Research Article

Research on Intelligent Tourism Information System Based on Data Mining Algorithm

Jia Du

Lanzhou University of Arts and Science, Lanzhou 730000, China

Correspondence should be addressed to Jia Du; 1000247@luas.edu.cn

Received 28 June 2021; Revised 26 July 2021; Accepted 3 September 2021; Published 23 September 2021

1. Introduction

Smart travel is developed on traditional sightseeing services which is new way of gathering information [1]. It encompasses the most modern computer technology in terms of software and hardware. In the fields of e-commerce, online banking, tourism, medical care, etc., we momentarily witness situations in which data mining is successful [2]. Many data have been gathered in the tourism sector with the popularization of tourist information. Data mining and visualization technology is, therefore, required to process tourist information to assess tourism effects thoroughly and in detail and give efficient support for tourism industry. In the field of scientific literature, [3] the concept of intelligent tourism destinations has become a hot topic, and constructive arguments are not just in tourism. This notion is largely linked to the development of modern IT, which resulted in the changing of tourism purchasing behavior, leading to the implementation of new information systems for destination management organizations (DMOs), to govern the destination efficiently. To exchange knowledge and skills, the task is to provide visitors in real time a personal service and link all parties. In smart travel, it is not a passive way for users to be provided with what information they browse. Rather, it is to provide users with correct scenic spot information through a personalized recommendation method, which is very important [4].

In literature, a number of articles [4, 5] have proposed different smart travel system and guidance. For instance, in [6], the author provides examples of this aspect to help the user find the information if the user has insufficient knowledge of a certain aspect of tourism resources. Similarly in [7], the authors describe that how the knowledge-based recommendation system replicates depending on the user’s operational behavior and addresses trip knowledge. The author also suggests an improvement strategy; however, it is very difficult to apply it because this knowledge base needs often to be updated. Moreover, the literature [8–11] conducted an in-depth study on the navigation of the intelligent itinerary based on research and design constraints. These systems use technologies such as WebGIS and Apriori algorithm. This is an interactive platform where users can change the travel schedule according to their wishes until they are satisfied.

In this research, we proposed an intelligent recommendation system for travel services and introduced the
analysis results to users using big data technology. The study focuses on the function of large data algorithms in the intelligent tourism system. We also improve the visualization of big data technologies on the intelligent tourist information system. The functional structure of the intelligent tourist information system is developed according to current requirements. The proposed system performance is assessed in combination with experiments to enhance the marketing effect of intelligent tourism and the user experiences. The major contributions of the paper are as follows to improve the big data-based algorithms with the k-means algorithm.

(i) The proposed model examines the smart tourism system using digital technology when traveling
(ii) The proposed model improves the visualization of big data technologies
(iii) The proposed model analyzes a huge amount of data with experiments to enhance the marketing effect of intelligent tourism
(iv) The performance of the suggested model is thoroughly assessed using the powerful computing abilities of the spark cluster to handle the intelligent tourism system efficiency and quality

The rest of the paper is organized as follows. In Section 2, a proposed system model is designed for smart tourism. The big data application analytics analysis is conducted in Section 3. The experimental results and discussion are further summarized in Section 4. Finally, Section 5 concludes the paper with a summary and future research directions.

2. Tourism Distribution Model

In this section, we elaborate the framework of the smart tourism platform, the improved mining algorithm for visualized data of tourism, and the paradigm of communication competence in the communicative capability of tourist skills.

2.1. Framework of Smart Tourism. The smart travel ecosystem is built around the smart travel platform. The overall framework of smart tourism is shown in Figure 1. From Figure 1, the framework is composed of five parts i.e., smart tourism platform, tourism management department, tourism enterprise, traveler, and security system. The tourist industry and the department for tourism manage the smart tourism platform and manipulate tourism businesses using the appropriate information received from tourism companies on the platform. Tourism businesses, as the market’s major body, receive different useful information from smart tourism platforms to give visitors tailored tourism services and product launches. Travelers, as customers, may easily receive information and services supplied by travel firms via smart travel platforms. The security system enables the creation and subsequent safe and stable functioning of the smart tourist system in terms of policies, money, skills, and procedures.

Technical framework of the smart tourism platform is shown in Figure 2. From Figure 2, the technical framework of the smart travel platform is divided into four levels i.e., data collection layer, network communication layer, data center layer, and business application layer. The data collection layer is the direct source of data information, which is composed of cameras, sensors, smart phones, and other information input terminals. The network communication layer is the transmission path of data information and makes the transmission of data information between layers possible. The data center layer is responsible for the preservation of data information, classifying the data, and further mining and analysis. The business application layer uses the valuable information provided by the data center layer to provide rich business functions to different users.

2.2. Improved Mining Algorithm of Visualized Data of Smart Tourism. In the traditional coordinate system, all the axes cross each other. Figure 3 shows the parallel coordinates in the 6-dimensional space. Before displaying, the algorithm first converts the source data. Generally, the data types can be divided into two types: sequence data and actual data. The sequence data is represented by real numbers. It can also be represented by an integer, but in this case, the meaning of continuity is different from the definition of mathematical perspective. The view allows users to roughly identify the data distribution of each attribute, especially different types of data are displayed in different colors, and can more clearly indicate the differences between different types of data.

Furthermore, parallel coordinates according to their features are suited for continuous data and cannot be used for discrete data. In the event of real and discrete data, the processing of the data in the actual number type has to be continuous. The so-called parallel coordination concept consists of the mapping of N-dimensional reservoirs in parallel N axes to line segments. The N-points that cross the line and the N-axis represent the N-dimensional data of the data point. A linear independent equation [12] can represent the polyline produced by this line segment:

\[
\frac{x_1 - a_1}{u_1} = \frac{x_2 - a_2}{u_2} = \cdots = \frac{x_n - a_n}{u_n}.
\]

Among them, the slope is

\[
m_i = \frac{u_{i+1}}{u_i}.
\]

The following formula represents the intercept on the \( x_{i+1} \) axis in the \( x_{i+1}x_i \) plane [13]:

\[
b_i = (a_{i+1} - m_i a_i).
\]

Here, if the distance between parallel axes is 1, then the coordinates of the intersection point are

\[
\left(\frac{i}{1 - m_i}, \frac{b_i}{1 - m_i}\right), i = 1, 2, \ldots, n - 1.
\]

The point in the rectangular coordinate of the plane is mapped to the parallel coordinate as a line segment. In
planar rectangular coordinates, multiple points on a straight line are mapped to parallel coordinates by multiple line segments that intersect one point so that a point in parallel coordinates can correspond to a straight line in planar rectangular coordinates. The mathematical relationship between rectangular coordinates and parallel coordinates, mutual transformation, distance comparison, etc., has been proved very early. In short, the above two-dimensional situation can be expanded in a multidimensional space, and these mapping relationships provide a basis for understanding the multidimensional dataset in the parallel coordinate system [14]. Correspondence between parallel coordinates and rectangular coordinates is shown in Figure 4.

2.3. Intelligent Information System for Smart Tourism. The advantages of the parallel coordinate method are mainly obvious. Users can quickly convert from the traditional rectangular coordinate system to the parallel coordinate system and can better use the two-dimensional plane to display multidimensional data, and compared with other data visualization methods, such as scatter plots, it is simple and intuitive for users to receive. Data preprocessing is an important step of knowledge discovery in the database, which helps to improve the quality of data and ensure the accuracy and performance of the results. Data preprocessing is a broad field that includes different strategies and technologies related in many complex ways. In view of the ongoing research and the unique characteristics of the Iris data, the focus is on solving the data conversion problem. Data conversion is the conversion of all values used for variables. In other words, the variable value of the object is used for the conversion of each object. Normalization and standardization are general methods of variable transformation. Moreover, the general normalization method is to linearly transform the data, map the corresponding value to a new interval, and perform the following minimum and maximum normalization, as shown in the following formula [15]:

![Figure 1: The overall framework of smart tourism.](image-url)
\[ x'_{ij} = \frac{x_{ij} - \text{Min}_j}{\text{Max}_j - \text{Min}_j} \cdot (\text{newMax} - \text{newMin}) + \text{newMin}. \quad (5) \]

In the formula, \( x_{ij} \) is the original data, \( \text{Max}_j \) is the maximum value of the \( j \)th column in the array, \( \text{Min}_j \) is the minimum value in column \( j \), \( \text{newMin} \) and \( \text{newMax} \) are the new maximum and minimum values, respectively, which are two constants, and \( x'_{ij} \) is the value in the new interval after conversion. In the data mining samples, each variable represents various properties of the sample, and the measurement unit is different. In this way, variables with large absolute values are affected by it, while variables with small
absolute values may disappear. In order to ensure that the variables are in the same position in the analysis, the data can be centralized and standardized conversion [16]. Taking P as the total number of variables and n as the total number of samples (records), the average value of the jth variable is as follows:

\[ x_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}, \quad j = 1, 2, \ldots, p. \quad (6) \]

The center transformation of the nth data of the jth variable:

\[ x'_{ij} = x_{ij} - \bar{x}_j, \quad i = 1, 2, \ldots, n. \quad (7) \]

Through this conversion, the average value of each variable is 0. In other words, the value of each variable is the same base point. Standardization is based on centralization and transformation to make the range of changes of various variables equal. This paper mainly adopts standard deviation standardization. The standard deviation of the ith variable is as follows [17]:

\[ S_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_{ij} - \bar{x}_j)^2}. \quad (8) \]

Deviations standardization:

\[ x'_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j}, \quad i = 1, 2, \ldots, n. \quad (9) \]

After the conversion, the average value of each variable is 0, and the standard deviation is 1.

The cluster separation point also substantially impacts the average value and standard deviation, such that the above conversion has to be adjusted. First, the average is substituted with the median value. The standard difference is then substituted with the absolute default. In the following, you will find the median value of the jth variable:

\[ \mu_j = x_{ij}, \quad i = \frac{n}{2} + 1; \quad N \text{ is an odd number}, \]

\[ \mu_j = \frac{1}{2} (x_{ij} + x_{i+1}), \quad i = \frac{n}{2}; \quad N \text{ is even}. \quad (10) \]

In a nutshell, if there are odd values, the central value is the middle value. If there is an even value, the middle value is the average of the middle two number s [18]. The average absolute baseline deviation of the jth variable is shown below:

\[ \sigma_j = \frac{1}{n} \sum_{i=1}^{n} |x_{ij} - k\mu_j|. \quad (11) \]

In the formula, \( \mu_j \) is the mean or median. The multiplication factor \( k \) is used to stretch the distribution range near the central value of the data, and since the number of even numbers is even in the normalization of the data using Iris, using \( k = 2 \) has a better effect. For the data n of the jth variable, the median absolute standard deviation is standardized [19]:

\[ x'_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \quad i = 1, 2, \ldots, n. \quad (12) \]

2.4. Application Systems of Smart Tourism. In addition, typical techniques of converting standardization include very bad standardization and very poor standardization. Some of the following typical methods for data conversion are commonly used for visual mining. Moreover, similar replacement conversions are needed for Iris data. Otherwise, you will not get the intended outcome. It should be mentioned that when other sets of data are different, the standard conversion of data preparation should be changed by the circumstances of a given collection of data. In this experiment, the dataset is utilized to conduct a unified digital process in a category, while assessing the clustering algorithm, and to differentiate the indistinguishable categories. Generally, such processing does not impact clustering method implementation and the visualization effect might be extremely obvious. Similarly, this column’s value attribute is prepared to prevent undue impact on other values of the attribute. At this time, the processing procedure cannot employ standardization. The smallest normalization process will have a great impact on other dimensional attributes, so it is necessary to use standard deviation conversion or absolute standard deviation standard conversion. At the same time, it must be ensured that the conversion method has the least impact on other attributes. Of course, in the classification algorithm, the category attribute is only used to identify the category, so it does not require preprocessing [20]. Since the quality of a cluster analysis process depends on the choice of measurement standards, the measurement standards must
be chosen carefully. The distance between classes is defined to illustrate the relationship between the two types.

(1) The shortest distance method is as follows [21]:

\[ D_s(C_a, C_b) = \min \{d(x, y) | x \in C_a, y \in C_b \}. \]  \hfill (13)

(2) The group average method is as follows:

For the group average version of hierarchical clusters, the closeness of two clusters is defined as the average closeness of all points of different clusters:

\[ D_i(C_a, C_b) = \frac{\sum d(x, y)}{m_i \cdot m_j} \{x \in C_a, y \in C_b \}. \]  \hfill (14)

Among them, \( m_i \) and \( m_j \) are the size and number of classes \( C_a \) and \( C_b \), respectively. In addition to the above two methods, there are methods for explaining the distance between classes, including the sum of squares method, the longest distance method, the center method, and the class average method. Of course, when the data is complex, the differential sum of the squares method can be considered [22]. There are two most common methods for measuring this proximity i.e., Euclidean distance and cosine similarity:

\[ d(x_i, y_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - y_{jk})^2}. \]  \hfill (15)

\[ \text{Sim}(x_i, y_j) = \frac{\sum_{k=1}^{n} x_{ik} y_{jk}}{\sqrt{\sum_{k=1}^{n} x_{ik}^2} \sqrt{\sum_{k=1}^{n} y_{jk}^2}}. \]  \hfill (16)

The nearest neighbor method is also called passive learning, and its advantages are as follows:

(1) The distribution of attribute values does not need to be known beforehand. Most alternative typing approaches nevertheless require previous knowledge of the distribution of the property value.

(2) Suitable for gradual education.

(3) The classification precision of the KNN is typically greater than other classification methods since there is no requirement to create explicit rules.

3. Big Data Application Analytics

The beginning point and destination of smart travel construction are various smart apps based on big data. Big data is useful only via its discovery and analysis, and companies can benefit only by applying big data to applications. As a result, following the construction of the data collecting layer, network communication layer, and data center layer, various smart apps are built to provide relevant services for travel management departments, travel firms, and travelers to realize genuine smart travel. It is the smart travel application system, as indicated in Figure 5. The overall structure of the user recommendation system is shown in Figures 6 and 7, which show the process of interacting with the user, and the B/S mode is adopted.

4. Experimental Results and Discussion

During the construction of the aforementioned system, this article validates the performance of the big data smart tourism system built in this paper and performs system tests based on the actual scenario of smart tourism. Furthermore, as the system input, this paper collects huge tourist data from the Internet to assess the data mining effect of the smart tourism system developed in this research. Furthermore, this article divides the data into 59 sets and assesses the influence of system data mining in the form of expert grading. The
acquired findings are presented in Table 1 and Figure 8 below.

According to the study findings, the smart tourism system developed in this article has a significant impact on the smart mining of tourist data. On this premise, this paper combines the actual scenario with simulation research to validate the user experience satisfaction of the system created in this article. The acquired findings are presented in Table 2 and Figure 9.

According to the aforementioned results, the big data smart tourism system built in this article has good data mining effects, and the smart tourism system can effectively improve user experience, with high user satisfaction. This is also the unavoidable future path of tourism growth.
Table 1: Statistical table of evaluation of data mining effect of the smart tourism system.

<table>
<thead>
<tr>
<th>No.</th>
<th>Data mining</th>
<th>No.</th>
<th>Data mining</th>
<th>No.</th>
<th>Data mining</th>
<th>No.</th>
<th>Data mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>93.99</td>
<td>16</td>
<td>96.71</td>
<td>31</td>
<td>90.12</td>
<td>46</td>
<td>87.91</td>
</tr>
<tr>
<td>2</td>
<td>94.57</td>
<td>17</td>
<td>92.14</td>
<td>32</td>
<td>87.03</td>
<td>47</td>
<td>96.85</td>
</tr>
<tr>
<td>3</td>
<td>88.13</td>
<td>18</td>
<td>92.09</td>
<td>33</td>
<td>95.76</td>
<td>48</td>
<td>95.05</td>
</tr>
<tr>
<td>4</td>
<td>87.32</td>
<td>19</td>
<td>90.20</td>
<td>34</td>
<td>90.55</td>
<td>49</td>
<td>96.33</td>
</tr>
<tr>
<td>5</td>
<td>89.54</td>
<td>20</td>
<td>92.26</td>
<td>35</td>
<td>93.99</td>
<td>50</td>
<td>92.89</td>
</tr>
<tr>
<td>6</td>
<td>89.69</td>
<td>21</td>
<td>90.73</td>
<td>36</td>
<td>94.45</td>
<td>51</td>
<td>87.69</td>
</tr>
<tr>
<td>7</td>
<td>96.84</td>
<td>22</td>
<td>91.40</td>
<td>37</td>
<td>88.49</td>
<td>52</td>
<td>95.45</td>
</tr>
<tr>
<td>8</td>
<td>91.44</td>
<td>23</td>
<td>93.27</td>
<td>38</td>
<td>86.98</td>
<td>53</td>
<td>92.58</td>
</tr>
<tr>
<td>9</td>
<td>87.46</td>
<td>24</td>
<td>96.03</td>
<td>39</td>
<td>90.21</td>
<td>54</td>
<td>89.33</td>
</tr>
<tr>
<td>10</td>
<td>91.64</td>
<td>25</td>
<td>95.60</td>
<td>40</td>
<td>90.91</td>
<td>55</td>
<td>87.33</td>
</tr>
<tr>
<td>11</td>
<td>93.09</td>
<td>26</td>
<td>90.50</td>
<td>41</td>
<td>88.53</td>
<td>56</td>
<td>90.04</td>
</tr>
<tr>
<td>12</td>
<td>87.31</td>
<td>27</td>
<td>86.01</td>
<td>42</td>
<td>87.09</td>
<td>57</td>
<td>93.26</td>
</tr>
<tr>
<td>13</td>
<td>88.93</td>
<td>28</td>
<td>92.14</td>
<td>43</td>
<td>87.48</td>
<td>58</td>
<td>94.58</td>
</tr>
<tr>
<td>14</td>
<td>88.67</td>
<td>29</td>
<td>87.45</td>
<td>44</td>
<td>93.08</td>
<td>59</td>
<td>94.66</td>
</tr>
</tbody>
</table>

Figure 8: Statistical diagram of evaluation of data mining effect of the smart tourism system.

Table 2: Statistical table of user experience satisfaction of the smart tourism system.

<table>
<thead>
<tr>
<th>No.</th>
<th>Satisfaction</th>
<th>No.</th>
<th>Satisfaction</th>
<th>No.</th>
<th>Satisfaction</th>
<th>No.</th>
<th>Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>87.50</td>
<td>16</td>
<td>88.72</td>
<td>31</td>
<td>91.63</td>
<td>46</td>
<td>88.09</td>
</tr>
<tr>
<td>2</td>
<td>86.60</td>
<td>17</td>
<td>92.64</td>
<td>32</td>
<td>92.12</td>
<td>47</td>
<td>87.42</td>
</tr>
<tr>
<td>3</td>
<td>93.91</td>
<td>18</td>
<td>85.89</td>
<td>33</td>
<td>89.89</td>
<td>48</td>
<td>91.70</td>
</tr>
<tr>
<td>4</td>
<td>91.16</td>
<td>19</td>
<td>85.49</td>
<td>34</td>
<td>89.96</td>
<td>49</td>
<td>85.86</td>
</tr>
<tr>
<td>5</td>
<td>92.00</td>
<td>20</td>
<td>89.50</td>
<td>35</td>
<td>84.11</td>
<td>50</td>
<td>85.28</td>
</tr>
<tr>
<td>6</td>
<td>91.77</td>
<td>21</td>
<td>87.06</td>
<td>36</td>
<td>88.75</td>
<td>51</td>
<td>93.96</td>
</tr>
<tr>
<td>7</td>
<td>87.67</td>
<td>22</td>
<td>84.36</td>
<td>37</td>
<td>85.98</td>
<td>52</td>
<td>91.04</td>
</tr>
<tr>
<td>8</td>
<td>85.65</td>
<td>23</td>
<td>93.72</td>
<td>38</td>
<td>91.44</td>
<td>53</td>
<td>92.12</td>
</tr>
<tr>
<td>9</td>
<td>85.54</td>
<td>24</td>
<td>86.07</td>
<td>39</td>
<td>89.80</td>
<td>54</td>
<td>92.12</td>
</tr>
<tr>
<td>10</td>
<td>93.88</td>
<td>25</td>
<td>93.13</td>
<td>40</td>
<td>91.45</td>
<td>55</td>
<td>90.94</td>
</tr>
<tr>
<td>11</td>
<td>85.81</td>
<td>26</td>
<td>91.94</td>
<td>41</td>
<td>89.82</td>
<td>56</td>
<td>89.64</td>
</tr>
<tr>
<td>12</td>
<td>85.79</td>
<td>27</td>
<td>92.71</td>
<td>42</td>
<td>88.10</td>
<td>57</td>
<td>84.88</td>
</tr>
<tr>
<td>13</td>
<td>88.05</td>
<td>28</td>
<td>89.69</td>
<td>43</td>
<td>90.86</td>
<td>58</td>
<td>89.89</td>
</tr>
<tr>
<td>14</td>
<td>88.60</td>
<td>29</td>
<td>86.25</td>
<td>44</td>
<td>87.90</td>
<td>59</td>
<td>86.08</td>
</tr>
<tr>
<td>15</td>
<td>91.54</td>
<td>30</td>
<td>91.76</td>
<td>45</td>
<td>91.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. Conclusion

In this paper, we propose a new tourism system model using big data technology and create the large-scale tourist ability mining model. We provide the new paradigm examination of the management for tourism development. In terms of data, traveling firms must rely on big data technologies if they are to become better. Tourism is an information-intensive sector, with information resources depending on the operation and administration of companies. Therefore, an unanticipated increase for companies by using big data technologies to investigate traveling data. Finally, the experimental investigation also shows that the system built in this work has some effect.

Data Availability

The data used to support the findings of the study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This study was supported by Lanzhou University of Arts and Sciences Distinguished Young Scientific Research Talent Cultivation Project (2018JCQN009) and Talent Innovation and Entrepreneurship Project of Lanzhou City (2018-RC-68).

References


[12] F. Femenia-Serra, B. Neuhofer, and J. A. Ivars-Baidal, “Towards a conceptualisation of smart tourists and their role...


