



Article

Using the TensorFlow Deep Neural Network to Classify Mainland China Visitor Behaviours in Hong Kong from Check-in Data

Shanshan Han ¹ , Fu Ren ^{1,2}, Chao Wu ¹, Ying Chen ¹, Qingyun Du ^{1,2,3,4,*} and Xinyue Ye ^{5,*} 

¹ School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China; hanshan@whu.edu.cn (S.H.); renfu@whu.edu.cn (F.R.); chaowu@whu.edu.cn (C.W.); 2014301130140@whu.edu.cn (Y.C.)

² Key Laboratory of Geographic Information Systems, Ministry of Education, Wuhan University, Wuhan 430079, China

³ Key Laboratory of Digital Mapping and Land Information Application Engineering, National Administration of Surveying, Mapping and Geoinformation, Wuhan University, Wuhan 430079, China

⁴ Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China

⁵ Department of Geography, Kent State University, Kent, OH 44242, USA; xye5@kent.edu

* Correspondence: qydu@whu.edu.cn (Q.D.); xye5@kent.edu (X.Y.); Fax: +86-27-6877-8893 (Q.D.)
Tel.: +86-27-8766-4557 (Q.D.); +1-419-494-7825 (X.Y.); Fax: +86-27-6877-8893 (Q.D.)

Received: 26 February 2018; Accepted: 19 April 2018; Published: 21 April 2018



Abstract: Over the past decade, big data, including Global Positioning System (GPS) data, mobile phone tracking data and social media check-in data, have been widely used to analyse human movements and behaviours. Tourism management researchers have noted the potential of applying these data to study tourist behaviours, and many studies have shown that social media check-in data can provide new opportunities for extracting tourism activities and tourist behaviours. However, traditional methods may not be suitable for extracting comprehensive tourist behaviours due to the complexity and diversity of human behaviours. Studies have shown that deep neural networks have outpaced the abilities of human beings in many fields and that deep neural networks can be explained in a psychological manner. Thus, deep neural network methods can potentially be used to understand human behaviours. In this paper, a deep learning neural network constructed in TensorFlow is applied to classify Mainland China visitor behaviours in Hong Kong, and the characteristics of these visitors are analysed to verify the classification results. For the social science classification problem investigated in this study, the deep neural network classifier in TensorFlow provides better accuracy and more lucid visualisation than do traditional neural network methods, even for erratic classification rules. Furthermore, the results of this study reveal that TensorFlow has considerable potential for application in the human geography field.

Keywords: check-in data; visitor behaviours; deep neural network; TensorFlow; Hong Kong

1. Introduction

In recent years, considerable research has focused on human mobility and travel behaviours using big data. These big data include Global Positioning System (GPS) data [1,2], mobile phone tracking data [3,4], and social media check-in data [5,6]. These data have been widely used to determine transportation patterns [3,4], urban daily commuting behaviours [7] and even dynamic movement trajectories in combined spatiotemporal analyses [8,9]. Specifically, studies of tourism management have explored the potential of applying big data to assess tourist behaviours. In 2015, the world tourism industry generated a revenue of \$1.5 trillion, with 1.2 billion international arrivals [10]. Therefore,

such studies of tourist management are essential to the tourism industry, which plays an important role in economic development in many countries and regions, particularly in popular tourist destinations. In recent years, tourism research has focused more on tourists than on tourism resources, particularly tourist movements and behaviours. However, human behaviours are complex; they may be prompted by intentions or habits; modified by skill, affect and attitude; and affected by physical and contextual conditions [11]. Many methods either are based on certain assumptions to simulate human behaviours or are unable to consider all the factors that influence human behaviour [12].

In this context, deep learning methods may provide state-of-the-art solutions to comprehensively understand human behaviours. Since Google's Artificial Intelligence programme AlphaGo became the first computer programme to beat a 9-dan professional human player without a handicap in March 2016 [13], people have paid increasing attention to deep learning and artificial intelligence and noted the powerful potential of their applications; moreover, deep neural networks have rapidly outpaced human beings' understanding of the nature of their solutions [14]. Specifically, Google DeepMind developed AlphaGo in 2015. In April 2016, DeepMind started using TensorFlow for future research and eventually moved completely to TensorFlow. TensorFlow is a deep learning open source library developed by Google Inc. Since the library was open sourced in November 2015, TensorFlow has been able to excel in image processing [15], including handwritten digit recognition [16,17], visual object recognition and detection [18] from images, and even dynamic object tracking from video [19]. In addition, the tool can be used in voice recognition, natural language processing [20], etc. A new study on DeepMind showed that deep neural networks can be successfully explained in a psychological manner [14], suggesting that deep neural networks can potentially understand and extract human behaviours in a more interpretable way. Many studies have already applied deep learning methods in various areas of human behaviour research, such as human action recognition [21] and human trajectory prediction [22]. Still, few attempts have been made to understand tourist behaviours. In this paper, we explore the possibility of applying the TensorFlow deep neural network to tourism geography to classify tourist behaviours and innovatively implement a deep learning method constructed using TensorFlow to classify behaviours of check-in users based on neural network theory.

The following paper is composed of five parts. Following an introduction, Section 2 briefly reviews existing studies about tourist research involving social media data and human behaviour research involving deep learning methods. Section 3 illustrates the research methodology. In this section, we provide a brief introduction to the theory underlying TensorFlow and neural networks and illustrate the data processing flow. Specifically, the data processing flow includes preprocessing and classification steps. Section 4 introduces the study area, explains the data sources and discusses the concrete data preprocessing steps. Section 5 presents the classification results and resulting analyses. Notably, we present the classification results and compare the accuracy and other metrics of the proposed method with that of other traditional neural networks. Moreover, we determine the proportion of each classification result and analyse the characteristics of visitors. Finally, Section 6 concludes the paper, discusses the strengths and limitations of the study, and offers future research directions.

2. Literature Review

2.1. User-Generated Big Data for Tourist Research

Traditional demographic, survey, and opinion poll data [23–25] have been used to assess tourist behaviour patterns. Additionally, these data have been combined with traditional multivariate statistical methods, such as logistic regression analysis [23] or principal component analysis [24]. However, these data require sampling, are limited in extent and are difficult to collect and update; therefore, it is difficult to comprehensively capture up-to-date tourist behaviours [26]. These limitations have been largely overcome with the emergence of big data. Big data sources provide dynamic and up-to-date data for studies of tourist behaviour and provide better insight into tourism preferences and tourism resource management than other sources do. Before 2010, Lau and McKercher attempted

to use geographic information systems (GIS) to explore tourist movement patterns [27]. Later, they used GPS recorders to produce highly accurate and fine-grained trajectory data and GIS analysis to identify 78 discrete movement patterns [28]. Leung et al. collected trip diaries from six different websites and used content and social network analyses to analyse and map overseas tourist patterns in Beijing during the Olympics [29]. Many studies have used geotagged photographs, such as those from Flickr, to mine the characteristics of tourist behaviours [26,30]. Uncovering tourist behaviours can contribute to tourist attraction prediction [2,31,32]. In addition, tourist behaviours can be combined with the morphological structures of tourist attractions using space syntax analysis to manage and protect tourism resources [33]. Specifically, identifying and classifying tourist behaviours can help tourism managers understand different tourist preferences and contribute to personalised tourist attraction recommendations. In the early 1980s, Plog et al. summarised eight tourist characteristics according to all existing typologies [34]. McKercher and his colleagues compared the behavioural patterns of first-time and repeat visitors to Hong Kong and found that the visitors adopted different travel patterns [1]. Padhi et al. demonstrated that there are three primary types of tourists: those who travel for business purposes, those who travel for leisure and those who travel to academic conferences [31]. Bianchi et al. classified travellers in Chile into short-haul travellers and long-haul travellers to investigate their respective intentions [35].

The results of the aforementioned studies suggest that over the past decade, big data have become more widely used than traditional data in tourist behaviour research, and an increasing number of big data processing methods have arisen. However, human behaviours are complex and diverse, and traditional models may struggle to learn and express patterns of human behaviour. Thus, deep learning methods that have developed rapidly and succeeded in many fields in recent years may provide a novel way to learn human behaviours.

2.2. Deep Learning Methods for Human Behaviours

Deep learning methods have recently become popular in solving supervised learning problems in many fields, such as image processing [36,37], speech recognition [38,39] and natural language processing [40]. Specifically, an increasing number of researchers have attempted to apply deep learning methods to the study of human behaviour. The reason is that human behaviour is complex; indeed, human behaviour may be prompted by intentions or habits; modified by skill, affect and attitude; and affected by physical and contextual conditions [11]. Hartford et al. built a deep neural networks model to predict human participants' behaviour in strategic settings, as most existing approaches either assumed participants to be perfectly rational or attempted to directly model each participant's cognitive processes based on insights from cognitive psychology and experimental economics [12]. In the field of human action recognition, Baccouche et al. proposed a fully automated model incorporating convolutional neural networks and recurrent neural networks to well classify human actions [21]. Fei-Fei Li and her team proposed a prediction model for human trajectory in a crowded space named "Social LSTM", which stands in contrast to traditional methods using hand-crafted functions. The difficulty of trajectory prediction in a crowded space is that not only should each individual trajectory be considered, but also, interactions among humans cannot be neglected, which in traditional methods may fail to be thoroughly considered [22]. Yao et al. predicted next locations in trajectories on a larger temporal and spatial scale with Twitter data and obtained satisfactory accuracy [41]. Compared with other traditional methods, such as the use of hidden Markov models, the proposed method's higher performance may be due to the ability of LSTM to make full use of contextual locations rather than merely relying on the last several locations. In general, deep learning methods have demonstrated much recent success in a variety of human behaviour research fields. Nevertheless, few studies have focused on tourist behaviour, which is of great importance to tourist management and economic development. The aforementioned studies motivate us to apply deep learning methods to tourism research.

3. Methodology

3.1. TensorFlow and Neural Networks

In this paper, TensorFlow is mainly used to classify tourist behaviours. TensorFlow is an open-source machine learning library created by the Google Brain Team's researchers and engineers. The library was originally developed for machine learning and deep neural network research and was open sourced in November 2015 by Google [42]. TensorFlow uses data flow graphs to represent all computational operations and data in the machine learning algorithm. In TensorFlow, nodes in the graph represent mathematical operations and the start of feeding in or the end of outputting information. Edges represent multidimensional data arrays (tensors) that communicate between nodes [43]. These tensors flow to all the nodes and ultimately complete the machine learning process. TensorFlow also provides a convenient visualisation tool called TensorBoard to easily display images of computational graphs.

Most algorithms in TensorFlow are based on neural networks. Neural networks, also called artificial neural networks or connectionist systems, were originally inspired by the biological neural networks that constitute animal brains. A neural network is a massively parallel distributed processor composed of simple processing units that has a natural propensity for storing experiential knowledge and making it available for use [44]. The basic elements of a neural network include a neuron, a set of synapses, an adder and an activation function. A neuron is an information-processing unit that is fundamental to the operation of a neural network. Each connection (synapse) between neurons is characterised by a weight or strength of its own and can transmit a signal to another neuron. An adder is used to sum the input signals, which are weighted by the respective synaptic strengths of the neurons. An activation function is used to limit the amplitude of the output of a neuron (Figure 1). We can describe the neuron k in Figure 1 mathematically.

$$u_k = \sum_{j=1}^m W_{kj} x_j \quad (1)$$

$$y_k = \varphi(u_k + b_k) \quad (2)$$

Neurons are typically organised in layers. Typically, there is an input layer, an output layer and a hidden layer. The input layer represents the first layer through which input signals travel before entering the network. The output layer is the final layer, and it outputs the result of the entire network. The hidden layer is the layer between the input and output layers. The more hidden layers that are present, the deeper the neural network architecture is.

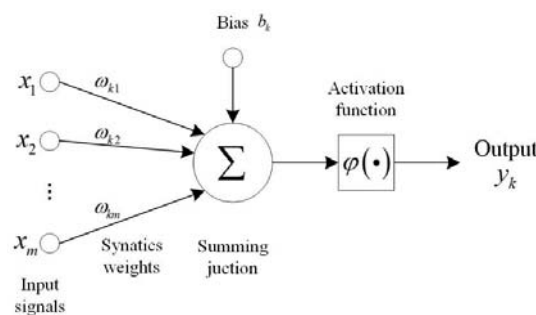


Figure 1. A general neuron model.

Specifically, `tf.contrib.learn`, one of the multiple application programme interfaces (APIs) in TensorFlow, enables users to easily increase the number of hidden layers and other parameters and rapidly build a model without massive duplicate codes, making it easy to configure, train, and evaluate a variety of machine learning models [45]. In this paper, DNNClassifier (Deep Neural Network

Classifier), a well-encapsulated and easy-to-use classifier model of the `tf.contrib.learn` API based on a deep neural network, is mainly used to classify user behaviours.

3.2. Data Processing

Social media check-in data are used in this study. Before inputting training data to TensorFlow and classifying tourist behaviours, some preliminary work is required to process the check-in data. Suppose there are k types of points of interest (POIs) that are reclassified according to visitor classification requirements. The dataset of users is $U = \{u_1, u_2, \dots, u_i, \dots, u_n\}$, and the dataset of POIs is $P = \{p_1, p_2, \dots, p_k\}$. The total number of check-in records for a user u_i is sum_i , and the number of check-in records that are created for each type of POI is represented by $P_i = \{p_{i1}, p_{i2}, \dots, p_{ik}\}$. Thus, the frequency of each type of POI is expressed as follows.

$$f_{ij} = \frac{p_{ij}}{sum_i}, j \in \{1, 2, \dots, k\}, \text{ and } \sum_{j=1}^k f_{ij} = 1. \quad (3)$$

Therefore, the dataset of each user checking in to each type of POI is $F_i = \{f_{i1}, f_{i2}, \dots, f_{ik}\}$. According to prior research, the check-in behaviours of users are generally assessed based on the types of POIs where user check-ins occur [5]. Consequently, we can classify a user's behaviour according to the largest check-in frequency and the combination of the frequencies of all types of POIs. Then, we categorise m types of visitors and establish corresponding classification rules according to the status quo. However, visitor behaviours are classified mainly based on the dominant categories and the combinations of check-in POIs. In this classification approach, visitors may not be strictly mutually exclusive (i.e., some visitors may be associated with the features of more than one classification and may be difficult to distinguish) because of the diversity and complexity of human activities. Such complications increase the difficulty of deep neural network classification using TensorFlow.

Because the TensorFlow DNNClassifier is a supervised neural network, an artificially classified training dataset is needed before training. Additionally, a test dataset is required to validate the accuracy of the neural network. Therefore, after establishing the classification rules, we must classify a portion of the users according to the established rules. Generally, the ratio of the training dataset to the test dataset is 80%:20%. Thus, we can construct a neural network classifier with `tf.contrib.learn.DNNClassifier`. Next, the parameters of the classifier are optimised, including the number of hidden layers, the number of units in each hidden layer, and the number of global iteration steps. The model is fit using the training data, and the test dataset is used to evaluate the accuracy of the model. The parameters are adjusted to improve the accuracy if necessary. After an optimal accuracy is reached, the remaining unclassified records are classified. The entire workflow of user behaviour classification is illustrated in Figure 2.

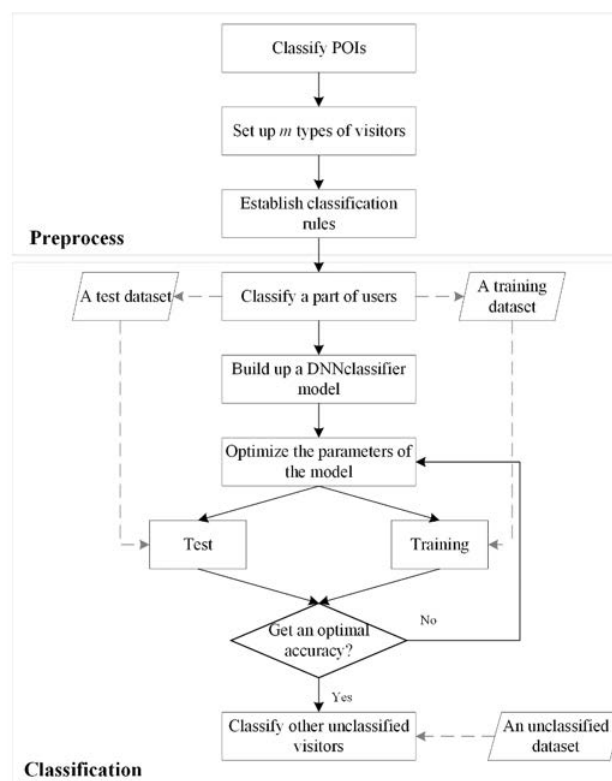


Figure 2. Workflow of user behaviour classification.

4. Materials

4.1. Research Case

Hong Kong is located on the southeast coast of China and is adjacent to Shenzhen City, Guangdong Province (Figure 3). The city is a highly prosperous international metropolis with a total area of over one thousand square kilometres and a population of over seven million people since 2014. Hong Kong is one of the most famous tourist cities in the world. The city is praised as a “shopping paradise”, “gourmet paradise” and the “oriental pearl”. However, few researchers in Mainland China have focused on tourism in Hong Kong compared with studies of other cities in Mainland China. According to the monthly visitor arrival statistics of the Hong Kong Tourism Board, the total number of visitor arrivals to Hong Kong in December 2015 reached 5 million, which increased by 5.4% in December 2016. Moreover, it is estimated that the number of visitor arrivals will continue to grow [46]. Additionally, 73% of visitors in December 2015 were from Mainland China, and the growth rate of visitors from Mainland China is 1.1 times higher than that of the total number of visitors. Thus, the tourism industry of Hong Kong is driven by tourists from Mainland China, and this number continues to grow.

Since the return of the sovereignty of Hong Kong to China in 1997, Hong Kong has been defined as the Hong Kong Special Administrative Region of the People’s Republic of China, and the “one country, two systems” policy has been implemented. Consequently, Hong Kong has its own social system, currency, tariff preference, etc. Some residents in Mainland China prefer the low-price merchandise, good education resources, etc., that Hong Kong offers. Consequently, many Mainland China visitors visit Hong Kong not only for tourism or vacation but also for other reasons, such as shopping and education. Residents in Mainland China must apply for an “Exit-Entry Permit for Travelling to and from Hong Kong” to enter the territory of Hong Kong and have a time limit on their stay in Hong Kong.

Based on the status quo, it is necessary to study the preferences and behaviours of visitors to Hong Kong from Mainland China to manage and improve tourism quality and estimate the influence of these visitors on Hong Kong.

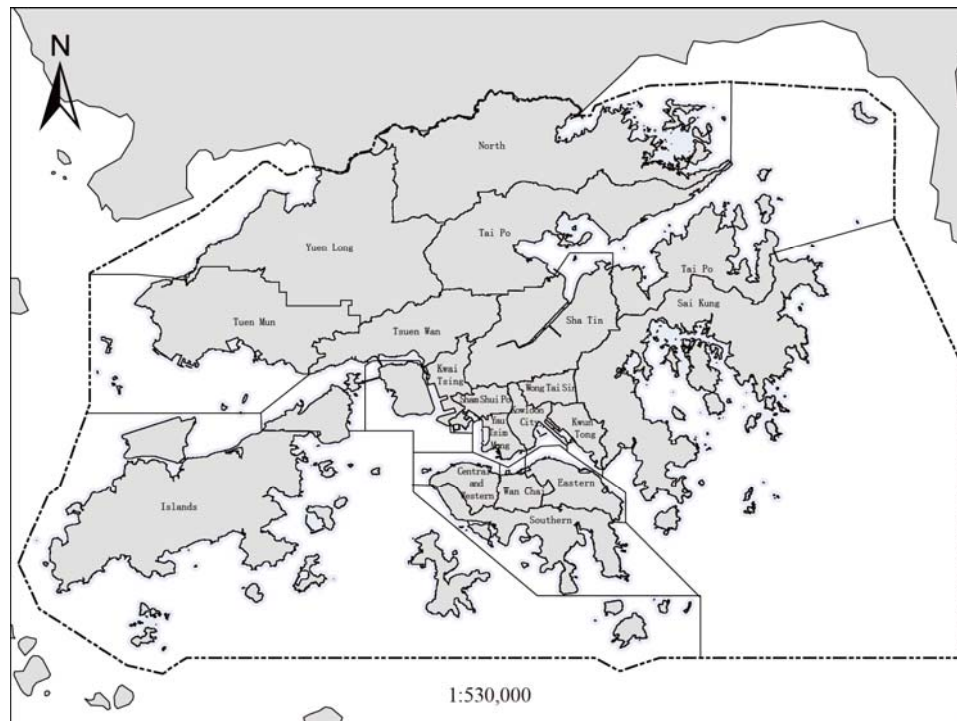


Figure 3. The case study city: Hong Kong.

4.2. Data Specification

Check-in data from Sina Weibo are used to extract and analyse the behaviours of visitors from Mainland China. Weibo is a famous social networking platform in China similar to Twitter. Users can create check-in records with location information and other forms of information, such as words, pictures and video. Many users utilise Weibo to record their daily lives. Therefore, these check-in data can reflect user activities to some extent. We focus on check-in data from Weibo in Hong Kong created between January 2014 and December 2014. In addition, to avoid ambiguities in judgement, we remove users who made no more than two check-ins in Hong Kong during this period. After removing these user records, we analyse 259,062 check-in records for more than 42,000 users with accounts registered in Mainland China. More than 9000 POIs are included in these records.

Because POIs in Weibo use the coordinate system of Gaode Map [47], POIs are obtained and categorised according to the Gaode POI Classification Code [48]. Therefore, based on the Gaode POI Classification Code and our research objectives, we first reclassify the aforementioned POIs into nine types (Table 1): common attractions, special event attractions, transport, hotels, catering, retail, education, residence, and other. In this context, common attractions are tourist attractions or scenic spots that tourists can visit whenever they want and are not affected by special events. For instance, theme parks, natural areas, etc., qualify as common attractions. Compared with common attractions, special event attractions are POIs with check-in records remarkably affected by special events, such as international conferences, exhibitions, and concerts. Correspondingly, these types of POIs include conference and exhibition centres, coliseums, etc. To distinguish among transit passengers and other visitors, transport POIs only include airports and wharfs, as well as their surrounding areas and other associated places. The hotels category includes hotels, family inns, youth hostels, etc. Catering includes restaurants, cafés, nosheries, bakeries, pubs, etc. The retail category includes retail stores,

shopping malls, commercial streets, night markets, etc. Education includes colleges and universities, adult education institutions, secondary schools, primary schools, kindergartens, public libraries and other relevant places. The residence category includes rental houses, villas, etc. Finally, other places are those excluded from the abovementioned categories. These POIs mainly include public spaces, such as hospitals, courts and post offices.

Table 1. Points of interest (POI) categories and check-in data.

No.	Type	Abbreviation	Check-in POI Count	Check-in Users	Number of Check-Ins
1	Common attractions	AFC	1200	30,717	40,599
2	Special event attractions	AFS	126	5251	9054
3	Transport	TRA	193	15,257	20,401
4	Hotels	HOT	677	19,722	32,882
5	Catering	CAT	2787	12,915	15,644
6	Retail	RET	1599	32,870	115,319
7	Education	EDU	378	3756	7469
8	Residence	RES	1021	3001	7199
9	Other	OTH	1563	6790	10,495

According to the status quo, the behaviours of Mainland visitors to Hong Kong are preclassified into the following types (Table 2):

Table 2. Classification of visitor behaviours.

No.	Classification	Classification Rules
1	Purchasing-oriented visitors	Most check-in records occur at shopping POIs (RET), with few at common tourist attractions (AFC) and hotels (HOT)
2	Tourism-oriented visitors	Most check-in POIs are common tourist attractions (AFC) and hotels (HOT), while some are restaurants (CAT), shopping POIs (RET), etc.
3	Special event-oriented visitors	Most check-in records occur at attractions for special events (AFS)
4	Education-oriented visitors	Check-in records occur mostly at education POIs (EDU), as well as some in residential areas (RES) and other (OTH) areas
5	En-route visitors	Check-in records only occur at airports and ports or surrounding areas (TRA)
6	Others	Check-in users that cannot be classified by the aforementioned classification rules

- (1) Purchasing-oriented visitors. Because of tariff preferences and monetary exchange rates, mainland residents, particularly residents living near Hong Kong (such as residents in Guangdong Province or other neighbouring provinces), are fond of buying in Hong Kong. Some people are even professional “daigou”, which means that they buy products in Hong Kong on behalf of mainland residents [49]. Therefore, the main purpose of visitors of this type is purchasing. Most of their check-in POIs are shopping malls, retail stores, etc.
- (2) Tourism-oriented visitors. Visitors of this type are typical tourists. Their check-in locations are mainly concentrated in tourist and scenic spots, as well as hotels. In addition, because Hong Kong is a famous “shopping paradise” and “gourmet paradise”, some of these visitors’ check-in locations are word-of-mouth shopping malls and restaurants.
- (3) Special event-oriented visitors. This type of visitor comes to Hong Kong for particular events, such as concerts, large international conferences and exhibitions. The majority of these visitors’

- check-in locations are conference and exhibition centres and coliseums. Additionally, those who participate in the same event have similar check-in records over a certain time at a certain place.
- (4) Education-oriented visitors. These visitors can be subdivided into two types. The first type is those who study and live in Hong Kong and can be regarded as temporary residents. Most of these visitors are undergraduates or postgraduates, and some are middle school students. The other type is those who are born and study in Hong Kong but live in Mainland China [50]. These students are called “Shenzhen-Hong Kong cross-boundary students”. These students are common because many mainland pregnant women give birth to children in Hong Kong, and their children, who do not have Hukou in mainland, cannot study in mainland public schools.
 - (5) En-route visitors. Visitors of this type merely pass through Hong Kong while travelling to other destinations. Notably, many international flights stop at Hong Kong International Airport. Additionally, there are ports in Hong Kong where ships can transfer passengers to other regions or to ships from other regions.
 - (6) Others. Other visitors are those who cannot be classified into the aforementioned categories.

5. Classification and Analysis Results

5.1. Classification Results of TensorFlow

In this analysis, we classify a training dataset of 4000 records and a test dataset of 1000 records artificially according to the classification rules. Then, we construct a neural network with four fully connected hidden layers, with 10, 20, 20 and 10 units in each layer. We leverage four metrics to evaluate the DNNClassifier, including classification accuracy, precision, recall and f-score. Accuracy is the percentage of the correctly classified results among all the classified results, which is a commonly used and easy-to-understand metric in measuring the quality of a classifier. Precision is the ratio describing how many classified samples are correct, and recall is the ratio describing how many actual labels were correctly classified [51]. Recall and precision are often traded off such that a very high precision often accompanies a low recall. Therefore, the f-score is also introduced to evaluate the classifier. The f-score is a combined metric that can balance recall and precision to measure the quality of a classifier. Only if both precision and recall are fairly high will the f-score reach a high value. The f-score is calculated as follows:

$$f\ score = 2 * precision * recall / (precision + recall) \quad (4)$$

In addition, we compare these metrics of the DNNClassifier with those of other traditional machine learning classification models, including back propagation neural networks (BPNNs), radial basis function neural networks (RBFNNs), random forest methods and support vector machines (SVMs). BPNNs are multilayer feedforward neural networks trained according to an error back propagation algorithm [52,53] and represent one of the most widely applied neural network architectures [54]. RBFNNs are also feedforward neural networks, but they feature three fixed layers: an input layer, a single hidden layer and an output layer [55]. Random forest models are not neural networks but a combination of tree predictors, such that each tree depends on the values in a random vector sampled independently and with the same distribution for all trees in the forest [56]. SVMs are state-of-the-art discriminative classifiers that incorporate statistical learning, maximum margin optimal hyperplane and other concepts [57].

In this comparison, the main parameters of the compared methods are set as follows: A BPNN is constructed with three hidden layers of 10, 20 and 10 units in each layer using the Levenberg–Marquardt algorithm as the training function [58]. The RBF has a single hidden layer with 10 units. In addition, the random forest model parameters include a maximum tree depth of 10, a minimum leaf size of 5 and 100 trees. The penalty parameter C and kernel function of the SVM are 1.0 and an RBF kernel, respectively. In addition, we repeat the training processing of all the aforementioned methods five times and obtain the average performance to avoid random errors. Table 3 shows the

performance of all the methods and the bold fonts in each column represent the best performance of each metric. TensorFlow DNNClassifier outperforms the other methods in accuracy, recall and f-score. The accuracy of DNNClassifier can reach 92.43%, which is 2.18% to 5.83% higher than that of the other models, followed by the accuracies of the BPNN and SVM. For precision, although the random tree model provides the highest precision, it has a low recall and therefore a low f-score, indicating some minor classes may be easily misclassified. Both the recall and f-score of DNNClassifier are considerably high compared with those of the other methods, denoting that DNNClassifier is a comparatively well-performing classifier when addressing human behaviour classification problems.

Table 3. Comparison of various models (%).

No.	Neural Network Model	Accuracy	Precision	Recall	f-Score
1	Back propagation neural network	90.25	87.22	87.94	87.58
2	Radial basis function neural network	86.6	80.04	71.34	75.44
3	Random forest	88.87	92.05	73.76	81.90
4	Support vector machine	89.75	87.78	82.16	84.88
5	DNNClassifier	92.43	88.17	89.94	89.05

In addition to the improved performance, another advantage of TensorFlow is its powerful monitoring and visualisation tools. Without any monitoring or logging information, the classification training can be considered a black box approach. We monitor every 100 global steps in the DNNClassifier model and visualise them in TensorBoard. Through the graph visualisation, we can view the entire computational graph of the model (Figure 4) and the expansion of the DNN layer (Figure 5). Additionally, the scalar summary illustrates the progression of the accuracy (Figure 6a) and loss values (Figure 6b). As the number of global steps increases, the loss value decreases sharply, particularly at 1000 global steps, and then decreases slowly over subsequent iterations. The accuracy value continuously increases and remains steady after approximately 1500 global steps. The final accuracy and loss values are 92.3% and 0.19, respectively, after 3000 global steps.

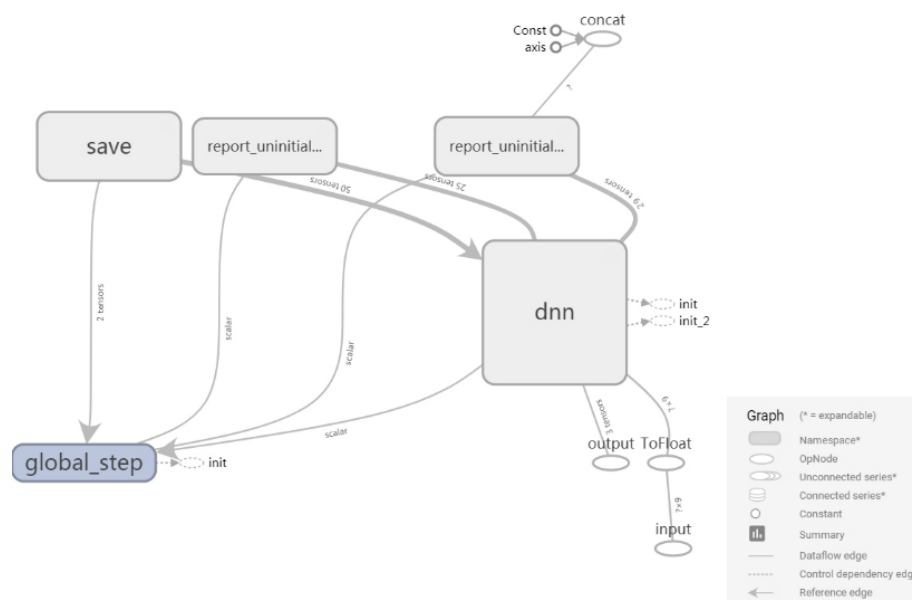


Figure 4. The complete Main Graph of the model.

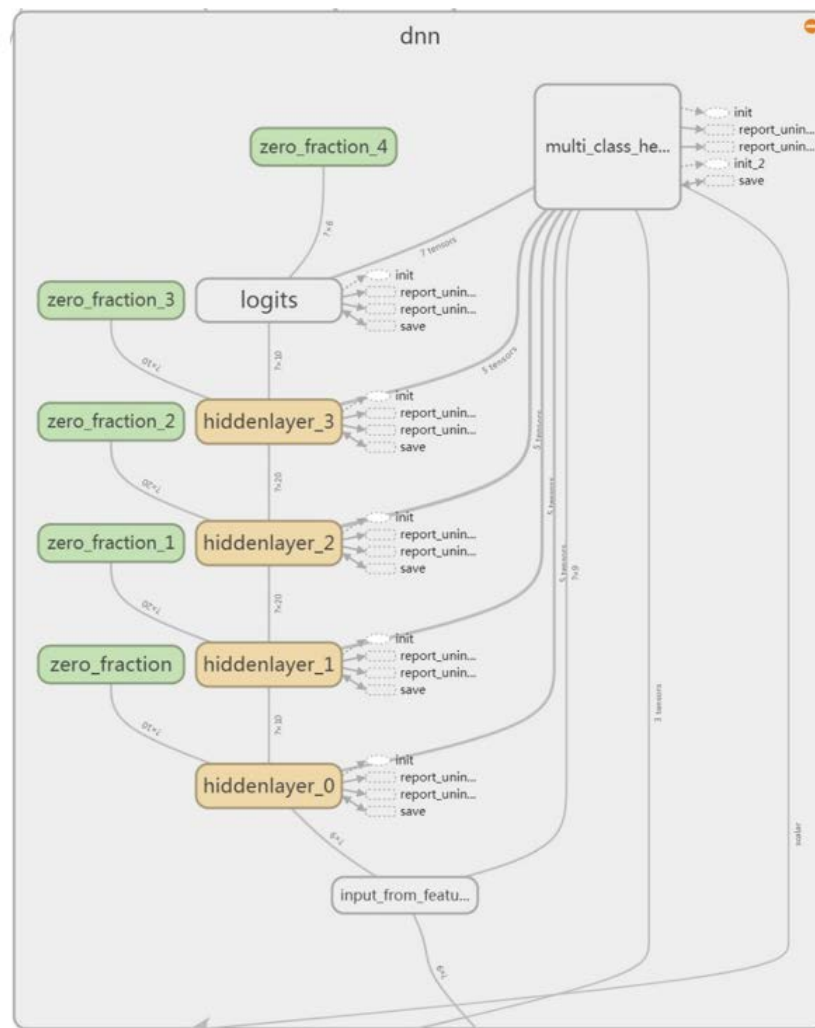


Figure 5. The expansion of the DNN layer.

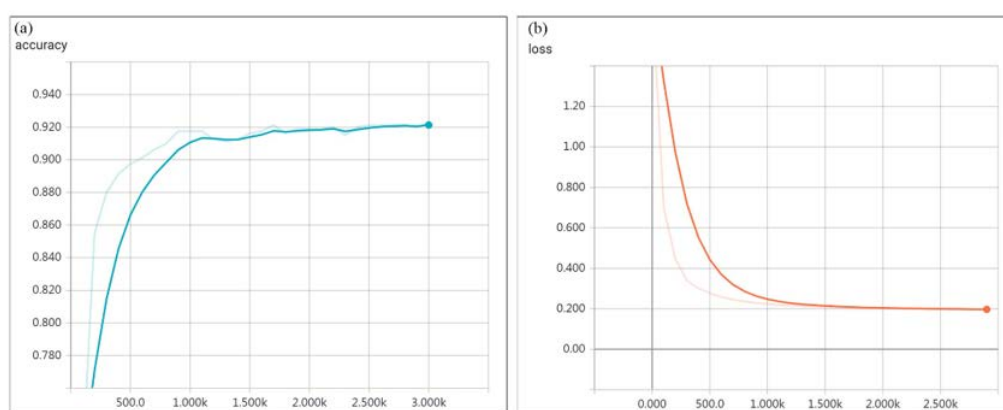


Figure 6. Scalar visualisation of the DNNClassifier model: (a) accuracy value (b) and loss value.

5.2. Proportions and Characteristics of Visitor Behaviours

The proportions of different classification types in the final classification results are shown in Table 4. In this section, we summarise and depict certain characteristics of visitor behaviours.

In addition, we analyse the spatial, temporal and other features of major classification to verify the results.

Table 4. Proportion of each classification type.

No.	Classification	Sum	Percentage (%)
1	Purchasing-oriented visitors	5831	13.8
2	Tourism-oriented visitors	27,404	64.9
3	Special event-oriented visitors	1875	4.4
4	Education-oriented visitors	1577	3.7
5	En-route visitors	2357	5.6
6	Others	3210	7.6

5.2.1. Tourism-Oriented Visitors

As shown in Table 4, tourism-oriented visitors account for the largest proportion of any class, at 64.9%, which is similar to the official value of 63% reported by visitors from Mainland China who vacation in Hong Kong [59]. This result demonstrates that the majority of Mainland China visitors to Hong Kong travel for tourism or vacation.

(1) Force-directed graph of the top 20 tourist and scenic spots

A force-directed graph is created using the Yifan Hu algorithm [60], as shown in Figure 7. The graph shows the popularity of the 20 most popular tourist attractions as well as the relationships among attractions. These results cover 70% of the top ten favourite scenic spots listed in the Visitor Profile Report 2014 of the Hong Kong Tourism Board [59]. Each node represents one scenic spot, and the size of the node represents the weight of the check-in frequency. An edge connects two nodes and reflects that a tourist chose to go to both POIs during their visit. The size of an edge is associated with the number of tourists who visited the two POIs.

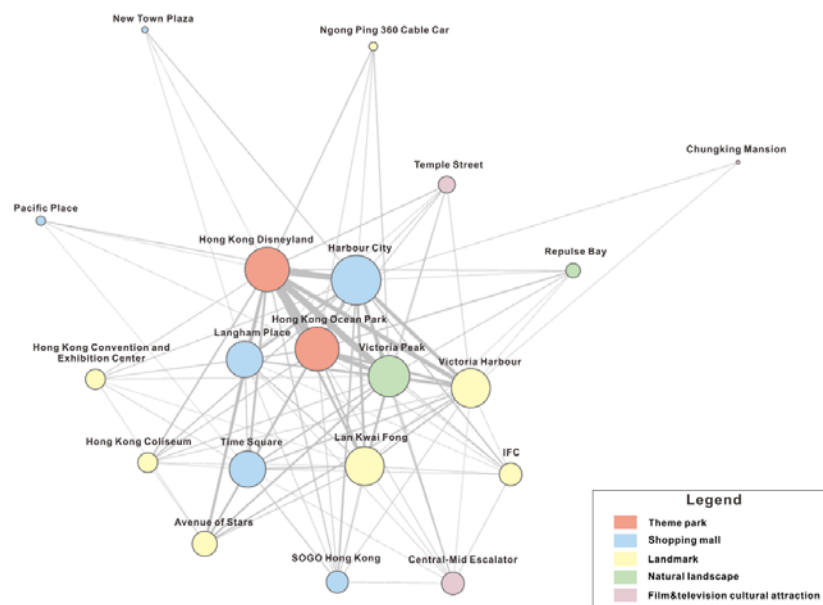


Figure 7. Force-directed graph of tourist attractions of class 2 visitors.

Based on the graph, we can conclude that the major tourism pattern in Hong Kong is “theme park” + “shopping mall”. Hong Kong Disneyland and Hong Kong Ocean Park play core roles as major tourist spots in Hong Kong. Although the category of shopping malls cannot be strictly regarded as a tourist or scenic category, the large number of check-in records associated with shopping malls

suggests that Hong Kong deserves its reputation as a “shopping paradise”, and shopping is appealing to tourists. Other popular POIs include landmark scenic spots, such as Victoria Harbour, the Avenue of the Stars, and Lan Kwai Fong. In addition, because of the rich and profound film and television culture in Hong Kong, many places have become popular due to the filming of a certain movie or television show. For instance, Central-Mid Escalator and Chungking Mansion, two locations where scenes from the classic movie “Chungking Express” directed by Kar Wai Wong were filmed, plotted in the top 20 scenic spots graph, although Chungking Mansion is not a typical or official scenic spot.

(2) Monthly visit analysis

We compare the number of check-in records of tourism-oriented visitors per month with the official statistics regarding the number of Mainland China tourists per month (Figure 8). Although certain small-scale trends fluctuate, the overall trend is approximately consistent. Specifically, the line plot of check-in records shows that July, August and December are major tourist months.

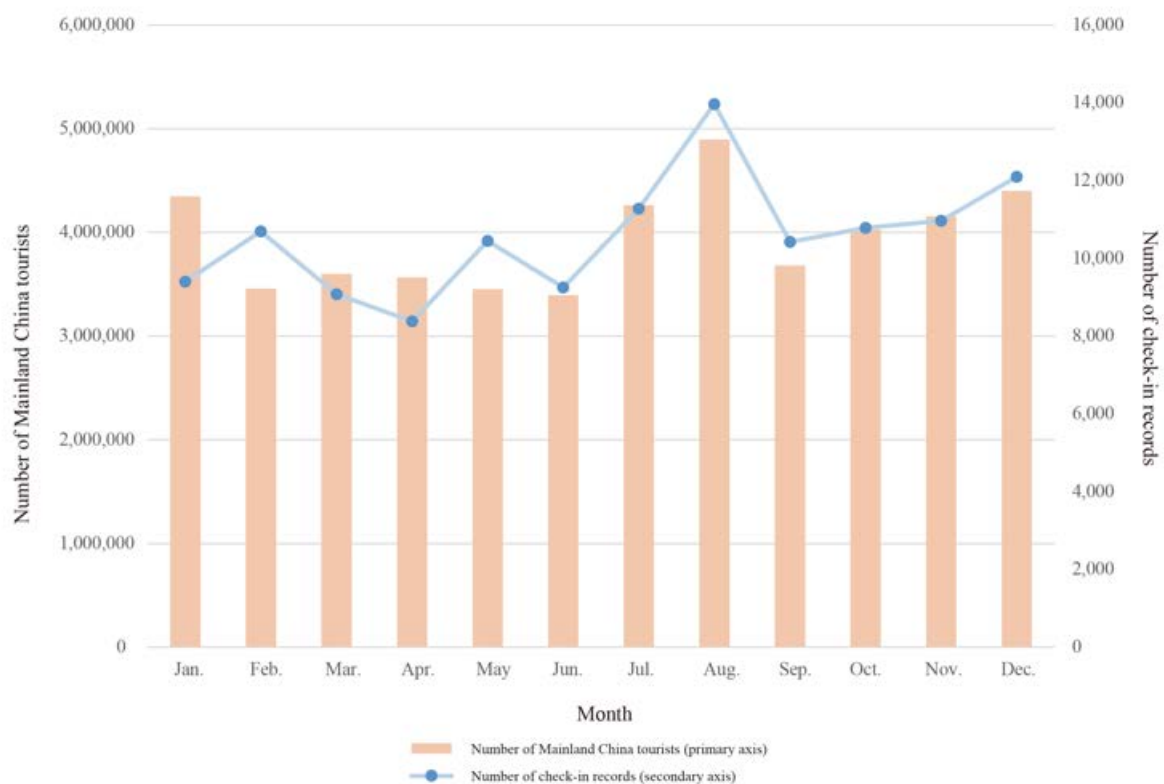


Figure 8. Number of visitors based on check-ins and number of actual tourists.

5.2.2. Purchasing-Oriented Visitors

(1) Proportion of visitors from different sources

In this section, we analyse the registration locations of users in class 1 and compare the proportions of class 1 visitor sources with those of all visitors. We find that among all visitors, visitors from Guangdong Province, the province closest to Hong Kong, accounted for the largest proportion of total visitors at 30.8%. The following provinces are also neighbouring provinces: Shanghai (11.4%), Beijing (10.4%), Fujian (6.8%), Zhejiang (5.5%), Jiangsu (5.4%), Sichuan (4.3%), and Hubei (3.8%). In addition, visitors from Shenzhen account for 37.1% of visitors from Guangdong Province, followed by visitors from Guangzhou (30.2%), Foshan (4.9%), Dongguan (4.2%) and other cities in Guangdong (Figure 9). For class 1 visitors, the major provinces of origin were nearly the same, but the proportion of

Guangdong Province visitors increased to 55.0%. Moreover, the percentage of visitors from Shenzhen increased to 55.2% (Figure 10), and the percentages of visitors from Guangzhou (19.4%), Dongguan (4.5%), Foshan (3.3%) and other cities also varied. The statistical results are consistent with the actual situation, as most residents in Guangdong Province, particularly in cities adjacent to Hong Kong, such as Shenzhen, Guangzhou and Dongguan, visit Hong Kong often to purchase items due to the low transportation costs and convenient procedure of applying for an “Exit-Entry Permit for Travelling to and from Hong Kong”.

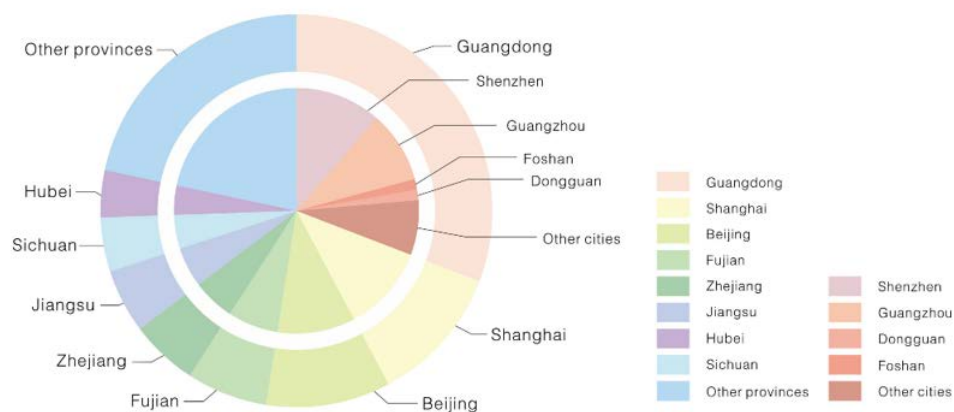


Figure 9. Proportions of visitor sources for all visitors.

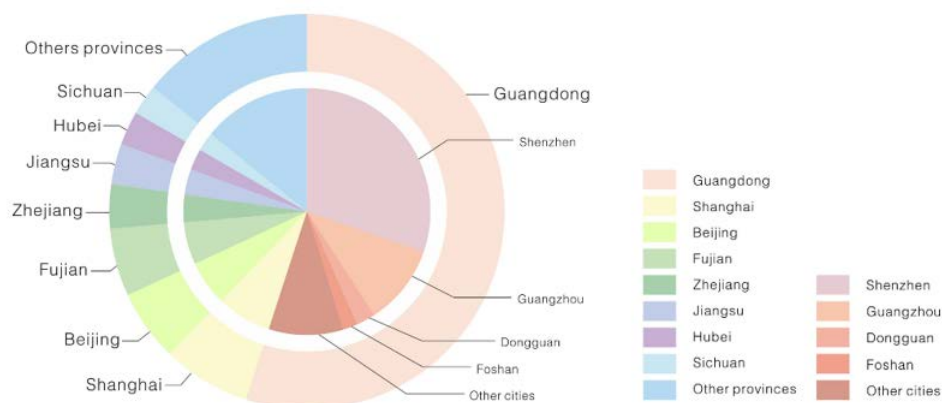


Figure 10. Proportions of visitor sources for class 1 (purchasing-oriented visitors).

(2) Kernel density analysis of visitor check-in locations

In this section, we analyse all check-in records of purchasing-oriented visitors and conduct a kernel density analysis. We find that the hot spots visited by these visitors are concentrated near the following locations: ① shopping malls close to Shenzhen-Hong Kong ports, such as shopping malls in Tuen Mun (close to Shen Bay Port) Sheung Shui and Fanling (close to Futian Port and Luohu Port); ② shopping malls near subway stations, particularly the subway line that starts at the Shenzhen-Hong Kong ports, such as shopping malls near Tai Wo Station and Tai Po Market Station along the East Rail Line and Long Ping Station and Yuen Long Station along the West Rail Line; and ③ popular and concentrated shopping areas, such as Mong Kok, the most congested shopping district in Hong Kong (Figure 11). This result is consistent with the actual situation. Notably, most visitors who only do some shopping in Hong Kong shop along the subway lines and close to subway stations and the Shenzhen-Hong Kong ports to reduce transportation costs. In some cases, visitors travel farther to shop to enjoy more choices of commodities.

(3) Visit patterns

In this section, we compare the visit patterns of purchasing-oriented visitors and tourism-oriented visitors, including their average stay times, proportions of same-day trips (i.e., trips without staying overnight) and average numbers of trips to Hong Kong in a year. Table 5 shows that the average stay time of purchasing-oriented visitors is shorter than that of tourism-oriented visitors, while the average numbers of trips to Hong Kong of purchasing-oriented visitors is greater than that of tourism-oriented visitors. In addition, the proportion of same-day trips for purchasing visitors is greater than 75%, while that of tourism-oriented visitors is less than 50%. This result reveals that tourism-oriented visitors do not visit Hong Kong as frequently as purchasing-oriented visitors do and are inclined to stay in Hong Kong overnight to visit multiple scenic spots during one trip. Conversely, purchasing-oriented visitors tend to visit Hong Kong many times throughout the year and generally stay in Hong Kong for only one day without staying overnight.

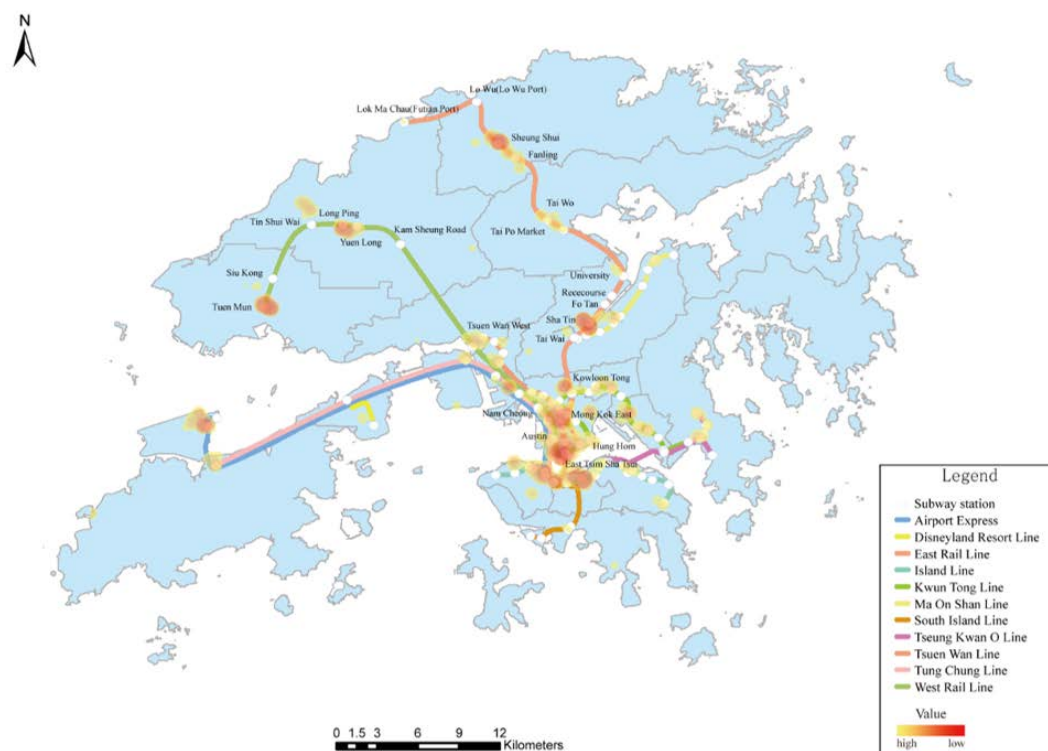


Figure 11. Kernel density of shopping-oriented visitor check-in records.

Table 5. Comparison of the visit patterns of purchasing-oriented visitors and tourism-oriented visitors.

Type	Average Stay Time (Days)	Average Number of Visits to Hong Kong (No.)	Proportion of Same-Day Trips (%)
Purchasing-oriented visitors	1.6	3.3	76.1
Tourism-oriented visitors	2.3	1.5	46.4

5.2.3. Special Event-Oriented Visitors

Previous studies have shown that social media check-in data can be used to detect urban events [61,62]. Although we do not use a text-mining method in this study, it is still possible to detect some special events that large numbers of people attend based on check-ins at a certain place within a short period. We regard more than one check-in record on a certain date and at a certain place as a supposed event (for example, May 22nd at Hong Kong Coliseum and May 22nd at the

AsiaWorld Expo). The check-in records of special event-oriented visitors indicate the occurrence of 366 supposed events based on 3820 check-in records. By looking up the event histories posted on the official websites of these places, we find that 251 of these events did occur and covered 88.0% of check-in records. In Table 6, we list the 20 most frequent check-in dates and corresponding POIs and events based on 36.8% of the check-in records of special event-oriented visitors. The popularity of a star or the influence of a conference or an exhibition results in the clustering of people at a certain place within a short period. In return, the check-in frequency can reflect the popularity of the star or the influence of the conference or exhibition to some extent.

5.2.4. Education-Oriented Visitors

The classification results obtained for education-oriented visitors can be used to subdivide these visitors into the following subclasses according to their major check-in locations (Table 7). Specifically, 95.3% of these visitors are university students, and 2.3% of them are secondary school students. These two types of visitors can actually be regarded as temporary residents. Only 0.3% of education-oriented visitors are primary school students, and unfortunately, we cannot detect the existence of “Shenzhen-Hong Kong cross-boundary students”, as expected. This detection issue may be associated with the limited user scope of the social network, as the age range of major users is between 13 and 35 [63].

Table 6. The twenty most frequent check-in dates and the corresponding special events.

No.	Date	Check-in Frequency	Place	Event
1	30 May 2014	145	Hong Kong Coliseum	Mayday Just Rock it!! Hong Kong Concerts
2	24 May 2014	121	Hong Kong Coliseum	
3	29 May 2014	109	Hong Kong Coliseum	
4	23 May 2014	81	Hong Kong Coliseum	
5	22 May 2014	74	Hong Kong Coliseum	
6	26 May 2014	65	Hong Kong Coliseum	
7	27 May 2014	57	Hong Kong Coliseum	
8	1 Jun. 2014	120	AsiaWorld Expo	EXO FROM. EXOPLANET #1—THE LOST PLANET IN HONGKONG
9	16 Aug. 2014	106	AsiaWorld Expo	2014 JYJ Asia Tour Concert THE RETURN OF THE KING IN HONG KONG
10	30 Aug. 2014	94	AsiaWorld Expo	JYP NATION “ONE MIC” 2014 Concert
11	3 Dec. 2014	91	AsiaWorld Expo	Mnet Asian Music Awards 2014 MAMA in Hong Kong
13	24 Nov. 2014	60	AsiaWorld Expo	Opus2 Jay 2014 WORLD TOUR
14	23 Nov. 2014	46	AsiaWorld Expo	
15	22 Nov. 2014	42	AsiaWorld Expo	
16	1 Jun. 2014	44	Hong Kong Coliseum	Della In Concert-Hong Kong
17	15 Mar. 2014	39	Hong Kong Convention and Exhibition Centre	C3 in Hong Kong
18	25 Jul. 2014	39	Hong Kong Coliseum	Stefanie Sun <Kepler> 2014 World Tour-Hong Kong
19	27 Jul. 2014	35	Hong Kong Coliseum	
20	30 Aug. 2014	38	Hong Kong Coliseum	Father of Concert Glory SHOW

Table 7. Sub-classifications and respective proportions of education-oriented visitors.

Subclassification	Above University	Secondary School	Primary School	Others
Proportion (%)	95.3	2.3	0.3	2.1

6. Discussion

Identifying and classifying tourist behaviours can help tourism managers understand different tourist preferences, recommend personalised tourist attractions, develop tourism products and manage tourism resources. However, human behaviours are diverse and complex; therefore, it is difficult to

assess and classify these behaviours. Deep learning methods may provide state-of-the-art solutions to these issues. Notably, deep learning methods can surpass the abilities of human beings in many fields, and deep neural networks can be explained in a psychological manner. The results suggest that the deep neural network in TensorFlow can be used to process the complex and erratic classification rules of user behaviour classification problems and yield results with satisfactory accuracy. In addition, the deep neural network in TensorFlow is not a “black box” due to its powerful monitoring and visualisation tools. In this paper, we process location-based social network data using a deep learning method. As data volumes increase, traditional data processing methods will increasingly struggle to process big data. Thus, the use of deep learning methods to process big data is trending. In this study, we use the deep learning method in TensorFlow to analyse tourism geography. Although TensorFlow has been available as an open-source product for nearly two years, it has been applied in few studies of human behaviour assessment. As deep learning methods become more popular, they should be applied in both natural science and social science to provide state-of-the-art solutions to social and human problems.

In future work, we will attempt to apply recurrent neural networks (RNNs) and the Word2Vec technique for the text mining of tweets to produce accurate information and classification results and further integrate deep learning methods and human geography.

Acknowledgments: We thank Dawei Gui, Pengfei Ning, Yuan Wei and Zekun Li for their help. In addition, this study was supported by the National Key Research and Development Program of China (2016YFC0803106) and the National Natural Science Foundation of China (Project No. 41571438).

Author Contributions: Shanshan Han, Fu Ren, Chao Wu, Ying Chen, Qingyun Du and Xinyue Ye worked collectively. Specifically, Qingyun Du and Xinyue Ye proposed the original idea and organised the content; Fu Ren helped us to complete the study; Shanshan Han and Chao Wu analysed the data and designed the experiments; Ying Chen drew the statistical figures; and Shanshan Han wrote the paper. All of the co-authors drafted and revised the article collectively, and all authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- McKercher, B.; Shoval, N.; Ng, E.; Birenboim, A. First and repeat visitor behaviour: Gps tracking and gis analysis in Hong Kong. *Tour. Geogr.* **2012**, *14*, 147–161. [CrossRef]
- Zheng, W.; Huang, X.; Li, Y. Understanding the tourist mobility using gps: Where is the next place? *Tour. Manag.* **2017**, *59*, 267–280. [CrossRef]
- Phithakkitnukoon, S.; Horanont, T.; Witayangkurn, A.; Siri, R.; Sekimoto, Y.; Shibasaki, R. Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan. *Pervasive Mob. Comput.* **2015**, *18*, 18–39. [CrossRef]
- Asakura, Y.; Iryo, T. Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transp. Res. Part A* **2007**, *41*, 684–690. [CrossRef]
- Cao, J.; Hu, Q.; Li, Q. A study of users’ movements based on check-in data in location-based social networks. In Proceedings of the International Symposium on Web and Wireless Geographical Information Systems, Seoul, Korea, 29–30 May 2014; Springer: Berlin, Germany, 2014; pp. 54–66.
- Min, L.I.; Wang, X.C.; Zhang, J.; Liu, Z.J. Study on check-in and related behaviors of location-based social network users. *Comput. Sci.* **2013**, *40*, 72–76.
- Li, L.; Yang, L.; Zhu, H.; Dai, R. Explorative analysis of wuhan intra-urban human mobility using social media check-in data. *PLoS ONE* **2015**, *10*, e0135286. [CrossRef] [PubMed]
- Huang, Q.; Cao, G.; Wang, C. From where do tweets originate? A gis approach for user location inference. In Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, Dallas/Fort Worth, TX, USA, 4–7 November 2014; ACM: New York, NY, USA, 2014; pp. 1–8.
- Wu, L.; Zhi, Y.; Sui, Z.; Liu, Y. Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE* **2014**, *9*, e97010. [CrossRef] [PubMed]
- World Tourism Organization. Annual Report 2015. Available online: http://cf.cdn.unwto.org/sites/all/files/pdf/annual_report_2015_lr.pdf (accessed on 23 August 2017).

11. Salah, A.A.; Lepri, B.; Pianesi, F.; Pentland, A.S. Human behavior understanding for inducing behavioral change: Application perspectives. In Proceedings of the International Workshop on Human Behavior Understanding, Amsterdam, The Netherlands, 16 November 2011; Springer: Berlin, Germany, 2011; pp. 1–15.
12. Hartford, J.S.; Wright, J.R.; Leyton-Brown, K. Deep learning for predicting human strategic behaviour. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2016; pp. 2424–2432.
13. YouTube. Match 1—Google Deepmind Challenge Match: Lee Sedol vs. Alphago. Available online: <https://www.youtube.com/watch?v=vFr3K2DORc8&t=1h57m> (accessed on 4 July 2017).
14. Ritter, S.; Barrett, D.G.T.; Santoro, A.; Botvinick, M.M. Cognitive psychology for deep neural networks: A shape bias case study. *arXiv*, 2017.
15. TensorFlow. Image Recognition. Available online: https://www.tensorflow.org/tutorials/image_recognition (accessed on 25 June 2017).
16. TensorFlow. Mnist for ml Beginners. Available online: <https://www.tensorflow.org/versions/r0.7/tutorials/mnist/beginners/index.html> (accessed on 25 June 2017).
17. Kovalev, V.; Kalinovsky, A.; Kovalev, S. Deep learning with theano, torch, caffe, tensorflow, and deeplearning4j: Which one is the best in speed and accuracy? In Proceedings of the XIII International Conference on Pattern Recognition and Information Processing, Minsk, Belarus, 3–5 October 2016.
18. Duc, H.H.; Jung, K. Applying tensorflow with convolutional neural networks to train data and recognize national flags. In *Advanced Multimedia and Ubiquitous Engineering: Mue/Futuretech 2017*; Park, J.J., Chen, S.-C., Raymond Choo, K.-K., Eds.; Springer: Singapore, 2017; pp. 367–373.
19. Ferri, A. *Object Tracking in Video with Tensorflow*; Universitat Politècnica de Catalunya: Barcelona, Spain, 2016.
20. Dean, J. Large-scale deep learning for intelligent computer systems. In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, San Francisco, CA, USA, 22–25 February 2016.
21. Baccouche, M.; Mamalet, F.; Wolf, C.; Garcia, C.; Baskurt, A. Sequential deep learning for human action recognition. In Proceedings of the International Workshop on Human Behavior Understanding, Amsterdam, The Netherlands, 16 November 2011; pp. 29–39.
22. Alahi, A.; Goel, K.; Ramanathan, V.; Robicquet, A.; Li, F.F.; Savarese, S. Social lstm: Human trajectory prediction in crowded spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 961–971.
23. Oh, J.Y.-J.; Cheng, C.-K.; Lehto, X.Y.; O’Leary, J.T. Predictors of tourists’ shopping behaviour: Examination of socio-demographic characteristics and trip typologies. *J. Vacat. Mark.* **2004**, *10*, 308–319. [[CrossRef](#)]
24. Johns, N.; Gyimóthy, S. Market segmentation and the prediction of tourist behavior: The case of bornholm, denmark. *J. Travel Res.* **2002**, *40*, 316–327. [[CrossRef](#)]
25. Morrison, A.M.; Braunlich, C.G.; Cai, L.A.; O’Leary, J.T. A profile of the casino resort vacationer. *J. Travel Res.* **1996**, *35*, 55–61. [[CrossRef](#)]
26. Vu, H.Q.; Li, G.; Law, R.; Ye, B.H. Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tour. Manag.* **2015**, *46*, 222–232. [[CrossRef](#)]
27. Lau, G.; Mckercher, B. Understanding tourist movement patterns in a destination: A GIS approach. *Tour. Hosp. Res.* **2006**, *7*, 39–49. [[CrossRef](#)]
28. Mckercher, B.; Gigi, L. Movement patterns of tourists within a destination. *Tour. Geogr.* **2008**, *10*, 355–374. [[CrossRef](#)]
29. Leung, X.Y.; Wang, F.; Wu, B.; Bai, B.; Stahura, K.A.; Xie, Z. A social network analysis of overseas tourist movement patterns in beijing: The impact of the olympic games. *Int. J. Tour. Res.* **2012**, *14*, 469–484. [[CrossRef](#)]
30. Yuan, Y.; Medel, M. Characterizing international travel behavior from geotagged photos: A case study of flickr. *PLoS ONE* **2016**, *11*, e0154885. [[CrossRef](#)] [[PubMed](#)]
31. Padhi, S.S.; Pati, R.K. Quantifying potential tourist behavior in choice of destination using google trends. *Tour. Manag. Perspect.* **2017**, *24*, 34–47. [[CrossRef](#)]
32. Clements, M.; Serdyukov, P.; de Vries, A.P.; Reinders, M.J.T. Using flickr geotags to predict user travel behaviour. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland, 19–23 July 2010; ACM: New York, NY, USA, 2010; pp. 851–852.
33. Yuan, L.; Xiao, L.; Yu, Y.; Xu, W.; Law, A. Understanding tourist space at a historic site through space syntax analysis: The case of Gulangyu, China. *Tour. Manag.* **2016**, *52*, 30–43.
34. Plog, S.C.; Ritchie, J.R.B.; Goeldner, C.R. Understanding psychographics in tourism research. In *Understanding Psychographics in Tourism Research*; CABI: Wallingford, UK, 1987; pp. 203–213.

35. Bianchi, C.; Milberg, S.; Cúneo, A. Understanding travelers' intentions to visit a short versus long-haul emerging vacation destination: The case of Chile. *Tour. Manag.* **2017**, *59*, 312–324. [CrossRef]
36. Chen, K.; Yan, Z.J.; Huo, Q. A context-sensitive-chunk bptt approach to training deep lstm/blstm recurrent neural networks for offline handwriting recognition. In Proceedings of the International Conference on Document Analysis and Recognition, Johannesburg, South Africa, 12–13 January 2016; pp. 411–415.
37. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Available online: <https://arxiv.org/abs/1409.1556> (accessed on 25 June 2017).
38. Graves, A.; Jaitly, N.; Mohamed, A.R. Hybrid speech recognition with deep bidirectional LSTM. In Proceedings of the Automatic Speech Recognition and Understanding, Olomouc, Czech, 8–12 December 2013; pp. 273–278.
39. Graves, A.; Jaitly, N. Towards end-to-end speech recognition with recurrent neural networks. In Proceedings of the International Conference on Machine Learning, Beijing, China, 21–26 June 2014; pp. 1764–1772.
40. Graves, A. Generating Sequences with Recurrent Neural Networks. Available online: <https://arxiv.org/abs/1308.0850> (accessed on 25 June 2017).
41. Yao, D.; Zhang, C.; Huang, J.; Bi, J. SERM: A recurrent model for next location prediction in semantic trajectories. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; ACM: New York, NY, USA, 2017; pp. 2411–2414.
42. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv* **2016**, arXiv:1603.04467.
43. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M. Tensorflow: A system for large-scale machine learning. *arXiv* **2016**, arXiv:1605.08695.
44. Haykin, S.S. *Neural Networks and Learning Machines*; Pearson: Upper Saddle River, NJ, USA, 2009; Volume 3.
45. TensorFlow. Tf.Contrib.Learn Quickstart. Available online: https://www.tensorflow.org/get_started/tflearn (accessed on 25 June 2017).
46. Hong Kong Tourism Board PartnerNet. Visitor Arrival Statistics. Available online: https://securepartnernet.hktb.com/china/sc/research_statistics/research_publications/index.html?id=3631 (accessed on 5 July 2017).
47. Sina Tech. Gaode Unites Sina to Launch the Social Network Map Platform. Available online: <http://tech.sina.cn/?sa=t84v44d21223704&pos=108&vt=4> (accessed on 12 March 2017).
48. Gaode Open Platform, Web Services API and Related Downloads. Available online: <http://lbs.amap.com/api/webservice/download/> (accessed on 12 March 2017).
49. Kuah-Pearce, K.E. *Chinese Women and the Cyberspace*; Amsterdam University Press: Amsterdam, The Netherlands, 2008.
50. Wikipedia. Shenzhen-Hong Kong Cross-Boundary Students. Available online: https://en.wikipedia.org/wiki/Shenzhen%E2%80%93Hong_Kong_cross-boundary_students (accessed on 5 May 2017).
51. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008; Volume 1.
52. Li, J.; Cheng, J.-H.; Shi, J.-Y.; Huang, F. Brief introduction of back propagation (bp) neural network algorithm and its improvement. In *Advances in Computer Science and Information Engineering*; Springer: Berlin, Germany, 2012; pp. 553–558.
53. Buscema, M. Back propagation neural networks. *Subst. Misuse* **1998**, *33*, 233–270. [CrossRef]
54. Hecht-Nielsen, R. Theory of the backpropagation neural network. *Neural Netw.* **1988**, *1*, 445–448. [CrossRef]
55. Powell, M.J.D. *Radial Basis Functions for Multivariable Interpolation: A Review*; Clarendon Press: Wotton-under-Edge, UK, 1987.
56. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
57. Ukil, A. Support vector machine. *Comput. Sci.* **2002**, *1*, 1–28.
58. Ranganathan, A. The levenberg-marquardt algorithm. *Tutor. Algorithm* **2004**, *11*, 101–110.
59. Hong Kong Tourism Board PartnerNet. Visitor Profile Report—2014. Available online: http://securepartnernet.hktb.com/filemanager/intranet/ir/ResearchStatistics/paper/Visitor-Pro/Profile2014/visitor_profile_2014_0.pdf (accessed on 15 July 2017).
60. Hu, Y. Efficient, high-quality force-directed graph drawing. *Math. J.* **2005**, *10*, 37–71.
61. Weng, J.; Lee, B.-S. Event detection in twitter. *ICWSM* **2011**, *11*, 401–408.

62. Xu, Z.; Zhang, H.; Hu, C.; Mei, L.; Xuan, J.; Choo, K.K.R.; Sugumaran, V.; Zhu, Y. Building knowledge base of urban emergency events based on crowdsourcing of social media. *Concurr. Comput. Pract. Exp.* **2016**, *28*, 4038–4052. [[CrossRef](#)]
63. Weibo. Weibo Users Development Report—2014. Available online: <http://data.weibo.com/report/reportDetail?id=215> (accessed on 27 July 2017). (In Chinese)



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).