





Digital Transformation in Landscape Design of Rural Characteristic Town Based on Big Data Technology

Ting Wan¹ and Shuo Tian^{2*}

^{1,2}School of Landscape, Northeast Forestry University, Harbin, Heilongjiang 150040, China
harbin23222@163.com, tingwang2222@aliyun.com

Corresponding Author: Shuo Tian, harbin23222@163.com

Abstract. In order to improve the landscape design effect of rural characteristic town, this paper combines big data technology to mine and analyze the current situation of landscape design of rural characteristic town, and proposes a fuzzy C-means clustering algorithm based on incomplete data of pseudo-nearest neighbor interval. Moreover, this paper uses the pseudo-nearest neighbor rule to describe the missing attribute value of an incomplete sample as an interval number, and the complete attribute value of the sample is described as an interval number with equal values at both ends. In addition, this paper transforms the numerical data set into an interval data set, and uses the fuzzy C-means clustering algorithm to cluster the interval data set of rural characteristic town landscapes. Finally, this paper verifies the effectiveness of this method through experimental research. Through experimental research, it can be known that the landscape design method of rural characteristic towns based on big data technology proposed in this paper has a certain effect.

Keywords: Big data; rural areas; characteristic towns; landscape design; Digital Transformation

DOI: <https://doi.org/10.14733/cadaps.2024.S2.19-38>

1 INTRODUCTION

With the continuous improvement of the income and living standards of urban and rural residents, the fast-paced and high-intensity of urban life makes people more yearning for open and fresh pastoral scenery, which has led to the rapid development of rural tourism. On the other hand, the construction of characteristic towns is inseparable from the development of characteristic tourism. The focus of the tourism construction of characteristic agricultural towns lies in the development of rich agricultural tourism activities [22].

An effective way to achieve coordinated development between rural areas and cities is to accelerate the realization of agricultural modernization to optimize my country's economic structure [16]. Cultivating agricultural characteristic towns should make good use of their own resources such as labor, scientific and technological strength, and various natural resources to develop their own industries according to the actual local conditions. This can greatly stimulate domestic demand to increase the income of local farmers and improve the level of development in rural areas. Therefore, the emergence of the concept of characteristic towns is a new rural development path that has emerged today when traditional urbanization is unsustainable. It can strengthen the connection between rural areas and cities and continuously narrow the gap between urban and rural areas, realize the overall development of urban and rural areas, and continuously optimize the economic structure of our country [17].

The realization of agricultural modernization has always been one of the most important goals of my country's economic and social development. But to realize the modernization of agriculture and promote the process of urbanization, the rapid development of agriculture is required. Nowadays, under the policy environment of rural revitalization, rural areas are not only carrying out pure agricultural production, but the more important function is to maintain the ecological environment and other social functions. Building a modern agriculture with green ecology and high technological content is an effective way to achieve healthy and rapid development in rural areas. But on the other hand, with the development of the times, many people's consumption concepts are also constantly changing. Now people have mostly got rid of the state of pursuing food and clothing, and more of them want spiritual pursuits. Being close to nature, green ecology is a more popular way for modern people to play. Places can tap natural resources, traditional culture and local customs with their own characteristics to develop the tertiary industry. Let people fully experience the local humanities in the beautiful local scenery. This model has a lot of room for development and can realize the coordinated and healthy development of primary, secondary and tertiary industries.

This article combines big data technology to excavate and analyze the current situation of rural characteristic town landscape design, and this article formulates rural characteristic town landscape design strategies with the support of big data technology.

2 RELATED WORK

The planning of rural villages and towns must clearly define the main body of the plan and understand the objects of the plan [15]. In order to cater to the taste of the government and maximize the benefits, we cannot ignore the dominant status of farmers, the lifestyle and living habits of local residents, the particularity of the culture of the villages and towns, and the deep-seated reasons for the development of the villages and towns. If we directly transfer the urban planning method to the villages and towns to solve the construction problems in the villages and towns, ignore the natural texture and the historical context of the villages and towns, blindly demolish and build and pursue a novel, modern, and lavish image effect, the phenomenon of "one thousand villages and one aspect" will appear after the "one aspect of a thousand cities" [6]. In order to advance the progress of the project, some leaders implemented the strategy of forced demolition and construction, which even led to bloodshed. This has led to the disappearance of the government's credibility and the consciousness of serving the people, and the relationship between the people and the government has become increasingly tense [20].

Villages and towns in the new era must have modern and convenient conditions, but they must not lose the original charm of villages and towns [8]. At present, the vast majority of people cannot melt into the fast-paced life, and cannot adapt to the cold steel concrete in the city to block the warmth of human affection [1]. When they returned to their homeland, they had found that the villages and towns had lost their previous taste. As a result, "homesickness" appeared, and a series of activities such as prosperous village and town experience, ecological agriculture sightseeing,

farmhouse entertainment, etc., also explained people's expectation of returning to nature and retaining "homesickness"[10].

The natural characteristics of the landscape of villages and towns are an important aspect that distinguishes them from cities. Through the precipitation of history and the natural gathering and dispersal of communities, they form their own geomantic texture. Therefore, village and town planners should lower their own posture, abandon the model experience, fully respect nature, and carry out development guidance with awe [19]. Since there are more natural and cordial things in villages and towns, the planning of villages and towns needs more emotion. In the planning of villages and towns, it is necessary to pay full attention to the maintenance of natural texture, and preserve the natural landscape full of historical sentiment in the villages and towns [21].

The traditional history and culture of the villages and towns are the soul of the villages and towns, as well as the foundation of the development of the villages and towns. Therefore, excavating and carrying forward the traditional historical culture of villages and towns not only preserves the spiritual memory of human beings, but also provides inexhaustible resources for tourism in villages and towns [2]. The traditional culture of villages and towns is the accumulation and precipitation of long-term lifestyles and habits of the working people, and is closely related to modern life [4]. In the protection of villages and towns, we should not only consider the inheritance of traditional culture, but also pay attention to its integration with modern culture and ideas, so that it can be accepted by the people nowadays. In the planning of villages and towns, it is necessary to protect the rural cultural characteristics of the villages and towns, correctly guide the construction and development of the villages and towns, find the soul for the villages and towns, and make the traditional historical and cultural texture of the villages and towns better develop [13].

3 LANDSCAPE DATA MINING ALGORITHM FOR RURAL CHARACTERISTIC TOWNS

In order to give a description of the pseudo-nearest neighbor interval for missing attribute values, the definition of pseudo-similarity is first given below.

For sample set $\bar{X} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\} \subset R^{m \times s}$, before calculating (complete or incomplete) the pseudo-neighbor relationship between samples \bar{x}_k and \bar{x}_l , first, their attribute values are rearranged according to the following rules.

- If the sample attribute values \bar{x}_{kj} and \bar{x}_{lj} ($1 \leq j \leq s$) are missing at the same time, they will be moved to the rightmost end of \bar{x}_k and \bar{x}_l respectively, and indicated by "*" at the same time;
- If the sample attribute values \bar{x}_{kj} and \bar{x}_{lj} ($1 \leq j \leq s$) are not missing, they are moved to the leftmost end of \bar{x}_k and \bar{x}_l respectively;
- 3.If one and only one of the sample attribute values \bar{x}_{kj} and \bar{x}_{lj} ($1 \leq j \leq s$) is missing, their position remains unchanged, and the missing attribute value is represented by "#";

The pseudo-similarity of samples \bar{x}_k and \bar{x}_l is defined as follows[12]:

$$S_p(\bar{x}_k, \bar{x}_l) = d \sum_{i=1}^d \left(\frac{\bar{x}_{kj}}{\sqrt{\sum_{i=1}^d \bar{x}_{ki} \bar{x}_{ki}}} \frac{\bar{x}_{li}}{\sqrt{\sum_{i=1}^d \bar{x}_{li} \bar{x}_{li}}} \right) \quad (1)$$

Among them, the coefficient d represents the completeness of the attribute values of the two samples. The following is an example.

We set samples $\bar{x}_k = [12, 20, ?, 9, 15]$ and $\bar{x}_l = [10, 18, ?, 14, ?]$. Among them, "?" means that the attribute value of the corresponding position of the sample is missing, and they can be expressed as $\bar{x}_k = [12, 20, 9, 15, *_{k3}]$ and $\bar{x}_l = [10, 18, 14, \#_{l5}, *_{l3}]$ after rearranging them according to the above rules. Formula (1) can be used to calculate the pseudo-similarity between samples x and x as:

$$S_p(\bar{x}_k, \bar{x}_l) = 3 \times \frac{12 \times 10 + 20 \times 18 + 9 \times 14}{\sqrt{12^2 + 20^2 + 9^2} \times \sqrt{10^2 + 18^2 + 14^2}} = 2.92$$

It can be seen that when there is $d = s$, that is, when none of the sample attribute values are missing, except that the pseudo-similarity is multiplied by a constant coefficient d , the definitions of pseudo-similarity and cosine similarity are consistent. When the sample has missing attribute values, the pseudo-similarity is positively correlated with the completeness of the attribute value, that is, the pseudo-similarity takes into account the cosine similarity between two samples and the number of their complete attribute values. When the sample has more missing attribute values, the uncertainty of the sample is greater, and the coefficient d in front of the pseudo-similarity strengthens the effect of the complete attribute value of the sample.

According to the definition of pseudo-similarity above, for sample \bar{x}_l , if there is $k = \operatorname{argmax} S_p(\bar{x}_l, \bar{x}_k), 1 \leq k \leq n, k \neq l$, sample \bar{x}_k is called the pseudo-nearest neighbor sample of \bar{x}_l .

We set y as the missing attribute value of the incomplete sample x , and the sample set $y = \{y_1, y_2, \dots, y_q\} \subset \bar{x}$ is the q pseudo-nearest neighbor samples of the sample \bar{x}_k . Among them,

there is $S_p(\bar{x}_k, y_1) \geq S_p(\bar{x}_k, y_2) \geq \dots \geq S_p(\bar{x}_k, y_q)$. From the perspective of nearest neighbors, the

missing attribute value x_{kj} of a sample can be replaced with the corresponding attribute value $y_{i,j}$ of its nearest neighbor sample. In view of the uncertainty of missing attribute values, this paper uses the interval composed of the corresponding attribute values of the q pseudo-nearest neighbor samples of the incomplete sample to describe the missing attribute value, which is called the pseudo-

nearest-neighbor intervals (PNNI) of the attribute value of the incomplete sample, that is $\left[\tilde{x}_{kj}^-, \tilde{x}_{kj}^+ \right]$

. Among them, there is $\tilde{x}_{kj}^- = \min \{y_{1,j}, y_{2,j}, \dots, y_{qj}\}$ and $\tilde{x}_{kj}^+ = \max \{y_{1,j}, y_{2,j}, \dots, y_{qj}\}$ [11].

For the known attribute value x_{kj} , the interval $\left[\tilde{x}_{kj}^-, \tilde{x}_{kj}^+ \right]$ can also be used to describe. Among them, there is $\tilde{x}_{kj}^- = \tilde{x}_{kj}^+ = x_{kl}$. Therefore, the FCM algorithm based on interval numbers can be used for cluster analysis.

Obviously, the choice of q value cannot be ignored in PNNI. If the q value is too small, it is easy to cause deviation, and the obtained interval may not contain the true value of the missing attribute value; if the q value is too large, the range of the interval will become larger, and at the same time, it will confuse the attribute characteristics and directly affect the clustering results. In particular, when q is the number of the entire sample, the interval description range of the missing attribute value is within the value range of the entire attribute space, and such an interval is meaningless. On the basis of the q -value selection strategy, this article is improved, based on the incomplete data set, artificially assume that 1 rather than z complete attribute values are missing, so as not to damage the distribution of sample missing attribute values as much as possible. , So as to calculate the pseudo-nearest neighbors as true as possible, and then analyze the ratio of the assumed missing attribute values in the determined pseudo-nearest neighbor interval when $q=2,3,\dots$. The inflection point q of the change curve is used as the value point. When $q < q_0$, the ratio of missing attributes falling in the corresponding nearest neighbor interval is small, and the interval estimation deviation is large. When $q > q_0$, the ratio does not change much, and $q=q_0$ As the number of pseudo-nearest neighbor samples of the incomplete sample in the incomplete data set, the confidence that the interval contains the true value of the missing attribute is also relatively high based on the reasonable range of the interval.

A fuzzy C-means clustering algorithm for incomplete data based on pseudo-nearest neighbor interval (FCM-PNNI) is proposed. The algorithm first uses the pseudo-nearest neighbor rule to describe the missing attribute values of incomplete samples as interval numbers, and the complete attribute values of the samples are described as two The number of intervals with the end value equal, so as to transform the numerical data set into an interval data set, and finally use the fuzzy C-means clustering algorithm to cluster the interval data set. The algorithm uses pseudo-similarity to calculate the neighbor relationship of incomplete samples, and at the same time considers the cosine similarity between samples and the completeness of sample attribute values: description of the interval number of missing attribute values, considering that the missing attribute values are corresponding in pseudo-nearest neighbor samples The distribution information of the attribute value also reflects the uncertainty of the missing attribute value to a certain extent. The clustering finally obtains the hyper-convex polyhedron formed by the interval cluster center, which can reflect the shape and sample distribution of the sub-category to a certain extent[9].

For the incomplete data set $\bar{X} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\} \subset R^{m \times s}$, first, according to the q value strategy described in section 1.2, the curve of the interval accuracy with the q value is obtained, and the q_0 value at the inflection point is selected as the q value. Then, according to the algorithm described in Table 1, the incomplete data set is transformed into an interval data set. We set the sample set $\bar{X} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\} \subset R^{m \times s}$ as a transformed interval data set, each sample has s -dimensional attributes and the interval number is used to describe the sample attributes, that is $\bar{x}_k = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n] \subset R^{m \times s}$. Then, the FCM-PNNI algorithm minimizes the following objective function under the constraints of formula (2)[7]:

$$\begin{aligned}
J(U, V) &= \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|\tilde{x}_k - \tilde{v}_i\|_2^2 \\
&= \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \left[\left(\tilde{x}_k^- - \tilde{v}_i^- \right) \left(\tilde{x}_k^- - \tilde{v}_i^- \right)^T + \left(\tilde{x}_k^+ - \tilde{v}_i^+ \right) \left(\tilde{x}_k^+ - \tilde{v}_i^+ \right)^T \right]
\end{aligned} \tag{2}$$

$$\tilde{v}_i = \left[\tilde{v}_i^-, \tilde{v}_i^+ \right], 1 \leq i \leq c$$

Among them, \tilde{v}_i is the interval cluster center of the i -th subcategory. In this way, the problem of clustering incomplete data sets is transformed into the goal of clustering a complete interval data set. The specific derivation of interval data set clustering can be found in the literature. The necessary conditions for the objective function (2) to obtain the minimum value are directly given here: as shown in formulas (3), (4) and (5).

$$\tilde{v}_i^- = \frac{\sum_{k=1}^n u_{ik}^m \tilde{x}_k^-}{\sum_{k=1}^n u_{ik}^m}, i = 1, 2, \dots, c \tag{3}$$

$$\tilde{v}_i^+ = \frac{\sum_{k=1}^n u_{ik}^m \tilde{x}_k^+}{\sum_{k=1}^n u_{ik}^m}, i = 1, 2, \dots, c \tag{4}$$

$$u_{ik} = \left[\sum_{i=1}^c \frac{\left(\left(\tilde{x}_k^- - \tilde{v}_i^- \right) \left(\tilde{x}_k^- - \tilde{v}_i^- \right)^T + \left(\tilde{x}_k^+ - \tilde{v}_i^+ \right) \left(\tilde{x}_k^+ - \tilde{v}_i^+ \right)^T \right)^{\frac{1}{m}}}{\left(\left(\tilde{x}_k^- - \tilde{v}_i^- \right) \left(\tilde{x}_k^- - \tilde{v}_i^- \right)^T + \left(\tilde{x}_k^+ - \tilde{v}_i^+ \right) \left(\tilde{x}_k^+ - \tilde{v}_i^+ \right)^T \right)^{\frac{1}{m}}} \right]^{-1}, i = 1, 2, \dots, c \tag{5}$$

In a special case, if there is $\exists k, h, 1 \leq k \leq n, 1 \leq h \leq c, \forall j: \tilde{x}_{kj} \subset \tilde{v}_{kj}$, in short, if there is an interval of sample \tilde{x}_k that is located inside the convex polyhedron formed by the interval cluster center \tilde{v}_h , the sample belongs to this sub-category completely. From the perspective of degree of membership, there are:

$$u_{ik} = \begin{cases} 1, i = h \\ 0, i \neq h \end{cases} \tag{6}$$

The FCM-PNNI algorithm is described as:

- According to the q value in section 1.2, the strategy is selected to select the q value:
- For any k, j , according to the PNNI algorithm in Tab.1, the attribute value \tilde{x}_{kj} is changed to

the interval number $\left[\tilde{x}_{kj}^-, \tilde{x}_{kj}^+ \right]$:

- We set the number of clustering categories c , the fuzzy weighting index m , and the iteration stop threshold ε . The membership matrix $U^{(0)}$ is initialized and satisfies the constraints of formula (2);
- When the number of iteration steps is l ($l = 1, 2, \dots$), formulas (3), (4) and $U^{(l-1)}$ are used to update the cluster center $V^{(l)}$;
- If there is $\exists k, h, 1 \leq k \leq n, 1 \leq h \leq c, \forall j: \tilde{x}_{kj} \subset \tilde{v}_{kj}$, formula (6) is used to update $U^{(l)}$, otherwise, formula (5) and $\tilde{v}_i^{+(l)}$ and $\tilde{v}_i^{-(l)}$ are used to update $U^{(l)}$;
- If the condition $\|U^{(l)} - U^{(l-1)}\| \leq \varepsilon$ is met, the algorithm stops iteration and outputs $U^{(l)}$ and $V^{(l)}$, otherwise there is $l = l + 1$, and it returns to step (4) for the next iteration.

Euclidean distance is the most commonly used distance, but the concept of distance itself is much broader. As long as the following four properties are satisfied, it can be called distance, but it may not all have meaning.

For any vector a, b, c , there are[14]:

- Non-negativity: $d(a,b) \geq 0$
- Reflexivity: if and only if $a=b$, $d(a,b)=0$
- Symmetry: $d(a,b)=d(b,a)$
- Triangle inequality: $d(a, b)+d(b,c) \geq d(a,c)$

Obviously Euclidean distance satisfies the above four properties. For the distance between incomplete samples, due to the uncertainty of the missing attribute value, the distance between the incomplete sample and the cluster center should also be uncertain. Starting from the complexity of objective things, the description form of interval numbers is more in line with the true colors of things.

As shown in Figure 1, x_1, x_2, x_3, x_4 , and x_5 are the representations of 4 two-dimensional samples in space. The true Euclidean distance between point x and point x is:

$$d(x_1, x_2) = \sqrt{(2-1)^2 + (6-1)^2} = \sqrt{26}$$

If it is assumed that some attribute values of point x_1 and point x_2 are missing. However, if the distances related to points x_1 and x_2 are known: $d(x_1, x_3) = \sqrt{2}$, $d(x_1, x_4) = \sqrt{45}$, $d(x_1, x_5) = \sqrt{17}$, $d(x_2, x_3) = 4$, $d(x_2, x_5) = \sqrt{5}$, $d(x_2, x_5) = \sqrt{5}$, the triangle inequality properties used in the triangle surrounded by points x_1, x_2, x_3 are:

$$|d(x_1, x_3) - d(x_2, x_3)| < d(x_1, x_2) < d(x_1, x_3) + d(x_2, x_3)$$

$d(x_1, x_2) \in (4 - \sqrt{2}, \sqrt{4} + \sqrt{2})$ is calculated, and the properties of the triangle inequality used in the triangle surrounded by points x_1 , x_2 , and x_3 are:

$$|d(x_1, x_4) - d(x_2, x_4)| < d(x_1, x_2) < d(x_1, x_4) + d(x_2, x_4)$$

$d(x_1, x_2) \in (\sqrt{45} - \sqrt{5}, \sqrt{45} + \sqrt{5})$ is calculated, and the properties of the triangle inequality used in the triangle surrounded by points x_1 , x_2 , and x_3 are:

$$|d(x_1, x_5) - d(x_2, x_5)| < d(x_1, x_2) < d(x_1, x_5) + d(x_2, x_5)$$

$$d(x_1, x_2) \in (\sqrt{17} - \sqrt{5}, \sqrt{17} + \sqrt{5}) \text{ is calculated.}$$

Based on the above results, $d(x_1, x_2) \in (\sqrt{45} - \sqrt{5}, \sqrt{45} + \sqrt{5})$ can be obtained. This interval

estimate is quite close to the true value $\sqrt{26}$ and is better than any one of the interval estimates. Obviously, when there are enough triangles involved in the calculation, a more accurate interval estimate will be obtained. At the same time, it can be seen that point x_5 does not play an effective

role in the estimation of the final $d(x_1, x_2)$. From an intuitive point of view, because the size of $d(x_1, x_5)$ and $d(x_2, x_5)$ are close, the difference between the two is close to 0, and the effect of estimating the lower bound of the interval value is very poor. At the same time, the values of $d(x_1, x_5)$ and $d(x_2, x_5)$ are both large, resulting in a large sum of the two, and the upper bound of the interval value is also not well estimated. Therefore, choosing a suitable sample point is very important to the distance estimation result. If the point adjacent to the point x_1 or x_2 is selected as the third point to form a triangle calculation with x_1 and x_2 , the difference between it and the two sides formed by x_1 and x_2 is larger, and the sum of the two sides is smaller. Therefore, a reasonable estimate of the upper and lower bounds of the distance interval can be made at the same time[5].

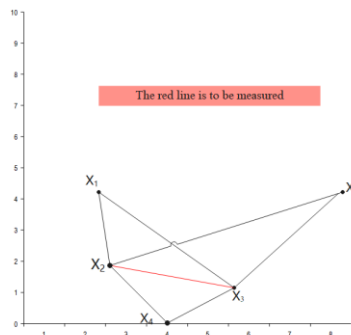


Figure 1: Triangle inequality is used to find the distance between point x_1 and point x_2 .

We set $\bar{x}_k = [x_{ik}, x_{2k}, \dots, x_{sk}]^T$ as an incomplete sample in the incomplete data set $\bar{X} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\} \subset R^{m \times s}$. The distance between \bar{x}_k and the i -th cluster center v_i , can be described by TID, as shown below[18]:

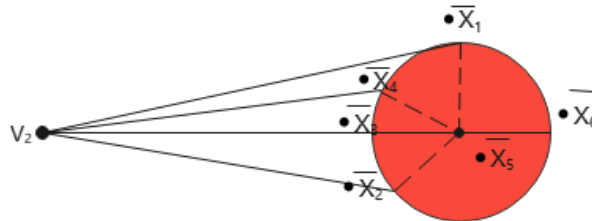


Figure 2: The triangular distance between \bar{x}_k and v_i .

In Figure 2, $\bar{x}_1 \sim \bar{x}_5$ is the five adjacent samples of x in the attribute space. Among them, $\bar{x}_1 \sim \bar{x}_5$ may be a complete sample or an incomplete sample. In order to be able to be on the original data set instead of the normalized and compressed Euclidean Calculate the neighbor samples in the space, and use the pseudo-similarity defined in section 1.1 to describe the neighbor relationship. Obviously, due to the incompleteness of sample \bar{x}_k , the distance between \bar{x}_k and v_i cannot be directly calculated. As shown in Figure 2, each sample point \bar{x}_p ($1 \leq p \leq 5$) and \bar{x}_k and v_i near \bar{x}_k can form a triangle. In special cases, there are three points and one line. According to the triangle inequality, the distance $d(\bar{x}_k, v_i)$ between \bar{x}_k and v_i satisfies the following two conditions[3]:

$$d(\bar{x}_k, \bar{x}_p) + d(\bar{x}_p, v_i) \geq d(\bar{x}_k, v_i) \quad (7)$$

$$\left| d(\bar{x}_k, \bar{x}_p) - d(\bar{x}_p, v_i) \right| \leq d(\bar{x}_k, v_i) \quad (8)$$

The interval type distance based on the triangle inequality is defined as follows:

$$d_{TID}(\bar{x}_p, v_i) = [d_{ki}^-, d_{ki}^+] \quad (9)$$

Among them, there is,

$$d_{ki}^- = \max \{c_p \mid c_p = |d(\bar{x}_k, \bar{x}_p) - d(\bar{x}_p, v_i)|, 1 \leq p \leq \lambda\} \quad (10)$$

$$d_{ki}^+ = \min \{s_p \mid s_p = |d(\bar{x}_k, \bar{x}_p) + d(\bar{x}_p, v_i)|, 1 \leq p \leq \lambda\} \quad (11)$$

In the formula, λ is the number of pseudo-nearest neighbor samples of sample x . Obviously, the larger the value of λ , the smaller the interval range of the interval-type distance $d_{TID}(\bar{x}_p, v_i)$. Compared with the local distance used in the PDS strategy, the TID distance proposed in this section makes full use of the distribution information of the pseudo-nearest neighbor samples of incomplete samples. At the same time, the interval description of distance also reflects the uncertainty of missing attributes to a certain extent. When the uncertainty distance $d(\bar{x}_k, v_i)$ is calculated, if the pseudo-nearest neighbor sample \bar{x}_p ($1 \leq p \leq 5$) of the incomplete sample \bar{x}_k is a complete sample, The TID distance can be used to convert a larger uncertainty distance into a solution about a smaller uncertainty distance $d(\bar{x}_k, \bar{x}_p)$ and a relatively large certainty distance $d(\bar{x}_k, v_i)$, which is more reasonable than simply using a local distance.

The following describes the four properties that the proposed TID distance versus distance should satisfy. First, the TID distance obviously satisfies symmetry and is defined based on the triangular inequality; secondly, for non-negativity, since either the absolute value or the summation expression is used in the calculation, as shown in formulas (10) and (11), Satisfy the nature of non-negativity; finally, for reflexivity, since the starting point of the proposed TID distance itself is used to calculate the distance from the incomplete sample to the cluster center, there is no situation where the incomplete sample and the complete cluster center are the same, so In clustering, TID distance and PDS distance can also be used to calculate the distance between incomplete samples and cluster centers.

We set $\bar{X} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\} \subset R^{m \times s}$ as an incomplete data set, and the fuzzy C-means clustering algorithm (TIFCM) for incomplete data based on TID distance minimizes the following objective function:

$$J(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ij}^m d_{kj}^2 \quad (12)$$

In the formula, c is the number of sub-categories divided, m is the fuzzy coefficient, and $V = \{v_1, v_2, \dots, v_n\} \subset R^{s \times c}$ is the cluster center matrix. At the same time, \bar{x}_k is the membership degree of the k -th sample x belonging to the i -th cluster center v , there is $\forall i, k : u_{ik} \in [0, 1]$, and it satisfies the following constraints:

$$\sum_{i=1}^c u_{ik} = 1, i \leq k \leq n \quad (13)$$

The distance d_{ki} between \bar{x}_k and v_i is defined as follows:

$$d_{ki} = \begin{cases} \|\bar{x}_k - v_i\|_2, & \text{if } \bar{x}_k \in \bar{X}_w \\ d_{TID}(\bar{x}_k, v_i) = [d_{ki}^-, d_{ki}^+], & \text{if } \bar{x}_k \in \bar{X}_L \end{cases} \quad (14)$$

The following formula derives the necessary conditions for minimizing the above formula (12). According to formula (14), the objective function (12) is rewritten as follows:

$$J(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ij}^m [A_k d_{ki}^2 + (1 - A_k) d_{TID}^2] \quad (15)$$

Among them, there is

$$A_k = \begin{cases} 1, & \text{if } \bar{x}_k \in \bar{X}_W \\ 0, & \text{if } \bar{x}_k \in \bar{X}_L \end{cases}$$

Lagrangian operators are introduced:

$$J(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ij}^m [A_k d_{ki}^2 + (1 - A_k) d_{TID}^2] + \sum_{k=1}^n \lambda_k \left(\sum_{i=1}^c u_{ik} - 1 \right) \quad (16)$$

In order to facilitate the derivation, we set $\bar{d}_{ik}^2 = A_k d_{ki}^2 + (1 - A_k) d_{TID}^2$ and derive the derivation of each u_{ik} in the above formula. The derivation process is simplified here, $k = 1, i = 1, 2, \dots, c$ is used as an example to derivate u_{i1} , and the following c formulas are obtained:

$$\begin{cases} mu_{11}^{m-1} \bar{d}_{11}^2 + \lambda_1 = 0 \\ mu_{21}^{m-1} \bar{d}_{21}^2 + \lambda_1 = 0 \\ \dots \\ mu_{c1}^{m-1} \bar{d}_{c1}^2 + \lambda_1 = 0 \end{cases} \quad (17)$$

Formula (17) is simplified to:

$$\begin{cases} u_{11} = \left(\frac{-\lambda_1}{m \bar{d}_{11}^2} \right)^{1/(m-1)} \\ u_{21} = \left(\frac{-\lambda_1}{m \bar{d}_{21}^2} \right)^{1/(m-1)} \\ \dots \\ u_{c1} = \left(\frac{-\lambda_1}{m \bar{d}_{c1}^2} \right)^{1/(m-1)} \end{cases} \quad (18)$$

The above formula (18) is substituted into formula (13) to obtain:

$$(-\lambda_1)^{1/(m-1)} = \frac{1}{\sum_{i=1}^c (m \bar{d}_{i1}^2)^{1/(1-m)}} \quad (19)$$

When substituting the above formula (19) into each formula in formula (18), We can obtain the necessary conditions that u_{ik} ($k = 1, i = 1, 2, \dots, c$) meets when the objective function of equation (14) achieves the minimum value. We use $k = 1, i=1$ as an example to substitute formula (19) into the first formula in formula (18), as follows:

$$u_{11} = \left(\frac{-\lambda_1}{m\bar{d}_{11}^2} \right)^{1/(m-1)} = \frac{(m\bar{d}_{11}^2)^{1/(1-m)}}{\sum_{i=1}^c (m\bar{d}_{i1}^2)^{1/(1-m)}} = \frac{(\bar{d}_{11}^2)^{1/(1-m)}}{\sum_{i=1}^c (\bar{d}_{i1}^2)^{1/(1-m)}} \quad (20)$$

If there is $\bar{x}_k \in \bar{X}_w$, that is $A_k = 1$, $d_{ik}^2 = A_k d_{ki}^2 + (1 - A_k) d_{TID}^2 = d_{ki}^2$ is substituted into (10) to simplify and get the following formula:

$$u_{11} = \frac{(\bar{d}_{11}^2)^{1/(1-m)}}{\sum_{i=1}^c (\bar{d}_{i1}^2)^{1/(1-m)}} = \frac{(d_{11}^2)^{1/(1-m)}}{\sum_{i=1}^c (d_{i1}^2)^{1/(1-m)}} \quad (21)$$

If there is $\bar{x}_k \in \bar{X}_w$, that is $A_k = 1$, in the same way, $d_{ik}^2 = A_k d_{ki}^2 + (1 - A_k) d_{TID}^2 = d_{TID}^2$ is substituted into (20) to simplify and get the following formula:

$$u_{11} = \frac{(\bar{d}_{11}^2)^{1/(1-m)}}{\sum_{i=1}^c (\bar{d}_{i1}^2)^{1/(1-m)}} = \frac{(d_{TID}^2)^{1/(1-m)}}{\sum_{i=1}^c (d_{TID}^2)^{1/(1-m)}} \quad (22)$$

According to the above derivation, in the same way, we can obtain the necessary conditions that u_{ik} satisfies when the objective function takes the minimum value when $k = 1, i \neq 1$ and $k \neq 1, i = 1, 2, \dots, c$, as follows:

$$u_{ik} = \begin{cases} \frac{(\bar{d}_{11}^2)^{1/(1-m)}}{\sum_{i=1}^c (d_{i1}^2)^{1/(1-m)}}, 1 \leq i \leq c, 1 \leq k \leq n, \text{if } \bar{x}_k \in \bar{X}_w \\ \frac{(\bar{d}_{TID}^2)^{1/(1-m)}}{\sum_{i=1}^c (d_{TID}^2)^{1/(1-m)}}, 1 \leq i \leq c, 1 \leq k \leq n, \text{if } \bar{x}_k \in \bar{X}_L \end{cases} \quad (23)$$

Since the TID distance is described by the number of intervals here, for the convenience of calculation, the values at both ends of the interval are used to calculate the average and the results are organized as follows:

$$u_{ik} = \begin{cases} \left[\sum_{i=1}^c \left(\frac{d_{ki}}{d_{ki}} \right)^{\frac{2}{m-1}} \right]^{-1}, 1 \leq i \leq c, 1 \leq k \leq n, \text{if } x_k \text{ is complete sample} \\ \frac{(u_{ik} + \bar{u}_{ik})}{2}, 1 \leq i \leq c, 1 \leq k \leq n, \text{if } x_k \text{ is not complete sample} \end{cases} \quad (24)$$

$$\underline{u}_{ik} = \left[\sum_{i=1}^c \left(\frac{d_{ki}}{d_{kt}} \right)^{\frac{2}{m-1}} \right]^{-1} \quad 1 \leq i \leq c, 1 \leq k \leq n \quad (25)$$

$$\bar{u}_{ik} = \left[\sum_{i=1}^c \left(\frac{d_{ki}^+}{d_{kt}^+} \right)^{\frac{2}{m-1}} \right]^{-1} \quad 1 \leq i \leq c, 1 \leq k \leq n \quad (26)$$

$$d_{TID}(\bar{x}_k, v_i) = [d_{ki}^-, d_{ki}^+] \quad (27)$$

$$d_{TID}(\bar{x}_k, v_t) = [d_{kt}^-, d_{kt}^+] \quad (28)$$

According to the above derivation, the conditions that the cluster center meets when the objective function is minimized are as follows:

$$v_{ij} = \frac{\sum_{k=1}^n u_{ik}^m I_{kj} x_{kj}}{\sum_{k=1}^n u_{ik}^m I_{kj}} \quad 1 \leq i \leq c, 1 \leq j \leq s \quad (29)$$

Among them, there is:

$$I_{kj} = \begin{cases} 0, & \text{if attribute value } x_{kj} \text{ is missing} \\ 1, & \text{if the attribute value } x_{kj} \text{ is not missing} \end{cases}$$

The TIFCM algorithm is described as:

1. The algorithm sets the number of clustering categories c , the fuzzy weighting index m , and the iteration stop threshold ε , and the membership matrix $U^{(0)}$ is initialized and satisfies the constraints of formula (13);
2. When the number of iteration steps is l ($l = 1, 2, \dots$), the algorithm uses equation (29) and $U^{(l-1)}$ to update the cluster center $V^{(l)}$;
3. The algorithm uses formula (15) and $V^{(l)}$ to update the membership matrix $U^{(l)}$;
4. If condition $\|U^{(l)} - U^{(l-1)}\| \leq \varepsilon$ is met, the algorithm stops iterating and outputs the membership matrix and clustering center. Otherwise, there is $l = l + 1$, and the algorithm returns to step (2) for the next iteration.

4 LANDSCAPE DESIGN OF RURAL CHARACTERISTIC TOWN BASED ON BIG DATA TECHNOLOGY

According to different levels of space, the planning of small towns with rural characteristics can be divided into five levels: regional planning, overall planning, zoning planning, regulatory detailed planning, and constructive detailed planning. Due to the differences in planning objects, the functions of different stages of planning, the performance of the planning concept at different stages has its own focus, which is mainly reflected in the three aspects of town system planning, urban overall planning and detailed planning (as shown in Figure 3).

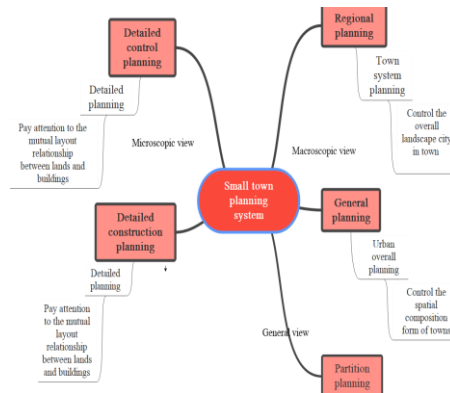


Figure 3: The embodiment of the landscape design of agricultural characteristic towns in all levels of planning.

This paper uses data mining to analyze the landscape design data of rural characteristic towns. The classification map of conventional town patterns can be divided into several types as shown in Figure 4.

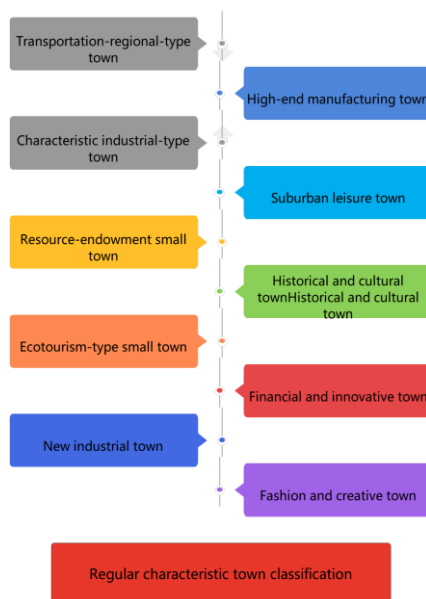


Figure 4: Classification map of the conventional town model.

Modern agricultural industry is the foundation for the industrial development of the park, as well as the foundation for the economic and environmental development of the park. In traditional agriculture, agricultural products must at least go through the processes of seedling purchase, crop production, product transportation, and product sales before they can finally reach consumers. Too many production links lead to a gradual increase in product value, and farmers are always at the bottom of the value chain, and the economic benefits are not high (Figure 5). Through the development of the whole industry chain agriculture to integrate technology research and development, product production, product processing, and product distribution into one, intermediate links can be eliminated and product quality can be improved. Moreover, products are delivered directly to consumers through modern logistics and distribution methods or the way consumers buy in the park (Figure 6). At the the same time, by using the unique agricultural environment to develop tourism and leisure industries and other life industries, the vitality and economic benefits of the park's life can be improved. In addition, the industrial integration of the park shortens the communication distance between various industries, reduces the overall input cost of the industry, and improves the efficiency of agricultural production and the added value of products.

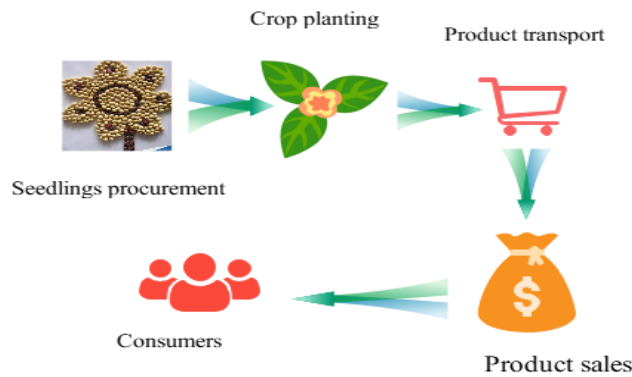


Figure 5: Landscape design model of the agricultural full-industry chain.

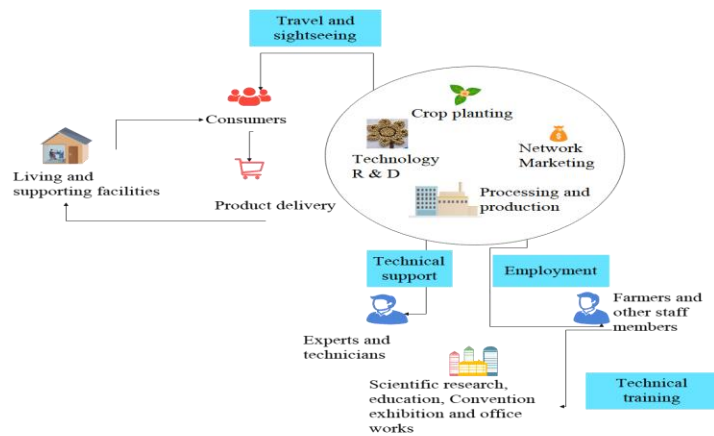


Figure 6: The full-industry chain model of agricultural parks in characteristic small towns.

Through the above methods, the landscape design of rural characteristic towns can be realized. On this basis, the effectiveness of this method is verified through experimental research. First, this paper verifies the data mining effect of the algorithm proposed in this paper. After that, this paper obtains relevant data on the landscape planning of rural characteristic towns through the Internet, verifies the effect of data mining, and verifies the results through expert scoring methods, and the results are shown in Table 1 and Figure 7.

<i>NO</i>	<i>Data processing</i>	<i>NO</i>	<i>Data processing</i>	<i>NO</i>	<i>Data processing</i>
1	95.58	23	86.15	45	90.23
2	94.15	24	86.40	46	96.31
3	96.71	25	93.52	47	89.47
4	90.30	26	88.80	48	86.77
5	90.80	27	91.53	49	95.67
6	94.75	28	87.17	50	89.66
7	91.62	29	93.21	51	89.15
8	86.39	30	88.03	52	91.30
9	89.23	31	96.86	53	92.03
10	90.80	32	95.15	54	93.63
11	95.56	33	95.72	55	92.03
12	96.28	34	95.04	56	90.56
13	91.59	35	86.37	57	93.19
14	86.37	36	88.76	58	86.96
15	94.30	37	86.74	59	86.39
16	90.95	38	88.96	60	95.89
17	87.53	39	89.60	61	91.88
18	90.03	40	87.69	62	88.64
19	94.04	41	95.47	63	96.58
20	90.88	42	88.34	64	94.23
21	94.50	43	94.04	65	86.47
22	87.79	44	96.52	66	87.67

Table 1: Statistical table of the data mining effect of the landscape design of rural characteristic towns.

The above experiments verify that the data mining algorithm can play an important role in the data mining of the landscape design of rural characteristic towns. On this basis, this paper evaluates the effect of the landscape design of rural characteristic towns, and the results are shown in Table 2 and Figure 8. From the shown experimental research, it can be seen that the landscape design method of rural characteristic towns based on big data technology proposed in this paper has certain effects.

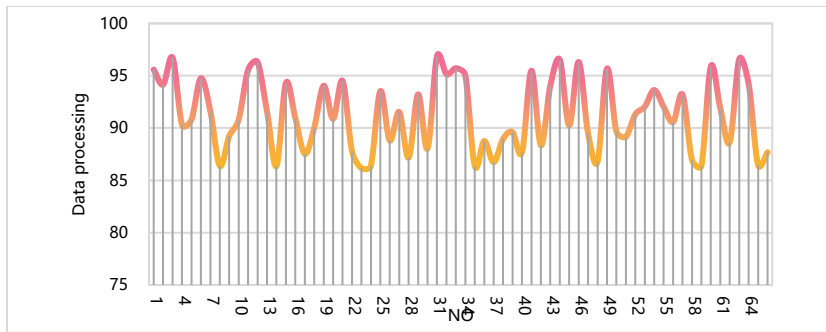


Figure 7: Analysis of the effectiveness of data mining algorithms.

NO	Design effect	NO	Design effect	NO	Design effect
1	76.70	23	85.39	45	86.03
2	85.30	24	76.03	46	88.50
3	87.17	25	88.08	47	83.88
4	87.96	26	75.65	48	80.40
5	91.56	27	89.10	49	76.39
6	82.72	28	81.07	50	89.60
7	78.03	29	84.37	51	81.47
8	84.63	30	83.81	52	83.22
9	90.69	31	77.78	53	84.71
10	85.98	32	78.06	54	74.92
11	76.96	33	74.45	55	75.75
12	87.90	34	75.60	56	83.68
13	86.86	35	87.74	57	83.00
14	86.86	36	80.73	58	77.93
15	86.49	37	83.38	59	78.13
16	88.80	38	79.36	60	81.36
17	91.71	39	90.32	61	81.51
18	83.21	40	79.89	62	91.89
19	79.38	41	77.70	63	84.39
20	79.65	42	86.06	64	77.46
21	79.17	43	88.63	65	76.16
22	75.45	44	84.13	66	80.96

Table 2: Evaluation of the effect of landscape design in rural characteristic towns.

5 CONCLUSION

With the development of traditional urbanization, the dual development of urban and rural areas is becoming more and more unbalanced, and the gap between urban and rural social, economic and cultural differences is getting bigger and bigger.

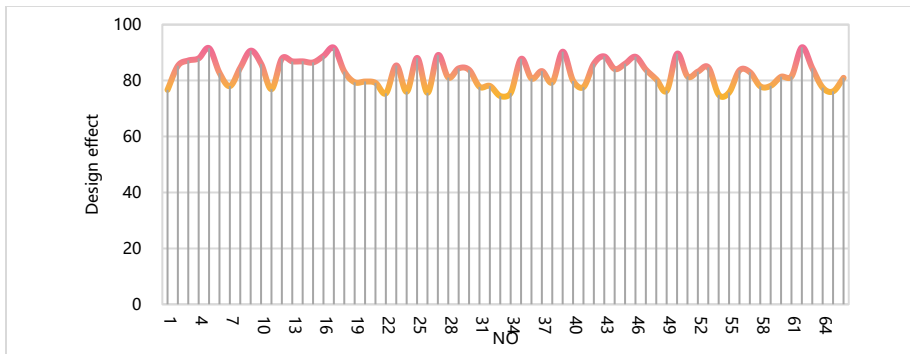


Figure 9: Landscape design effects of rural characteristic towns based on data mining.

Although in recent years, the expansion of urbanization and the great development of high-tech have concealed this to a certain extent, there are still many problems that need to be solved urgently. If the rural areas cannot keep up with the development of the cities, then the gap between them will only get bigger and bigger, which will eventually bring more problems. As one of the important directions for the development of characteristic towns in the country, the construction of characteristic agricultural towns is of great significance to Hubei Province, a large agricultural province. Moreover, agricultural development is the core of agricultural characteristic towns, so whether agriculture can be developed well determines whether agricultural characteristic towns can be successfully constructed. The development model of characteristic towns is crucial to the development of the industry. This paper combines the big data technology to carry out the research on the landscape design of rural characteristic towns. The experimental research shows that the landscape design method of rural characteristic towns based on big data technology proposed in this paper has a certain effect. the development of characteristic agricultural towns is vital for addressing the urban-rural divide in Hubei Province. By incorporating big data technology and digital art into landscape design, these towns can not only bridge the development gap but also create unique and appealing environments that celebrate their agricultural heritage and cultural identity.

Ting Wan, <https://orcid.org/0009-0007-8971-0840>

Shuo Tian, <https://orcid.org/0009-0004-2257-4432>

REFERENCES

- [1] Abreu, I.; Nunes, J. M.; Mesias, F. J.: Can rural development be measured? design and application of a synthetic index to portuguese municipalities, *Social Indicators Research*, 145(3), 2019, 1107-1123. <https://doi.org/10.1007/s11205-019-02124-w>
- [2] Antic, M.; Santic, D.; Kašanin-Grubin, M.; Malic, A.: Sustainable rural development in Serbia-relationship between population dynamics and environment, *Journal of Environmental Protection and Ecology*, 18(1), 2017, 323-331.
- [3] Beyene, A. D.; Mekonnen, A.; Randall, B.; Deribe, R.: Household level determinants of agroforestry practices adoption in rural Ethiopia, *Forests, Trees and Livelihoods*, 28(3), 2019, 194-213. <https://doi.org/10.1080/14728028.2019.1620137>
- [4] Cecchini, M.; Zambon, I.; Pontrandolfi, A.; Turco, R.; Colantoni, A.; Mavrakis, A.; Salvati, L.: Urban sprawl and the 'olive'landscape, *Sustainable land management for 'crisis' cities*, *GeoJournal*, 84(1), 2019, 237-255. <https://doi.org/10.1007/s10708-018-9848-5>

- [5] Chen, C.; LeGates, R.; Fang, C.: From coordinated to integrated urban and rural development in China's megacity regions, *Journal of Urban Affairs*, 41(2), 2019, 150-169. <https://doi.org/10.1080/07352166.2017.1413285>
- [6] Das, M.; Das, A.: Dynamics of Urbanization and its impact on Urban Ecosystem Services, A study of a medium size town of West Bengal Eastern India, *Journal of Urban Management*, 8(3), 2019, 420-434. <https://doi.org/10.1016/j.jum.2019.03.002>
- [7] de Waroux, O. L. P.; Cohuet, S.; Ndazima, D.; Kucharski, A. J.; Juan-Giner, A.; Flasche, S.; Edmunds, W. J.: Characteristics of human encounters and social mixing patterns relevant to infectious diseases spread by close contact, a survey in Southwest Uganda, *BMC Infectious Diseases*, 18(1), 2018, 1-12. <https://doi.org/10.1186/s12879-018-3073-1>
- [8] Drummond, F.; Snowball, J.: Cultural Clusters as a Local Economic Development Strategy in Rural Small Town Areas, The Sarah Baartman District in South Africa, *Bulletin of Geography Socio-economic Series*, 43(43), 2019, 107-119. <https://doi.org/10.2478/bog-2019-0007>
- [9] Ferretti, V.; Gandino, E.: Co-designing the solution space for rural regeneration in a new World Heritage site, A Choice Experiments approach, *European Journal of Operational Research*, 268(3), 2018, 1077-1091. <https://doi.org/10.1016/j.ejor.2017.10.003>
- [10] González Díaz, J. A.; Celaya, R.; Fernández García, F.; Osoro, K.; Rosa García, R.: Dynamics of rural landscapes in marginal areas of northern Spain, Past, present, and future, *Land Degradation & Development*, 30(2), 2019, 141-150. <https://doi.org/10.1002/ldr.3201>
- [11] Herrero-Jáuregui, C.; Arnaiz-Schmitz, C.; Herrera, L.; Smart, S. M.; Montes, C.; Pineda, F. D.; Schmitz, M. F.: Aligning landscape structure with ecosystem services along an urban-rural gradient, Trade-offs and transitions towards cultural services. *Landscape Ecology*, 34(7), 2019, 1525-1545. <https://doi.org/10.1007/s10980-018-0756-3>
- [12] Hulko, W.; Hovanes, J.: Intersectionality in the lives of LGBTQ youth: Identifying as LGBTQ and finding community in small cities and rural towns, *Journal of Homosexuality*, 65(4), 2018, 427-455. <https://doi.org/10.1080/00918369.2017.1320169>
- [13] Kaindoa, E. W.; Finda, M.; Kiplagat, J.; Mkandawile, G.; Nyoni, A.; Coetzee, M.; Okumu, F. O.: Housing gaps, mosquitoes and public viewpoints: a mixed methods assessment of relationships between house characteristics, malaria vector biting risk and community perspectives in rural Tanzania, *Malaria journal*, 17(1), 2018, 1-16. <https://doi.org/10.1186/s12936-018-2450-y>
- [14] Keane, M.; Chen, Y.: Entrepreneurial solutionism, characteristic cultural industries and the Chinese dream, *International Journal of Cultural Policy*, 25(6), 2019, 743-755.
- [15] Lin, C. L.: Establishing environment sustentation strategies for urban and rural/town tourism based on a hybrid MCDM approach, *Current Issues in Tourism*, 23(19), 2020, 2360-2395. <https://doi.org/10.1080/13683500.2019.1642308>
- [16] Long, H.; Zhang, Y.; Tu, S.: Rural vitalization in China: A perspective of land consolidation, *Journal of Geographical Sciences*, 29(4), 2019, 517-530. <https://doi.org/10.1007/s11442-019-1599-9>
- [17] Molnárová, K. J.; Skřivanová, Z.; Kalivoda, O.; Sklenička, P.: Rural identity and landscape aesthetics in exurbia, Some issues to resolve from a Central European perspective, *Moravian Geographical Reports*, 25(1), 2017, 2-12. <https://doi.org/10.1515/mgr-2017-0001>
- [18] Ngugi, H. N.; Mutuku, F. M.; Ndenga, B. A.; Musunzaji, P. S.; Mbakaya, J. O.; Aswani, P.; LaBeaud, A. D.: Characterization and productivity profiles of *Aedes aegypti*, breeding habitats across rural and urban landscapes in western and coastal Kenya, *Parasites & Vectors*, 10(1), 2017, 1-12. <https://doi.org/10.1186/s13071-017-2271-9>
- [19] Pinto-Correia, T.; Almeida, M.; Gonzalez, C.: Transition from production to lifestyle farming: new management arrangements in Portuguese small farms, *International Journal of Biodiversity Science, Ecosystem Services & Management*, 13(2), 2017, 136-146. <https://doi.org/10.1080/21513732.2017.1329753>

- [20] Ramsey, D.; Malcolm, C. D.: The importance of location and scale in rural and small town tourism product development: The case of the Canadian Fossil Discovery Centre, Manitoba, Canada. *The Canadian Geographer/Le Géographe Canadien*, 62(2), 2018, 250-265. <https://doi.org/10.1111/cag.12409>
- [21] Yi, Z.; Fei, Z.; Ruiqin, L.: The Path of Rural Revitalization in Rapidly Urbanizing Areas: Taking Southern Jiangsu as an Example, *China City Planning Review*, 27(4), 2018, 24-33.
- [22] Zhao, T.; Zhao, Y.; Li, M. H.: Landscape performance for coordinated development of rural communities & small-towns based on ecological priority and all-area integrated development, Six case studies in East China's Zhejiang Province, *Sustainability*, 11(15), 2019, 4096. <https://doi.org/10.3390/su11154096>