



A big-data analytics method for capturing visitor activities and flows: the case of an island country

Shah Jahan Miah¹ · HuyQuan Vu² · John Gammack³

Published online: 28 May 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Understanding how people move from one location to another is important both for smart city planners and destination managers. Big-data generated on social media sites have created opportunities for developing evidence-based insights that can be useful for decision-makers. While previous studies have introduced observational data analysis methods for social media data, there remains a need for method development—specifically for capturing people’s movement flows and behavioural details. This paper reports a study outlining a new analytical method, to explore people’s activities, behavioural, and movement details for people monitoring and planning purposes. Our method utilises online geotagged content uploaded by users from various locations. The effectiveness of the proposed method, which combines content capturing, processing and predicting algorithms, is demonstrated through a case study of the Fiji Islands. The results show good performance compared to other relevant methods and show applicability to national decisions and policies.

Keywords Big data · Decision making · Smart city initiatives · Data analytics · Location flows

1 Introduction

Location-based analytics techniques offer contemporary concepts to understand spatial activities, patterns, and groupings in geographically referenced phenomena [1]. Such phenomena enable new sources of insight for decision makers in various fields. Fixed sensors or cameras, combined with geo-located mobile devices, provide rich and voluminous data sources potentially tracing the behaviours and flows of populations of interest. Applications are ranging from tracking individuals of interest, to smart transport and infrastructure planning, to tourism destination planning and

management, are all potentially enhanced by location-based analytics [2].

Big data is a term that suggests growth and availability of data characterised by its large volume, velocity, rapid emergence and varied formats. In addition to these, other characteristics have been mooted such as Variability (over time and diversity of sources); and Volatility (inconsistent levels of production) [3]. Many studies have reported technical and social issues involved in big data, and searching, analysing, sharing, storage, visualising, and privacy remain current challenges [4]. These researchers recognise, along with previous authors, that new research must focus on designing new technologies to uncover hidden value from large, diverse and unstructured datasets that may help in defining or redefining geographically referenced phenomena. Big data analytics methods and tools have begun to be developed to do just that.

Big-data analytics research has grown rapidly in recent years, due to the increased availability of large amounts of people-generated (actively or passively) data sources. For instances, camera and sensor data from locations within cities can be used dynamically to manage traffic flows, energy demands and other infrastructure. Various inexpensive technologies and platforms, including smartphones, and RFID, help form an Internet (or Web) of Things, making

✉ Shah Jahan Miah
shah.miah@vu.edu.au

HuyQuan Vu
quan@tulip.org.au

John Gammack
John.Gammack@zu.ac.ae

¹ Victoria University Business School, Victoria University, Footscray Park Campus, PO BOX 14428, Melbourne, VIC 8001, Australia

² Department of Information Systems and Business Analytics, Deakin University, Burwood, Australia

³ Zayed University, Abu Dhabi Campus, Abu Dhabi, UAE

smart environments for living or workplaces [5]. Smart city research addresses new opportunities and challenges faced in meeting contemporary demands [6], and initiatives around the globe evolve to ultimately meet with citizen lifestyle expectation, demand of population growth in urban areas, modernisation of organisational strategies, and technological improvements. Although many initiatives are at the conceptual and exploratory phases, the realisation of benefits has been sufficiently widespread for considerable investment in contemporary research development. As a result, many successful smart applications are emerging—including Intelligent Traffic, Smart Logistics [7], smart homes [8], and smart healthcare [9].

Numerous challenges still remain: leveraging value and quality from volume and quantity continues to test organisations, and the current reality is well short of potential [10]. One challenge lies in identifying specific patterns of individual behaviour within large and varied data sets: patterns that can be used for fine grained targeting, such as tracking a person of interest through a wealth of CCTV footage, or micro-marketing applications. At present, capturing people's activities, behavioural and movement details for such purposes can be a significant task for authorities due to lack of (or abundance of!) data availability, evidence-based sources supporting a single view, and undifferentiated categories due to over-aggregation of collected data. Smart location research should explore further issues in big data collection, generation, and processing [1, 5].

For identifying concurrent details of personal activities and movements, big data generated in social media has been considered both as a valuable and reliable source [11]. However, this line of work, using rapidly-growing user-generated content, remains underexplored [5] e.g. specifically for understanding spatial activities. In addition, Sivarajah et al's [12] profiling of 20 years of big data research identifies “a need to develop and understand (Big data and big data analytics) in an intensive way using case studies” (p. 279).

Traditional managerial decision making within city management authorities in any country context is based on internal data, systems, process and procedures, and frequently decision-making ignores external datasets that may have significant value. Big data analytic methods are now emerging [13] with examples that include text mining of online user reviews [14]; hospitality management [15]; a visual analytics dashboard reflecting news and media sentiments associated with destinations and analysing user generated content for branding and location oriented marketing [16]. Such studies used social media as the primary data source. For example, in analogous work Huang [17] used Twitter check-ins effectively to identify connected business clusters in Los Angeles County with strategic implications for transportation, urban design and land use. Despite the potential of social media to suggest such insights, to date few studies

provide decision support for people monitoring and resource allocation purposes.

Following the design research paradigm¹ in which artefact design and evaluation are the central activities of research, the fundamental question of this paper is *How can we develop a new information systems artefact to assist in exploring general human activities and flows in various locations?* We focus on the design and evaluation of a general method relevant particularly to emerging big-data application areas, smart city initiatives (e.g. a collection of smart application development research for improving smart management through people tracking and recording of their activities), using the case study of Fiji Islands, a country for which tourism is a critical economic sector. Utilising social media content, we design and evaluate a new IT artefact called *People-Tracker*. Our artefact is classified in design science terms as a *method* that comprises intelligent techniques for exploring content collection, as well as for transition and temporal pattern exploration. The *People-Tracker* captures, classifies, transforms, processes and visualises patterns of people movement and contexts including visits, sightseeing and attending events. Our *People-Tracker* provides insights in terms of capturing country-wide flows and activities of people in various locations. The immediate practical contribution of the system is providing government authorities and destination management organizations with access to detailed records of travel routes, densities, and demographics to support decision making and planning for location development and building transportation infrastructure.

This paper is structured as follows. The next section provides background from existing relevant studies and describes our case study context. The Sect. 3 includes methodological details of our study, conducted following design science guidelines. The Sect. 4 then presents details of the proposed solution artefact followed by the design and evaluation insights. The final section discusses and concludes the paper with emphasis on the contributions of the study including consideration of the study's delimitations and directions for further work.

2 Related background works

The burgeoning of smart city, and by extension smart tourism destination, [20], initiatives lead to a massive increase in data generation and potential processing. Consequently, such enormous volumes of data are at the core of any smart

¹ Design science research is a pragmatic research paradigm that promotes the development and evaluation of information systems artefacts to solve real world issues [18, 19].

Table 1 Existing studies on locational analytics methods for big-data processing

Sources	Key techniques	Purpose of application
[24]	Social media data analytics method	Revealing insights of tourist behaviour in different tourists spots
[25]	Spatial and statistical analyses	Tracking numbers of tourists in a particular tourist destination
[22]	Space-syntax technique	Analysing tourist movement and outlining patterns based on cell phone location data
[45]	Image processing and spatial analysis	Understanding of Tourists' experiences on locations
[47]	Multimodal spatial–temporal	Analysing landmark photos to understand learn general, local, and temporal themes in locations
[48]	Kernel vector quantization (KVQ) technique	Extracting location-based information to organize photo content
[49]	Spatial clustering technique	Analysing geotagged photos, the aim is to discover popular tourist attractions in a city for assessing popularity level of tourist attraction
[50]	Path Scalar Matrix based technique	A personalized route recommendation algorithm is designed for tourist's personal preferences
[51]	Keyword extraction techniques	For discovering travel experiences to facilitate trip planning

application design for improving place-focussed services e.g. for development of public transportation and routes.

Smart Cities are a new style of city providing sustainable growth and designed to encourage healthy economic activities that reduce the burden on the environment while improving Quality of Life (QoL) – such as housing, economy, culture, social and environmental conditions [21]

Although cities have been the traditional research focus, similar issues are faced at regional and national levels, particularly where there are few large cities, and this aspect is relatively under researched. Falconer and Mitchell [6] noted that whilst Smart City research has triggered considerable theoretical and technology-led discussion, insufficient progress has been made in implementing related initiatives, specifically for smart data management and its processing for decision making. In this domain, different stakeholders have naturally different priorities: city engineers and technology companies view the city as a complex system with multiple layers while the architects and non-governmental organizations view the city in terms of people, social inclusion, and a sense of space [5]. Government decision makers, on the other hand, view the city in terms of economic growth and improved citizen's services supported by policy initiatives designed to effect change. All continuously require useful data from external sources such as social media sites in order to make their decisions.

In this context, the big data generated by individuals, in the form of online materials, may hold interesting and useful insights, reflecting interest and the affordances of locations. Such big data is available via various social media sites used for photo sharing (Flickr, Instagram), video sharing (YouTube), immediate comment and response sharing (Twitter), and hybrid photo, comment and discussion sharing (Facebook). However such social media data are rarely collected by government or location management authorities, despite

the fact that they might offer important insight into people's behaviour, activities and preferences. Approaches using social media big-data for producing insights related to activities and movements are, however, becoming evident across the tourism sector, such as for understanding tourist space [22] and for enabling real-time decisions in relation to people movement, their density or any other organizations' strategies [23]. Some focus on capturing visitor activities and generating behavioural insights in different locations [24], or use geo-tagged tweets to characterize the spatial, temporal and demographic features of tourist flows across Italian regions [25]. More importantly, these latter two studies offer the opportunity of gaining evidence-based insights into a variety of attractions and people's activities. Table 1 illustrates some examples of research relevant to activity oriented insights.

The examples illustrated in Table 1 represent solutions that are mostly geared to tourism management. Although the technologies have been developed for various relevant operations, little attention has been paid to developing a comprehensive technological solution such that nation-wide people movements can be monitored. Such technologies suggest processes that help key stakeholders and authorities to understand resident, visitor and traffic flows among cities, define patterns of activities, and make resource allocation decisions. Generally, such issues are highlighted in the case study context of the Fiji Islands, described in the next section. This motivates us to develop a comprehensive method that could capture, process and produce analytical insights to support an authority's effective decision making.

2.1 Case study background: visitor movement in Fiji

The Fiji Islands, located in the Pacific a 3 h flight north of New Zealand, have long been an exotic destination, with a "tropical paradise" image. Over 300 islands, many uninhabited, offer a range of water and land based activities

together with other indigenous attractions. Tourism is Fiji's major source of foreign income, contributing around 40% directly or indirectly to GDP and is a large-scale employer, providing more than one in every three jobs currently [26]. As such, sustainable tourism is critical to the national economy, and managing Fiji itself as a destination is a vital role for its government.

For continued growth in foreign visitor numbers, the resource must be maintained and enhanced, and decisions impact upon, and are entangled with, policies on environmental sustainability, transport and built environment infrastructure, and social, economic and business development. Naturally, with more visitor arrivals anticipated, accommodation, transport and human resource capacities are affected, and the identification of areas prioritised for tourism development requires informed consideration in the context of wider economic and infrastructure development.

Tourism development in Fiji is a whole-of-government activity, and has been managed with reference to a series of master plans which latterly have increased the strategic focus on sustainability, with shorter planning timeframes than before, coupled with annual reviews. The latest plan, *Fijian Tourism 2021*, [27] is out for consultation at the time of writing, and is framed for integrated development involving “coordinated activities and investment from Government, non-government organisations and private sector stakeholders” (p. 2). This implies a need for good information on where visitors actually go, how long they stay before moving on, and what they do at places of interest, to help inform long term infrastructure and investment decisions, whether public or private.

Part of the strategy concerns modernising for robust and effective data collection, along with improving market intelligence. A major source of tourism intelligence is the quarterly International Visitor Survey (IVS), which reports such information as the origin country of visitors, regions visited, and the main reason given for the visit. This is triangulated with the Fiji Bureau of Statistics' own figures, which uses slightly different categories. Major markets such as Australia and New Zealand are reliably identified, and a dominant reason given is typically “for holiday” [27]. It provides some qualitative data also, along with some demographic details and satisfaction levels.

The IVS however, as with many widely used tourism instruments, provides aggregated statistics which do not provide close insight into individual travel patterns and interests, and moreover does not record the flows involved or the amounts of time spent at specific attractions. For generating deeper insights into tourist behaviour, complementary tools are needed and this was the motivation for our design, described in the next section.

3 Methodology

3.1 Research approach

We adopt a design science research (DSR) methodology for designing and evaluating the *People Tracker* artefact. DSR has become a major paradigm in IT solution design, gaining enormous attention by design professionals and researchers because it promotes creation of innovative artefacts that reflect practical context and professional relevance [28]. Baskerville, Kaul and Storey [28] summarised DSR as building and evaluating an IT artefact as an outcome of a research project; producing new knowledge from the practices of design and development activities; and communicating research by reporting. This implies that DSR is more than just a methodology for developing design artefacts, but an approach that enables researchers to learn from solution artefact.

In order to differentiate DSR from other design approaches, we note that DSR addresses “either an unsolved problem in a unique and innovative way or a solved problem in a more effective or efficient way” [29, p. 143]. Gregor and Hevner [18], following March and Smith [30] described how design research can produce five distinct types of artefacts: constructs, models, methods, instantiations and theories. In our study, we design the *People Tracker* as a method such that a construction process can be replicated and applied beyond a simple IT development. The problem space to which the method applies is scoped, and problems as well as the professional relevance for which solutions were previously difficult or impossible are addressed.

Peppers et al. [31] proposed a particular DSR methodology that comprises six activities: namely 1: Problem identification and motivation; 2: Define the objectives for a solution; 3: Design and development; 4: Demonstration; 5: Evaluation; and 6: Communication. Gregor and Hevner [18] suggested that Peppers's et al. research methodology, building on other approaches, offers a useful synthesized general model and we find this model to be compatible with our underlying objectives.

In terms of the Peppers et al. [31] activity map, our research process included realizing a problem situation (activity 1), analysing published literature for similar methods (activity 2), developing a prototype of *People Tracker* and testing its quality attributes in practice through a case study analysis (activity 3). A proof-of-concept of the proposed *People Tracker* was demonstrated later to obtain quality feedback from a knowledgeable audience and to explore the fidelity of the design work (activity 4). Descriptive evaluation occurred via seminars with participant feedback (activity 5). A total system was finally communicated to a practical audience (activity 6) (Table 2).

Table 2 Different phases of the project

Phases	DSR design activities	Our artefact design
Identifying business problems and artefact types	Activity 1 and 2	There is a need to develop appropriate and efficient analytical methods to leverage unstructured location tagged content. Most current approaches are incapable of producing insights for this purpose. This is the essence of the research gap addressed by our DSR study The study has produced an analytics method designed to generate activities of people traveling and gathering in locations
Artefact creation and evaluation	Activity 3 and 5	Data mining methods employed in other domains were adapted for use in this study. The design process was iterative in order to cope with much of the uncertainty inherent in the problem space The location analytics artefact was constructed using established data mining techniques and proven algorithms and evaluated using fundamental and commonly-accepted research methods (experiment, focus group and case study) To demonstrate artefact effectiveness, experiments were conducted, with focus groups coupled to a case scenario analysis using test datasets
Research contributions of the artefact and communication of results	Activity 4 and 6	The method is theoretically informed through the framework of Hashem et al. [5]. We applied the principles for inventing a new method and the experimental outcomes have shown a clear effectiveness The artifact has been presented to government authorities and academic professionals who evaluated it for usability and efficacy in a focus group setting. The academic report is intended for publication following recommended practice and schema

3.2 Proposed techniques

This section presents our geotagged social media analytics method to support area-specific decision-making in monitoring and providing support services. Various issues need addressing for effective processing and analysis of geo-tagged social media. Firstly, most social media platforms (e.g. Facebook, Instagram, Twitter, and Flickr) allow users to upload and share visual content, such as photos, typically demonstrating their interests or activities. Photos are an important asset that allows decision makers to gain insights into people's behaviour [32]. Although comments or keywords can be posted together with photos, many photos are posted without being accompanied by textual description or tags. It is time consuming for users to provide such meta-data for a large number of photos. Textual meta-data, if available, are often in the form of short text, or keywords, inadequate comprehensively to describe what is presented in a photo. The social media analytic method should be able to directly process the photos and provide insightful information to decision makers without having to rely on textual description or tags. Secondly, transportation management and location transitions are key issues in country management, with a strong influence on economic growth [33]. Whereas transitions are dynamic, geotagged social media data are static, as the raw GPS data attached to each post represents a unique location. The social media analytic method should be able to capture and reveal transitional information.

Thirdly, the data available on social media are treated as historical data, while decision making should be made to accommodate future situations. Trend prediction from geotagged social media will help support decision makers. Having considered these challenges, our analytic method comprises four stages: (1) Social Media Data Collection; (2) Photo Theme Exploration; (3) Transition Exploration; and (4) Temporal Pattern Exploration.

3.3 Social media data collection

The first stage is to extract data from social media for subsequent analysis. Most social media platforms provide an Application Programming Interface (API), allowing developers to build applications that interact with their system through a registered account. Rate limits apply to free accounts on various social media platforms, with higher limits available for a fee. The API functions vary depending on the platforms, and are often well-documented, e.g. Facebook Graph API, Twitter REST API, and Flickr API. Users can develop programs in various programming languages to communicate using service requests, responded to with appropriate data according to given parameters. Here, we make use of Flickr's API documented at www.flickr.com/services/api/, and which provides free access to its entire database. *PhotoSearch* functions of Flickr API allow users to specify location and time to extract geotagged photos data. Actual photos and meta-data such as geotags, time taken and owner

information can be extracted. The locations for photo extraction are specified by a bounding box, whose parameters are given by la_{min} , la_{max} , lo_{min} , and lo_{max} for minimum latitude, maximum latitude, minimum longitude and maximum longitude respectively. Further detail on data collection is given in the case study section later.

3.4 Photo theme exploration

This stage provides decision makers with a supporting tool to facilitate the analysis and understanding of a large photo collection. The tool should directly apply to the actual photos, and automatically identify photos with similar content or themes, to make it convenient for analysis. The raw photo data need to be converted into a suitable representation that effectively captures their visual characteristics. Themes are effectively high level concepts covering various visual contents. No prior knowledge is assumed about what themes are available in a photo collection, or how to determine the similarity between photos. The supporting tool needs to work in an unsupervised manner and provide decision makers suggestions on potential photo themes. We address these challenges by adopting advanced techniques in computer vision and deep learning to capture visual features and automatically learn potential themes in photo collections, as detailed below.

3.4.1 Visual feature representation

Suppose, $P = \{p_1, p_2, \dots, p_m\}$ is a photo collection that we extracted from social media platforms and would like to explore its contents. The first step is to convert each photo p_i into a vector of features \mathbf{v}_i that represent its visual characteristics. We adopt a set of visual descriptors, proposed by Motion Picture Expert Group (<https://mpeg.chiariglione.org/>), to describe the visual features of the contents in photos: MPEG-7 has been an international standard for multimedia content description [34]. Although many descriptors are available, we select the following to describe various photo contents, which cover colour, texture and shape. Namely:

Scalable Color Descriptor (SCD): is computed based on Haar-transform applied onto values of a color histogram of a photo in hue-saturation-value (HSV) color space. The color space of HSV is uniformly quantized into 256 bins with 16 levels in H, 4 levels in S, and 4 levels in V. The histogram values are mapped onto 4-bit integer representation, assigning higher significance to smaller values, and then encoded using Haar transform. The basic unit of the transform consists of a sum and a difference operation, where the sum operation is equivalent to computing a histogram with half

number of bins, which results in subsets of the coefficients equivalent to histograms of 128 bins.

Color Layout Descriptor (CLD): is a compact and size-invariant representation of colour spatial distribution in a photo, by applying Discrete Cosine Transformation (DCT) on a two dimensional array of location representative colours in YCbCr color space. Y is the luma component and CB and CR are the blue-difference and red-difference chroma components. A photo is first divided into 64 blocks, and then a single dominant colour is extracted from each block by averaging pixel colours. This process results in a tiny image icon of size 8×8 , which is then transformed using 8×8 DCT into three set of 64 DCT-coefficients. They are zig-zag scanned and the first few coefficients are non-linearly quantized with 12 coefficients, 6 for Y, 3 for Cb and 3 for Cr.

Homogeneous Texture Descriptor - (HTD): is a robust, effective and easy to compute descriptor to characterize texture content in photo quantitatively. A photo is filtered by a bank of orientation and scale sensitive filters. The outputs in the frequency domain are described by mean and standard deviation as the first 2 coefficients. The frequency space is then partitioned into 30 channels of equal angular direction. The channels are modelled using Gabor functions to compute textual energy and energy deviations, which are then logarithmical, scaled to obtain a total of 60 coefficients.

Edge Histogram Descriptor (EHD): describes the spatial distribution of edges in a photo. A photo is first divided into 16 blocks, whose edge histograms are computed. The edges are grouped into five categories including vertical, horizontal, 45 diagonal, 135 diagonal and isotropic (non-orientation specific). Each local histogram has five bins corresponding to the five categories which produce 80 coefficients in total. Detailed description on the computation of the above descriptor is available from [35].

Besides the above mentioned descriptors, we also represent photo content using bag of visual words (BoW), a powerful technique used in many photo classification tasks [36]. Local region descriptors, named Speeded Up Robust Features (SURF) [37], are first extracted from each photo. *K-means* clustering is then applied to a large set of random SURF features to construct word vocabulary. Each visual word is defined by the centre of a cluster; the number of clusters determines the number of words. The value for k was set to 200 in our case, as a small number of visual words was sufficient for image categorization task, [38] and to save computation time. The SURF descriptors in each photo are then vector quantized into the visual words in the vocabulary.

The bag of visual words feature presents the count of visual words in the photos. After the feature extraction process, each photo p_i is represented as 482-dimensions vector feature, $v_i = \{128 - SCD, 12 - CLD, 62 - HTD, 80 - EHD, 200 - BoW\}$.

3.4.2 Photo theme identification

Having the visual features extracted from the photo, the next step is to apply a learning algorithm to automatically explore the potential themes. We adopt a technique in deep learning, named Deep Believe Network (DBN) [39], for this task. Deep learning is a relatively new brand of machine learning, which allows for learning high level abstractions of data in an unsupervised manner [40]. DBNs comprise unsupervised neural networks such as Autoencoder, stacked onto each other layer-by-layer. Auto-encoder is similar to a feed-forward artificial neural network, that maps sets of input data onto a set of outputs, which has an input layer, an output layer and one or more hidden layers connecting them. But the output layer has the same number of nodes as the input layer, as the purpose of Auto-encoder is to reconstruct its own input instead of predicting target values. In DBNs, each layer is trained in an unsupervised manner, and the hidden layer of Auto-encoder servers is the visible layer to the next.

An Auto-encoder consists of two parts, the encoder and the decoder, which can be defined by transition functions α and β . Let denote $x \in R^n = X$ as the input data, $y \in R^m = Y$ as code or latent variables, and $\sigma(\cdot)$ as activation function such as sigmoid function. The encoder function $\alpha : x \rightarrow y$ for one hidden layer auto encoder takes the form:

$$y = \sigma_1(Wx + b)$$

Here, y is usually referenced as code. σ is an activation function. W and b are weight and bias parameters respectively. Then, the decoder function $\beta : y \rightarrow \hat{x}$ maps y to the reconstruction \hat{x} of the same shape as x , which takes the form:

$$\hat{x} = \sigma_2(W'y + b')$$

The learning process of Auto-encoder aims to estimate values for $W, W', b,$ and b' such that the reconstruction representation \hat{x} is as close to the original input data x as possible, which minimizes reconstruction errors:

$$L(x, x') = \|x - x'\|^2 = \|x - \sigma_2(W'(\sigma_1(Wx + b)) + b')\|^2$$

At the first layer, the input data are the vectors of photo visual features $v \in R^{482}$. After training each auto encoder layer of DBNs, the value of feature vector y is used as input to train the next layer. The feature vector of the last layer encodes high abstract level concepts, which we regard as prospective themes. Since sigmoid function is used, the values of $y \in R^k$

are bound between 0 and 1. The value for k is provided by the user to determine the number of potential themes in the photo collection. In order to select the most representative photo for analysis, we measure the relevance of each photo to each potential theme using cosine similarity.

$$similarity(y, t) = \frac{y \cdot t}{\|y\| \|t\|}$$

where t is a theme representative vector, having the same dimension as y , but with 1 at the location of the current theme, and 0 elsewhere. If a photo is closely related to a theme then its encoded feature has high similarity to the representative vector of such theme. Users can inspect photos with a strong response to each prospective theme to determine the content covered within each theme.

3.5 Transition exploration

This section describes a method based on Markov Chain for modelling travel flows of people between geographical areas for traffic management [41]. Let $L = \{l_1, l_2, \dots, l_m\}$ be a set index referring to specific locations under consideration: each location may be a city or a geographical area determined by a user. If the social media data are tagged with geographical information (GPS latitude and longitude coordinates), we can determine the visited locations. The travel path of a person from time $t = 1$ to time $t = k$ is represented as $T = \{l_1^{t_1}, l_1^{t_2}, \dots, l_1^{t_k}\}$, here, l_i can be any location in L . The likelihood of people moving to a location l_i is defined by:

$$P(l_i^{t_n} | l_i^{t_{n-1}}, l_i^{t_{n-2}}, \dots, l_i^{t_0}) = P(l_i^{t_n} | l_i^{t_{n-1}})$$

The above equation implies the conditional independent assumption of Markov Chain, in order to model likelihood of travel from one location to another. The transition probability between location l_i at time $t = n$ and location l_j at time $t = n + 1$ ($i \neq j$), is computed by:

$$P(l_j^{t_{n+1}} | l_i^{t_n}) = \frac{P(l_j^{t_{n+1}} \cap l_i^{t_n})}{P(l_i^{t_n})}$$

where the numerator is the probability of travel paths to both locations l_i and l_j , with l_j being visited after l_i . The denominator is computed by:

$$P(l_i^{t_n}) = \sum_{j=1}^m P(l_i^{t_n} \cap l_j^{t_{n+1}})$$

The transition probability for all possible travel paths among different locations can be presented in one step transition probability matrix:

$$\mathbf{P} = \begin{pmatrix} 0 & P(l_2^{t_{n+1}} | l_1^{t_n}) & \dots & P(l_m^{t_{n+1}} | l_1^{t_n}) \\ P(l_1^{t_{n+1}} | l_2^{t_n}) & 0 & \dots & P(l_m^{t_{n+1}} | l_2^{t_n}) \\ \vdots & \vdots & \ddots & \vdots \\ P(l_1^{t_{n+1}} | l_m^{t_n}) & P(l_2^{t_{n+1}} | l_m^{t_n}) & \dots & 0 \end{pmatrix}$$

each entry $P_{ij} \in \mathbf{P}$ reflects how likely people are to travel from location l_i to location l_j .

3.6 Temporal pattern exploration

This section presents a general method to extract temporal information from social media data with time series modeling for trend estimation. Each post on social media platforms is automatically tagged with temporal information of the post or the actual time when a photo was taken. The temporal information, combined with geographical information, implicitly capture the footprints of users. Decision makers can track the presence of people at specific locations and at a given point in time. These insights are vital in crowd management to avoid location overload and minimize environmental impact [42]. We construct a time series data representing presence at a specific location. Let $S_i = \{s_{i1}, s_{i2}, \dots\}$ be a set of social media items posted during a time unit i as day, week or month. The time series data is denoted as $O = \{o_{t_1}, o_{t_2}, \dots, o_{t_N}\}$, each value o_{t_i} of the time series is defined by the number of users owning the item S_i .

The trend of the time series can be estimated using a fitting function (e.g. linear, exponential or quadratic) [43]. The use of parametric functions allows for smooth trend curves to be estimated, representing the overall tendency to infer future trends. The choice of fitting function is experimentally determined using the mean absolute error (MAE), a metric to measure model performance in time series analysis:

$$MAE = \frac{\sum_{i=1}^N |o_{t_i} - \hat{o}_{t_i}|}{N}$$

where \hat{o}_{t_i} is the estimated trend value. The model that produces lowest MAE is selected for trend estimation. Beside trend, seasonal demand, as reflected in monthly visit patterns are important for decision makers to tailor their strategy more appropriately. The time series data O is constructed according to month, where values of the same months are added up. The seasonal pattern is measured by the likelihood of users present in each month. We demonstrate the application of this approach in the case study later.

3.7 People-tracker artefact

The People Tracker artefact comprises four main components: data collection to extract the necessary data for analysis, photo theme exploration to identify people's activities, transition exploration to model travel behaviour, and temporal pattern exploration to model the presence of people at locations (e.g. for crowd management). These components are required to capture, reveal and monitor people's activity information to support country-wide decision-making. The advantage of our approach is the ability to extract photo themes in an unsupervised manner from photo collections without the need for textual meta-data, so that all available photos can be utilized. The geotag and photo taken time data was utilized to model the spatial and temporal behavioural patterns of people, which are difficult to capture using traditional surveying and statistical analysis methods (Fig. 1).

In consideration of the functionality of an application, the end-user only needs to select one or more social media platforms as data sources. Relevant parameters can be provided such as geographical area to survey and time period. The system will automatically retrieve the relevant data and perform the computation based on the built-in algorithms. Three sets of output can be obtained including prospective photo themes that may turn into a projected activity represented through some sample photos, flow likelihood represented on top of an integrated map, and graphs showing temporal patterns.

4 Artefact evaluation

For evaluating information systems artefacts the state-of-the-art literature distinguished qualitative and quantitative evaluation approaches in general, whilst Hevner et al. [19, p. 77] noted that "a mathematical basis for design allows many types of quantitative evaluations of an IT artefact, including optimization proofs, analytical simulation, and quantitative comparisons with alternative designs. The further evaluation of a new artefact in a given organizational context affords the opportunity to apply empirical and qualitative methods". Hevner et al. [19] proposed five evaluation methods (i.e. observational, analytical, experimental, testing, and descriptive) and suggested techniques to follow such as: experimental, observational, computer and lab simulations, field experiments, lab experiments, analytical processes, testing methods, case studies, surveys and field studies. Evaluation processes should be conducted during development stage to understand prospective problems and also should be conducted after the development stage. Gregor and Hevner [18] suggested that DSR artefacts must be evaluated for reliability, validity, and utility through data from the case studies that were used to inform the artefact's design. We conducted

Fig. 1 Conceptual framework of people-tracker artefact

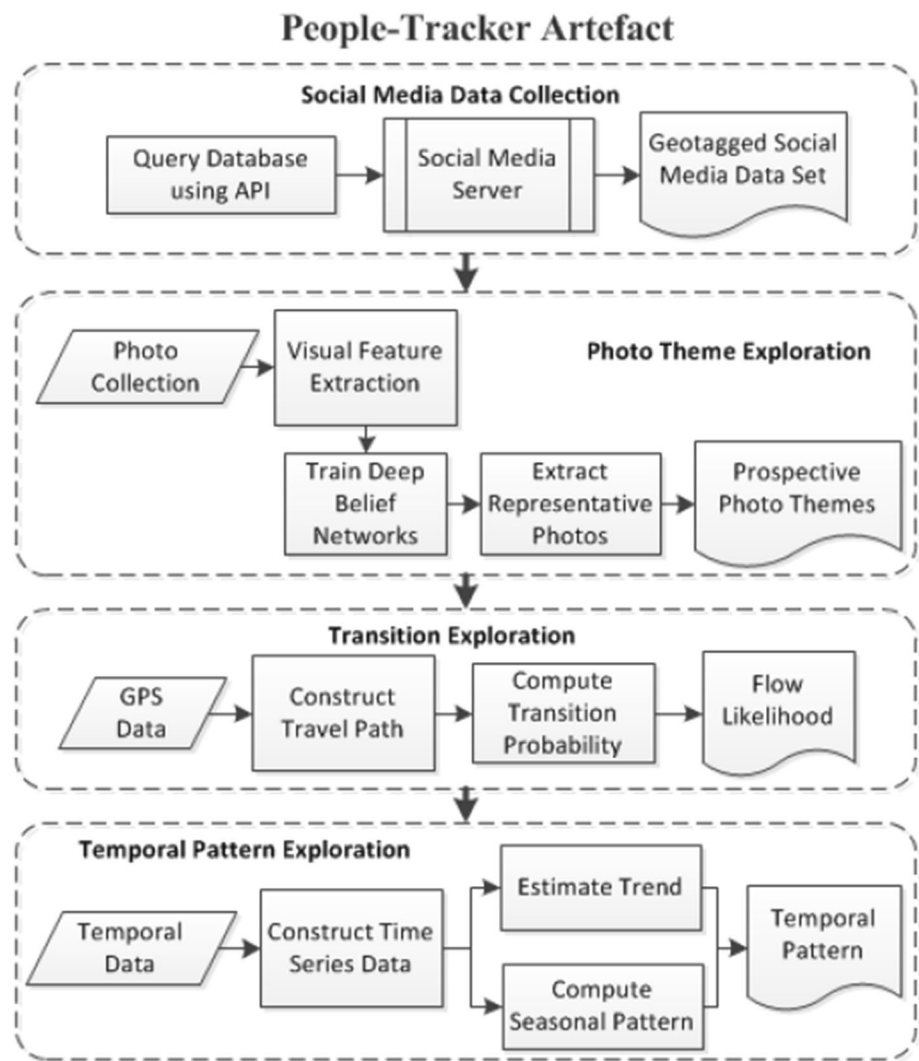


Table 3 Overall evaluation activities

Processes from the framework of Venable et al. [44].	Description
(1) Explicate the objectives	Our aim is to identify utility aspects
(2) Choose the evaluation strategies	We adopt the descriptive and observational approaches [19]
(3) Determine the properties to evaluate	We aim to demonstrate the accuracy, relevance and usability aspects of the proposed artefact
(4) Design the individual evaluation cases	We describe case details and practical demand of the government service providers through focus group discussions

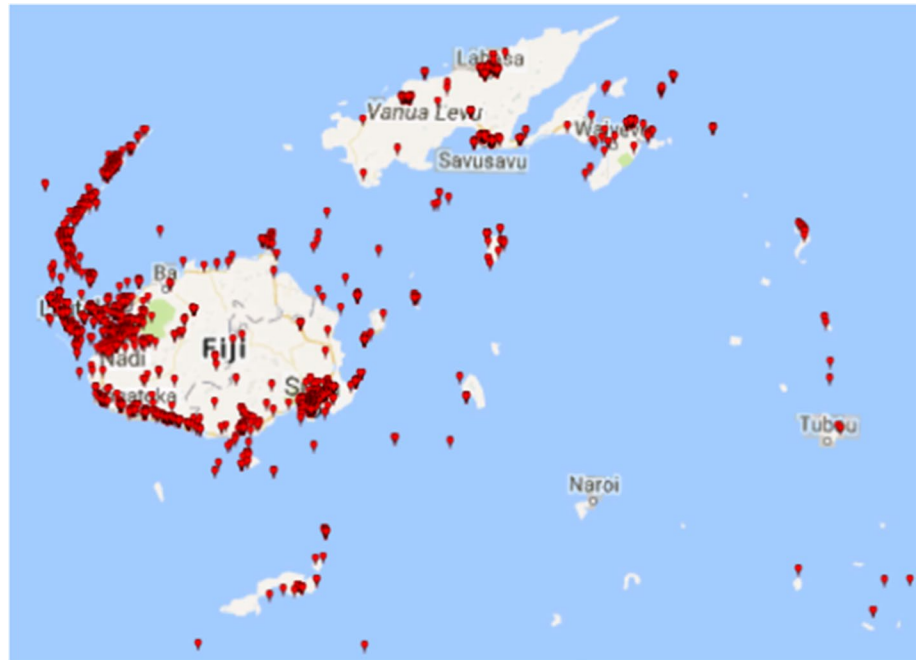
the qualitative evaluations after the development stage in this study, with internal quantitative testing during development.

In our study, we employed experimental and testing methods to evaluate the qualities of the proposed artefact using Venable et al.’s [44] framework for conducting the entire evaluation activities, as Table 3 illustrates.

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. We also adopted the descriptive approach by using case data (Fiji islands) as representative of the

Table 4 Parameters of bounding boxes

Parameter	Bounding box 1	Bounding box 2	Description
la_{min}	-19.788693	-19.788693	Minimum longitude of the bounding box
la_{max}	-15.391481	-15.391481	Minimum latitude of the bounding box
lo_{min}	176.456902	-180	Maximum longitude of the bounding box
lo_{max}	180	-177.929084	Maximum latitude of the bounding box

Fig. 2 Location of the geotagged photos in Fiji

country's residents and visitors, and which is readily validated against justifiable, accepted industry knowledge and independent travel statistics.

4.1 Data collection

We performed data collection of geotagged photos in Fiji based on the Stage 1 of the People Tracker Artefact. We need first to design a bounding box to cover the geographical area of Fiji for data extraction. Note that the longitude range is from -180 to 180, where the longitude lines -180 and 180 are at the same location on earth's surface and cross the middle of Fiji, so it is necessary to specify two bounding boxes (Table 4) as input parameters into the Photo Search function of Flickr API to cover the entire geographical area of Fiji.

No time limit was set as we wished to collect as much data as possible. In total, 11,130 photos were collected from 556 users: a sizable data set for a low population country like Fiji. The locations of the collected photos are recorded in form of GPS coordinates (latitude and longitude). We visualize the photo locations on a map using Google Earth, as shown in Fig. 2. We can see that the majority of the photos

were taken in the western region of the country, where major cities are located, such as Suva, Nadi and nearby islands.

4.2 Photo themes analysis

This section presents analysis exploring the common themes in the photos taken by tourists in Fiji, based on the Stage 2 of the people tracker artefact. Various visual features were used to capture as much detail from the photos as possible. The visual features of the photos are first extracted, where each photo is represented by a vector of 482 features (Scalable Color Descriptor, Color Layout Descriptor, Homogenous Texture Descriptor, Edge Histogram Descriptor, Region-Based Shape Descriptor, and Bag of Visual Word Features). We then apply the Auto-Encoder algorithm to learn hidden concepts in the photo collections so as to explore common photo themes. The number of nodes in the first layer was set to 400 to learn the high level concepts from the low level visual features, and then another round of Auto-Encoder was applied to the high level features to learn the themes. The number of nodes in the second layer is provided by users to determine the number of themes in the photo collection to be analysed.

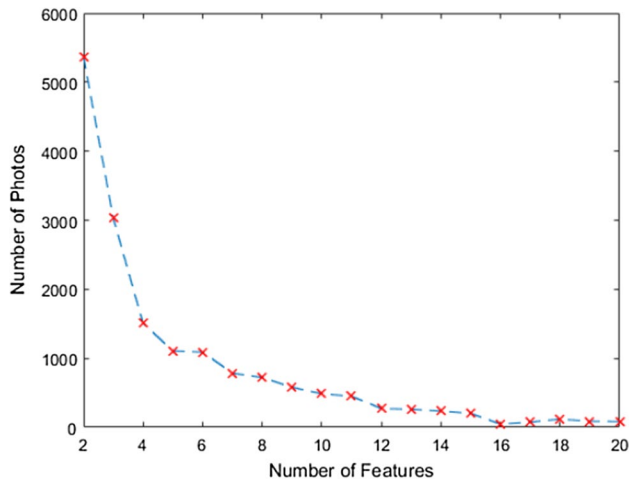


Fig. 3 Number of photos in the last theme with different value of k

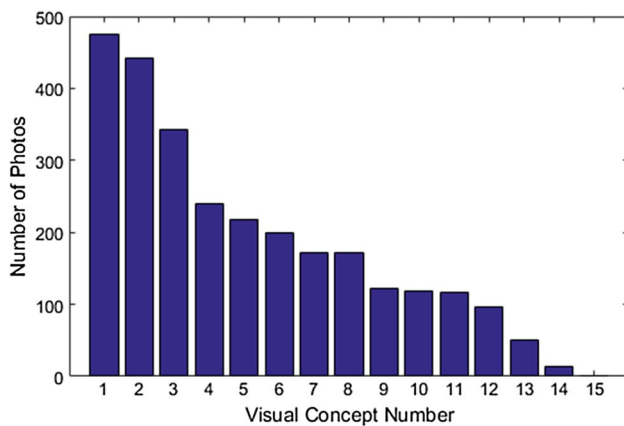


Fig. 4 Number of photos in all themes when $k=15$

4.2.1 Model evaluation

Since, there is no prior knowledge on how many themes are available; we performed an evaluation of the model on different numbers of nodes of the second layer to determine the most suitable one. We set the value for k (number of potential themes) from 2 and increased it by one iteratively. We counted the number of photos belonging to each prospective theme with different k values. Here, a photo is considered as belonging to a prospective theme, if its cosine similarity to the vector representing such theme is 0.7 or more. If a theme has a high number of photos, it is *popular*: if a theme has few photos, it is *insignificant*. We examine the number of photos in the theme having fewest photos, as shown in Fig. 3 with k from 2 to 20. We can see that the numbers drop as k increases, because more themes are used to represent the photo collections, lowering the number of photos. We notice that when k is set to 16 or higher, the number of photos in

the last theme is almost 0. This means that there are only 15 significant themes in the photo collection. Thus, we pick 15 as the value for k in our analysis.

4.2.2 Photo theme identification

In order to make it easy for users to determine the themes, we only extract photos which have strong similarity to the prospective themes. We set a cosine threshold of 25 degrees. Figure 4 shows the number of photos which are close to their corresponding theme, ordered from the most to the least photos. The common themes are those with the most photos associated with them. For instance, themes 1 and 2 have more than 400 photos.

Business decision makers can manually examine the photos in each theme for insights into the captured contents. Table 5 shows some representative photos for the top 10 themes and their prospective name. We examine the photos in each theme to identify the general content.

Theme one contains photos showing people visiting sea sites and watching oceans and waves, a common leisure activity to help refresh the mind. Theme two shows general scenes of land including hills, mountains, and agricultural crops. Themes three and seven have common content around visitor activities in Fiji, such as camping and traveling to other locations. Daytime site-seeing activity such as coach tours or parades is distinguished from ad hoc activities, such as evening beach walks. The differences between these two themes is background colour; theme three photos have darker backgrounds, more likely to be taken near dawn or dusk or when the sky is cloudy with less sunlight; theme seven photos have bright backgrounds, probably taken during day time with clear skies. Theme four shows people enjoying various day time water sports, such as diving, water skating, or swimming. Theme five shows people's activities in various gathering occasions such as marriage ceremonies, music, barbeque, food and others. Theme six contain photos showing sunset and sunrise at various locations close to the sea. Views with coconut trees, which are iconically tropical, also received considerable interest as indicated in theme eight. Theme nine shows underwater activities such as diving for creature watching. Theme ten shows people gathering for shopping and wandering in crowds or market places.

We are aware that the distinction between some themes may seem trivial. In addition to statistically derived groups, in application contexts more or less granularity may be relevant, with clusters merged or differentiated based on frequency of items, or semantic criteria from external classifications or human judgements. Nevertheless, for our chosen settings, the algorithm effectively identified relatively different themes from the photo collection, and people who understand Fiji can clearly understand the differences.

Table 5 Representative photos in top 10 themes































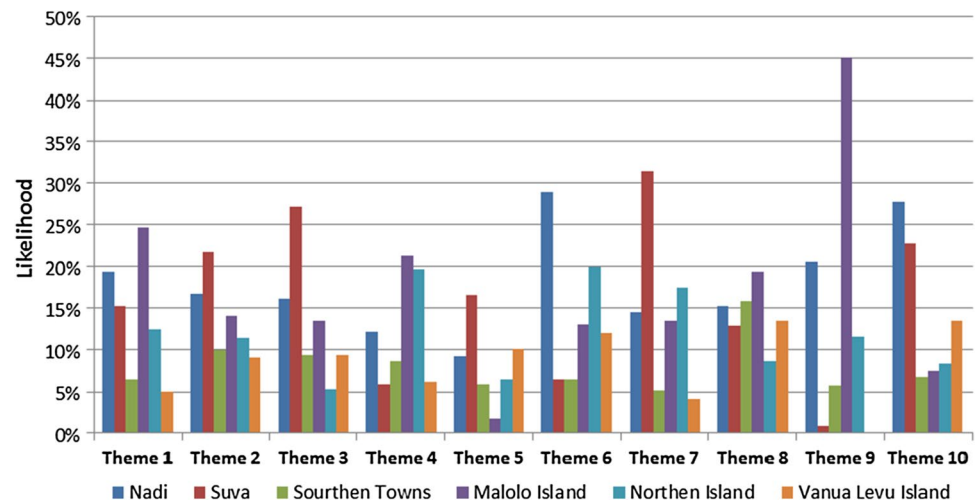
Theme	Description	Sample photos		
1	Sea focussed scenes			
2	Land focussed scenes			
3	Land based activity			
4	Sea based activity			
5	Organised social events			
6	Sun set and sun rise scenes			
7	Daytime site seeing			
8	Coconut tree scenes			
9	Under water activity			
10	Everyday crowds			

Fig. 5 Likelihood of photo themes by location

4.2.3 Photo theme distribution

We further investigate the activities distribution across their locations in Fiji. We grouped the photos in each theme according to different areas in Fiji. For instance, Nadi, a city on the North West part of Fiji's main island, is the main transportation hub. Suva is the capital of Fiji, with special characteristics such as broad avenues, lush parks, grand British colonial buildings and various tourist attractions. Southern towns include a number of small cities and towns along the south coast of Fiji's main island. Malolo Island is located in the north west of Fiji, with numerous famous resources for leisure and tourism. Northern Islands contains various islands from Nalauwaki to Yasawa-i-Rara, located further to the north of Fiji's main island. Vanua Levu Island is the second largest island of Fiji, located to the North East.

We computed the likelihood of the photo according to these areas, shown in Fig. 5, and found the following. Human activity photos in themes three and seven were taken more in Suva than other places with above 25% and 30% respectively. Suva is the capital city, where scenes with human activities are popular. People gathering at shopping and market place photos in theme ten appear mostly in Suva and Nadi, the two most crowded cities in Fiji. Sea activity photos in theme four are taken mostly in Malolo and Northern Islands. It is interesting that sunset and sunrise photos in theme six were taken mostly in the northern part including Nadi and Northern Islands. Underwater activity photos in theme nine are most popular in Malolo Island with 45%, probably due to the developed underwater tourism in this area.

4.3 Flow analysis

This section analyses the flow of residents reflected through the trajectory of geotagged photos. The photos are arranged

in sequential time order, and the photo locations are connected by a line on a map representing the movement (Fig. 6). The majority of movement is within the main island and to the nearby islands.

In order to examine the flow of people, Markov Chain technique, described in the Stage 3 of the People Tracker artefact, is applied to compute the transition probability between the areas. For present purposes, only the transition probabilities ≥ 0.22 are shown in Fig. 7.

- Residents from Malolo Island, Vanua Levu and Suva are most likely to flow to Nadi (probabilities > 0.4). Especially for residents in Vanua Levu, Nadi is the most connected area with travel flow of around 0.6, consistent with the fact that Nadi is an important transportation hub, which was captured by the geotagged photo data.
- People are very likely to travel between Southern towns and Suva with probabilities around 0.4, indicating the high need of residents to move between these areas.
- People are likely to travel from Malolo directly to Northern Islands with probability of 0.32, while from Northern Islands; they are likely to travel to Nadi with a probability of 0.39, much more than their transition of 0.26 back to Malolo. Transportations for the northern areas can thus be adjusted to accommodate the travel need based on these patterns.
- The probability of 0.22 from Northern Islands to Southern Towns indicates a need for directly commuting between these areas. Fiji government may consider creating directly lines of transportation between Northern Islands and Southern towns.

Fig. 6 Raw travel trajectories in Fiji

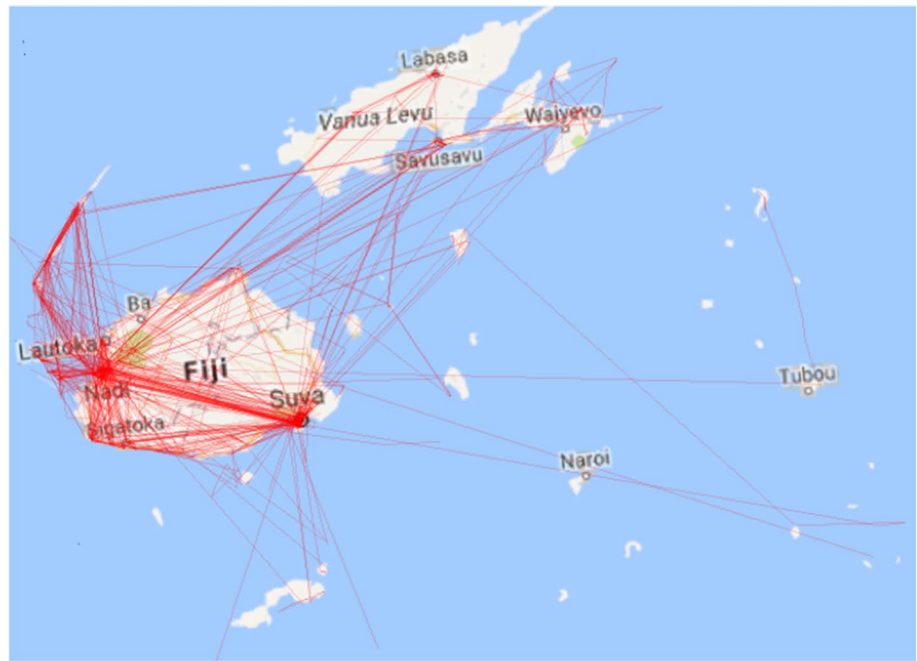
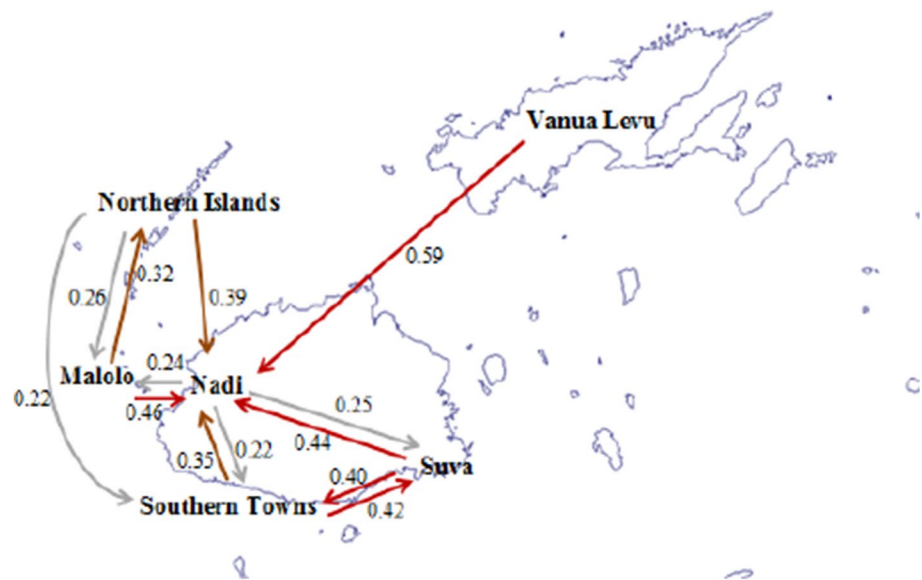


Fig. 7 Flow probabilities between different areas of Fiji



4.4 Tourism prediction

Tourism is the largest industry of Fiji, so it is important for country management to understand the trends and arrival patterns of tourists so that country-wide development can be aligned with the development of tourism industry. This section uses temporal pattern exploration techniques, based on the Stage 4 of the People Tracker artefact, to reveal such patterns from geotagged-photo data. We firstly identify *tourists* from Flickr users who stated their location of origin as outside Fiji. While some users did not provide their location (this is not required to register for a Flickr account) in

total 556 tourists were identified, whose photo collections representing their temporal presence in Fiji were processed to construct time series data by month.

We firstly examine the time series to predict the future trend of tourist arrival. Since no prior knowledge is available about which fitting function is best to model the available data, we carried out an initial evaluation to determine the most suitable. Three popular functions were examined: Quadratic Model, Linear Model, and Exponential Model [43]. The majority of the photos were taken between 2009 and 2016. We used the data from 2009 to 2015 as training

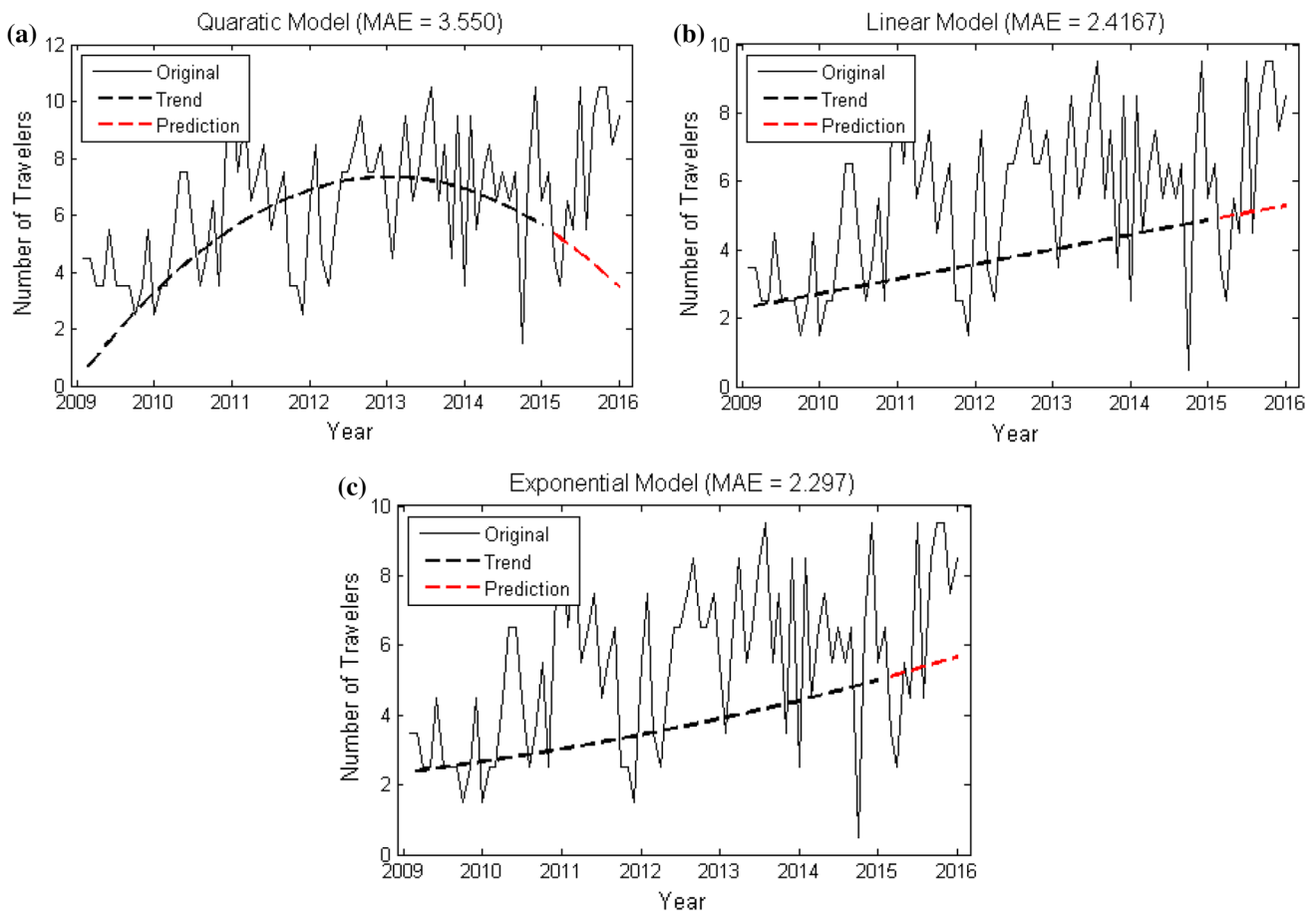


Fig. 8 Trend estimation of different models

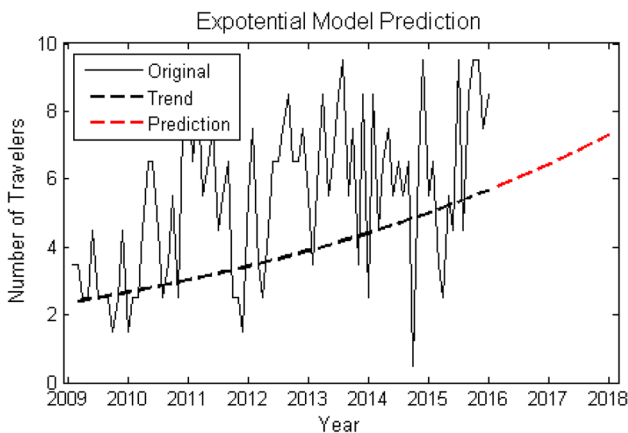


Fig. 9 Trend prediction using exponential model

data for model fitting, while the time series data for 2015 to 2016 was used for testing.

Figure 8 shows the original time series data, estimated trend and prediction value of the evaluated models. The original time series data represents the number of travellers

by month. Due to space limitation, only the labels for year are displayed. The MAE values were computed on the testing data, shown in Fig. 8’s titles. The Exponential model has the prediction error of 2.297, which is lowest among evaluated models. We thus select the exponential model for the prediction task, where the exponential model is fitted on the entire time series and plot prediction values for 2017 and 2018 in Fig. 9. The prediction from our data set clearly shows an increasing trend in the number of tourists. More and more tourists come to Fiji and share their photos on social media platform like Flickr. This pattern also reflects the increasing popularity of social media among tourists, and government agencies should utilize such data sources better to capture information about visitors’ behavioural patterns to support their decision making.

We are aware that different groups of tourists may have different behavioural patterns. Thus, we further examine their seasonal pattern reflected in the time series data. The majority of tourists came from countries in Oceania (158 users), North America (119 users) and Europe (92 users). This is consistent with the fact that Australia and New Zealand in Oceania are geographically located closed to Fiji,

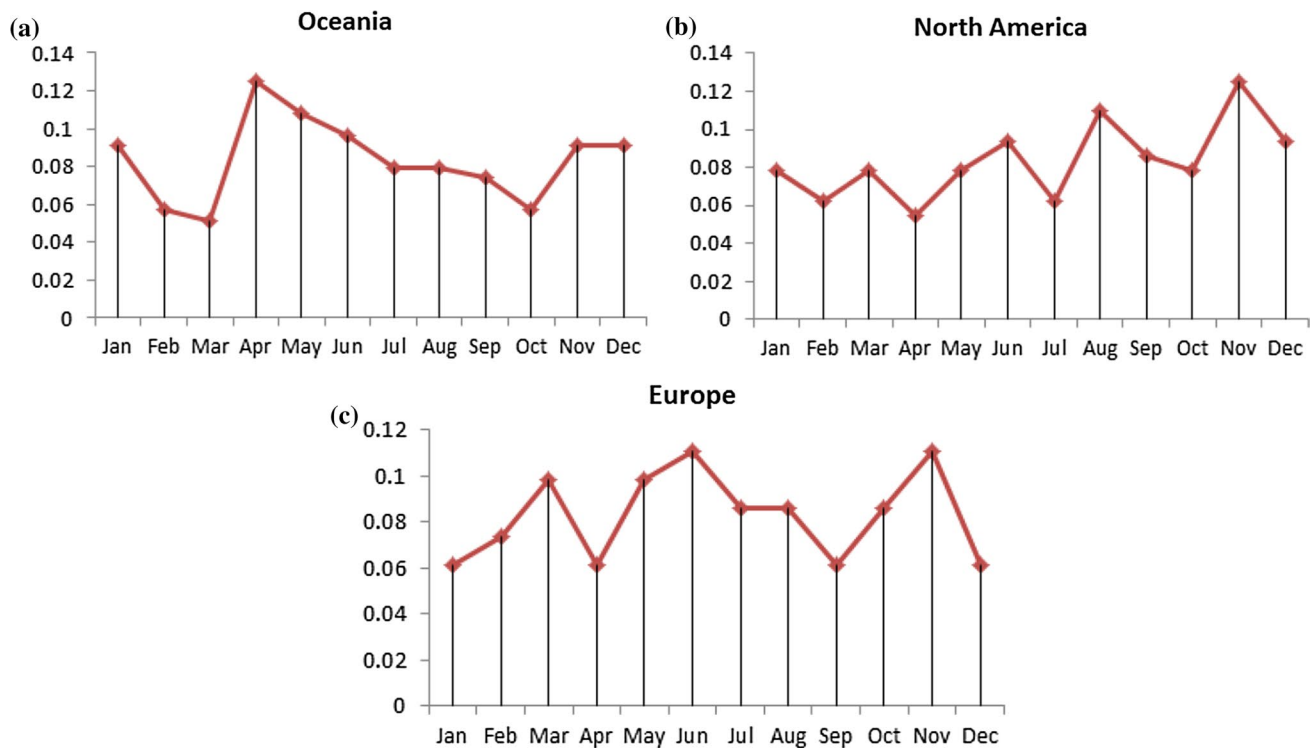


Fig. 10 Seasonal patterns

thus people from these countries are more likely to visit Fiji than people from other continents. Interestingly, very few tourists from Asian countries were found in our data set despite their locations being relatively close to Fiji in comparison to European or North American countries.

For the purpose of seasonal pattern analysis, we group tourists based on their continent of origin. We fit the exponential function to the time series data of each tourist group. The seasonal patterns, (Fig. 10), are obtained by subtracting the original data from the trend component. The fluctuation of the seasonal component is of interest rather than the actual values. We can see that tourists from Oceania tend to visit Fiji in April with a clear peak, distinct from other months. March and October are the low season for Fiji travel of this tourist group. North Americans tend to visit Fiji toward the end of the year with peaks in August and November. Tourists from Europe tend to visit Fiji in March, June and November. Country management organizations in Fiji, especially those in tourism sectors, can tailor their travel package and services better to suit targeted tourist groups at different times of the year to improve satisfaction, important in developing the tourism industry.

5 Discussion

Beyond the internal data from various organisational resources, data generated by social media provide neutral details of individuals' experiences and expression with time-stamped, demographic and evidence-based insights that contributes, with external validity, to the authorities' decision support. Traditional analytics methods and specialised techniques have been limited for tourist management purposes, although they focus on analysing huge and unstructured social media datasets in diversified data formats, the growth of this data is massive. Previous studies introduced analytics solutions for automatically detecting people's behaviour and city preferences (e.g. [45]) but were not designed for comprehensive decision support for country wide people monitoring, nor justified through independent validation, considering user and their decision support situations. The advantage of our approach is the ability to directly process the photo content using modern visual features and deep machine learning techniques to identify human activities, which could not be obtained effectively using traditional manual approaches. Intra-country travel flows of people were also identified effectively by applying Markov Chain technique on geographical information of the photos. Our approach performs analysis and produces predictions based on temporal information extracted from social media data that enhances validity of other data sources such as surveys.

Whilst interest in Smart City applications design have has triggered, the proposed design study has contributed in terms of a practical work that has been made in implementing related initiatives, specifically for smart data management and its processing for different stakeholders or decision makers such as city engineers and other solution design professionals. As the requirements of design research must provide either an unsolved problem in an innovative way or a solved problem through a more efficient method, the proposed IS artefact can be viewed as an innovative systematic solution for exploring on people activities for country or city wised demand forecasting when considering about modern surveillance for new inside generation in terms of people movements, social inclusion, and a sense of space [5].

Following the DSR approach in which the six activities [31] were utilised for assisting the proposed artefact design in three distinctive phases, we have gone beyond existing big data analytics methods in automatically detecting people's interest in objects, particular spots and clusters, along with detailed insights on collective behavioural flows and nationality profiles. Although the analytics based artefact has developed and evaluated in the form of a general method for generating meaningful supporting details and predictive insights from a selected social media content such as geotagged photos, the three distinctive design phases that we introduced in the study would be reused for similar artefact design studies. Results show that our solution method is able to detect key patterns and trends for a representative case context in a country perspective.

In a recent smart city research initiative, Singh et al. [46] introduced a method for monitoring road conditions using data collected from the built-in sensors of smart phones. Singh experienced issues such as (1) sensitivity of sensors vary from phone to phone; (2) detection rate of smartphones differ; (3) positioning a smartphone on moving vehicles may impact accuracy. In our research, we offered a new method for detecting social media big-data that provides opportunity for gaining evidence-based insights. Our method uniquely integrates component technologies such as clustering with temporal pattern extracting techniques comprehensively to convert data into effective decision support information regarding people's movement and flows.

Social media provides a source of valuable, but unstructured data. One limitation of the study, however, concerns the completeness of the collected data, so comparing unique photographer numbers with known visitor numbers is required to indicate sample representativeness. Another limitation is the profile of those posting photos, which may be assumed to be younger and tech-savvy. In time though, it is anticipated that older people will increasingly be taking photos and uploading them to social media. A general limitation is that photos are being used as a proxy for identifying future interest, and past data is not a guaranteed predictor of

the future. For strategic decisions however, it is likely that results from this analysis would be triangulated with results based on other or internal data sources, and that the results provided by the approach described here will add value to existing instruments.

6 Conclusion

This study promoted a smart city research initiative, through which an application of social media data were demonstrated for their utilities in providing decision support benefits for city authorities. A new method called *People Tracker* for analysing social-media based big-data to provide decision support for government authorities was presented. We followed an established DSR methodological guideline for the design, development and dissemination of the generated artefact, namely a *method*, one of the four types of design artefact recognised in the DSR literature. We utilised *MATLAB* as a numerical computing environment, together with Google maps, a desktop web mapping service, as the technical testing environment to design and evaluate the proposed solution method. Under the smart city research paradigm, the benefit of the design is to improve strategic decision support so that local authorities can smartly control and monitor people's activities across locations. A comprehensive understanding of people's activities, interests, visited locations, experiences are explored and future demands are predicted to support decision making of smart city stakeholders. The study showcased how data generated by social media could provide neutral details of individuals' experiences and expression in order to produce evidence-based insights that have effective contributions, with external validity going beyond the internal analysis, to the authorities' decision support. An immediate extension of this work will be to apply the proposed method in other countries to explore people behaviour and their preferences in different social and business contexts. Photos on other social media platforms could also be incorporated for generating more comprehensive understanding that would be leading to appropriate managerial or strategic decision support.

References

1. Pick JB, Turetken O, Deokar AV, Sarkar A (2017) Location analytics and decision support: reflections on recent advancements, a research framework, and the path ahead. *Decis Support Syst* 99:1–8
2. Lozano MG, Schreiber J, Brynielsson J (2017) Tracking geographical locations using a geo-aware topic model for analyzing social media data. *Decis Support Syst* 99:18–29
3. Power DJ (2015) 'Big data' decision making use cases. Springer, Cham. https://link.springer.com/chapter/10.1007%2F978-3-319-18533-0_1. Accessed 20 Nov 2018

4. Song H, Liu H (2017) Predicting tourist demand using big data. In: Xiang Z, Fesenmaier DR (eds) *Analytics in smart tourism design—concepts and methods*. Springer, Cham
5. Hashem IAT, Chang V, Anuar NB, Adewole K, Yaqoob I, Gani A, Ahmed E, Chiroma H (2016) The role of big data in smart city. *Int J Inf Manag* 36:748–758
6. Falconer G, Mitchell S (2012) Smart city framework: a systematic process for enabling smart+ connected communities. Cisco Internet Business Solutions Group (IBSG), Cisco IBSG, Cisco, 09/12, pp 1–10
7. Mangharam R, Reyerson M, Viscelli S, Balakrishanan H, Bayen A, Amin S, Richards L, Bagley L, Pappas G (2017) MOBILITY21: strategic investments for transportation infrastructure & technology, CCC Led Whitepapers. <http://cra.org/ccc/resources/ccc-led-whitepapers/>. Last accessed 12 April 2017
8. Chen L, Nugent C, Mulvenna M, Finlay D, Hong X (2009) Semantic smart homes: towards knowledge rich assisted living environments. In: McClean S, Millard P, El-Darzi E, Nugent C (eds) *Intelligent patient management*. Springer, pp 279–296
9. Demirkan H (2013) A smart healthcare systems framework. *IT Prof* 15(5):38–45
10. Harvey C (2017) Big data challenges, datamation. <http://www.datamation.com/big-data/big-data-challenges.html>. Accessed 20 Nov 2018
11. Thelwall M, Goriunova O, Vis F, Faulkner S, Burns A et al (2016) Chatting through pictures? A classification of images tweeted in one week in the UK and USA. *J Assoc Inf Sci Technol* 67(11):2575–2586
12. Sivarajah U, Kamal MM, Irani Z, Weerakkody V (2017) Critical analysis of big data challenges and analytical methods. *J Bus Res* 70:263–286
13. Xiang Z, Fesenmaier DR (2017) *Analytics in smart tourism design: concepts and methods*. Springer. <http://www.springer.com/gp/book/9783319442624>. Accessed 20 Nov 2018
14. Marine-Roig E (2017) measuring destination image through travel reviews in search engines. *Sustainability* 9(8):1425. <https://doi.org/10.3390/su9081425>
15. Xiang Z, Schwartz Z, Gerdes JH, Uysal M (2015) What can big data and text analytics tell us about hotel guest experience and satisfaction? *Int J Hosp Manag* 44:120–130. <https://doi.org/10.1016/j.ijhm.2014.10.013>
16. Marine-Roig E, Clavé SA (2015) Tourism analytics with massive user-generated content: a case study of Barcelona. *J Destin Mark Manag* 4:162–172. <https://doi.org/10.1016/j.jdmm.2015.06.004>
17. Huang A (2016) TweetDay: a better visualization for your Twitter timeline. <https://medium.com/@andyoyellow/tweetday-a-better-visualization-for-your-twitter-timeline-49e95daf4ae>. Accessed 20 Nov 2018
18. Gregor S, Hevner AR (2013) Positioning and presenting design science research for maximum impact. *MIS Q* 37(2):337–355
19. Hevner A, March S, Park J, Ram S (2004) Design science in information systems research. *MIS Q* 28(1):75–105
20. Buhalis D, Amaranggana A (2014) Smart tourism destination. In: *Proceedings of 21st information and communication technologies in tourism*. https://doi.org/10.1007/978-3-319-03973-2_40. Accessed 20 Nov 2018
21. Japan Smart City Portal (2013) Japan moving forward on smart cities. <https://www.greenbiz.com/news/2013/02/25/japan-moving-forward-smart-cities>. Accessed 20 Nov 2018
22. Li Y, Xiao L, Ye Y, Xu W, Law A (2016) Understanding tourist space at a historic site through space syntax analysis: the case of Gulangyu, China. *Tour Manag* 52:30–43
23. Cheng M, Edwards D (2015) Social media in tourism: a visual analytic approach. *Curr Issues Tour* 18(11):1080–1087
24. Miah SJ, VU HQ, Gammack J, McGrath GM (2017) A big-data analytics method for tourist behaviour analysis. *Inform Manage* 54(6):771–785
25. Chua A, Servillo L, Marcheggiani E, Moere AV (2016) Mapping Cilento: using geotagged social media data to characterize tourist flows in southern Italy. *Tour Manag* 57:295–310
26. WTTC (2015) World travel & tourism council, travel & tourism economic impact 2015 Fiji. <https://www.wttc.org/-/media/files/reports/economic%20impact%20research/countries%202015/fiji2015.pdf>. Accessed 20 Nov 2018
27. Kalouniviti M (2016) Tourism report data supports bureau statistics, The Fiji Times ONLINE. <http://www.fijitimes.com/story.aspx?id=383648>. Accessed 22 Nov 2017
28. Baskerville RL, Kaul M, Storey VC (2015) Genres of inquiry in design-science research: justification and evaluation of knowledge production. *MIS Q* 39(3):541–564
29. Geerts LG (2011) A design science research methodology and its application to accounting information systems research. *Int J Account Inf Syst* 12:142–151
30. March S, Smith G (1995) Design and natural science research on information technology. *Decis Support Syst* 15:251–266
31. Peffers K, Tuunanen T, Rothenberger M, Chatterjee S (2008) A design science research methodology for information systems research. *J Manag Inf Syst* 24(3):45–77
32. Kisilevich S, Krstajic M, Keim D, Andrienko N, Andrienko G (2010) Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections. In: *Proceedings of the 14th international conference information visualisation*. <https://doi.org/10.1109/IV.2010.94>. Accessed 20 Nov 2018
33. Andrijic E, Haimes Y, Beatley T (2013) Public policy implications of harmonizing engineering technology with socio-economic modeling: application to transportation infrastructure management. *Transp Res Part A Policy Pract* 50:62–73
34. ISO (2002) *Electronic (ISO2002) ISO/IEC 15938-1:2002 Information technology—multimedia content description interface—part 1: systems 2002*
35. Manjunath BS, Ohm JR, Vasudevan VV, Yamada A (2001) Color and texture descriptors. *IEEE Trans Circuits Syst Video Technol* 11:703–715
36. Csurka G, Dance CR, Fan L, Willamowski J, Bray C (2004) Visual categorization with bags of keypoints. *Workshop on statistical learning in computer vision*, pp 1–22
37. Bay H, Ess A, Tuytelaars T, Gool LV (2008) SURF: speeded up robust features. *Comput Vis Image Underst* 110:346–359. <https://doi.org/10.1016/j.cviu.2007.09.014>
38. Li FF, Perona P (2005) A Bayesian hierarchical model for learning natural scene categories. In: *Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition*, pp 524–531. <https://dl.acm.org/citation.cfm?id=1069129>. Accessed 20 Nov 2018
39. Hinton GE (2009) Deep belief networks. *Scholarpedia* 4(5):5947. <https://doi.org/10.4249/scholarpedia.5947>
40. Deng L, Yu D (2014) Deep learning: methods and applications. *Found Trends Signal Process* 7(3–4):197–387. <https://doi.org/10.1561/20000000039>
41. Xia J, Zeepongsekul P, Arrowsmith C (2009) Modelling spatio-temporal movement of tourists using finite Markov chains. *Math Comput Simul* 79:1544–1553
42. Haghighi PD, Burstein F, Li H, Wang C (2013) Integrating social media with ontologies for real-time crowd monitoring and decision support in mass gatherings. In: *Pacific Asia conference on information systems*, p 64. <https://aisel.aisnet.org/pacis2013/64>. Accessed 20 Nov 2018
43. Cooray T (2008) *Applied time series: analysis and forecasting*. Alpha Science Intl Ltd., Oxford
44. Venable J, Pries-Heje J, Baskerville R (2016) FEDS: a framework for evaluation in design science research. *Eur J Inf Syst* 25(1):77–89

45. Zhou X, Xu C, Kimmons B (2015) Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Comput Environ Urban Syst* 54:144–153
46. Singh G, Bansal D, Sofat S, Aggarwal N (2017) Smart patrolling: an efficient road surface monitoring using smartphone sensors and crowdsourcing. *Pervasive Mobile Comput* 40:71–88. <https://doi.org/10.1016/j.pmcj.2017.06.002>
47. Min W, Bao BK, Xu C (2014) Multimodal spatio-temporal theme modelling for landmark analysis. *IEEE Multimedia* 21(3):20–29
48. Spyrou E, Mylonas P (2016) Analyzing Flickr metadata to extract location-based information and semantically organize its photo content. *Neurocomputing* 172:114–133
49. Deeksha SD, Ashrith HC, Bansode R, Kamath S (2016) A spatial clustering approach for efficient landmark discovery using geo-tagged photos. In: 2015 IEEE international conference on electronics, computing and communication technologies (CONECCT). Bangalore, India
50. Xu Y, Hu T, Li Y (2016) A travel route recommendation algorithm with personal preference. In: 2016 12th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD), Changsha, China
51. Wen YT, Cho KJ, Peng WC, Yeo J, Hwang SW (2016) KSTR: keyword-aware skyline travel route recommendation. In: 2015 IEEE international conference on data mining, 2016, Atlantic City, NJ, USA

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.