

Article

A Human-Guided Machine Learning Approach for 5G Smart Tourism IoT

Rongqun Peng ¹, Yingxi Lou ¹, Michel Kadoch ^{2,*} and Mohamed Cheriet ² 

¹ School of Computer Science and Technology, Shandong University of Technology, Zibo 255000, China; pengrq@sdut.edu.cn (R.P.); louyxi@163.com (Y.L.)

² Department of Electrical Engineering, ETS, University of Quebec, Montreal, QC H3C 3J7, Canada; mohamed.cheriet@etsmtl.ca

* Correspondence: michel.kadoch@etsmtl.ca

Received: 5 May 2020; Accepted: 2 June 2020; Published: 8 June 2020



Abstract: With the continuous development of tourism, the integration of the Internet of Things (IoT) into tourism projects is considered a very promising technology. Smart tourism aims to use the IoT to maximize information communication; that is, the IoT technology will become an important element to meet the needs of a new generation of tourists. Therefore, in this study, we propose a human-guided machine learning classification method based on tourist selection behavior. This classification method can effectively help tourists make a decision in choosing a certain tourist destination. The results obtained from the cross-validation experiments and performance evaluation prove the effectiveness of this method.

Keywords: smart tourism; the internet of things; machine learning; human-guided

1. Introduction

The Internet of Things (IoT) is a part of the future Internet ecosystem that will have a significant impact on the development of healthcare and e-tourism services [1]. Location and geographic context-aware services have become a new area of rapid business development. These services play a key role in the development of IoT scenarios, smart spaces, and proactive solutions. Among the above, smart tourism is one of the most fascinating application areas [2]. In many cities, various social problems, such as non-security, fraud, and lack of appropriate resource information, are the biggest obstacles to tourism. The design of smart tourism can help travelers capture and process all of the data involving users, turning them not only into useful information but also into personalized knowledge [3]. Figure 1 shows the application of the IoT in smart tourism. Our system is designed to develop the tourism industry so as to attract different customer groups and provide users with innovative and ideal communication, information, and entertainment applications and services in a new business model. According to the needs of cloud-based services [4], we designed a new method of searching for travel information [5]. By bringing together smartphones, Global Positioning System (GPS), Google Maps, Augmented Reality (AR), and Worldwide interoperability for Microwave Access (WiMAX) networks [6], we provide mobile users with a whole new travel experience [7]. At present, the 5th Generation Wireless System (5G) is the world's leading technology, and its applications are gradually spreading in our daily lives. The development direction and prospects of 5G smart tourism are very optimistic. Integrating 5G technology [8] into the tourism field and combining it with AR, Artificial Intelligence (AI), social sharing, and other applications while simultaneously benefitting from its high speed, large bandwidth, and low-latency features will give rise to an intelligent upgrade of tourist attractions. Promoting the application and development of 5G smart tourism is an important

way to boost tourism in all regions of the country and to facilitate the intelligent development of scenic spots. It will also definitely bring significant benefits to tourist attractions.

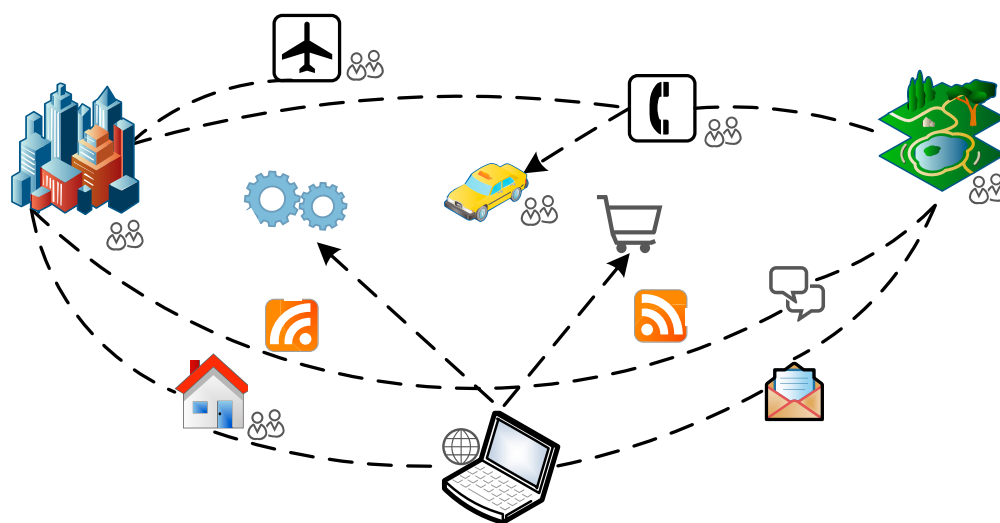


Figure 1. Smart tourism using the IoT.

Tourism is one of the most prosperous economic activities in the world; in fact, it is one of the main components of economic growth in many countries [9]. Many countries in the world use tourism to revitalize their national economies, as tourism can greatly promote the development of various domains, such as the regional economy, as well as increase employment opportunities [10]. In addition, it can play an important role in promoting mutual understanding with other countries.

The tourism experience is the core product of the tourism industry, directly affecting tourists' satisfaction [11]. However, traditional tourism approaches may not attract enough tourists to revitalize the economy. The World Tourism Organization (UNWTO) has introduced the concept of smart tourism and defined it as clean, green, ethical, and high-quality tourism [12]. The realization of smart tourism is an important and novel way to transform the tourism industry. Recently, the tourism industry has undergone a shift to a more user-friendly intelligent system based on "smart tourism" information technology [13]. Smart tourism is considered customer-centric and is designed to fully satisfy tourists' dietary, accommodation, travel, shopping, and entertainment needs. It basically means applying smart technology at all stages of travel to improve the visitor experience.

The purpose of smart tourism is to heighten the tourist experience and enhance the competitiveness of tourist destinations [14]. In addition, it promotes the sustainable development of the country. To upgrade the travel experience in a certain city, smart technology enables users to download smart applications on their phones. Such smart applications can enable tourists to take advantage of local cuisines, stories, customs, amongst other things, for a complete immersion in the local culture. Intelligent tourism systems should provide tourists with relevant and meaningful information based on big data analysis, personal information, behavior patterns, etc., to help them easily make decisions before or during a trip. Therefore, smart tourism should be able to meet the requirements of short-term economic needs and long-term sustainable development.

Based on the above reasons, we propose a method for predicting the behavior of smart travel users. This method is a K-Nearest Neighbor (KNN) user behavior prediction method based on human-guided machine learning. Compared with the traditional KNN algorithm, it can realize the analysis and response of fuzzy and uncertain problems by integrating human perception, cognitive capabilities, machine computing capabilities, and storage capabilities. Cross-validation experiments and performance evaluation methods show that the technology can classify the decision-making behaviors of travel users. It also has a good prediction ability. The classification results can quickly predict whether the travel user will choose the destination in the short term.

The rest of this article is organized as follows. Section 2 introduces the role and application of the IoT in smart tourism. Section 3 details the role of human-guided machine learning in dealing with smart tourism issues. Section 4 introduces the KNN user behavior prediction method based on human-guided machine learning. Finally, Section 5 summarizes this article.

2. System Architecture and Method

2.1. IoT Technology

The IoT, as an integrated part of the future development of the Internet, is defined as a dynamic global network infrastructure with self-configuring capabilities based on standards and interoperable communication protocols. The basic idea of the IoT concept is to address the ubiquity of various objects, such as Radio Frequency Identification (RFID) tags, sensors, actuators, mobile phones, etc., enabling them to interact with each other through unique addressing schemes to achieve common goals, such as the realization of intelligence identification, location tracking, surveillance, and management functions [15]. In the IoT, everything can be exchanged without human intervention. In fact, the IoT contains two meanings: first, the IoT is the Internet, and its extension, core, and foundation are still the Internet; second, the scope of the IoT has been extended to everything to achieve information exchange. The architecture of the IoT is shown in Figure 2. The IoT is regarded as a new generation of information and communication technology. It has three very important basic elements, namely, information collection, information transmission, and information processing; the most critical technology is information collection [16]. By collecting and processing information, we can achieve the real-time detection of various information.

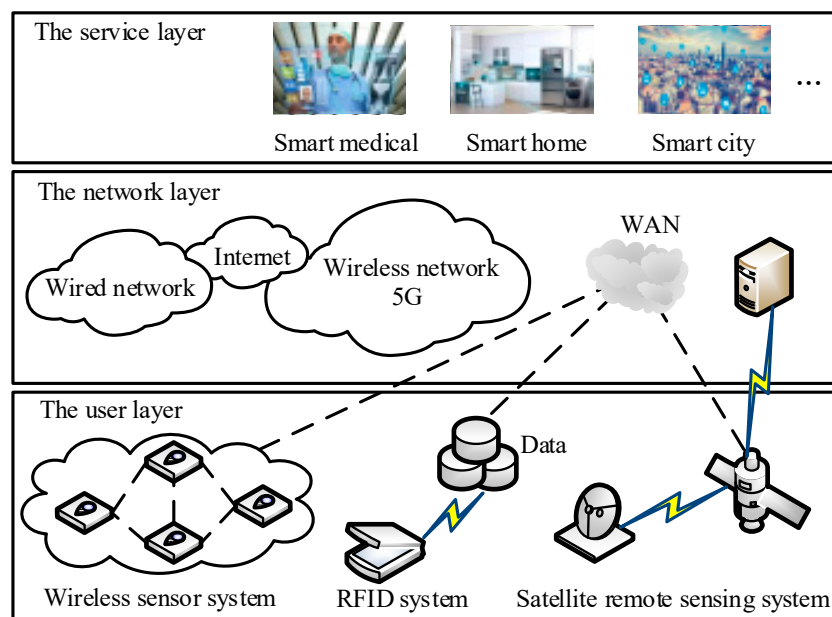


Figure 2. The architecture of 5G IoT.

With the IoT, we can provide users with a variety of new services and benefits, as their infrastructure allows for a more effective use of information about the user's context, such as geographic location, weather conditions, availability and functionality of various surrounding objects, history records before use, feedback from other users, and so on [17]. This allows users to create a personalized service space, configure the environment, and start services to represent themselves, thereby providing maximum personal comfort to all users at any time and any place.

2.2. 5G IoT for Smart Tourism

With the development of Internet-based information acquisition technology, the IoT will start the third wave of industrialization [18]. The development and application of the IoT is not only one of the important strategic measures to solve economic problems, but also a new trend of globalization, which will lead us into the new information age. The same holds true in its application in the travel and entertainment business. It can help improve the depth and breadth of information perception in each link and provide reliable support for the intelligentization of the tourism industry. The application of smart tourism is shown in Figure 3. People can easily access and use services intuitively. This means that comprehensive and well-designed services will play an important role in future cloud services. The scenic area management system is the most important link to realize tourism automation and high efficiency. The application of the IoT in tourist attractions not only conforms to the concept of “low carbon”, but can also greatly improve management efficiency [19]. Potential applications of the IoT in tourism include wildlife monitoring and tracking, marine monitoring, bird and plant species surveillance, tourism information services, hotel services, tourism marketing, mountain climbing, and weather surveillance. The IoT can monitor and track wildlife in many ways, for instance via sensors implanted in animals, by transmitting data to a server for workers to access and track the animals’ movements. The application of the IoT in smart tourism can implement functions such as tourism management, ticket management, passenger flow management, information collection, security monitoring, and environmental monitoring. Moreover, statistical analysis, statistical reporting, and integrated management are easier to implement for IoT systems.

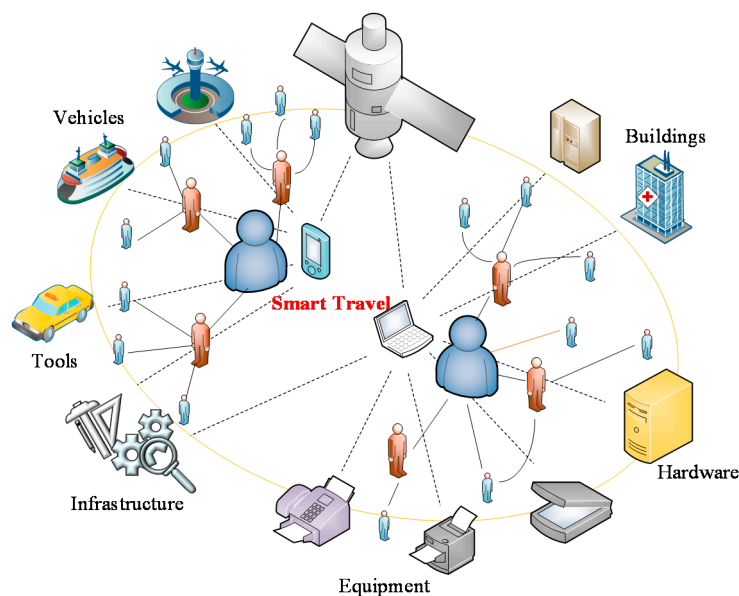


Figure 3. The application of smart tourism.

The combination of 5G, VR, AI, and other technologies allows tourists to experience the beauty and culture of their destination in an immersive manner regardless of the time and place. The early transmission of tourist attraction photos through 5G is also an advantage. Thus, users can enjoy the beautiful scenery without leaving home. Using 5G and voice interaction technology, intelligent robots can actively serve tourists, respond quickly to demand, and improve service efficiency in scenic spots. The intelligent hawk-eye system throughout the scenic area provides an analysis of the AI functions of edge computing through the 5G network, forming an invisible protection net in the scenic area, which can realize an unconscious service for tourists and a refined management. Users can also take photos and videos to generate corresponding interactive videos and high-definition video panorama travel notes, reproduce beautiful memories in three dimensions, and comprehensively record what they have seen. The scenic area can simultaneously generate a panoramic travel experience for each

tourist and then enable him or her to immediately receive and share this experience upon leaving the scenic area.

The implementation of smart tourism requires the combination of the IoT with a variety of easy-to-use applications, relying on four core information and communication technologies, i.e., the IoT, mobile communications, cloud computing, and AI technologies, to maximize the use of the IoT for information communication [20]. Therefore, the IoT technology will become an important element in the smart tourism solution to meet the needs of the new generation of tourists.

For further analysis, we expound how smart tourism services can achieve essential functions by choosing a destination and searching for suitable travel arrangements. These services provide a personalized representation of a destination, with a focus on attracting users' attention to inspire them to learn more about the place and ultimately visit it. This is accomplished by providing users with useful and interesting details, such as information about the experiences of tourists on the go (photos, reviews, experiences, etc. that people show on social networks), to help them choose the best travel arrangements for them [21]. The destinations may also be divided according to age groups, so as to improve efficiency when users are selecting destinations.

Users can also share publicly transmitted data instantly from their smartphones [22]. Traffic information (whether it is a public transport company or a private transport company) can immediately improve travelers' schedules and provide them with up-to-date information. It can differentiate users or travelers at different levels and provide useful information that is easy to understand. For example, when a user utilizes a smartphone to log into our service platform to read interactive devices such as RFID, Quick Response (QR) codes, or AR pictures, our system can help filter most of the information and display what the user likes or what suits him or her. The intelligent tourism management system based on the IoT has completely changed the traditional tourism management method and improved efficiency [23].

3. Human-Guided Machine Learning

3.1. Human-Guided Machine Learning Introduction

Intelligent machines have become close companions of human beings. The interaction and cooperation between humans and intelligent machines will be indispensable in the formation of our future society [24]. However, many problems faced by humans are often highly complex, uncertain, and open ended. As humans are the arbiters of the ultimate goal of intelligent machines, human intervention in machines has been consistent throughout the development of these systems [25]. In addition, even if sufficient data resources or unlimited data resources are provided for the AI system, the possibility of human intervention in the intelligent system cannot be ruled out. Many problems of the AI system need to be solved, such as, for example, how to understand the nuances and ambiguities of human language in the face of human–computer interaction systems and how to avoid the risks and even the possible harm caused by the limitations of AI. The human-guided machine learning system combines machine and human intelligence to overcome the shortcomings of existing AI systems [26].

As shown in Figure 4, in human-guided machine learning systems, human intelligence can be integrated into the AI systems to complement the machine functions throughout their life cycle. Human-guided machine learning systems can share computing tasks with humans as needed to overcome the shortcomings of the AI systems. Human participation can prevent errors and failures that can be caused by AI systems working alone, and human feedback can lead to a benign improvement cycle, allowing the system to continuously learn [27].

Human-guided machine learning systems can also easily solve problems that are not easy to classify or that could not be integrated in machine learning. For different fields, different human-guided machine learning systems should be constructed. Human-guided machine learning models integrate machine learning and human decision making. They make use of machine learning (supervised and unsupervised) to create a model from the training data or a small number of samples and use the

model to predict new data [28]. In the hybrid learning framework, when the system is abnormal or the computer prediction error rate is large, it should be determined whether the prediction needs to be adjusted manually or if manual intervention is required. The system's knowledge base is then automatically updated. Human intervention in the algorithm can improve the accuracy and credibility of the system. Of course, human-guided machine learning requires minimal personnel involvement so that the computers can do most of the work. The intelligence of the hybrid learning model can greatly expand the scale and efficiency of tasks that humans can complete.

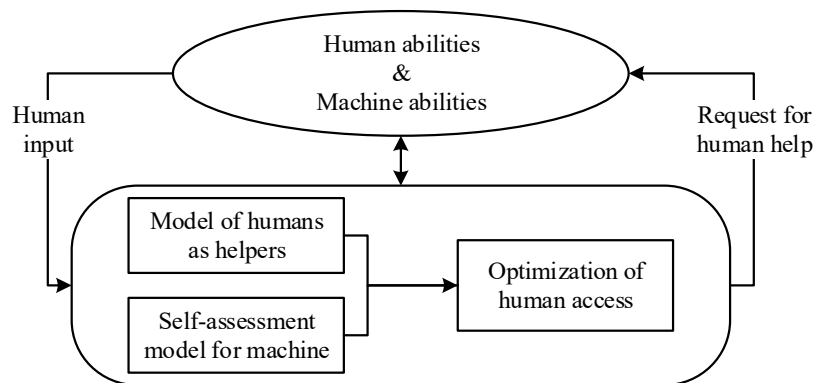


Figure 4. Reasoning capabilities for human-guided machine learning systems.

3.2. Tourist Behavior Introduction

With the rising trends in travel, most travel websites provide the following services: travel information, traffic information, flight or hotel reservations, etc. A suitable travel time and a large number of historical orders will become the final destination for new tourist users. If we analyze the multiple behavior indicators of users, we expect to use these existing data to predict user behavior information, so that tourists can make decisions based on historical order information or browsing information.

3.3. Use Human-Guided Machine Learning to Analyze Tourist Behavior

By analyzing the behavior of travel users, we can help users quickly determine their travel destinations. Human-guided machine learning is an effective tool for data analysis. By using AI, we can cluster and classify user behaviors and predict user decision behaviors.

However, for users with personal information or fewer historical orders, using AI methods to analyze them will not achieve good results and will even increase the error of the entire system. Therefore, we use human-guided machine learning methods to analyze user behavior. For special users, we can remove or tag them using manual methods.

4. Experimental Analysis and Simulation

We used travel datasets published by the Data Castle (DC) contest to analyze user behavior. The dataset includes the information of more than 50,000 users in the travel app (some of them completed the order after browsing and some only viewed the historical order on the app), including personal information. Their millions of browsing records and corresponding historical order records also contain user reviews of historical orders.

By analyzing the data, we learned that, for most users, their choice of a tourist destination has a few regularities. For example, among the users distributed in 31 provinces in mainland China, about 61.78% come from Beijing, Shanghai, and Guangzhou. The destinations of these tourists are distributed in 51 countries worldwide, with about 54.64% of them choosing Asia. In order to predict a user's decision-making behavior, we mainly consider the following aspects when selecting variables as the user's decision-making object:

1. The user browses travel information in order to determine whether to travel to a certain tourist destination.
2. Judging the practical value of a desired tourist destination based on information such as user historical orders.

Combining the above points, we will train and predict whether users will travel to tourist destination A within a short period of time, which greatly saves users' preparation time in the early stages of travel.

4.1. Dataset Labeling Method

We need to label the travel user's choice of destination. The basis for labeling is the positioning technology for users. Positioning technology is mainly divided into two categories: GPS and Location-Based Service (LBS) [29]. In the GPS-based method, the Mobile Station (MS) receives and measures signal parameters from at least four of the 24 GPS satellites in the current network. Therefore, the GPS system has a relatively high accuracy and can provide very accurate location information about the user. However, in practical applications, embedding a GPS receiver in a mobile device will cause a large increase in cost, battery consumption, etc., and to a certain extent, expose the personal information of the user's geographic location.

The LBS technology obtains the location information of mobile terminal users through wireless communication networks of telecommunications and mobile operators [30]. Relying on Base Stations (BS) to collect user data, this technology can pinpoint the user's location accurately under the condition that there is no dead zone in the mobile phone network coverage area. It will not consume the power of the user's mobile terminal equipment, and there is no requirement for the hardware configuration of the mobile terminal equipment. Compared with GPS, this method is more feasible. Therefore, the positioning method used in this paper is the LBS BS positioning technology. The method is used to measure the signal parameters received by the mobile user on the network BS. Using this technique, the BS measures the signal sent from the MS and relays it to the central site for further processing and data fusion to provide an estimate of the MS's position [31,32].

The data fusion step combines measurement data from different BSs to obtain an estimate of the MS position. As shown in Figure 5, let (x_m, y_m) represent the MS position coordinates in the Cartesian coordinate system. Let the coordinates of BSs (BS₁, BS₂, and BS₃) be (x_1, y_1) , (x_2, y_2) , and (x_3, y_3) . For simplicity, only the x and y coordinates are considered in the derivation. The z coordinates are ignored. This corresponds to the case where BSs and mobile users are located on a relatively flat plane. Without loss of generality, the origin of the Cartesian coordinate system is set to BS₁, that is, $(x_1, y_1) = (0, 0)$. The most common signal parameters are the arrival time, angle, and amplitude of the MS signal. Here, we use the Time Difference of Arrival (TDOA) method to determine the approximate location of the user. For example, in Figure 5, the MS_m detects that its coordinate position is within the coordinate range of the travel destination through the BS. It is considered to have reached the destination within a short time and can be marked as "1". However, if the MS_m detects that its coordinates are not within the coordinates of the attractions, it is considered not to have reached the destination within a short time and can be marked as "0".

The TDOA data fusion method is estimated for the combined MS signals when three different BSs are reached [33]. Since the wireless signal travels at the speed of light ($c = 3 \times 10^8$ m/s), the distance between MS and BS_i is:

$$r_i = (t_i - t^0)c \quad (1)$$

Among them, t^0 is the time point when the MS starts transmission, and t_i is the MS signal at BS_i. The distance (x_m, y_m) can be estimated by solving the following set of equations (r_1, r_2, r_3) :

$$r_1^2 = x_m^2 + y_m^2 \quad (2)$$

$$r_2^2 = (x_2 - x_m)^2 + (y_2 - y_m)^2 \quad (3)$$

$$r_3^2 = (x_3 - x_m)^2 + (y_3 - y_m)^2 \quad (4)$$

Without loss of generality, we can assume that $r_1 < r_2 < r_3$.

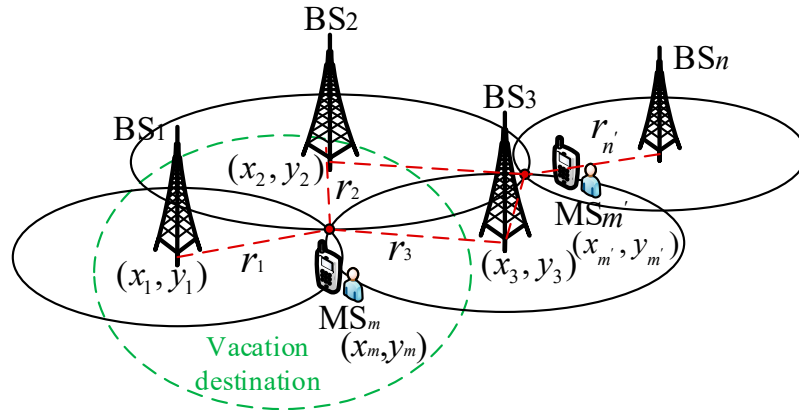


Figure 5. TDOA data fusion using multiple BSs.

Next, we define the difference in distance between the MS and different BS_i values as:

$$r_{i1} = r_i - r_1 = (t_i - t^0)c - (t_1 - t^0)c = (t_i - t_1)c \quad (5)$$

Equation (3) can be rewritten as the TDOA measurement value r_{21} as:

$$(r_{21} + r_1)^2 = K_2^2 - 2x_2x_m - 2y_2y_m + r_1^2 \quad (6)$$

where:

$$K_i^2 = x_i^2 + y_i^2 \quad (7)$$

Expanding and rearranging the terms yields:

$$-x_2x_m - y_2y_m = r_{21}r_1 + \frac{1}{2}(r_{21}^2 - K_2^2) \quad (8)$$

Similarly, Equation (4) results in:

$$-x_3x_m - y_3y_m = r_{31}r_1 + \frac{1}{2}(r_{31}^2 - K_3^2) \quad (9)$$

Rewriting these equations in matrix form gives:

$$\mathbf{H}\mathbf{x} = r_1\mathbf{c} + \mathbf{d} \quad (10)$$

where:

$$\mathbf{x} = \begin{bmatrix} x_m \\ y_m \end{bmatrix}, \mathbf{c} = \begin{bmatrix} -r_{21} \\ -r_{31} \end{bmatrix}, \mathbf{d} = \frac{1}{2} \begin{bmatrix} K_2^2 - r_{21}^2 \\ K_3^2 - r_{31}^2 \end{bmatrix} \quad (11)$$

Equation (10) can be used to solve the unknown r_1 of x :

$$\mathbf{x} = r_1\mathbf{H}^{-1}\mathbf{c} + \mathbf{H}^{-1}\mathbf{d} \quad (12)$$

Substituting this intermediate result into Equation (2), we can obtain the quadratic equation in r_1 . Solve r_1 and substitute the positive root into Equation (12) to get the final solution of x .

If the MS position involves more than three BSs, then Equation (10) remains.

$$\mathbf{H} = \begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_3 \\ \vdots & \vdots \end{bmatrix}, \mathbf{c} = \begin{bmatrix} -r_{21} \\ -r_{31} \\ -r_{41} \\ \vdots \end{bmatrix}, \mathbf{d} = \frac{1}{2} \begin{bmatrix} K_2^2 - r_{21}^2 \\ K_3^2 - r_{31}^2 \\ K_4^2 - r_{41}^2 \\ \vdots \end{bmatrix} \quad (13)$$

which yields the following least-squares intermediate solution:

$$\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (r_1 \mathbf{c} + \mathbf{d}) \quad (14)$$

Combine this intermediate result with Equation (2) again to get the final estimate of $\hat{\mathbf{x}}$. If the second-order statistics of TDOA measurement errors are known, a more accurate solution can be obtained, as in [34].

According to the calculation results, if the coordinates of the user's MS are within the range of the attractions, it proves that the user selects a certain travel destination in the short term, and it is marked as "1"; otherwise, it is marked as "0".

4.2. KNN Classification for Tourist Behavior

In this simulation, the dataset we used had information from more than 50,000 users. We first needed to go down the dataset and filter out the data with more meaningful simulation results. We used the TDOA method to determine the approximate location of the user and detect whether the coordinate position was within the coordinate range of the travel destination. By comprehensively considering the multiple classification algorithms in machine learning, as shown in Table 1, the classification ideas in this manuscript are basically consistent with the ideas of the KNN algorithm. KNN is simple and easy to use when compared with other algorithms. The model training time is fast, the prediction effect is good, and it is not sensitive to outliers, so KNN was chosen in this study for the simulation.

Table 1. Advantages and Disadvantages of each Algorithm.

Algorithm	Advantages	Disadvantages
Decision Tree	There is no need to generalize the data like for other algorithms, such as removing redundant or blank attributes.	(1) Difficult to use in dealing with missing data. (2) Leads to the emergence of overfitting problems. (3) Ignores the correlation between attributes in the dataset.
Bayes	The estimated parameters required are few, and the algorithm is simple.	(1) The prior probability needs to be known. (2) An error rate in the classification decision.
Support Vector Machine (SVM)	It is advantageous when the sample data capacity is small.	(1) No universal solution to nonlinear problems. (2) Lack of sensitivity to data.
KNN	It is simple, effective, and suitable for the automatic classification of class domains with relatively large sample sizes.	When the samples are unbalanced, the prediction accuracy of rare categories is low.

Next, we will use the KNN method to classify user behavior. KNN is a supervised learning method that classifies new data by measuring the distance between the new data point and the remaining labeled data points [35]. This is a distance-based algorithm that can determine the category of new data

based on the category of the nearest neighbor [36]. We usually use the Euclidean method to calculate the distance between data points. Then, we will select the training data of the first k distances closest to the new data. The category with the most training data is the category of the new data. There is no fixed experience in choosing K values. Generally, a smaller value is selected based on the distribution of the samples, and a suitable K value can be selected through cross-validation.

4.3. Human-Guided Machine Learning-Based KNN Classification for Tourist Behavior

The traditional KNN algorithm does not involve the participation of AI, and there are inaccuracies when processing the edge information of some more complex data. Introducing human intelligence into the loop of intelligent systems can achieve a close coupling between advanced cognitive mechanisms for analysis and response in fuzzy and uncertain problems and machine intelligent systems. Therefore, the two adapt to and cooperate with each other, forming a two-way information exchange and control. Such “1 + 1 > 2” hybrid enhanced intelligence can be achieved by integrating human perception, cognitive ability, machine computing ability, and storage ability. Next, we apply the KNN algorithm, combined with human-guided machine learning, to the classification task for users.

In order to observe the classification error rate of KNN, we define the training error rate as the ratio of KNN training sample tags to input tags. The error rate is expressed as:

$$\frac{1}{K \sum_{x_i \in N_K(x)} I(y_i \neq c_i)} = 1 - 1/K \sum_{x_i \in N_K(x)} I(y_i = c_j) \quad (15)$$

Thus, the proper k can be maximized:

$$1/K \sum_{x_i \in N_K(x)} I(y_i = c_j) \quad (16)$$

in the training set.

We use grid search and cross-validation to choose the appropriate K value to verify the accuracy of the classification [37]. First, we screened the data. After screening, there were 40,307 data in total. The data was divided into six groups. The first five groups were used for cross-validation. Each group had 6717 data, for a total of 33,585 data, while the sixth group had 6722 data. The data were used for the final accuracy test.

In the cross-validation, we used four sets of data as the training set and the remaining one set of data as the test set. Figure 6 shows the accuracy of the different numbers of neighbors of the training set and test set; when $K = 8$, the training error rate is minimal. Finally, the remaining 6722 data were checked as the final verification set. The correct rate was 84.06%.

In the field of machine learning, a confusion matrix is a specific matrix used to visualize the performance of an algorithm [38]. In predictive analysis, the confusion matrix is a table consisting of False Positives (FP), False Negatives (FN), True Positives (TP), and True Negatives (TN). By using an obfuscation matrix, it is easy to see if the learning machine will obfuscate two similar classes [39]. The “confusion matrix” of the classification results in this experiment is shown in Figure 7.

In order to comprehensively observe the generalization performance of the model, the Receiver Operating Characteristic (ROC) curve can be drawn to better observe the characteristics of the model’s distinguishing types [40]. The ROC curve is usually more effective when the class distribution is not uniform. The Area Under Curve (AUC) is the area under the ROC curve, which is an indicator used to judge the performance of the learning machine [41]. The AUC-ROC curve is a performance measure for classification problems under various threshold settings. ROC is a probability curve, and AUC represents the degree or measure of separability. It embodies the distinguishing ability of a model. The higher the AUC, the higher the model’s ability to distinguish between 0 and 1, that is, the more it can reflect whether the model can correctly predict the user’s decision.

As shown in Figure 8a, the ROC curve is plotted with TPR versus FPR, where TPR is on the y-axis and FPR is on the x-axis. Points A, B, and C are the AUCs of three different classification results with different parameters. After comparing these three points, we can ascertain that the AUC-A has the best performance.

The ROC curve of the classification model we established is shown in Figure 8b. As shown in the figure, the curve is close to a semi-ellipse and the AUC = 0.809, indicating that the classification effect of our model is good. Therefore, using this model, we can classify whether travel users will choose destination A or not. Based on these classification results, corresponding services are provided to users.

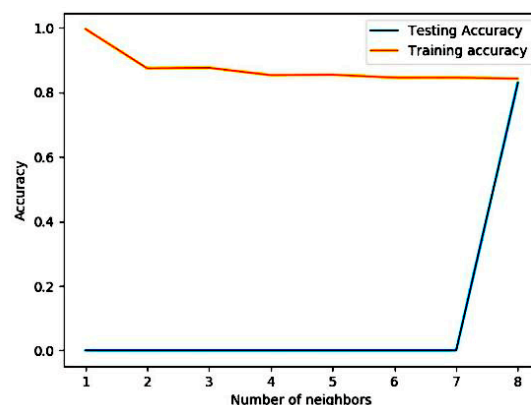


Figure 6. The correct rate of cross-validation.

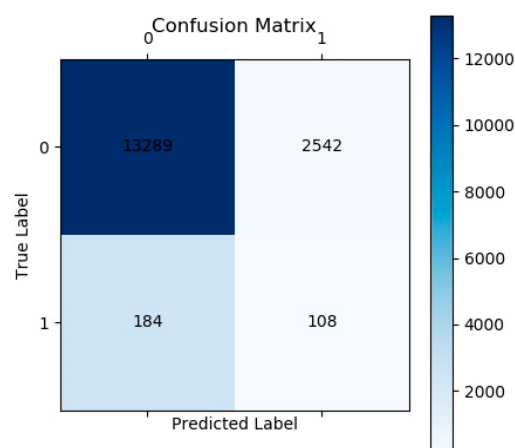


Figure 7. The confusion matrix.

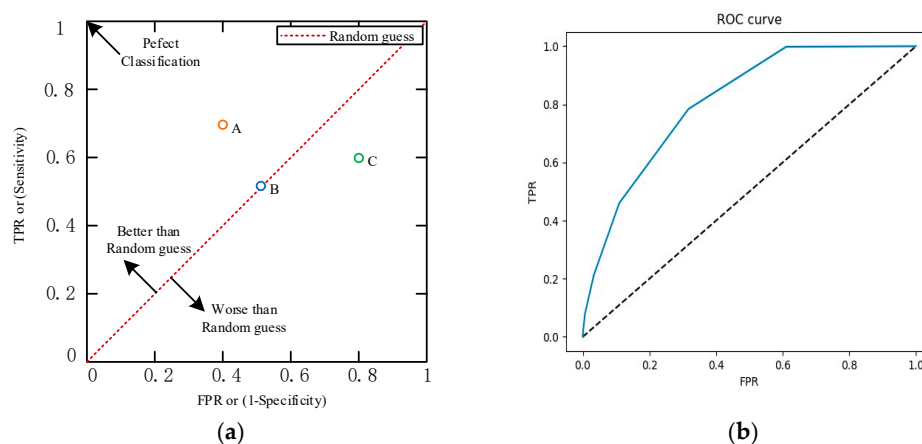


Figure 8. ROC-AUC curves. (a) Comparison of different ROC-AUC curves. (b) The ROC curve of the proposed model.

5. Conclusions

With the continuous development of the tourism industry, integrating the IoT into tourism projects is a very promising technology. Smart tourism uses the IoT to maximize information analysis and information integration, as well as to achieve a fast and convenient information exchange between users. In other words, the IoT technology will become an important element to meet the needs of the new generation of tourists. Based on this, we propose a tourist behavior decision-making method based on human-guided machine learning called KNN classification. This classification method can effectively help tourists make a decision on whether to choose a certain travel destination based on historical order data and historical browsing information. The results obtained from the cross-validation experiments and performance evaluation prove the effectiveness of the method.

Author Contributions: R.P. contributed the central idea, analyzed most of the data, and wrote the initial draft of the paper. Y.L. contributed the idea for the simulation experiment and provided constructive suggestions. M.K. contributed in the revision of the paper and polishing of the language. All authors discussed the results and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: Partial results of this work have been published in the International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI) 2018. The authors would like to thank the anonymous reviewers for their helpful comments.

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

References

1. Solmaz, G.; Wu, F.-J.; Cirillo, F.; Kovacs, E.; Santana, J.R.; Sánchez, L.; Sotres, P.; Munoz, L. Toward Understanding Crowd Mobility in Smart Cities through the Internet of Things. *IEEE Commun. Mag.* **2019**, *57*, 40–46. [[CrossRef](#)]
2. Tripathy, A.K.; Tripathy, P.K.; Ray, N.; Mohanty, S.P. iTour: The Future of Smart Tourism: An IoT Framework for the Independent Mobility of Tourists in Smart Cities. *IEEE Consum. Electron. Mag.* **2018**, *7*, 32–37. [[CrossRef](#)]
3. Zanella, A.; Bui, N.; Castellani, A.; Vangelista, L.; Zorzi, M. Internet of Things for Smart Cities. *IEEE Internet Things J.* **2014**, *1*, 22–32. [[CrossRef](#)]
4. Wu, Y.; Rong, B.; Salehian, K.; Gagnon, G. Cloud Transmission: A New Spectrum-Reuse Friendly Digital Terrestrial Broadcasting Transmission System. *IEEE Trans. Broadcast.* **2012**, *58*, 329–337. [[CrossRef](#)]
5. Mora, L.; Bolici, R.; Deakin, M. The First Two Decades of Smart-City Research: A Bibliometric Analysis. *J. Urban Technol.* **2017**, *24*, 3–27. [[CrossRef](#)]
6. Rong, B.; Qian, Y.; Lu, K.; Chen, H.-H.; Guizani, M. Call Admission Control Optimization in WiMAX Networks. *IEEE Trans. Veh. Technol.* **2008**, *57*, 2509–2522. [[CrossRef](#)]
7. Lacinák, M.; Ristvej, J. Smart City, Safety and Security. *Procedia Eng.* **2017**, *192*, 522–527. [[CrossRef](#)]
8. Sun, S.; Kadoch, M.; Gong, L.; Rong, B. Integrating network function virtualization with SDR and SDN for 4G/5G networks. *IEEE Netw.* **2015**, *29*, 54–59. [[CrossRef](#)]
9. Hung, J.C. The smart-travel system: Utilising cloud services to aid traveller with personalised requirement. *Int. J. Web Grid Serv.* **2012**, *8*, 279. [[CrossRef](#)]
10. Gretzel, U.; Sigala, M.; Xiang, Z.; Koo, C. Smart tourism: Foundations and de-velopments. *Electron. Mark.* **2015**, *25*, 179–188. [[CrossRef](#)]
11. Gautam, B.P.; Asami, H.; Batajoo, A.; Fujisaki, T. Regional Revival through IoT Enabled Smart Tourism Process Framework (STPF): A Proposal. In Proceedings of the 2016 Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS), Sapporo, Japan, 25–28 August 2016; pp. 743–748.
12. Balandina, E.; Balandin, S.; Koucheryavy, Y.; Mouromtsev, D.; Ekaterina, B.; Sergey, B.; Yevgeni, K.; Dmitry, M. Innovative e-Tourism Services on Top of Geo2Tag LBS Platform. In Proceedings of the 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Bangkok, Thailand, 23–27 November 2015; pp. 752–759.

13. Balandina, E.; Balandin, S.; Koucheryavy, Y.; Mouromtsev, D. IoT Use Cases in Healthcare and Tourism. In Proceedings of the 2015 IEEE 17th Conference on Business Informatics, Lisbon, Portugal, 13–16 July 2015; Volume 2, pp. 37–44.
14. Hung, J.C.; Hsu, V.; Wang, Y.-B. A Smart-Travel System Based on Social Network Service for Cloud Environment. In Proceedings of the 2011 Third International Conference on Intelligent Networking and Collaborative Systems, Fukuoka, Japan, 30 November–2 December 2011; pp. 514–519.
15. Muthuraman, S.; Al Hazi, M. Smart Tourism Destination—New Exploration towards Sustainable Development in Sultanate of Oman. In Proceedings of the 2019 5th International Conference on Information Management (ICIM), Cambridge, UK, 24–27 March 2019; pp. 332–335.
16. Gcaba, O.; Dlodlo, N. The internet of things for South African tourism. In Proceedings of the 2016 IST-Africa Week Conference, Durban, South Africa, 11–13 May 2016; pp. 1–8.
17. Tan, L.; Wang, N. Future internet: The internet of things. In Proceedings of the 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), Chengdu, China, 20–22 August 2010; Volume 5.
18. Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. *Futur. Gener. Comput. Syst.* **2013**, *29*, 1645–1660. [[CrossRef](#)]
19. Lee, I.; Lee, K. The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Bus. Horizons* **2015**, *58*, 431–440. [[CrossRef](#)]
20. Lei, Y.; Ma, P.; Zhao, L. The Internet of Things brings new wave of the information industry. *Int. J. Comput. Sci. Netw. Secur.* **2011**, *11*, 15–21.
21. Werthner, H.; Klein, S. *Information Technology and Tourism: A Challenging Relationship*; Springer-Verlag Wien: Vienna, Austria, 1999.
22. Lin, Y. The Application of the Internet of Things in Hainan Tourism Scenic Spot. In Proceedings of the 2011 Seventh International Conference on Computational Intelligence and Security, Hainan, China, 3–4 December 2011; pp. 1549–1553.
23. Guo, X.; Zeng, T.; Wang, Y.; Zhang, J. Fuzzy TOPSIS Approaches for Assessing the Intelligence Level of IoT-Based Tourist Attractions. *IEEE Access* **2018**, *7*, 1195–1207. [[CrossRef](#)]
24. Kamar, E. Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 4070–4073.
25. Zheng, N.; Liu, Z.-Y.; Ren, P.-J.; Ma, Y.-Q.; Chen, S.-T.; Yu, S.-Y.; Xue, J.-R.; Chen, B.-D.; Wang, F.-Y. Hybrid-augmented intelligence: Collaboration and cognition. *Front. Inf. Technol. Electron. Eng.* **2017**, *18*, 153–179. [[CrossRef](#)]
26. Liang, H.; Yang, L.; Cheng, H.; Tu, W.; Xu, M. Human-in-the-loop reinforcement learning. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017.
27. Barnes, M.J.; Chen, J.Y.; Jentsch, F.; Redden, E.; Light, K. An Overview of Humans and Autonomy for Military Environments: Safety, Types of Autonomy, Agents, and User Interfaces. In Proceedings of the 10th international conference on Engineering Psychology and Cognitive Ergonomics: Applications and Services, Las Vegas, NV, USA, 21–26 July 2013; Springer: Berlin/Heidelberg, Germany, 2013.
28. Dellermann, D.; Lipusch, N.; Ebel, P. Integrating Ecosystem Intelligence with the Hybrid Intelligence Accelerator. *SSRN Electron. J.* **2017**, *61*, 637–643. [[CrossRef](#)]
29. Sayed, A.H.; Yousef, N.R. Wireless location. In *Wiley Encyclopedia of Telecommunications*; Proakis, J., Ed.; Wiley: New York, NY, USA, 2003.
30. Gustafsson, F.; Gunnarsson, F. Mobile positioning using wireless networks: Possibilities and fundamental limitations based on available wireless network measurements. *IEEE Signal Process. Mag.* **2005**, *22*, 41–53. [[CrossRef](#)]
31. Spirito, M. On the accuracy of cellular mobile station location estimation. *IEEE Trans. Veh. Technol.* **2001**, *50*, 674–685. [[CrossRef](#)]
32. Weiss, A. On the accuracy of a cellular location system based on rss measurements. *IEEE Trans. Veh. Technol.* **2003**, *52*, 1508–1518. [[CrossRef](#)]
33. Sayed, A.H.; Tarighat, A.; Khajehnouri, N. Network-based wireless location: Challenges faced in developing techniques for accurate wireless location information. *IEEE Signal Process. Mag.* **2005**, *22*, 24–40. [[CrossRef](#)]

34. Chan, Y.; Ho, K.C. A simple and efficient estimator for hyperbolic location. *IEEE Trans. Signal Process.* **1994**, *42*, 1905–1915. [[CrossRef](#)]
35. Quek, Y.; Woo, W.L.; Logenthiran, T. DC equipment identification using K-means clustering and kNN classification techniques. In Proceedings of the 2016 IEEE Region 10 Conference (TENCON), Singapore, 22–25 November 2017.
36. Zhou, L.; Wang, L.; Ge, X.; Shi, Q. A clustering-Based KNN improved algorithm CLKNN for text classification. In Proceedings of the 2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010), Wuhan, China, 6–7 March 2010; Volume 3.
37. Sun, S.; Huang, R. An adaptive k-nearest neighbor algorithm. In Proceedings of the 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, Yantai, China, 10–12 August 2010; Volume 1.
38. Keller, J.M.; Gray, M.R.; Givens, J.A. A fuzzy K-nearest neighbor algorithm. *IEEE Trans. Syst. Man Cybern.* **1985**, *4*, 580–585. [[CrossRef](#)]
39. Yang, L. K-Nearest Neighbor Classification Based on Semantic Distance. *J. Softw.* **2005**, *16*, 2054. [[CrossRef](#)]
40. Bradley, A. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [[CrossRef](#)]
41. Slaby, A. ROC Analysis with Matlab. In Proceedings of the 2007 29th International Conference on Information Technology Interfaces, Cavtat, Croatia, 25–28 June 2007.



© 2020 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/>).