www.cadanda.com Computer-Aided Design AND APPLICATIONS

Intelligent Tourism Development in Ethnic Art Areas Based on Embedded Data Mining Technology

Li Zhang[1*]

Shanxi Finance & Taxation College, Shanxi, Taiyuan, 030024, China

Corresponding author: Li Zhang, ZLL5685681@163.com

Abstract. In order to improve the development effect of intelligent tourism in ethnic areas, this paper combines data mining technology to construct an intelligent tourism development system in ethnic areas. Moreover, this paper uses a formula to illustrate the direct and indirect relationship of users from the perspective of network topology.When calculating the strength of social interaction, this paper considers the reciprocity and commonality of social behaviors, and calculates the magnitude of perceived interaction strength from both sides of the interaction. In addition, this paper introduces time slice and time decay factor to dynamically model the intensity of social interaction, and constructs an intelligent tourism development system in ethnic areas based on data mining technology.The experimental analysis results show that the intelligent tourism development system in ethnic areas based on data mining proposed in this paper can play an important role in the development of intelligent tourism in ethnic areas.

Keywords: data mining; ethnic areas; intelligence; tourism development
DOI: https://doi.org/10.14733/cadaps.2024.S8.13-30

1 INTRODUCTION

Tourism is a typical information-intensive industry, and information technology is of great significance to tourism. From traditional tourism (Tourism) to electronic tourism (e-Tourism) relying on electronic information technology, and then to smart tourism (Smart-Tourism) relying on the integration of technology, the Internet and information technology continue to promote the progress and development of tourism. The essence of smart tourism is the application of intelligent technology mainly based on information and communication technology in the tourism industry.

Intelligent tourism technology has been widely used in the tourism industry. It not only runs through the entire process of tourism activities, but also involves various industries in the tourism industry chain. With the application of the Internet in tourism and the continuous development of intelligent tourism technology, the impact of the use of intelligent tourism technology on the overall tourism experience and satisfaction of tourists cannot be ignored. Through the use of intelligent tourism technology, tourists can obtain the food, accommodation, travel, shopping and entertainment information they need during their travels, conduct transaction activities such as ticket booking, room reservation, and meal ordering, and experience the convenience brought by smart devices. As the main body of tourism activities, the sense of experience and satisfaction generated by tourists in the perception and use of intelligent tourism technology is very important.

The intelligent tourism system in ethnic areas under big data is a new type of system that requires high-end technical support. The system relies on big data to accurately grasp the needs of tourists, adjust and output service products in a timely manner during the service process, quickly respond to tourists' evaluations after the trip, and further strengthen tourism destination marketing [19]. To sum up its operation process, the following characteristics can be summarized. First, the technology is highly dependent, and the initial investment and maintenance costs are large. To realize the application of big data, it is necessary to rely on high-end technical support, such as real-time monitoring of passenger flow in scenic spots, precise positioning, real-time transmission of information, etc., so that complex functions can be completed. The second is that the provider of intelligent tourism in ethnic areas is also the served [13].For example, tourism administrative departments provide scenic spot security services in conjunction with traffic management departments, and they also rely on passenger flow monitoring to issue safety warnings and divert tourists. Third, big data puts forward higher requirements for the security service part of the intelligent tourism system in ethnic areas. Public services under big data are different from traditional public services. From data mining, analysis, to online and offline resource integration and distribution, it all involves the information security of public service providers and tourists. This puts forward higher requirements for the confidentiality qualification of service providers [21].

This paper combines data mining technology to construct an intelligent tourism development system in ethnic areas, enhance the intelligent development of tourism in ethnic areas, and improve the development effect of tourism in ethnic areas.

2 RELATED WORK

Literature [10] expounds the application of tourism big data in smart tourism through the analysis of enterprises, governments and tourists. First of all, in the six categories of food, housing, travel, shopping and entertainment, suppliers that can make full use of big data include shops, airlines, travel agencies, and scenic spots. On the one hand, it can realize the informatization of the internal management of enterprises and further implement the supervision system. On the other hand, based on the cooperation with major search engines and OTAs, it can grasp the needs of tourists in time and provide a basis for formulating marketing strategies. Literature [3] forms tourism strategies and survey reports through classification and sorting, and provides more accurate product positioning and marketing strategies for the creation of tourism products. Secondly, the use of big data by the government can not only realize the macro management of the industry, but also grasp the initiative of public services. For the government itself, it can realize cross-department, cross-industry, cross-regional resource integration and sharing, strengthen the horizontal cooperation of various departments, establish a tourism big data exchange platform, and finally form a data exchange and sharing mechanism [8]. Only when the government improves its macro-leading ability and relies on big data technology to improve public opinion monitoring and dynamic analysis can it provide tourists and tourism companies with comprehensive and effective services. Thirdly, tourists are consumers of tourism products and the core value of smart tourism. The key target of tourism big data mining is tourists. By analyzing their consumption behavior, we can formulate more popular and distinctive tourism products, which can not only provide tourists with a new experience of rice, but also further stimulate consumption. Through evaluation and feedback, word-of-mouth marketing of tourist attractions can be better carried out [2].

In the tourism industry, IoT ZigBee technology is mainly used in intelligent tour guides and tourist positioning. Reference [9], on the basis of studying ZigBee technical specifications, network topology, positioning technology (two positioning technologies based on ranging and distance-independent), routing performance of AODV algorithm, scenic spot division and tourist positioning principle, etc. The ZigBee RF module of CC2430 builds a hardware platform, and uses IAR development tools to compile software programs for nodes such as communities, tourists and tour guides, and realizes a positioning system for tourists in scenic spots. Reference [7] studies ZigBee technical specifications, routing protocols, and other wireless Based on the comparison of technologies (Bluetooth, UWB, WiFiand infrared), the scheme design and working principle of the ZigBee-based audio guide system, etc., an improved energy-balanced routing algorithm and elimination of voice over the existing ZigBee-based audio guide system are proposed. Jitter buffer delay algorithm, compared and studied three kinds of voice compression coding technology, such as waveform coding, parameter coding and hybrid coding, and also proposed to use the ZigBee RF module of JN5139 and the audio acquisition module of WM8510 to build a hardware platform, and use the unique JN- The SW-4031-SDK-Toolchain-v1.1.exe development toolchain realizes ZigBee networking communication, audio acquisition driver, ADPCM voice compression coding, buffer delay and other software functions, and points out the improvement of the system through functional and performance tests The communication distance is improved, and the communication flexibility is improved. Literature [1] proposes a solution based on the study of ZigBee technical characteristics, network topology, protocol architecture, network establishment process, network method of tour guide system, and scheme design of intelligent tour guide system. Use the ZigBee RF module of CC2430, the MP4 terminal node module of JZ4755, the RJ45 module of NePort, etc. to build the hardware platform, use the development platforms such as IAR, RD4755CETUS, etc. to design the software program of MP4 terminal node, routing node and gateway node, use MFC language, VC++ to develop Environment, PHP language, SQL Server database and Apache to realize the software design of server-side and monitoring webpage. Literature [12] proposed a new weighting algorithm based on RSSI value to solve the problem of inaccurate positioning caused by inaccurate measurement of RSSI value. Finally, an intelligent multimedia tour guide system with functions such as precise positioning through the MP4 terminal, timely playback of the video of the routing node, timely receipt of the alarm information from the MP4 terminal, and determination of the alarm location at the monitoring terminal is realized during the tour.

The characteristics of tourism resources are: diversity, uniqueness, variability and sustainability. Diversity is manifested in the types and characteristics of resources; including natural and humanistic; landscape and cultural; both ancient and modern [5]; both physical and observable. The uniqueness is shown in that each tourism resource is different from the usual customs, cultural background and living environment of tourists. The greater the difference, the more unique and attractive it will be for tourists [6]. Variability is manifested in the fact that some things were not regarded as tourism resources at first, but with the change of tourists' needs and the influence of the surrounding environment, it becomes an attractive tourism resource [20]. In the process of sustainable development, most of the tourism resources have the characteristics of infinite reuse and regeneration. As the main tourism resources such as sightseeing, vacation and special tourism resources, the tourists themselves cannot take them away, they only take away their various Impressions and Feelings [11]. As long as it is properly protected, tourism resources can be reused for a long time, and a new round of development based on the previous development can still be treated as a resource [17].

3 INTELLIGENT TOURISM ALGORITHM BASED ON DATA MINING

Since interaction data can obtain objective data from social networks, user relationship strength is a potential data that cannot be directly obtained. Therefore, when building the model, the relationship strength is regarded as a latent variable, and the interaction strength is regarded as a dependent variable. The specific latent variable regression model is as follows:

$$I = \beta_0 + \beta_1 R + \varepsilon \tag{1}$$

For the measurement of interaction strength, the method considers three social behaviors of commenting, forwarding, @, and directional factors, and the calculation formula is defined as follows:

$$I_{pq} = \frac{f_{pq} \times d_{pq}}{f_{pq} + d_{pq}} \times \lg\left(\frac{i_{pq}}{\lg(avg)} \right) \tag{2}$$

In order to avoid the phenomenon that the number of user interactions is 0 affecting the accuracy of the model, the interaction strength is smoothed, which is defined as d_{pq}, as follows[17]:

$$d_{pq} = \frac{i_{\text{total}}}{2n \times (n-1)} \tag{3}$$

Three types of characteristic indicators of users of the intelligent tourism system are selected: the user characteristic index $U_i\left(U_{t1}, U_{i2}, \ldots, U_{il}\right)$ represents the forwarding volume of the intelligent tourism system of user i. Intelligent tourism system feature $T_j\left(T_{j1}, T_{j2}, T_{j3}\right)$ represents intelligent tourism system j forwarding, collection, @ quantity. When selecting the network topology index, five variables are mainly considered as whether the nodes belong to the same community, the aggregation degree of the community, the degree centrality, the average distance close centrality and the common neighbor nodes as the user association feature indicators, which are denoted by $R_{y_y} = \left(r_{y_1}, r_{y_2}, r_{y_3}, r_{y_4}, r_{y_5}\right)$, and the specific quantification method is as follows[18]:

$$r_{ij_1} = \begin{cases} 1, i, j \text{ Belong to the same community} \\ 0, i, j \text{ Do not belong to the same community} \end{cases} \tag{4}$$

$$r_{ij_2} = \frac{\sum_{k}^{n}\left(c_{\max^j} - c_k^j\right)}{\max\left[\sum_{k}^{n}\left(c_{\max^j} - c_k^j\right)\right]} \tag{5}$$

Among them, $c_{\max^j}$ is the maximum centrality of the community to which node j belongs, and c_k^j refers to the degree centrality of the k-th node of the community to which node j belongs.

$$r_{ij_3} = \frac{\sum_{k}^{n} a\left(p_j, p_k\right)}{\mathrm{n}-1} \tag{6}$$

Among them, $a\left(p_j, p_k\right)\begin{cases}1 \text{ Node K is the adjacent node of node j} \\ 0 \text{ Node K is not an adjacent node of node j}\end{cases}$.

$$r_{ij_4} = \left[\frac{\sum_{k}^{n} d\left(p_j, p_k\right)}{n-1}\right]^{-1} \tag{7}$$

Among them, $d\left(p_j, p_k\right)$ is the shortest path between node j and another community node k[4].

$$r_{ij_5} = \left|NB_i \cap NB_j\right| \tag{8}$$

Among them, NB_i and NB_j respectively represent the set of neighbor nodes between node i and node j.

This method selects the typical classification algorithm logistic regression in the linear classification method to construct the classifier, and uses the example of whether the forwarding behavior of the intelligent tourism system occurs or not to quantify the influence relationship between users, and the quantification method is as follows:

$$RT_{ij} = \begin{cases}1 & \text{User I forwarded microblog J} \\ 0 & \text{User I did not forward microblog J}\end{cases} \tag{9}$$

The hypothesis function $h\left(U_i, T_j, R_{ij}\right)$ is constructed to represent the probability that the forwarding behavior RT_{ij} is 1, then the conditional dependencies when $RT_{ij}=1$ and $RT_{tj}=0$ are[16]:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{n}\sum_{j=1}^{m} RT_{ij} \times \log\left(h_\theta\left(U_i, T_j, R_{ij}\right)\right) + \left(1-RT_{tj}\right)\times\log\left(1-h_\theta\left(U_i, T_j, R_{iy}\right)\right)\right]$$
$$p\left(RT_y = 0 \middle| \; U_1, T_j, R_{yy}\right) = 1 - h_\theta\left(U_t, T_j, R_y\right) \tag{10}$$

$$\theta_m \leftarrow \theta_m - \alpha\frac{\partial}{\partial U} J\left(\theta_m\right) \tag{11}$$

For the value of θ, the model adopts the stochastic gradient descent method to solve, and the specific iterative update is[14]:

$$\theta_m \leftarrow \theta_m - \alpha \frac{\partial}{\partial U} J\left(\theta_m\right) \tag{12}$$

Among them, θ_m is the weight of each feature index component, $J(\theta)$ is the loss function, and the calculation formula is:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{n}\sum_{j=1}^{m} RT_{ij} \times \log\left(h_\theta\left(U_1, T_j, R_{\eta j}\right)\right) + \left(1 - RT_{yj}\right) \times \log\left(1 - h_\theta\left(U_i, T_j, R_{\eta j}\right)\right)\right] \tag{13}$$

A probabilistic graphical model was established, as shown in figure 1. In the directed graph, $x^{(i)}$ and $x^{(j)}$ are used to represent the attribute vector of user i and user j, $y_t^{(y)}$ is used to represent the interaction between user i and user j at time $t = 1, 2, 3, \ldots, m$, and $z^{(y)}$ is used to represent the relationship strength between user i and user j , the direction of the arrow is used to indicate the direction of influence.The upper part is the discriminant that can be represented by $(p(z/x))$, and the lower part is the generative expression, which can be represented by $(p(y,z))$, and the joint distribution is:

$$p\left(z^{(ij)}, y^{(i)} / x^{(i)}, x^{(J)}\right) = p\left(z^{(ij)} / x^{(i)}, x^{(j)}\right)\prod_{t=1}^{m} p\left(y_t^{(jj)} / z^{(ij)}\right) \tag{14}$$

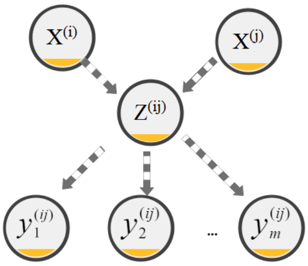


Figure 1: Probabilistic Graphical Model.

The similarity in this model is denoted by $s_k\left(x^{(1)}, x^{(j)}\right)(k = 1, 2, 3, \ldots, n)$, and the classical Gaussian distribution is used to represent the dependency between the strength of the relationship and the similarity, where w is an n-dimensional weight vector, v is the variance of the Gaussian model as follows:

$$p\left(z^{(jj)} \mid x^{(i)}, x^{(j)}\right) = \mathrm{N}\left(w^T s\left(x^{(l)}, x^{(j)}\right), v\right) \tag{15}$$

As shown in figure 2, for the lower part of the modified model, in order to improve the accuracy of the model, a set of user auxiliary variables $a_{t1}^{(v)}, a_{t2}^{(v)}, \ldots, a_{t_i}^{(v)}$ are introduced to further illustrate the user interaction at time T, and the logistic function is used to represent the relationship between the strength of the relationship and the interaction behavior. The calculation is as follows:

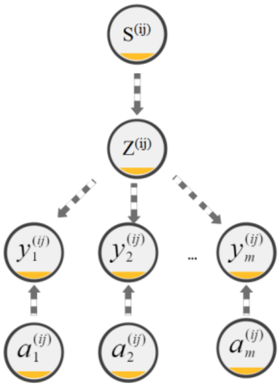


Figure 2: Description of the Lower Part of the Probability Graph Model.

$$p\left(y_t^{(ij)} = 1 / z^{(ij)}, a_t^{(ij)}\right) = \frac{1}{1 + e^{-\left(\theta_{11} a_{11}^{(v)} + \theta_{12} a_{t2}^{(v)} + \ldots + \theta l a_{1l}^{(v)} + \theta l_{l+1} z^{(y)} + b\right)}} \tag{16}$$

In the experiment, $D = \{(i_1, j_1), (i_2, j_2), \ldots, (i_n, j_n)\}$ is used to represent the sample data in N groups of interactive objects, and the three parameters $x^{(v)}$, $y^{(v)}$, and $a_t^{(v)}$ can obtain objective data. In summary, the joint probability of this model is:

$$\begin{aligned} & p(D / w, \theta) p(w, \theta) \\ & = \left(\prod_{(i,j) \in D} p\left(z^{(v)}, y^{(v)} / x^{(i)}, x^{(J)}, w, \theta\right) \right) p(w) p(\theta) \\ & = \prod_{(t,J) \in D} p\left(z^{(v)} / x^{(t)}, x^{(J)}, w\right) \prod_{t=1}^{m} p\left(y_t^{(v)} / z^{(v)}, \theta_t\right) p(w) p\left(\theta_t\right) \\ & \propto \prod_{(t,j) \in D} \left(e^{\frac{1}{2v}\left(w^T s^{(v)} - z^{(o)}\right)^2} \prod_{t=1}^{m} \frac{e^{-\left(\theta_t^T u^{(v)} + b\right)\left(1 - y_t^{(v)}\right)}}{1 + e^{-\left(\theta_t^T u_t^{(j)} + b\right)}} \right) \cdot e^{-\frac{\lambda_w}{2} w^T w} \prod_{t=1}^{m} e^{-\frac{\lambda_\theta}{2} \theta_t^T \theta_t} \end{aligned} \tag{17}$$

In order to effectively illustrate the research task of this paper, the relevant concepts are first defined.

This definition 1 is: In this paper, the social network of the intelligent tourism system is defined as a set of tuples, $G=\{V,E,UA,UI,RS\}$.

Based on the above definition, the problem studied in this paper can be instantiated. That is, given a social network user set information, we predict the strength of the last known user relationship by connecting the edge set information, the user feature attribute set information, the social interaction behavior set information and the corresponding labeled user relationship strength, as shown in the formula (18).

$$f:G=\left(V,E,UA,UI,RS^{l}\right)\rightarrow RS^{u} \tag{18}$$

This definition 2 is a seed user. In order to simplify the experiment, 20 users of the intelligent travel system were selected as the main subjects of the study, and the strength of their relationship with 20 randomly selected interaction objects (including 10 fans and 10 followers) was calculated.

This definition 3 is a concern relation. We assume that user U pays attention to user V, that is, $U\rightarrow V$, then user U is the analysis of user V, and user V is the attention object of user U. There are four possible attention relationships among users, as shown in figure 3.

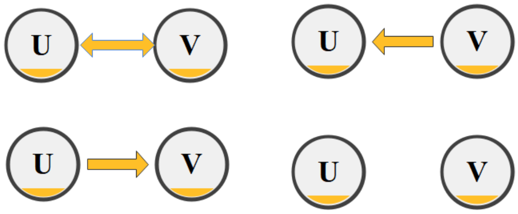


Figure 3: Schematic Diagram of Attention Relationship.

This paper uses the number of users' fans, the number of ratings, the number of original smart travel systems, the average forwarding, average comments, and average likes of each smart travel system to measure the user's node influence, which is defined as $UE=\left(u_{\text{fan},\, u\_\text{exp},\, u\_\text{rep},\, u\_\text{ret},\, u\_\text{com},}\ u\_lik\right)$. Since Mahalanobis distance is not affected by dimension and excludes the interference of variable correlation, this paper uses this formula to calculate the difference of node influence between users. Among them, $\text{sim}_{ue}(u,v)$ represents the user's influence similarity, and S^{1} is the inverse of the covariance matrix, which is calculated as follows:

$$\text{sim}_{ue}(u,v)=\sqrt{(\overrightarrow{UE(u)}-\overrightarrow{UE(v)})^{T}S^{-1}(\overrightarrow{UE(u)}-\overrightarrow{UE(v)})} \tag{19}$$

After calculating the weight of the feature words by using the TFIDF formula, the final short text feature words of the intelligent tourism system are selected, which is defined as

$UT(u) = (w_0, w_1, w_2,, w_n)$. In this paper, the classical cosine formula is used to calculate the short text similarity $\mathrm{sim}_{uu}(u,v)$ of the user intelligent travel system, as shown in formula (20).

$$\mathrm{sim}_{ut}(u,v) = \frac{UT(u) \cdot UT(v)}{\| UT(u) \| \times \| UT(v) \|} \tag{20}$$

To sum up, in this paper, $\mathrm{sim}_{us}(u,v)$ is used to represent the similarity of user network state attributes, and the calculation formula is as follows:

$$\mathrm{sim}_{us}(u,v) = \mathrm{sim}_{ue}(u,v) + \mathrm{sim}_{ut}(u,v) \tag{21}$$

This paper uses the Jaccard formula to calculate the similarity of users' personal background data, as shown in formula (22):

$$\mathrm{sim}_{ub}(u,v) = \frac{|UB(u) \cap UB(v)|}{|UB(u) \cup UB(v)|} \tag{22}$$

The similarity of user feature attributes is calculated as:

$$\mathrm{SIM}(u,v) = \lambda_1 \times \mathrm{sim}_{ub}(u,v) + \lambda_2 \times \mathrm{sim}_{us}(u,v) \tag{23}$$

Since the intelligent tourism system is a complex network with directionality, there are four cases of common neighbor nodes in the intelligent tourism system, as shown in figure 4.

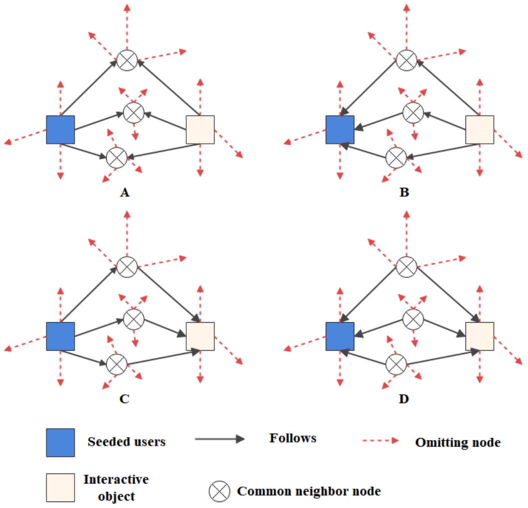


Figure 4: User Common Neighbor Node Graph.

As shown in figure 5, the circle in the upper left position represents the followers of user U, the circle in the lower left position represents the following objects of user U, the circle in the upper right position represents the fans of user V, and the circle in the lower right position represents the attention of user V. Area A is the intersection of user U's attention and user V's attention, B area is the intersection of user U's fans and user V's fans, and C area is the intersection of user U's attention, the attention of users and user V, and the intersection of fans. In this paper, $\mathrm{cnp}(u,v)$ is used to represent the proportion of common neighbor nodes among users, and the calculation is as follows:

$$\mathrm{cnp}(u,v)=\frac{|\varphi(u)\bigcap\varphi(v)|+|\omega(u)\bigcap\omega(v)|-|\omega(u)\bigcap\varphi(v)\bigcap\varphi(u)\bigcap\omega(v)|}{|\varphi(u)\bigcup\varphi(v)\bigcup\omega(u)\bigcup\omega(v)|} \tag{24}$$

Among them, $\varphi(u)$ represents the follower list of seed user U, $\varphi(v)$ represents the follower list of interactive object V, $\omega(v)$ represents the follower list of interactive object V, and $\omega(u)$ represents the follower list of seed user U.

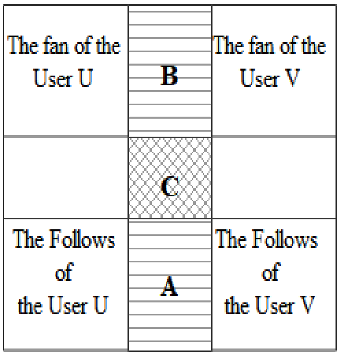


Figure 5: Schematic Diagram of Calculation of Common Neighbor Nodes.

In order to facilitate the calculation and description, this paper defines the tightness of the direct connection between nodes as $dlt(u,v)$, and the calculation is as follows:

$$\mathrm{dlt}(u,v)=cnp(u,v)\times cnl(u,v) \tag{25}$$

Among them, $cnp(u,v)$ is the number of common neighbor nodes, and $cnl(u,v)$ is the number of connecting edges between common neighbor nodes.

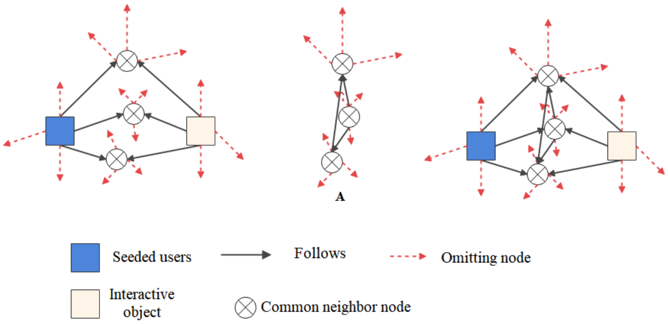


Figure 6: Schematic Diagram of User Common Neighbor Nodes Connecting Edges.

This paper uses Katz centrality to measure the magnitude of this indirect relationship. The Katz centrality method calculates the sum of all possible paths between nodes, and assigns a larger weight to the short path, which is calculated as follows:

$$\mathrm{ilt}_{\mathrm{Katz}}(u,v)=\sum_{l=1}^{\infty}\beta^{l}\cdot\left|p_{u,v}^{l}\right| \tag{26}$$

Among them, $ilt_{Kazz}(u,v)$ is the degree of indirect connection between users, $p_{u,v}^{l}$ is the set of all paths of length l between nodes U and V, and $\beta(0<\beta<1)$ is the attenuation coefficient, which is taken as 0.0005 in this paper with reference to the literature. According to the principle of third degree of influence, a node is considered to have influence not only on its neighbors (first degree) and neighboring nodes (second degree), but also even on neighboring nodes of neighboring nodes (third degree), while for nodes beyond this range, it is considered to have almost no influence. Therefore, in order to simplify the calculation and experimental process, this paper only considers the case of $l\le 3$.

In summary, the network structure connection strength is calculated as:

$$LS(u,v)=\lambda_3\times dlt(u,v)+\lambda_4\times ilt(u,v) \tag{27}$$

For the study of social reciprocity behavior, this paper considers four situations such as likes, reposts, comments, and mentions among users, which are defined as $UI=\{u\_b_1,u\_b_2,u\_b_3,u\_b_4\}$.

In order to describe the interaction of users in different stages, this paper divides the interaction behavior into time slices, as shown in figure 7. The black dots in the figure represent the seed user U, the white dots represent the interactive objects, and the black lines and numbers represent the interactive behaviors and the corresponding times between the user U and the interactive objects in the direction of the arrow during time T. Therefore, the interaction strength among social network users changes dynamically, and the time slice factor needs to be considered in the process of calculating the reciprocity of social behaviors.

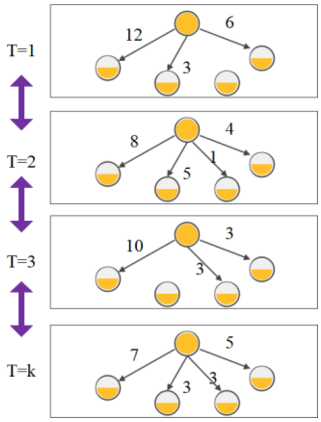


Figure 7: Schematic Diagram of Dynamic Changes in Social Behavior.

First, when T=k, the algorithm first perceives the magnitude of the interaction intensity from the perspective of user U, and calculates the degree of user V's response to a certain social behavior b_I of user U, which is denoted by $resp_{u\leftarrow v,b_1}^{t_k}(u,v)$, the direction is $u \leftarrow v$, and the calculation is as follows :

$$\mathrm{resp}_{u\leftarrow v,b_t}^{t_k}(u,v) = \frac{n_{u\leftarrow -v,b_i}^{t_k}}{n_{v\leftarrow -u,b_t}^{t_k}+1} \tag{28}$$

Among them, $n_{u\leftarrow v,b_1}^{t_k}$ represents the number of times that user V responds to a certain social behavior b_l to user U, $n_{v\leftarrow u,b_t}^{t_k}$ represents the number of times the interaction object U sends out the behavior to user V, and the denominator is increased by 1 to avoid the phenomenon that the number of interactions is 0. If $resp_{u\leftarrow v,b_1}^{t_k}(u,v) > 1$, it means that in this relationship, user V shows higher initiative and enthusiasm in social behavior b_1 than user U, and contributes more to the formation and maintenance of the relationship. If $0 < \mathrm{resp}_{u\leftarrow -v,b_1}^{t_k}(u,v) < 1$, it means that user V shows lower initiative and enthusiasm in social behavior b_l than its interacting object U, and contributes less to the formation and maintenance of relationship. If $resp_{u\leftarrow -v,b_1}^{t_k}(u,v) = 1$, it means that user V and its interacting object U show the same initiative and enthusiasm in social behavior b_1, and they are in the same position in the process of relationship formation and maintenance.

Secondly, considering the overall response degree of the user V to all interacting objects in the social network, $resp^{t_1}(v)$ is defined, and it is calculated as follows:

$$\mathrm{resp}^{t_k}(v) = \frac{n_{v\rightarrow,b_t}^{t_k}}{n_{v\leftarrow,b_t}^{t_k}+1} \tag{29}$$

Among them, $n_{v\rightarrow,b_1}^{t_k}$ represents the number of social behaviors b_I generated by user V to all interactive objects in the time slice T=k, and $n_{v\leftarrow,b_1}^{t_k}$ represents the number of times that user V receives the behavior. The operation of adding 1 is to avoid the phenomenon that the denominator is 0. If $resp^{t_k}(v) > 1$, it means that user V shows high responsiveness in the social process as a whole, and he is a user with strong initiative and altruism. If $resp^{t_k}(v) < 1$, it means that the overall response of user V is low, and he is in a passive state in the formation of the relationship.

Reciprocity behaviors in social networks usually manifest as users' friendly responses to their interacting objects.

Therefore, after comprehensively considering the response degree of user V to user U and the overall response degree of user V in the social network, at time T=k, the reciprocity size $re_{u\leftarrow r}^{t_k}$ of user V to user U is:

$$\mathrm{rec}_{u\leftarrow v}^{t_k} = \frac{\mathrm{resp}_{u\leftarrow v}^{t_k}(u,v)}{\mathrm{resp}^{t_k}(v)+1} \tag{30}$$

The operation of adding 1 to the denominator on the right side of the formula is to prevent the phenomenon that the overall response of the user is 0. At this time, if $rec_{u\leftarrow v}^{t_i} > 1$, it means that the degree of response of user V to user U is higher than the overall level of user V in the social network, and user V attaches great importance to this relationship. This is conducive to the maintenance of the relationship between users U and V. If $0 < rec_{u\leftarrow r}^{t_k} < 1$, it means that the degree of response of user V to user U is lower than the overall level of user V in the social network, and user V attaches less importance to this relationship than other relationships. This will affect U's perception of relationship strength, and it is also not conducive to the relationship maintenance between users U and V.

In this paper, we use the inverse interactive object frequency $IPF_{v,b_1}^{t_k}$ of the user V's behavior b_t to the interactive object to measure whether the user's high reciprocity behavior is common at time T=k, as shown in formula (31):

$$IPF_{v,b_s}^{t_k} = \log\left(\frac{|p|}{\left|\{j : b_1 \in p_J\}\right|+1}\right) \tag{31}$$

Among them, |p| refers to the total number of people who have social reciprocity behaviors b_1 with user V at time T=k. $\left|\{j : b_1 \in p_J\}\right|$ is the number of people who have social reciprocity behavior b_1 with user V at this moment and the reciprocity size is greater than 1, and 1 is added to prevent the denominator from being 0 interfering with the experimental results.

Considering that the interaction intensity perceived by users is a historical cumulative value, and there will be a short-term smoothing phenomenon, a time decay factor is introduced when calculating the interaction intensity of social behaviors. At time T=k+1, if $N_{u\leftarrow v,b_1}$ is used to represent the number of times that user V actively performs a certain behavior b_i to user U, after comprehensively considering the degree of reciprocity of social behaviors, the commonness of reciprocal behaviors, and the frequency of social behaviors, the intensity of social interaction among users is as follows:

$$BS_{v\leftarrow u}^{T_{k+1}}(v,u) = \sum_{i=1}^{n} rec_{v\leftarrow u}^{t_{k+1}} \times IPF_{v\leftarrow u,b_t}^{k+1} \times N_{v\leftarrow u,b_t}^{k+1} + BS_{v\leftarrow u}^{T_k}(v,u) \times \lambda e^{-\lambda t} \tag{32}$$

This paper adopts a three-layer stage model to build a relationship strength calculation model, including a social network user data acquisition and analysis module, a strength calculation module, and a data mining module, as shown in figure 8.

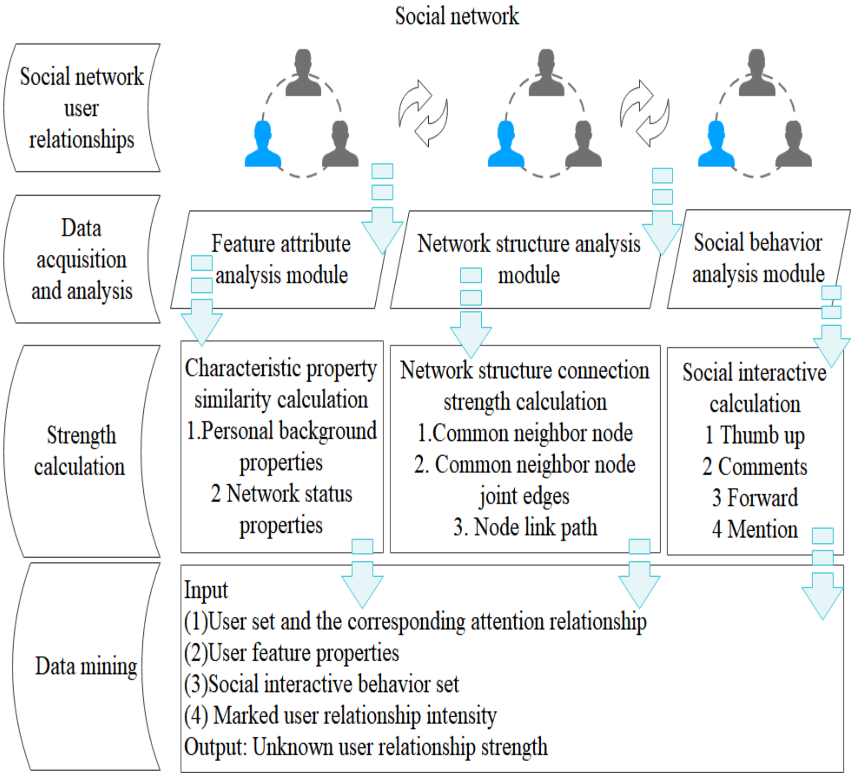


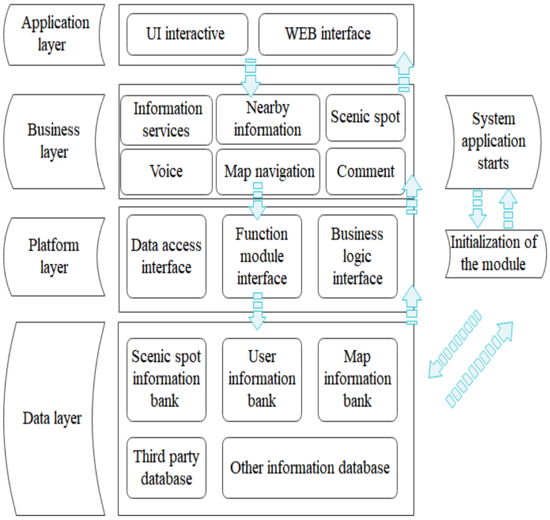
Figure 8: Research Model of User Directed Relationship Strength That Integrates Feature Attributes, Network Structure and Social Behavior.

After fully considering the influence of user feature attribute similarity, network structure connection strength, and social behavior interaction strength on relationship strength, this paper integrates these three dimensions and proposes a user relationship strength calculation model, as shown in figure 8. The formula for calculating the strength of the relationship is as follows, where $\alpha+\beta+\gamma=1, 0\le\alpha\le 1, 0\le\beta\le 1, 0\le\gamma\le 1$.

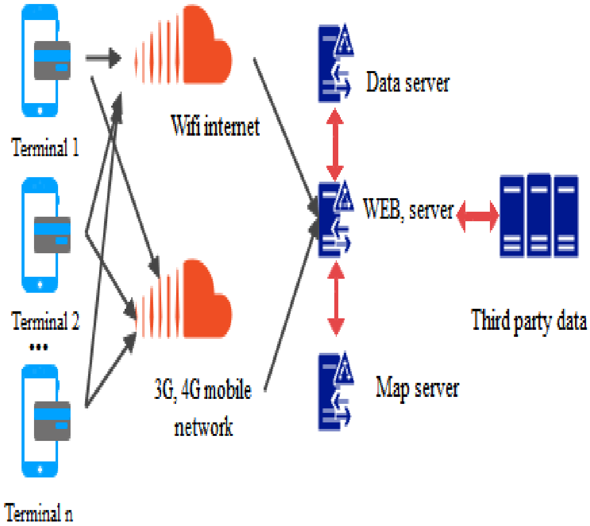
$$TS=\begin{cases}\alpha\times \mathrm{SIM}(u,v)+\beta\times LS(u,v)+\gamma\times BS_{u\leftarrow v}(u,v)\\ \alpha\times \mathrm{SIM}(u,v)+\beta\times LS(u,v)+\gamma\times BS_{v\leftarrow u}(v,u)\end{cases} \tag{33}$$

4 INTELLIGENT TOURISM DEVELOPMENT IN ETHNIC AREAS BASED ON DATA MINING TECHNOLOGY

When designing the logical structure of the intelligent tourism system in ethnic areas based on data mining technology, taking into account the characteristics of the Android platform used, the hierarchical design structure is adopted. The intelligent tourism system in ethnic areas based on data mining technology can be divided into four layers as a whole, and the logical structure of the system is shown in Figure 9(a).



(a) Logical Structure Diagram of Intelligent Tourism System in Ethnic Areas Based on Data Mining Technology



(b) Topological Structure of Intelligent Tourism System in Ethnic Areas Based on Data Mining Technology

Figure 9: Structure Diagram of Intelligent Tourism System in Ethnic Areas Based on Data Mining Technology.

When designing the system topology of the intelligent tourism system in ethnic areas based on data mining technology, the user's usage scenarios and usage needs are fully considered. Users are mainly obtained through mobile terminals, which requires the support of wireless networks (including mobile networks and the Internet), and the system topology is shown in Figure 9(b). Figure 10 shows an example of the intelligent scenery online recommendation of the system in this paper.

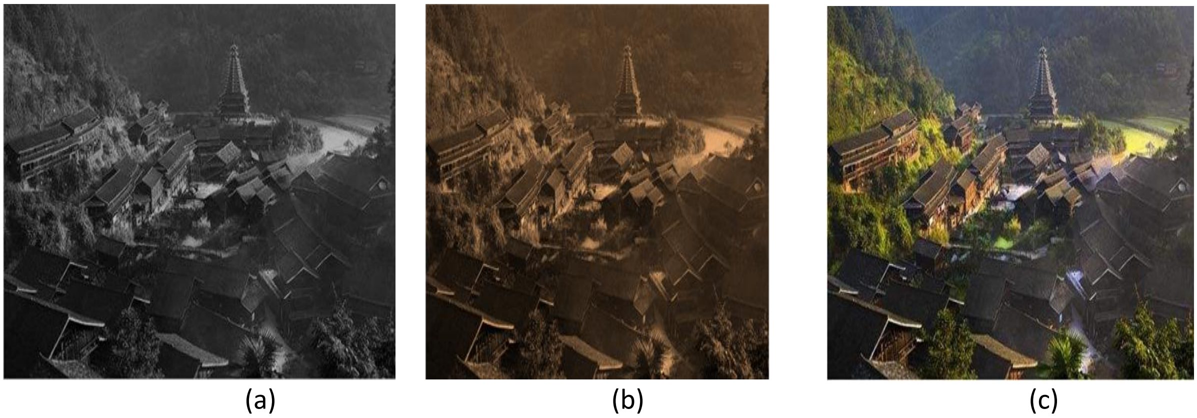


Figure 10: Example of Intelligent Scenery Online Recommendation:(A) Night Scenery,(B) Sunrise Scenery, (C) Daytime Scenery.

On the basis of the above research, the effect evaluation of the intelligent tourism development system in ethnic areas based on data mining proposed in this paper is carried out, and the evaluation results are shown in Table 1.

Number	*Tourism evaluation*	*Number*	*Tourism evaluation*	*Number*	*Tourism evaluation*
1	*82.46*	*23*	*82.15*	*45*	*78.71*
2	*81.09*	*24*	*86.13*	*46*	*80.20*
3	*84.72*	*25*	*83.46*	*47*	*80.23*
4	*86.24*	*26*	*81.68*	*48*	*81.61*
5	*85.13*	*27*	*82.36*	*49*	*79.50*
6	*79.72*	*28*	*81.93*	*50*	*78.16*
7	*80.36*	*29*	*80.66*	*51*	*79.98*
8	*80.20*	*30*	*79.59*	*52*	*84.65*
9	*85.65*	*31*	*82.89*	*53*	*83.81*
10	*84.58*	*32*	*83.07*	*54*	*86.39*
11	*80.68*	*33*	*86.66*	*55*	*79.56*
12	*85.39*	*34*	*80.23*	*56*	*87.66*
13	*80.62*	*35*	*77.77*	*57*	*79.32*
14	*81.09*	*36*	*84.11*	*58*	*81.70*
15	*78.02*	*37*	*84.80*	*59*	*86.60*
16	*78.22*	*38*	*80.48*	*60*	*86.16*
17	*78.98*	*39*	*82.69*	*61*	*80.20*
18	*83.45*	*40*	*78.17*	*62*	*79.11*
19	*86.05*	*41*	*84.55*	*63*	*81.90*
20	*79.64*	*42*	*80.83*	*64*	*79.97*
21	*80.07*	*43*	*83.91*	*65*	*78.84*
22	*80.77*	*44*	*81.39*	*66*	*84.08*

Table 1: Evaluation of the Effect of Intelligent Tourism Development System in Ethnic Areas Based on Data Mining.

From the above evaluation, it can be seen that the intelligent tourism development system in ethnic areas based on data mining proposed in this paper can play an important role in the development of intelligent tourism in ethnic areas.

5 CONCLUSION

Many cities combine their own characteristics and construction advantages to develop intelligent tourism systems in intelligent ethnic areas. However, due to the lack of access to various databases and the accumulation of a large amount of data, there is a gap between the supply of intelligent tourism in ethnic areas and the actual demand, and various problems emerge one after another. For example, there is waste or idleness in the construction of service facilities, and the planning structure is unreasonable.This is an urgent challenge for the current industrial transformation and upgrading of the tourism industry. However, the emergence of big data is precisely the opportunity to solve this problem. By promoting the application of Internet+ methods such as big data in intelligent tourism in ethnic areas, precise and scientific management and services can be achieved in passenger flow management, consulting services and experience optimization, so as to improve the intelligent tourism system in ethnic areas and further strengthen related effective management.This paper combines data mining technology to build an intelligent tourism development system in ethnic areas to improve the intelligent development of tourism in ethnic areas. The experimental analysis results show that the intelligent tourism development system in ethnic areas based on data mining proposed in this paper can play an important role in the development of intelligent tourism in ethnic areas.

Li Zhang, https://orcid.org/0009-0001-3107-0998

ACKNOWLEDGEMENT

Shanxi Provincial Philosophy and Social Science Planning Office：Research on exploitation of forest health tourism in Shanxi Province(2021YY116).

REFERENCES

[1] Alaei, A. R.; Becken, S.; Stantic, B.: Sentiment Analysis in Tourism: Capitalizing on Big Data, Journal of Travel Research, 58(2), 2019, 175-191. https://doi.org/10.1177/0047287517747753
[2] Anda, C.; Erath, A.; Fourie, P. J.: Transport Modelling in the Age of Big Data, International Journal of Urban Sciences, 21(sup1), 2017, 19-42. https://doi.org/10.1080/12265934.2017.1281150
[3] Bulchand-Gidumal, J.: M. Sigala, R. Rahimi, M. Thelwall (Eds.):Big Data and Innovation in Tourism, Travel, and Hospitality, Managerial Approaches, Techniques, and Applications, Zeitschrift für Tourismuswissenschaft, 13(2), 2021, 309-310. https://doi.org/10.1515/tw-2021-0017
[4] Chen, F.; Zhang, J.; Wang, Z.; Shi, S.; ; Liu, H.: Passenger Travel Characteristics and Bus Operational States: a Study Based on IC Card and GPS Data in Yinchuan, China, Transportation Planning and Technology, 42(8), 2019, 825-847. https://doi.org/10.1080/03081060.2019.1675796
[5] Chen, M.; Yang, J.; Hu, L.; Hossain, M. S.; Muhammad, G.: Urban Healthcare Big Data System Based on Crowdsourced and Cloud-Based Air Quality Indicators, IEEE Communications Magazine, 56(11), 2018, 14-20. https://doi.org/10.1109/MCOM.2018.1700571
[6] Del Vecchio, P.; Mele, G.; Ndou, V.; Secundo, G.: Creating Value from Social Big Data:

Implications for Smart Tourism Destinations, Information Processing & Management, 54(5), 2018, 847-860. https://doi.org/10.1016/j.ipm.2017.10.006
[7] Fukuda, D.: Innovative Travel Survey Methods and Behavior Modeling in the Era of Big Data, Asian Transport Studies, 5(3), 2019, 436-438.
[8] Gallego, I.; Font, X.: Changes in Air Passenger Demand as a Result of the COVID-19 Crisis: Using Big Data to Inform Tourism Policy, Journal of Sustainable Tourism, 29(9), 2021 , 1470-1489. https://doi.org/10.1080/09669582.2020.1773476
[9] Han, Y.; Kim, Y.: A Study of Measuring Traffic Congestion for Urban Network Using Average Link Travel Time Based on DTG Big Data, The Journal of The Korea Institute of Intelligent Transport Systems, 16(5), 2017 , 72-84. https://doi.org/10.12815/kits.2017.16.5.72
[10] Juanjuan, Z. H. A. O.; Chengzhong, X. U.; Tianhui, M. E. N. G.: Big Data-Driven Residents' Travel Mode Choice: A Research Overview, ZTE Communications, 17(3), 2019, 9-14.
[11] Li, B.; Kisacikoglu, M. C.; Liu, C.; Singh, N.; Erol-Kantarci, M.: Big Data Analytics for Electric Vehicle Integration in Green Smart Cities, IEEE Communications Magazine, 55(11), 2017 , 19-25. https://doi.org/10.1109/MCOM.2017.1700133
[12] Li, T.; Wang, J.; Huang, J.; Gao, X.: Exploring Temporal Heterogeneity in an Intercity Travel Network: A Comparative Study Between Weekdays and Holidays in China, Journal of Geographical Sciences, 30(12), 2020, 1943-1962. https://doi.org/10.1007/s11442-020-1821-9
[13] Llorca, C.; Molloy, J.; Ji, J.; Moeckel, R.: Estimation of a Long-Distance Travel Demand Model Using Trip Surveys, Location-Based Big Data, and Trip Planning Services, Transportation Research Record, 2672(47), 2018, 103-113. https://doi.org/10.1177/0361198118777064
[14] Nitu, P.; Coelho, J.; Madiraju, P.: Improvising Personalized Travel Recommendation System With Recency Effects, Big Data Mining and Analytics, 4(3), 2021, 139-154. https://doi.org/10.26599/BDMA.2020.9020026
[15] Peak, C. M.; Wesolowski, A.; zuErbach-Schoenberg, E.; Tatem, A. J.; Wetter, E.; Lu, X.; Bengtsson, L.: Population Mobility Reductions Associated with travel Restrictions During the Ebola Epidemic in Sierra Leone: Use of Mobile Phone Data, International journal of epidemiology, 47(5), 2018, 1562-1570. https://doi.org/10.1093/ije/dyy095
[16] Ren, J.; Luo, X.; Dong, L.; Dou, Y.; Zhang, N.; Li, Y.; Yao, S.: Analysis on Spatial-Temporal Features of Taxis' Emissions from Big Data Informed Travel Patterns: a Case of Shanghai, China, Journal of Cleaner Production, 142(2), 2017, 926-935. https://doi.org/10.1016/j.jclepro.2016.05.161
[17] Tian, H.; Presa-Reyes, M.; Tao, Y.; Wang, T.; Pouyanfar, S.; Miguel, A.; Iyengar, S. S.: Data Analytics for Air Travel Data: A Survey and New Perspectives, ACM Computing Surveys, 54(8), 2021, 1-35. https://doi.org/10.1145/3469028
[18] Torre-Bastida, A. I.; Del Ser, J.; Laña, I.; Ilardia, M.; Bilbao, M. N.; Campos-Cordobés, S.: Big Data for Transportation and Mobility: Recent Advances, Trends and Challenges, IET Intelligent Transport Systems, 12(8), 2018 , 742-755. https://doi.org/10.1049/iet-its.2018.5188
[19] Wang, C. J.; Ng, C. Y.; Brook, R. H.: Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing, Jama, 323(14), 2020, 1341-1342. https://doi.org/10.1001/jama.2020.3151
[20] Wang, H.; Tang, X.; Kuo, Y. H.; Kifer, D.; Li, Z.: A Simple Baseline for Travel Time Estimation Using Large-Scale Trip Data, ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 2019, 1-22. https://doi.org/10.1145/3293317
[21] Zhu, L.; Yu, F. R.; Wang, Y.; Ning, B.; Tang, T.: Big Data Analytics in Intelligent Transportation Systems: a Survey, IEEE Transactions on Intelligent Transportation Systems, 20(1), 2018, 383-398. https://doi.org/10.1109/TITS.2018.2815678