Guideline 1: Design of an Artifact.	The study has produced an innovative location analytics method designed to identify travelers' interests relevant to traveling marketing managers
Guideline 2: Problem Relevance.	There is a need to develop appropriate and efficient analytical methods to leverage unstructured location enabled meta-data, specifically in the domain of travellers' movements at targeted destinations.
Guideline 3: Design Evaluation.	To demonstrate artifact utility, experiments and focus groups have been conducted.
Guideline 5: Research Rigor.	The location analytics artifact was constructed using established data mining techniques and proven algorithms
Guideline 6: Design as a Search Process.	Data mining modelling methods were adapted for use in this study. The design process was iterative in order to cope with uncertainty inherent in the problem space and the method was designed to enable solutions.
Guideline 4: Research Contributions.	The method is theoretically innovative since current approaches do not specify efficient methods for producing detailed location specific insights.

Guideline 7: Communication of Research.	The artifact has been presented to travel industry professionals who evaluated it for usability and efficacy in a focus group setting, and the experimental outcomes have shown a clear benefit to representative target decision makers as reported in academic literature.

3.2 Key aspects of decision making in travel marketing

Travel destinations may be considered as complex products with various tangible and intangible elements (Blázquez et al., 2012). Due to resource limitations, it is crucial for agencies to identify what key areas to focus on and to promote these in packages and related products. An issue with existing approaches to tourist behavior analysis is that outbound package designers and other travel professionals usually predetermine what should be included. As such, important visitor interests may not be considered at all. Agencies have, to date, been unable to form a precise picture of what activities visitors have actually been involved in. For many tourism destinations (typically with a wide variety of different attractions), this information is crucial. Furthermore, traditional approaches to information gathering for management and planning purposes have previously relied heavily on surveying and questionnaires. This is time-consuming and not particularly effective. Agencies and marketing managers alike still face major difficulties in obtaining accurate answers to the following critical questions: What attracts travelers when visiting a destination? What locations do travelers visit to explore these attractions? What are the significant differences in travelers' interests at different destinations? What are the travelers' personal experiences at each of the visited destinations? By making use of user generated geo-content, a planner could obtain comprehensive information on travelers' activities, their experiences and personal reflections.

3.3 Data collection and processing Techniques

Our location analytics solution supports decision making both of business managers and travel agencies in tourism marketing. The solution artifact aims to extract insights into tourist behavior and interests from geotagged travel photos, which consists of four stages: 1) Travel Data Collection: geotagged photos available on photo-sharing platforms are extracted; 2) Meta-Data Processing: processing photo meta-data identifying tourist interests and visited attractions at different locations; 3) Emerging Pattern Mining: utilized to identify emerging attractions for different tourism markets; and

4) Location Representative Photo: image processing techniques identify popular and high quality photos useful for developing marketing material for destinations, and developing tailored tour products aimed at specific demographics.

3.3.1 Travel Data Collection

Geotagged photos are available on photo-sharing platforms such as Flickr (www.flickr.com). As a proxy for chosen and directly experienced locations, we make use of the photo data on Flickr, because it is a reliable and representative data source for capturing traveler's interests at particular locations. The photos are extracted using *PhotoSearch* function, (documentation available at www.flickr.com/services/api.) We now describe this first process in more technical detail.

With Flickr it is impossible to identify individual users whose photos should be downloaded. Thus, we first search for seed photos within a location, where the interested people are likely to take a photo. A bounding box is defined to cover the targeted location, whose coordinates are defined by $min_{la}, min_{lo}, max_{la}$, and max_{lo} for minimum latitude, minimum longitude, maximum latitude and maximum longitude respectively. For instance, if we are interested in travel location for domestic Australia Travelers, then we search for geotagged photos available within Australia. Then, we identify Australian travellers based on data identifying the owner's location of origin. The *userID* of the selected photo users are used as search queries for *PhotoSearch* function to retrieve their photo collections respective to specific global locations. Since, we are interested in outbound travel behavior at locations; only photos taken outside of the specified bounding box are kept, as they are likely to be taken during outbound trips to other countries. In addition to the bounding box, *PhotoSearch* allow user to specify the search query by date, t_{min} and t_{max} for earliest and latest time. Only photos taken within the specified time range are returned. If the time information is not specified, all photos available in the photo collection of the users are returned. We extract the photos of the selected users and their associated meta-data such as tags, description, title, geographical information (latitude and longitude) and time of snapshot. These data will be used in the later stages for traveller's interest identification.

3.3.2 Meta-Data Processing

The textual metadata (tags, description and title) of the photos often contain specific keywords identifying things or objects of interest to the photographer. We apply a number of text processing technique to process the textual metadata in order to identify traveler's interests. Suppose a geotagged photo data set P was collected, with the textual metadata of each photo p_i denoted as $t_{(p)}$. The text processing is carried out as follows:

(1) For each photo, *p*, its textual meta-data $t_{(p)}$ is first loaded into a text *tokenization* algorithm to break the text stream into words, phrases, symbols or other meaningful elements called "tokens".

(2) A text *filter* is applied to the tokens to remove elements such as symbols and numbers. All letters are normalized into lower case.

(3) A stemming algorithm is applied to the remaining tokens to reduce inflected words to their stem, base, or root form. For example, the words "shopping" and "markets", are reduced to "shop" and "market".

(4) In the English lexicon, each word is associated with a set of tags indicating their types such as nouns, verb or adjective. We assume the vocabulary of noun type is used to refer to entities of interests and attractions. Therefore, a stemmed noun list was constructed from the data set and denoted as $N = \{ n_1, n_2, \dots \}$

(5) In the context of travel marketing, popularity of travelers' interests should be measured by the number of travelers at locations, rather than the number of photos. We measure the popularity of each noun n_j based on a support value $supp(n_j) = \frac{|n_j|}{|O|}$, where $|n_j|$ is the number of users, whose textual meta-data of photos contain the noun n_j, and $|O|$ is the total number of users in the data collection.

The advantage of the text processing technique is that the nouns mentioned by the travelers can be automatically collected. Business managers can inspect this list and select the nouns describing traveler's interest for further analysis. The support values help to identify the most popular interests. In application to outbound travel, the support values can be examined with respect to different travel locations so the most popular attractions for each location can be revealed. We demonstrate this in the case study.

3.3.3 Emerging Pattern Mining

In the context of traveling marketing, business managers are not only interested in the most popular attractions but also in the differences between travel locations. Specific marketing strategies can be developed that promote the uniqueness of each travel location. However, it is not practical to manually examine traveller's interests for all possible locations to identify significant differences. We adopt emerging pattern mining (EMP) technique to address this issue. Dong and Li (1999) defined emerging patterns as item sets, whose support increases significantly from one data set to another.

Let $I = \{ i_1, i_2, \dots, i_n\}$ be a set of items. Subset $A \subseteq I$ is called a k-item-set, where $k = |A|$. The support for an item set, A with respect to a group G_p, is denoted a $supp(A, G_p)$s, which reflects how frequently A appears in this group. The change in the support for A from a group G_p to a group G_q is measured by a growth rate metric defined as:

$$Growth(A, G_p, G_q) = \begin{cases} 0 \text{ if } supp(A, G_p) = 0 \text{ and } supp(A, G_q) = 0 \\ \text{inf if } supp(A, G_p) = 0 \text{ and } supp(A, G_q) \neq 0 \\ \dfrac{supp(A, G_q)}{supp(A, G_p)} \text{ otherwise} \end{cases}$$

Given $\delta > 1$ as a growth rate threshold, an item set A is called emerging pattern if $max_{(p,q)}\{Growth(A, G_p, G_q)\} \geq \delta$, where $max_{(p,q)}$ mean that group G_p and G_q can be of any order, but the largest growth rate is compared with δ.

In our scenario, traveler's interests are represented as stemmed nouns, while its popularity is measured by the support values. Thus, the stemmed noun list $N = \{ n_1, n_2, \dots\}$ can be treated as a set of items as $I = \{ i_{n_1}, i_{n_2}, \dots\}$. The photos taken in a specific location p is treated as a group G_p. Interpretation of an interest on its own, rather than as part of an item set containing many others, is easier and more meaningful. When the number of items $i \in I$ is large, the use of an item set A with one item $k = 1$ is suggested. The support of each item with respect to each group is computed as $supp\left(i_{n_j}\right) = \frac{|n_j|}{\left|O_{G_p}\right|}$, where $\left|O_{G_p}\right|$ is the number of owner, who visited tourism location p. The growth rate of the support values for the stemmed nouns between locations can be computed to identify the emerging interests. Note that, in the original computation of emerging pattern, only the ratio of the support values was accounted. There are cases where $\frac{supp(A,G_q)}{supp(A,G_p)}$ is large but the

$supp(A, G_q)$ and $supp(A, G_q)$are small values. Item sets with small support values are not useful for business managers in marketing as not many travelers are interested in them. Therefore, we define a support threshold γ to identify emerging patterns, which have significant support. Only emerging patterns whose difference satisfies $Diff(A, G_p, G_q) = |supp(A, G_q) - supp(A, G_p)| \geq \gamma$, are retained for further analysis.

3.3.4 Location representative photo identification

Travel products, brochures and online travel websites are the major means of communication between travel agencies and travelers. Visual assets such as photos are important components to develop effective marketing material and connote a location by an iconic image. This section presents our approach that identifies representative photos of locations according to tourist interests. In our artifact, we define representative photos as those photos whose contents appear most frequently in a set of photos. Suppose G_p is a group of photos taken at a travelling location *p*. For each specific interest i_{n_j}, a subset of the photos $S \subseteq G_p$, which contain the keyword n_j is selected. Since *p* can be at a large scale such as city or a country, it is necessary to identify specific points of interest before representative photos are identified. We address this task by applying a clustering technique, named P-DBSCAN (Kisilevich et al., 2010), to the GPS data of the photos in *S*. Two parameters are required for the clustering algorithm, *r* for neighborhood radius and N_O for an owner number threshold. For each photo $d \in S$, its neighboring photos, whose location is within the radius *r* from *d* and owned by a different owner of *d*, are identified. If the neighboring photos of *d* are owned by at least N_O number of users, *d* is selected to a current cluster; otherwise *d* is discarded from *S*. The process iterates for all photos in S and results in a set of clusters $C^{n_j} = \{ c_1^{n_j}, c_2^{n_j}, ... \}$ for the interest i_{n_j}. For instance, a traveler coming to Britain may be interested in visiting castles, however, there are many possible castles all over Britain. P-DBSCAN helps identify the most popular castles, where many travelers visited and took photos.

The next step is to identify representative photos at each point of interest. The photo content is then represented as a bag of visual words, a popular and effective approach to represent photo content in object recognition and image classification (Yang et al., 2007). The visual words are constructed by applying k-means clustering to local region descriptors of sample photos. Cluster centers are then used as vocabulary of visual words. Here, the local region descriptors are based on Speeded-Up Robust Features (SURF) (Bay et al., 2008), an advanced feature descriptor and robust to occlusions and spatial variations (Vailaya et al., 2001). The bag of visual words for each photo is denoted as $\mathbf{v}^d = \{v_1^d, v_2^d, \ldots, v_w^d\}$, where the value of each element v_z^d is the number of times the visual word v_z appear in photo d.

We assess the popularity of photo visual content using multivariate kernel density estimation (M-KDE) technique (Terrell and Scott, 1992), a non-parametric way to estimate the probability density of multiple random variables. Although the choice of the kernel function is not crucial to the accuracy of KDE (Bowman and Azzalini, 1997), the number of variables can affect its performance (Nagler and Czado, 2015). Therefore, we apply Multi-Dimensional Scaling (MDS) technique (Terrell and Scott, 1992) to the visual words before input into KDE. MDS helps to reduce the dimensionality of the data points, while still preserving the similarity or distance between them. Each bag of words v^d having *w* dimensions is transformed into a low dimensional vector $\widehat{v^d}$ having $\widehat{w}$ dimensions, where the chosen value of $\widehat{w}$ is small (2 or 3). M-KDE is then applied on to $\widehat{v^d}$, and the photos with highest probability densities are returned as representative. Whilst we have necessarily described this in technical terms to allow replication or future comparison with other techniques, the essential steps of the method they address should be intuitively clear to non-technical audiences, and we demonstrate this in practice in the next section.

4 ARTIFACT DESCRIPTIONS

The method we proposed integrates four sub-processes: 1) Travel Data Collection; 2) Meta-Data Processing; 3) Emerging Pattern Mining; 4) Location Representative Photo Identification, whose conceptual framework is shown in Figure 2. In brief, geotagged photos posted by specific groups of travelers are extracted from Flickr based on a user defined bounding box and *userID* identified from a set of seed photos. The photos are downloaded together with their associated metadata. The textual meta-data are then processed and combined with location information to identify the interests of travelers at different locations. Emerging pattern mining is applied to the interests at different tourism markets to identify emerging attractions based on user defined growth rate threshold δ and difference threshold γ. Finally, the geotagged photo clustering technique (P-DBSCAN) is combined with image processing techniques to identify popular and high quality photos which can be used for developing marketing material for destinations.

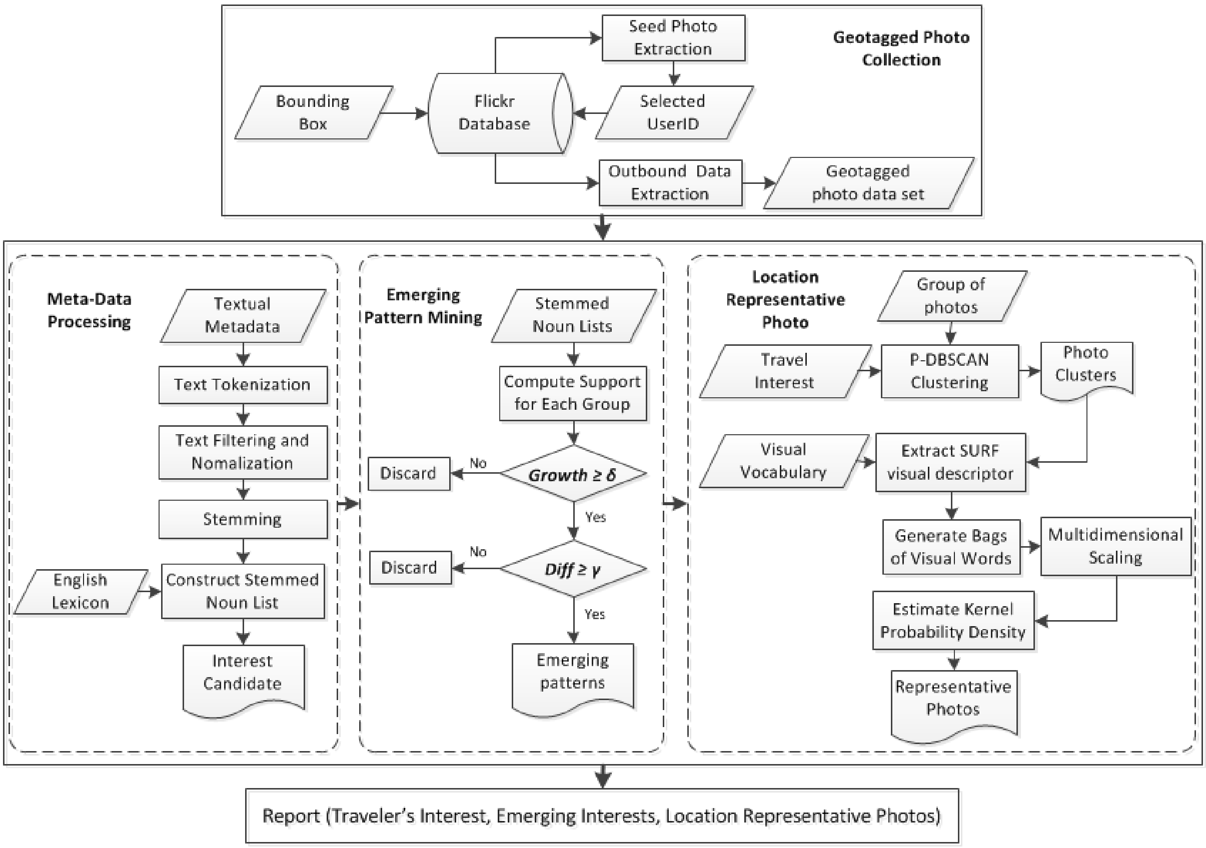


Figure 2: Architectural framework of the location analytics artifact

4.1 Design skeleton

In specifying the artifact we adopt the theoretical view of design presented by Gregor and Jones (2013) that defined eight structural components of IS design research. The components are (1) purpose and scope, (2) constructs, (3) principles of form and function, (4) artifact mutability, (5) testable propositions, (6) justificatory knowledge, (7) principles of implementation, and (8) an expository instantiation. These structural components also affirm and utilize the design science considerations for artifact design and implementation discussed by Iivari (2007). This study thus followed established guidelines of IT artifact design to develop the functionalities of the proposed artifact, detailed in table 3.

Table 3: Structural components of IS design research

The specification of the design theory	**Presented in this research**
The purpose and scope of the artifact are addressed	Designing a location analytics method for extracting location specific insights for managers' evidence-based decision support towards effective strategic directions. Sections 1 and 3 describe the purpose and scope of the study.
Principles of form and function incorporating underlying constructs are described	For the artifact, four sub-processes were used to architect the method. The principles related to representation of, and computational processes around, geotagged photos are detailed in Section 3 and illustrated in figure 2. To meet relevance requirements, we used iterative development and evaluation for the method.
Artifact mutability is addressed	The design artifact has been formulated to meet the need of strategic demand for travel agency product design and development. The method is specified generically to enable location specific analysis through geotagged data from social media. The architecture allows replacement of specific algorithms and techniques as these emerge, and the provision for alternative data sources and parameter settings allow a range of tailored applications to be developed in future.
Testable proposition of the design artifact is defined	Evaluation of the IT artifact with focus group for capturing qualitative data and case data demonstrated for relevance through comparative experiments in a laboratory setting are described (section 5).
Justificatory knowledge (kernel theory) is provided	The proposed method as an IT artifact is based on various sub processes and data mining techniques. The iterative development and evaluation process are also well established and are related to real and known issues within the context of problem domain (section 4).
Principles of implementation are presented	We described the sequenced integration of sub-processes for collecting, analyzing and processing a geotagged data set to provide more comprehensive understanding of travellers' interest on attractions. Section 3 gives technical details, and a worked case study in section 5 illustrates in practice.

An expository instantiation is given	The method is applied to representative outbound destinations for Australian travellers. The exposition and the generic value of the method is further validated by its presentation both to academic and industry audiences.

5. ARTIFACT EVALUATION USING CASE STUDY

Gregor and Hevner (2013) suggested that DSR artifacts must be evaluated for reliability, validity, and utility through data from the case studies that were used to inform the artifact's design. Hevner et al. (2004, p.77) noted that "a mathematical basis for design allows many types of quantitative evaluations of an IT artifact, including optimization proofs, analytical simulation, and quantitative comparisons with alternative designs. The further evaluation of a new artifact in a given organizational context affords the opportunity to apply empirical and qualitative methods". Therefore Hevner et al. (2004) offered five approaches for evaluating design artifacts: observational, analytical, experimental, testing and descriptive. We adopted two of these to evaluate the artifact both for internal and external validity. Firstly we used the experimental strategy to assess the internal validity of the proposed method through quantitative, comparative analysis (e.g. chi-square tests) and other internal assessments of comparative numeric settings and fitting models. The algorithms and techniques used, namely bag of visual words and P-DBSCAN are both stable and robust, and have been independently shown to be superior to alternatives in the literature already cited. We also adopted the descriptive approach by using case data (Melbourne City) as representative of the country's outbound travellers, and which is readily validated against justifiable, accepted industry knowledge and independent traveling statistics. Together these provide assurance both of the rigor of the method and its relevance to purpose. By reference throughout to ongoing stakeholder evaluation, specific validity and utility considerations were addressed during iterative development, which is established practice in system development and DSR.

5.1 Data Collection

We collected the geotagged photo data to capture outbound travel behavior of Australians following the method presented in Section 3.1. *PhotoSearch* function was used to extract a set of seed photo to identify photo owners. Input parameters into the search function are shown in Table 4. The bounding box was designed to cover the entire area of Australia as indicated on Google Map (www.google.com.au/maps). The photo taken timeframe was limited to recent years, from 2010 to 2016, to ensure the data was contemporaneous. All photos taken outside the bounding box in the specified time frame were extracted. Since, location of origin is a not mandatory to create Flickr accounts, some users did not provide such information. We selected only *userID*, whose location of origin is in Australia, as the target travel group of this case study. The selected *userID* was then inputted into the *PhotoSearch* function to extract their photos, taken outside of the bounding box. The same parameters for the timeframe were used. Ultimately, 320,596 outbound geotagged photos were extracted from 4,434 Australian travelers.

Table 4: Parameters used in the *PhotoSearch* function

Parameter	Value	Description
min_{la}	-45.966138	minimum latitude of the bounding box
min_{lo}	110.534693	minimum longitude of the bounding box
max_{la}	-10.076632	maximum latitude of the bounding box
max_{lo}	155.446799	maximum longitude of the bounding box
t_{min}	1/1/2010	earliest photo taken date
t_{max}	30/6/2016	latest photo taken date

5.2 Data Exploration

Based on GPS information from the geotagged photos, we identified the visited locations by countries. The country names were identified by mapping the GPS data of the photos onto Google maps. We count the number of unique travelers whose photos were taken in an identified country to represent the popularity of that country. Table 5 shows the top 20 countries, where Australian travelers visited most and shared photos on Flickr. The most popular country was United States of America with 27%, then UK with 21.99%. New Zealand is ranked third despite being much closer to Australia. We confirmed that the countries listed in Table 5 are consistent with identified popular destinations in recent surveys (ABS, 2016). Although a popular destination in ABS surveys, Fiji was not included in our list.

Travelers probably visit Fiji for relaxation rather than for traveling or sightseeing so may be less likely to take photos.

Table 5: Top 20 countries in the collected data set

Visited Country	# of Tourist	Percentage (%)
United States of America *	1199	27.04
United Kingdom*	975	21.99
New Zealand*	918	20.70
France	647	14.59
Italy	519	11.71
Singapore*	515	11.61
Japan*	496	11.19
Malaysia	463	10.44
Indonesia*	453	10.22
Thailand*	421	9.49
China*	417	9.40
Germany	394	8.89
Hong Kong	390	8.80
Canada	329	7.42
Spain	294	6.63
Viet Nam	261	5.89
India*	248	5.59
Netherlands	219	4.94
Switzerland	200	4.51
Ghana	177	3.99

**Popular Destinations according to Australian Bureau of Statistics (ABS, 2016)*

Figure 3 shows the number of Australian outbound travelers by year, as estimated by the Australian Tourism research authority (ABS, 2016). Our data however suggested a *decreasing* trend from 2010 to 2016, although at the time of writing 2016 data was incomplete. Terrorist attacks, ebola and zika viruses were all prominent in this period but a more likely explanation for this trend is due to the decrease of Australian Dollar strength (RBA, 2016), disinclining Australians to travel overseas. It should also be noted that the TRA's sampling methodology changed in this period, to reflect increased mobile and decreased landline use among the surveyed samples of travellers[2]. They observed different patterns among landline phone owners and mobile phone owners, and if the young are considered more likely to post photos to social media, but are usually less wealthy than older travelers, this might further explain the trend we found. Whether or not this is the correct explanation, it shows the value of a social media data source in helping to understand travelers' patterns.

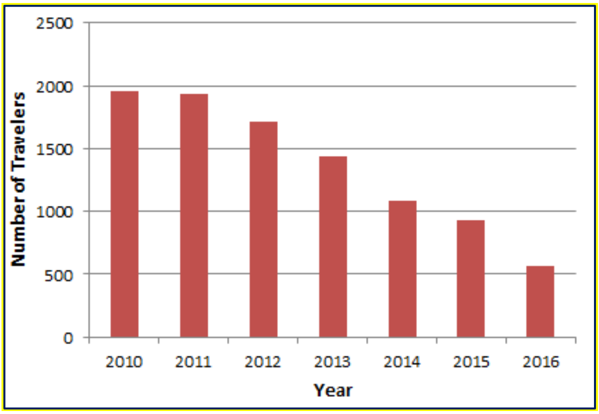


Figure 3: Number of Australian outbound travelers by year

5.3 Interest Identification

We applied the text processing technique described in Section 3.2 to the textual meta-data of the photos. A list of stemmed nouns was returned as potential interest candidates. In this section, we are interested in identifying the most popular interests of travelers; keeping only interests whose support values are greater than 0.1. This left us with a list of 133 interest candidates for further inspection. We refined the list to exclude general words such as "travel", "country", "overseas", "city", "destination", which do not describe specific meanings for tourist interests. The refined list of tourist interests is presented in Table 6, with higher levels of support shown in boldface. The interests are grouped into five categories including infrastructure, transportation, natural, cultural, and attraction, for ease of interpretation. In the infrastructure category, travelers are interested in not only outdoor scenes such as streets, bridges, roads and towers, but also indoor scenes of hotels. *Boat* is the most popular interest of transportation category. High support of more than 0.2 was found for natural scenes such as sunset,

[2] https://www.tra.gov.au/Research/Domestic-tourism-by-Australians/national-visitor-survey-methodology

beach, island, river and lake. General tourist attractions as such as park, garden, art, shop and market were identified in our interest list, and church, temple and museum proved popular cultural interests for Australian travelers overseas.

Table 6: Australian Traveler's Interests

	Interest	**Support**	**Interest**	**Support**
Infrastructure	street	**0.253**	hotel	0.170
	bridge	**0.201**	architecture	0.126
	road	0.189	castle	0.112
	tower	0.183		
Transport	boat	0.176	train	0.129
	car	0.137	station	0.125
Natural	sunset	**0.267**	lake	**0.203**
	beach	**0.235**	mountain	0.199
	island	**0.211**	rock	0.158
	river	**0.204**	snow	0.144
Cultural	church	0.144	museum	0.135
	temple	0.137		
Attraction	park	**0.228**	shop	0.143
	garden	0.158	market	0.135
	art	0.152	statue	0.115
Dining	food	0.129		

In the context of tourism marketing, it is important for managers to identify what interests travelers at each destination, so that appropriate product development and marketing strategies can be designed to suit those interests. Therefore, we incorporate the geographical information into the analysis of traveler's interests. Namely, we compute the support of each identified popular interest with respect to the popularly visited countries as show in Table 5. The support values are shown in Figure 4, which is generated by a map visualization technique (Krentzman et al., 2011). A dark cell represents high support value, while a light colored cell represents otherwise.

We can see that some countries have many popular interests, while some others have a few unique interests. For examples, travelers to the USA are interested in *bridge*, *architecture*, *beach*, *park* and *art*. Travelers to Indonesia are mostly interested in *beach*. By knowing the popular interests in each country, tourism marketers can recommend possible destination to Australian travelers according to what they are interested in. For example, if travelers are interested in natural scenes such as lakes and mountains, tourism marketers can recommend that they travel to Switzerland and design an appropriate package for them.

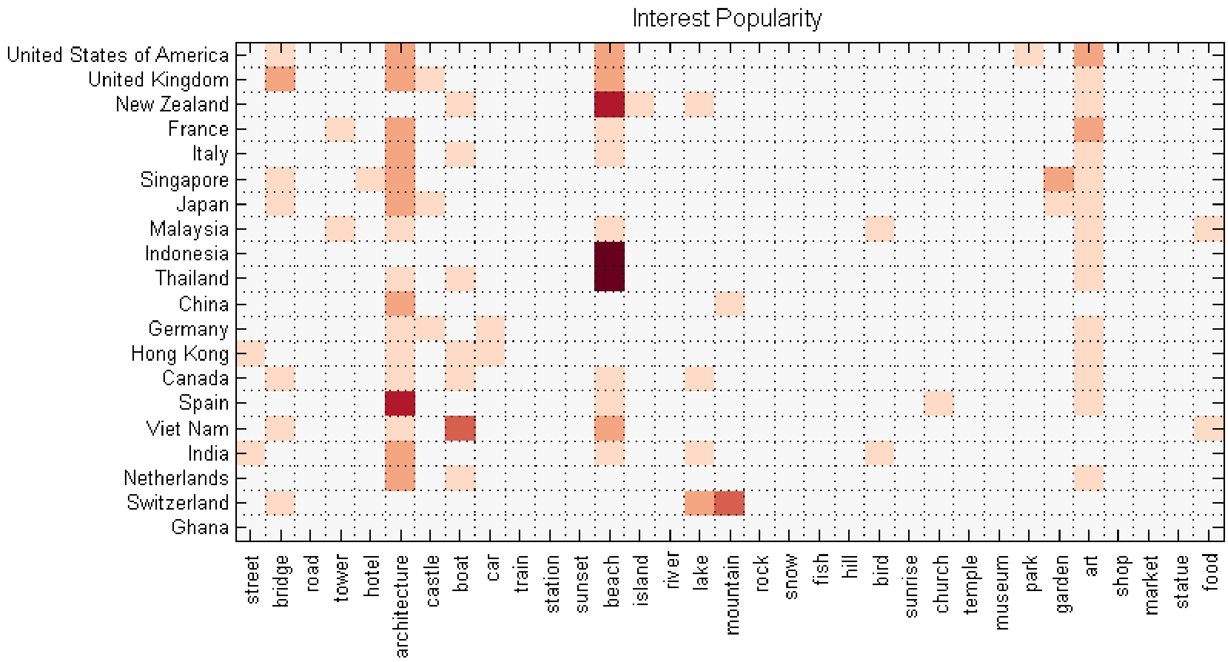


Figure 4: Interest popularity by country.

5.4 Emerging Interests

When making a decision to travel, travelers have choices among destinations, and options within those. They try to identify the special features about each destination to support their decision. Whether tailoring packages or providing advice it is important for tourism marketers and travel agents to identify any significant differences between products and destinations, so as to highlight the special features of a destination. For instance, US and UK are the most popular destinations for Australians. They seem to have relatively similar popular interests as indicated in Figure 3. So what are the special features about each country, their unique propositions? Here, planners are not interested in popular and common interests, but rather those interests showing significant difference between destinations. Aiming to identify such differences, our emerging pattern mining technique was applied to the stemmed noun lists of each country. All items in the stemmed noun list are input into the algorithm rather than only the popular items as identified in the previous section.

The growth rate threshold was set to $\delta = 2$, while the difference threshold was set to $\gamma = 0.05$. The thresholds mean that if the support value of an interest candidate in one group is at least twice that of the other group and the difference is at least 0.05, such interests were selected as differentiating interests between groups. This left us with 9 items as shown in Table 7, representing the most significant differences between UK and US.

Travelers are 6 times more likely to visit a castle in the UK than in the US. This is consistent with the UK's long history with many ancient castles, in contrast to the US. Churches and cathedrals are more than twice as likely to be visited in UK than in the US, while travelers have a significantly higher interest in US skylines than in the UK. Items such as *horse* can be further analyzed to identify whether eventing, trekking, or horse racing is particularly of interest to explain the difference between the UK and the US here. Using date information filters allows emerging patterns and growth areas to be finely determined.

Whilst this illustrative example may be unsurprising, the value is also shown at the lower end in identifying potentially niche propositions: for example the skiing in Bulgaria is less well known than the French Alps, and trekking in Kyrgyzstan are less mainstream options that can be discovered. Based on such findings, marketers can focus on highlighting such differences when promoting destinations.

Table 7: Identified Emerging Interests

Interests	**Support**		Difference	Growth	$\chi 2$	*p-value**
	US	**UK**				
castle	0.031	0.203	0.172	6.581	165.422	0.000
island	0.130	0.048	−0.082	2.699	42.609	0.000
church	0.064	0.146	0.081	2.268	39.352	0.000
cathedral	0.041	0.118	0.077	2.886	45.908	0.000
palace	0.033	0.108	0.075	3.259	48.742	0.000
skyline	0.103	0.039	−0.064	2.632	31.731	0.000
desert	0.073	0.011	−0.062	6.505	47.729	0.000

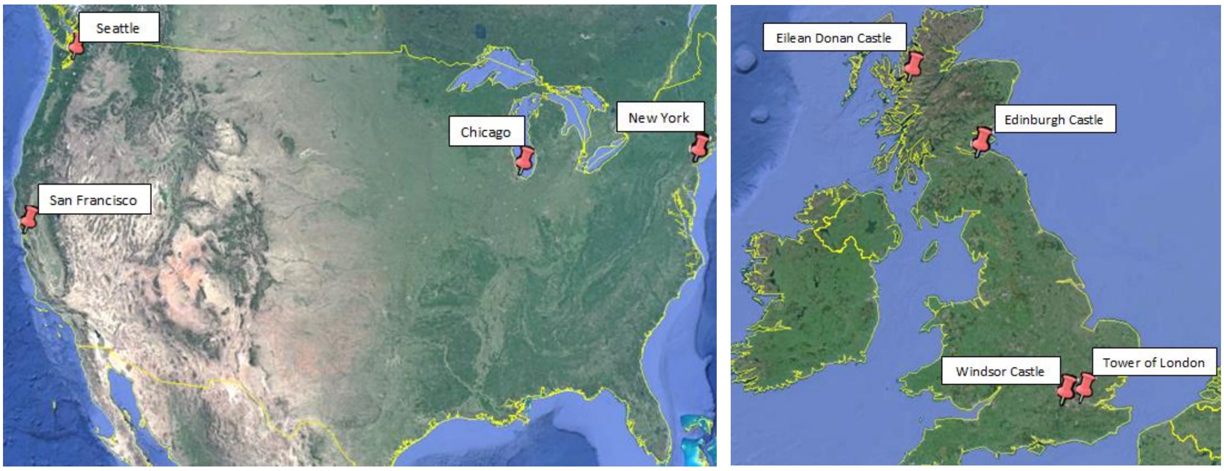
railway	0.025	0.080	0.055	3.197	34.249	0.000
horse	0.034	0.086	0.052	2.519	26.788	0.000

*significant at $p \leq 0.05$

5.5 Location Photos

Visual assets, such as photos are important in designing marketing material such as website, posters, and brochures. The photos should capture viewer's interest through its content. This section demonstrates the capability of the proposed system in identifying location-representative photos epitomizing traveler's interests. We selected *skyline* photos in US and *castle* photos in UK for the case study. P-DBSCAN clustering technique was applied to the photos collection to identify the representative location for each interest in those countries. Neighbor radius was set to r = 0.005, equivalent to approximately 500m. Minimum owner was set to $N_o = 10\%$ of the owner in each group of photos. Each cluster is a representative location for the interest under consideration.

Figure 5 and 6 shows identified representative locations for taking skyline photos in US and castle photos in UK. More specifically, the most popular locations for skyline interest in US are *Seattle, San Francisco, Chicago,* and *New York.* The most popular locations for castle interest in the UK are *Eilean Donan, Edinburgh, Windsor* and *London.* We then extract visual descriptors for the photo at each representative location and apply the Kernel Density Estimation to identify the photos with the most popular content as location representative photos. Figure 5a shows a representative location photo for skyline in New York, US, Figure 6a shows a representative location photo for Eilean Donan Castle in Dornie, UK. Such photos provide insights into the experience of travelers and reflect a prototypical view (for our method) of a destination, useful to tourism marketers as inspiration for marketing material.



a) b)

Figure 5: Clustering Result for a) skyline in US; b) castle in UK

a) Skyline (New York – US) b) Eilean Donan Castle (Dornie - UK)

Figure 6: Location Representative Photo.

5.6 Focus group evaluation

The final evaluation involved a confirmatory focus group of four travel professionals adopting the approach[3] of Tremblay et al. (2012) in that usability and efficacy were the central focus of qualitative data gathering for studying artifact evaluation.

[3] Tremblay et al. (2012) proposed a method of using focus group for exploratory and confirmatory purposes of artifact evaluation. However, we only adopted for the confirmatory purposes due to our project's requirements.

In our study, we employed descriptive and observational methods to evaluate the qualities of the proposed artefact using Venable et al.'s (2016) framework for conducting the entire evaluation activities in our study. The following table illustrates details of the steps.

Table 8: Overall evaluation activities

Processes captured form the framework of Venable et al. (2016)	*Descriptions*
(1) Explicate the objectives	Our aim is to capture utility and efficacy details of the artifact
(2) Choose the evaluation strategies	We adopt descriptive and observational approaches (see Hevner et al. (2004))
(3) Determine the properties to evaluate	We aim to demonstrate the accuracy, relevance and usability of the proposed artifact
(4) Design the individual evaluation cases	We describe the case details and practical demand of the travel service providers through focus group discussions

Participants were shown a demonstration of the system then asked questions on usability and efficacy. Table 9 summarizes their evaluation details including agreement ratings and relevant comments on the aspects of usability and efficacy.

Table 9: Evaluation outcome of confirmatory focus group

Addressing criteria	Outcome (Yes/No)				Practitioner's (P) comments
Usability Method produced useful information from social media data	Y	Y	Y	Y	"the method has a lot of practical application because it gives a lot more richer details" (P1) "produced info can be helpful in making plans and considering through insights" (P1) "we can get on what travelers do at destinations" (P2)
Method produced usable information to particular need of decision making	Y	Y	Y	Y	"useable information can be related to identifying popular locations and possible activities" (P3) "it gives me a more visual support for developing traveler packages specific to meeting our business demands" (P1)

Method was helpful for saving time	Y	Y	Y	Y	“because it is a pre-set model for seeking information” (P3) “saves time since I don’t have to browse profiles” (P1) “I can see a quick overview on people’s interests at locations” (P2)
Efficacy Positive changes in everyday practices if the proposed method is in use	N	Y	Y	Y	“I may not see changes because of my target market, but the method would be helpful for destination marketing for sure” (P1) “we can run the analytics every day to tap in opportunities” (P3) “changes can be seen based on real time data” (P4)
Can see possible benefits of proposed method	Y	Y	Y	Y	“we can identify most popular photos in a location where travelers like the most” (P3) “ I can design the best package for target market” (P4) “it will help to save time and resources” (P1)
Can the method be improved	Y	Y	Y	Y	“it would great if we have age component in results” (P1) “method should capture large number of samples for better accuracy” (P2) “the method should give seasonal based visitation details” (P3)
Can have possible disadvantages of the method	N	N	N	N	(None)

5.7 Discussion and validity analysis

The case study has demonstrated the performance of our solution artifact in answering critical questions for travel marketing managers. Namely, the identification of popular interests and their corresponding location (Section 5.1.3) provide insights about what attracts outbound travelers at a specific tourism destination. The use of emerging pattern mining (Section 5.1.4) highlighted the significant differences in travelers’ interests at different destinations. Personal travel experiences at visited destination are revealed by analysis of location photos (Section 5.2.5). These are valuable insights to support travel marketing decision marking, which has not been obtained using prior approaches.

Table 10: A comparison matrix of other analytics approaches that process geotagged photo content.

Sources	Key techniques	Main aspects	Experiments	Aims	Usability testing
Zhou et al. (2015)	Image processing and spatial analysis	Extraction of tourism destinations	Yes	Tracking of Interests of tourists	No
Hirota et al (2015)	Grid-Clustering algorithm	Extraction of relations of shooting spot and photographic subject	Yes	Identify hotspots	No
Su et al. (2016)	Spatial Regression, Variances Decomposition	Extraction of geographical preferences among tourists	Yes	Identify hotspots	No
Onder et al. (2016; 2017)	Polynomial regression analysis, Density-based clustering	Extraction of actual number of tourists in spots	Yes	Tracking of Tourists	No
Our method	**Meta data processing, emerging pattern mining, and image processing**	**Extraction of popular interests, location and representative photos**	**Yes**	**Tracking of travelers behavioral details**	**Yes**

Furthermore, our method provides an evidenced analysis complementary to the traditional forms of data collection and analysis used in the relevant industry. The proposed method addresses well-recognized issues as “historically, the Tourism Forecasts have typically overestimated visitors and visitor nights and underestimated outbound departures.” (Tourism Victoria, 2014, p. 6]. The figures produced by the Australian Bureau of Statistics (ABS) are at national level, and limited to count passenger movements, rather than the details of unique travelers’ interests that could offer more evidence-based support. As the data source is departure cards, often only specifying “Europe” as a destination, the source data is often “inadequately described”, ABS (2016) has similar lack of specificity applies to inbound travelers, where destination intentions may change or are not preidentified. This leads both to suppression of publication for analyses (where sampling variability is considered too high), and to breakdowns of interest being subject to sampling variability (for source ABS data that is collected but not fully enumerated). Extending the work of Dang et al. [11] for the purpose of locational data analysis our method provides fine-grained information and complementary insights on outbound destinations with the primary data of photographs providing evidence of actual experiences. The participants of the focus group expressed that the findings are richer than the official information published by Australian tourism data base and Australian Bureau of Statistics data, which affirms the validity of our method over existing approaches.

Although going beyond prior work as noted above, there are some limitations to the approach presented here. These include the reliance on metadata, which is stripped by some social media sites, and the validity of using photographs as a proxy for interests. This potentially limits the dataset available. Although *prima facie* using photos is reasonable, less photogenic or less iconic attractions may be under-represented in the set sampled. Equally we assume most travelers take and post photographs, although some demographics may be under or over-represented in the sample, requiring further tuning. The method is demonstrated and tested here with a single case study. We claim its general applicability due to the parameters that can be set for locations of interest although we have not reported its use in other trials we have made. Further examples would strengthen and test this claim. We used one photo-sharing site, Flickr.com, but others are becoming very popular now, and the data sources could become volatile or unsupported. For example Picasa, (a Flickr alternative) has closed in favor of Google photos. Our method also relies on an available API and a website's policy on photo usage and metadata, which could change.

6. CONCLUSION

In this paper we have described a method for extracting strategically valuable information from geotagged photos. Using a design science research approach we have developed a general location analytics method, then demonstrated and evaluated its utility, efficacy and usability in the domain of travel marketing industry through both a focus group and a representative case illustration. Participants' views on the proposed method were positive and their comments suggested direct relevance to their practical needs for developing location specific products (see table 9). The theoretical contribution of the study is the solution's effectiveness in identifying attractions, both established and emerging, along with innovative use of data analytic techniques that goes beyond those in the previous studies. By integrating various techniques our method overcomes particular shortfalls in the other approaches which, in themselves, are unable to provide detailed location insights into the visitation context and interests for each global location. By analyzing large, unstructured datasets of geotagged photos our method applies generically to relevantly targeted locations and generates useful results that both complement existing industry information and provide fine-grained insights for professionals. Our method can in principle be extended beyond its current scope of applicability to employ real time event detection systems for events management and monitoring.

In future we intend to investigate different popular social media platforms such as Facebook, Twitter, Instagram and Foursquare. For example, Twitter is a platform employed for 'instant' messaging, where users tend to post content relevant to very-current events and travel managers could usefully analyze location related sentiment. In our future research, we also would like to compare the capability of different social media platforms in supporting strategic travel decisions for a variety of travel destinations and data types.

REFERENCES

ABS (2016). 3401.0 - *Overseas Arrivals and Departures, Australia, Jun 2016*. <http://www.abs.gov.au/ausstats/abs@.nsf/products/961B6B53B87C130ACA2574030010BD05> , accessed on 17 September 2016

ArcNews (2012). Location Analytics: The Next Big Step in Business Analysis, <http://www.esri.com/news/arcnews/fall12articles/location-analytics-the-next-big-step-in-business-analysis.html>, Accessed on 23 Dec, 2016

Ayres, J, J Gehrke, T Yiu and J Flannick (2002). Sequential pattern mining using a bitmap representation. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Alberta, Canada , pp. 429 -435.

Baskerville, RL, M Kaul and VC Storey (2015). Genres of Inquiry in Design Science Research: Justification and Evaluation of Knowledge Production. *MIS Quarterly*, 39 (3), 541-564.

Baskerville, RL, K Lyytinen, V Sambamurthy and D Straub (2010). A Response to the Design-Oriented Information Systems Research Memorandum. *European Journal of Information Systems*, 20 (1), 11-15.

Bay, H, A Ess, T Tuytelaars and LV Gool (2008). Speeded-up robust features. *Computer Vision and Image Understanding*, 110(3), 346-359.

Blázquez, JJ, A Molina and A Esteban (2012). Key quality attributes according to the tourist product. *European Journal of Tourism Research*, 5(2), 166–170.

Bowman, AW and A Azzalini (1997). Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations. Oxford Statistical Science Series 18. Oxford University Press, New York, USA.

Capdevila J, M Arias and A Arratia (2016). GeoSRS: A hybrid social recommender system for geolocated data. *Information Systems*, 57, 111–128.

Chua, A, L Servillo, E Marcheggiani and AV Moere (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management*, 57, 295-310.

Dang, Y, Y Zhand, PJH Hu, SA Brown, Y Ku, JH Wang and H Chen (2014). An integrated framework for analyzing multilingual content in Web 2.0 social media. *Decision Support Systems*, 61, 126–135.

Dong, G and J Li (1999). Efficient mining of emerging patterns: discovering trends and differences. In *Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 43–52, San Diego, CA, USA.

Fotis J, D Buhalis and N Rossides (2012). Social media use and impact during the holiday travel planning process In: M. Fuchs, F. Ricci, & L. Cantoni (Eds.), *Information and communication technologies in tourism*, pp. 13–24. Vienna.

Furmanov, K, O Balaeva and M Predvoditeleva (2012). Tourism flows from the Russian federation to the European Union. *Anatolia - An International Journal of Tourism and Hospitality Research*, 23 (1), 17-31.

Gregor, S and AR, Hevner (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37 (2), 337-355.

Hevner, A, S March, J Park and S Ram (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28 (1), 75-105.

Huang, A, L Gallegos and K Lerman (2017). Travel analytics: Understanding how destination choice and business clusters are connected based on social media data. *Transportation Research Part C*, 77, 245–256

Hirota, M, M Endo, S Yokoyama and H Ishikawa (2015). Visualizing Shooting Spots using Geo-Tagged Photographs from Social Media Sites. *EGC 2015*, RNTI-E 28, 167-178.

Iivari, J (2007). A Paradigmatic Analysis of Information Systems as a Design Science. *Scandinavian Journal of Information Systems*, 19 (2), 39-64.

Jiang, S, H Chen, JF Nunamaker and D Zimbra (2014). Analyzing firm-specific social media and market: A stakeholder-based event analysis framework. *Decision Support Systems*, 67, 30-39.

Kim, HB, JH Park, SK Lee and S Jang (2012). Do expectations of future wealth increase outbound tourism? Evidence from Korea. *Tourism Management*, 33 (5), 1141–1147.

Krentzman, A, E Robinson, J Jester and B Perron (2011). Heat Maps: a technique for classifying and analyzing drinking behaviour. *Substance Use & Misuse*, 46, 687–695.

Kisilevich, S, F Mansmann and D Keim (2010). P-DBSCAN: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos. In *Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application*, Paper 38, Bethesda, MD, USA.

Li, Y, L Xiao, Y Ye, W Xu and A Law (2016). Understanding tourist space at a historic site through space syntax analysis: The case of Gulangyu, China. *Tourism Management*, 52, 30-43

Lim, C (2004). The major determinants of Korean outbound travel to Australia. *Mathematics and Computers in Simulation*, 64 (3-4), 477-485.

Lin, V S, A Liu and H Song (2015). Modelling and forecasting Chinese outbound tourism: An econometric approach. *Journal of Travel & Tourism Marketing*, 32 (1-2), 34-49.

Majid, A, L Chen, G Chen, HT Mirza, I Hussain and J Woodward (2012). A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science*, 27 (4), 662-684.

Naaman, M (2011). Geographic information from georeferenced social media data. *SIGSPATIAL Special*, 3(2), 54–61.

Nagler, T and C Czado (2015). Evading the curse of dimensionality in multivariate kernel density estimation with simplified vines. *Technical Report arXiv:1503.03305*, Cornell University.

Okuyama, K. and K. Yanai (2013). A travel planning system based on travel trajectories extracted from a large number of geotagged photos on the web. In Jin Xu and XU (ed), The era of interactive media (pp. 657-670). Springer New York.

Oku, K., F. Hattori, and K. Kawagoe (2015), Tweet-mapping method for tourist spots based on now-tweets and spot-photos, *Procedia Computer Science*, 60, 1318-1327

Onder, I, W Koerbitz and A Hubmann-Haidvogel (2016). Tracing Tourists by Their Digital Footprints: The Case of Austria. *Journal of Travel Research*, 55(5), 566–573.

Onder, I (2017). Classifying multi-destination trips in Austria with big data. *Tourism Management Perspectives*, 21, 54-58.

Park, SB, CM Ok and BK Chae (2016). Using Twitter Data for Cruise Tourism Marketing and Research. *Journal of Travel & Tourism Marketing*, 33(6), 885-898.

Prayag , G, SA Cohen and H Yan (2015). Potential Chinese travellers to Western Europe: segmenting motivations and service expectations. *Current Issues in Tourism*, 18(8), 1-19.

Provost , F and T Fawcett (2013). Data Science and Its Relationship to Big Data and Data-Driven Decision making. *Big Data*, 1(1), 51–59.

RBA (2016). Exchange Rate, Reserve Bank Australia, < http://www.rba.gov.au/statistics/historical-data.html>, accessed on 29 Nov 2016

Seetaram, N (2012). Estimating demand elasticities for Australia's international outbound tourism. *Tourism Economics*, 18 (5), 999-1017.

Shen, S, G Li and H Song (2011). Combination forecasts of international tourism demand. *Annals of Tourism Research*, 38 (1), 72-89.

Su, S., C Wan and S Su (2016). Characterizing geographical preference of international tourist and the local influential factors of China using geotagged photos on social media. *Applied Geography*, 73, 36-37.

Terrell , GR and DW Scott (1992). Variable kernel density estimation. *The Annals of Statistics*, 20(3), 1236-1265.

Thornton, P, A Williams and G Shaw (1997). Revisiting time-space diaries: An exploratory case study of tourist behaviour in Cornwall, England. *Environment and Planning A*, 29(10), 1847-1867.

Tourism Victoria (2014). Tourism Victoria Research User Guide, <https://www.tourism.vic.gov.au/component/edocman/?view=document&task=document.download&id=666>, accessed on 25 Dec 2016

TRA (2016). Outbound Tourism Statistics. Tourism Research Australia.. <https://www.tra.gov.au/research/australians-travelling-overseas/outbound-tourism-statistics/outbound-tourism-statistics > accessed on 7 April 2018

Tremblay, MC, AR. Hevner, and DJ Berndi (2012). Design of an information volatility measure for health care decision making. *Decision Support Systems*, 52, 331–341.

Vailaya, A, MA Figueiredo , AK Jain and HJ Zhang (2001). Image classification for content-based indexing. *IEEE Transactions on Image*, 10(1), 117-130.

Venable, J, J Pries-Heje and R Baskerville (2016). FEDS: a framework for evaluation in design science research. *European Journal of Information Systems*, 25 (1), 77–89.

Verkerke, G.J., EB van der Houwen, AA. Broekhuis, J. Bursa, G. Catapano, P. McCullagh, K. Mottaghy, P. Niederer, R. Reilly, V. Rogalewicz, P. Segers, and N. Verdonschot (2013). Science versus design; comparable, contrastive or conducive?, *Journal of the Mechanical Behavior of Biomedical Materials*, 21, 195-201

Warmerdam, K and A Pandharipande (2016). Location Data Analytics in Wireless Lighting Systems. *IEEE Sensors Journal*, 16(8), 2683-2690.

Wong, JY, HJ Chen, PH Chung and NC Kao (2006). Identifying valuable travelers and their next foreign destination by the application of data mining techniques. *Asia Pacific Journal of Tourism Research*, 11, 355- 373.

Xiang , Z and U Gretzel (2010). Role of social media in online travel information search. *Tourism Management*, 31(2), 179-188.

Yang, J, YG Jiang, AG Hauptmann and CW Ngo (2007). Evaluating bag of-visual-words representations in scene classification. In *Proceedings of International Workshop on Multimedia Information Retrieval*, pp. 197–206, Augsburg, Bavaria, Germany.

Yap, G (2013). The impacts of exchange rates on Australia’s domestic and outbound travel markets. *Mathematics and Computers in Simulation*, 93, 139-150.

Zhou, X, C Xu and B Kimmons (2015). Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Computers, Environment and Urban Systems*, 54, 144-153.

Zhuang, Y, S Fong, M Yuan, Y Sung, K Cho and RK Wong (2016). Location-based big data analytics for guessing the next Foursquare check-ins. *Journal of Supercomputing*, 73(7), 3112–3127.