www.gi.sanu.ac.rs, www.doiserbia.nb.rs
J. Geogr. Inst. Cvijic. 2020, 70(2), pp. 115–128

Original scientific paper

UDC: 911.372.7:711.3
https://doi.org/10.2298/IJGI2002115G

Received: May 8, 2020
Reviewed: July 2, 2020
Accepted: July 31, 2020

METHODS FOR CALCULATION THE SPATIAL DISTRIBUTION OF THE TERRITORY MEMBERSHIP TO THE URBANIZED BASED ON THE HYBRID NEURAL NETWORK

Bogdan Gats[1], Yaroslav Vyklyuk[1,2], Artur Horbovyy[3]*

[1]Bukovinian University, Faculty of Information Technologies and Economics, Department of Computer Systems and Technologies, Chernivtsi, Ukraine; e-mail: gatsbn@gmail.com
[2]Wenzhou University, College of Mechanical and Electrical Engineering, Institute of Laser and Optoelectronic Intelligent Manufacturing, Wenzhou, China; e-mail: vyklyuk@ukr.net
[3]National University of the State Fiscal Service of Ukraine, Institute of Information Technologies, Irpin, Ukraine; e-mail: horbovyy.Artur@gmail.com

Abstract: This paper presents the method for calculation of the spatial distribution of the territory membership to the urbanized (TMU)—the value that puts the degree of suitability for development (tourism development) in accordance with each node of the grid of the study area within (0, 1). The method is based on a hybrid neural network training. It allows us to assess all kinds of territories for tourist infrastructure development capabilities with relation to attractiveness and provide results visualization on the geographical information system (GIS) maps. The developed algorithm of picking and conversion of geospatial data from GIS for knowledge base generation uses attractors coordinates (vectors of roads, a city center, a railway station) and random points of the explored area, makes possible distance calculation between them (on the road to the attractor and Manhattan distance) and following conversion to the ASCII file, that allows unifying input parameters of the set of models for forecasting development of tourist infrastructure objects. The paper studied typical tourist towns of the Ukrainian Carpathians.

Keywords: spatial distribution of the territory membership to the urbanized; GIS; fuzzy logic

Introduction

Modeling and forecasting of the spatial form infrastructure evolution for various organizations, businesses, and institutions especially tourism-oriented are the key administrative goal not only for big cities but for different big and small settlements and territories. The necessity for solving this task is the following. Firstly, understanding of harmony, appropriateness, and efficiency of their evolution allows implementing scientific recommendations about the further development of the infrastructure objects. Secondly, tourist settlements are distinct by spontaneous stochastic development, i.e., infrastructure development occurs without general planning. Finally, the substantial role in tourist industry development is played by the territory type peculiarity as one of the most important factors of infrastructure development. The existence of attractive places and their usage as tourist attractions

*Corresponding author, e-mail: gatsbn@gmail.com

determines the necessity of the development of supporting infrastructure and investment in new facilities.

The solution to the problem of forecasting the development of the spatial form of tourist infrastructure objects is to use data obtained from geographic information systems (GIS) and on their basis to build appropriate mathematical models. For those purposes, manually getting data samples or developing a software add-on to GIS would implement the selection and processing of data. However, in this way, it is almost impossible to solve the task in a timely and effective manner. During the past years with the rapid technological development of remote sensing (RS), GIS and geospatial big data, numerous methods have been developed to identify urban land use.

Object-oriented classification (OOC) is used to extract urban land use patterns from high spatial resolution images through the physical features of ground objects such as spectral, shape and texture features (Deng, Zhu, He, & Tang, 2019; Lu et al., 2018). However, without considering spatial relationships among ground objects, OOC methods can only recognize land cover information with low-level semantic features. Points-of-interest (POIs) have been widely used to extract information pertaining to urban land use types and functional zones (Koohpayma, Tahooni, Jelokhani-Niaraki, & Jokar Arsanjani, 2019; Yu, Ai, He, & Shao, 2016). However, it is difficult to quantify the relationship between spatial distributions of POIs and regional land use types due to a lack of reliable models. Previous methods may ignore abundant spatial features that can be extracted from POIs. Yao et al. (2017) found an innovative framework that detects urban land use distributions at the scale of traffic analysis zones (TAZs) by integrating Baidu POIs, and a Word2Vec model was established. In the study of Ragagnin Pimentel and Pereira (2020), thematic cartography is used to explore and analyze tourism spatial distribution in Brazil.

Despite the various studies mentioned above, there is no specialized software for forecasting the spatial form of the infrastructure of tourist settlements. Obviously, the solution to this problem is possible with the use of advanced mathematical apparatus, modern methods of modeling socio-economic processes, including tourism, the use of GIS and the introduction of new information technologies (IT). Therefore, the development of methods for calculation of the spatial distribution of the territory membership to the urbanized, designed to predict the spatial development of tourist infrastructure, the spatial shape of tourist settlements and the dynamics of their growth, as well as determining the attractiveness of tourism development is an object of this study.

Methods of obtaining input parameters from geographic information systems

Features of determining the input parameters of the model of the spatial distribution of the territory membership to the urbanized

Predicting changes in the spatial shape of urban infrastructure is based on the use of appropriate models, where one of the parameters is the spatial distribution of the territory membership to the urbanized. It is a value that puts the degree of suitability for construction (tourism development) in accordance with each node of the grid of the study area within [0, 1]. For its calculation, GIS fuzzy logic is used, and the distances to existing infrastructure facilities are calculated. It is known that in an isotropic environment the development of a settlement infrastructure occurs symmetrically (Batty & Longley, 1994). Under real conditions the city is deformed by symmetry and the main determinants of anisotropy are roads and the existing infrastructure.

Considering the specifics of the Carpathian region with the absence of airports, motorways, small tourist settlements and the presence of access roads, the model of spatial distribution of territory membership to the urbanized can be represented as follows (1):

$$F_{ty}^{ur} : Ter \cup X \to Ty,\, Ty = \left\{\left(\xi_i, t_i\right)\right\}_{i=\overline{1,k_a}} \tag{1}$$

where ξ_i is an object for which an urbanization type is defined, $t_i \in \left\{0, 1, \ldots, N_{ty}\right\}$, N_{ty} is the number of urbanization types.

The selected parameters x_1—distance to the city center, x_2—distance to the nearest road, x_3—distance to the nearest railway station were chosen due to the fact that the distance to the attractors (ski lifts, historical centers) plays an important role in predicting the infrastructure of tourist settlements. Another development factor is transport networks (roads and railways). It is known that the development of settlement infrastructure (especially those specializing in tourism) tends to the roads. This is confirmed by numerous photographs of settlements taken from space. With the increasing distance to roads, the attractiveness of the territory development decreases (Romão, Kourtit, Neuts, & Nijkamp, 2018).

The feature of the Carpathian region is the presence of asphalt roads and soil tracks, which are important for the infrastructure of tourist settlements.

Locally-focused sites (healing waters, historical and cultural centers, ski lifts, parks, etc.) play the role of attractiveness centers (Munteanu et al., 2017). The input parameters for determining the metrics of attractiveness are not the coordinates of these objects, but the distance to them. By means of segments or multi-line long objects, e.g., roads, river channels are defined.

The above parameters can be obtained from GIS. If there is a minimum distance to the road, a perpendicular to each of the segments should be drawn and a minimum value among them will be found.

Let us take the equation of a line passing through two points (x_1, y_1), (x_2, y_2):

$$kx - y - kx_1 + y_1 = 0 \tag{2}$$

where $A = k$, $B = -1$, $C = y_1 - kx_1$. Then the distance from M (x_0, y_0) to a line passing through points (x_1, y_1), (x_2, y_2), will be the following:

$$d = \min_{1 \le i \le n} d_i = \min_{1 \le i \le n} \frac{\left|kx_0 - y_0 - kx_1 + y_1\right|}{\sqrt{k^2 + 1}} \tag{3}$$

The following transformations are required to determine the distance of the road to the attractor and the distance to the road. Let the roads of the studied territory be set by an array (Vyklyuk, 2009):

$$w_f\left(x_{f1}, y_{f1}, x_{f2}, y_{f2}\right), f = \overline{1,n} \tag{4}$$

where n is the number of road line segments in the study area, x_{f1}, y_{f1}, x_{f2}, y_{f2} are the coordinates of the line segment of the road.

Then the distance h for the point with coordinates x, y to the nearest road is determined according to the following considerations: consider a triangle with vertices A (x, y), B (x_{f1}, y_{f1}), C (x_{f2}, y_{f2}) (Figure 1). The length of its sides is defined as a, b, and c, respectively.

The height of the triangle on the line which the side a belongs to (segment of the road) is:

$$h_f' = \frac{2\sqrt{p(p-a)(p-b)(p-c)}}{a} \tag{5}$$

where p is the half-perimeter. The segments a_1 and a_2 that determine the distance between the base of altitude to vertices B and C the base of the base of altitude $\left(x_h', y_h'\right)$ are defined as:

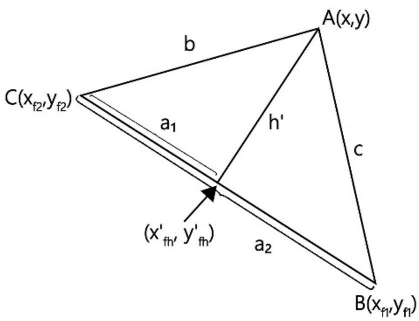

Figure 1. Triangle coordinates to determine the height and coordinates of the height basis.

$$a_1 = \sqrt{b^2 - h_f'^2} \tag{6}$$

$$a_2 = \sqrt{c^2 - h_f'^2} \tag{7}$$

$$x_{fh}' = x_{f1} - \left(x_{f1} - x_{f2}\right)\frac{a_1}{a} \tag{8}$$

$$y_{fh}' = y_{f1} - \left(y_{f1} - y_{f2}\right)\frac{a_1}{a} \tag{9}$$

The height (5) is the shortest distance to the road if the point $\left(x_{fh}', y_{fh}'\right)$ lies on a section of the road f. Otherwise, the shortest distance to the road will be determined as follows:

$$h_f = \begin{cases} a_1 + a_2 = a, h_f' \\ a_1 + a_2 > and < c, b \\ a_1 + a_2 > and > c, c \end{cases} \tag{10}$$

The coordinates of the intersection point are as follows respectively:

$$\left(x_{fh}, y_{fh}\right) = \begin{cases} a_1 + a_2 = a, \left(x_{fh}', y_{fh}'\right) \\ a_1 + a_2 > and < c, \left(x_1, y_1\right) \\ a_1 + a_2 > and > c, \left(x_2, y_2\right) \end{cases} \tag{11}$$

Then the distance to the nearest road is defined as follows:

$$h = \min_{f=1,n}\left(h_f\right) \tag{12}$$

In the case of vector coordinates describing point, linear or spatial objects, a series of modifications to the standard Pythagorean theorem can be used to determine distances. The original formula for determining the Euclidean or rectilinear distance between two points is (Figure 2a):

$$d_{ij} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2} \tag{13}$$

where $(X_i - X_j)$ is the distance along the X axis or longitude, $(Y_i - Y_j)$ is the distance along the Y axis or latitude, d_{ij} is the distance between two points.

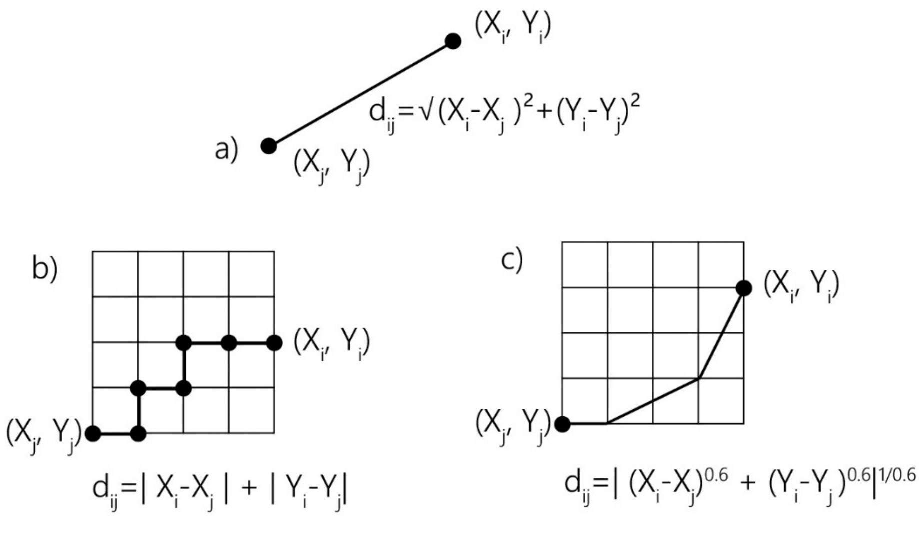

Figure 2. Distance measurement in a vector data model: a) Euclidean distance, b) Manhattan distance, c) the distance around the barrier.

If there is some obstacle that does not allow to move in a straight line, then the formula (13) can be generalized to a non-Euclidean form (Figure 2b):

$$d_{ij} = [(X_i - X_j)^k + (Y_i - Y_j)^k]^{\frac{1}{k}} \tag{14}$$

where power 2 is replaced by *k*, which is some numerical parameter.

In case of determining the distance between two points in a tourist town ("Manhattan distance"), where the houses and areas enclosed by a fence limit the traffic of segments at right angles to each other, the variable *k* is 1, and the formula for the distance takes the form:

$$d_{ij} = \left|X_i - X_j\right| + \left|Y_i - Y_j\right| \tag{15}$$

To determine the shortest distance around a barrier (lake, forest), the variable *k* takes the value 0.6 (Figure 2c). Any distance in a vector database can be determined by changing the parameter *k* in accordance with the existing conditions.

Methods for calculating and generating the spatial distribution of the territory membership to the urbanized

The methods for calculating the functional relationship between the parameters of model (1) and the objective function may be different: the weight matrix method (White & Engelen, 1993), the SLEUTH model (Crols et al., 2017), multiple criteria estimation (Serasinghe Pathiranage, Kantakumar, & Sundaramoorthy, 2018), logistic regression (El-Hallaq & Habboub, 2015), fuzzy networks (Wu, 2002), decision trees (Shafizadeh-Moghadam, Tayyebi, & Helbich, 2017).

Data Mining methods are used to find hidden dependencies in the raw data, which is the process of obtaining previously unknown, non-trivial, practically useful and accessible interpretations of the knowledge needed to make decisions in various fields of human activity. In view of this, the hybrid network ANFIS (Herold, Goldstein, & Clarke, 2003) and method for determining the degree of a territory's membership to urbanized based on fuzzy logic fuzzy inference systems (Chen et al., 2019) developed by the authors were used in this paper. The method is a sequence of the following steps.

Step 1. Forming a training set according to the following algorithm. A range of 3–5 kilometers was determined around the specified towns. An arbitrary point is chosen stochastically from this area, then its coordinates are determined, the distance to the roads, railway station, city center and the degree of urbanization, which is the initial variable of the training set. These values express the numerical characteristics of the study area; in the presence of the specified attribute, they take the value 1, and in the absence, they take 0. Such an algorithm is extended to the required number of times, resulting in a training set of any size. The training sample is a set and presented as a tuple $p = \langle x_1, x_2, x_3 \rangle$, where x_1 – distance to the city center, x_2 – distance to the nearest road, x_3 – distance to the railway station. The training set generation and calculation were performed using the software module written in MapBasic programming language for MapInfo Professional. The calculations used the training set containing 5,000 values. The training set is a numerical dimension matrix $m \cdot (n + 1)$, in which the number of rows *m* corresponds to the set volume, first *n* columns correspond to the values of the input variables of the model (1), and the last column corresponds to the value of the output variable of the territory membership to the urbanized of the model (1).

The base of training example were spatial structures of the main tourist towns of the Ukrainian Carpathians: Vorokhta village, Mygovo, Slavske, Polyanytsya, and Yaremche town. This choice was due to the fact that these towns are typical tourist centers of the Carpathian region. In addition, the cities belong to one region, are shaped and developed similarly to each other, are characterized by similar growth processes, the population has a similar mentality. The formation of the training set from several cities is aimed at finding the hidden dependencies in the dynamics of infrastructure development of the Ukrainian Carpathian tourist towns.

Several ski trails in the European scale were laid in Slavske (the main ones are Trostyan, Pogar, Polytech, FMI). They have one chairlift and 15 surface lifts. The trails are of any level of difficulty except the black ones (super complex). Slavske has a well-developed leisure infrastructure with numerous hotels, guesthouses, private estates.

Vorokhta is one of the main tourist centers of Ivano-Frankivsk region, both in summer and winter. There is a complex of four ski jumps that have the artificial surface and function throughout the year, as well as a cable lift (with the length of 2 km). Not far from the village there is a sports and tourist base "Zaroslyak", where the route to the highest peak of Ukraine, i.e., Hoverla, begins.

Yaremche is the most famous tourist center of Prykarpattya, has more than 40 tourist-recreational establishments and sanatoriums, more than 50 objects of green tourism. Climate treatments, mineral baths, etc. are used for spa therapy. The Carpathian National Nature Park, established in 1980 (the first in Ukraine), is located on almost half of the territory of the Yaremche City Council.

On the territory of Polyanytsya village there is Bukovel, the largest and most modern ski complex in Ukraine—16 lifts operate here, and it has about 60 kilometers of trails (blue, red and black ones). The capacity of all the lifts is about 30,000 people per hour. At the same, time on the slopes 12,000 people can be comfortably accommodated. The resort has developed the infrastructure (motels, multi-store, car parks, and restaurants).

In Mygovo the ski-tourist complex is located. The total length of the trails is more than 3,340 m, with three elevators and two multi-lifts function. Mygovo has a well-developed leisure infrastructure.

Step 2. Training a fuzzy hybrid neural network. In order to calculate the degree of the territory membership to the urbanized, a fuzzy Sugeno-type system was built using the ANFIS editor of the MATLAB package. Three input variables are described by three fuzzy terms and 27 fuzzy rules are formed. The fuzzy inference requires the fuzzy knowledge base:

$$\begin{gathered}\left(x_1 = \tilde{a}_{1j} \cap_j x_2 = \tilde{a}_{2j} \cap_j x_3 = \tilde{a}_{3j}\right) \Rightarrow \\ y = b_{j0} + b_{j1}x_1 + b_{j2}x_2 + b_{j3}x_3,\ j = \overline{1,27}\end{gathered} \tag{16}$$

where $\tilde{a}_{1j}$ is a fuzzy term which evaluates the variable x_i in j-th rule, $j = \overline{1,27}$, $\cap j$ is logical operation «AND» which binds the antecedents of the j-th rule, $\Rightarrow$ is a fuzzy implication, $b_{j0}, b_{j1}, b_{j2}, b_{j3}$ are some real numbers.

The rule conclusions are set by the linear function from the inputs:

$$d_j = b_{j0} + \sum_{i=\overline{1,3}} b_{ji} x_i \tag{17}$$

The degrees of membership for input vector $X^* = (x_1^*, x_2^*, x_3^*)$ to values $d_j = b_{j0} + \sum_{i=1}^{3} b_{ji} x_i$ are calculated as follows:

$$\mu_{d_j}(X^*) = \mu_j\left(x_1^*\right)\chi_j \mu_j\left(x_2^*\right)\chi_j \mu_j(x_3^*)\chi_j,\ j = \overline{1,27} \tag{18}$$

The output of the entire knowledge base is a fuzzy set $\tilde{y}$ corresponding to the input vector X^*:

$$\tilde{y} = \left(\frac{\mu_{d_1}(X^*)}{d_1}, \frac{\mu_{d_2}(X^*)}{d_2}, \frac{\mu_{d_3}(X^*)}{d_3}\right) \tag{19}$$

To obtain the output value of *y*, the fuzzy set is defuzzyficated and the weighted average is calculated:

$$y = \frac{\sum_{j=\overline{1.27}} \mu d_j(X^*) d_j}{\sum_{j=\overline{1.27}} \mu d_j(X^*)} \tag{20}$$

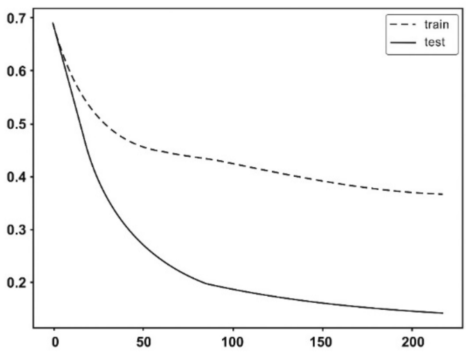

Figure 3. Train and test loss during training.

The criterion for the accuracy of training is the mean-square error. A data set was used for training, which is divided into a training and a test set in the ratio of 70% to 30%. The criterion for stopping learning is the point where the error of the test data set stops improving (Figure 3).

Step 3. Formation of the fuzzy inference systems. The fuzzy inference system contains three input variables, one output variable, each of which is described by three terms. To create the fuzzy inference, the grid partition method is used, according to which membership functions of fuzzy terms are evenly distributed within the data change range. The hybrid training method with an error tolerance of zero was used. The fuzzy system was trained during 25 iterations. In this work, the fuzzy inference systems with different types of membership functions were investigated: Gaussian, bell, triangular and trapezoidal. The Gaussian membership function is formed using the Gaussian (normal) distribution with the assumption that $\sqrt{2\pi}\sigma$ = 1 and is analytically represented as follows:

$$\left(x;\sigma,c\right)=e^{\frac{-(x-c)^2}{2\sigma^2}} \tag{21}$$

where σ and c are numerical parameters, the square of the first σ^2 is called the variance of the distribution and c is the mathematical expectation.

Bell-shaped membership function is given analytically as follows:

$$f(x;a,b,c)=\frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \tag{22}$$

where *a, b, c* are the numeric parameters that take arbitrary real values and are ordered by relation: $a < b < c$ moreover, the parameter $b > 0$.

The triangular membership function is analytically defined by the following expression:

$$f(x;a,b,c)=\begin{cases}0, x<a\\ \frac{x-a}{b-a}, a\le x\le b,\\ \frac{c-x}{c-b}, b\le x\le c,\\ 0, c\le x.\end{cases} \tag{23}$$

where *a, b, c* are the numeric parameters that take arbitrary real values and are ordered by relation: $a \le b \le c$.

The trapezoidal membership function is set analytically as follows:

$$f(x;a,b,c,d)=\begin{cases}0, x\le a\\ \frac{x-a}{b-a}, a\le x\le b,\\ 1, b\le x\le c,\\ \frac{d-x}{d-c}, c\le x\le d,\\ 0, d\le x.\end{cases} \tag{24}$$

where *a, b, c, d* are the numeric parameters that take arbitrary real values and are ordered by relation: $a \le b \le c \le d$.

To symbolically denote the basic terms of linguistic variables, we will use commonly accepted abbreviations (PS: positive small, PM: positive average, PB: positive large). The term-sets of the input variables "Distance to the city center", "Distance to the railway station", "Distance to the nearest road" is the set $T_1 = T_2 = T_3$ = {"close", "moderate", "far"} or symbolically $T_1 = T_2 = T_3$ = {"PS", "PM", "PB"}. The parameters and type of membership functions for the input variables of the model received in the training process are presented in tables 1–3.

Table 1
Parameters of membership function for input variable "Distance to the city center"

Column description	Term	Parameters of membership function
Gaussian	PS	[0.925 0.02424]
	PM	[0.9898 2.336]
	PB	[1 4.676]
Bell	PS	[1.12 2.007 0.02143]
	PM	[1.183 2.001 2.317]
	PB	[1.171 2 4.677]
Trapezoidal	PS	[−1.538 −0.617 0.7782 1.746]
	PM	[0.7602 1.7 3.067 3.991]
	PB	[3.008 3.988 5.372 6.293]
Triangular	PS	[−2.229 0.07656 2.326]
	PM	[0.07159 2.378 4.683]
	PB	[2.353 4.679 6.984]

Table 2
Parameters of membership function for input variable "Distance to the railway station"

Column description	Term	Parameters of membership function
Gaussian	PS	[2.14 0.0016]
	PM	[2.136 5.055]
	PB	[2.135 10.1]
Bell	PS	[1.171 2 4.677]
	PM	[2.525 2.002 5.053]
	PB	[2.522 2.001 10.1]
Trapezoidal	PS	[−1.128 −0.4727 0.513 1.173]
	PM	[0.5089 1.169 2.144 2.81]
	PB	[2.114 2.8 3.789 4.444]
Triangular	PS	[−1.62 0.02238 1.648]
	PM	[0.01566 1.661 3.3]
	PB	[1.529 3.294 4.936]

Table 3
Parameters of membership function for input variable “Distance to the nearest road”

Column description	Term	Parameters of membership function
Gaussian	PS	[0.7945 0.00844]
	PM	[0.8191 1.918]
	PB	[0.8629 3.81]
Bell	PS	[0.9475 2 0.005682]
	PM	[1.011 1.998 1.879]
	PB	[0.9778 1.996 3.814]
Trapezoidal	PS	[−1.3 −0.5405 0.5988 1.366]
	PM	[0.5917 1.358 2.493 3.261]
	PB	[2.469 3.252 4.396 5.155]
Triangular	PS	[−1.869 0.03289 1.952]
	PM	[0.02514 1.929 3.828]
	PB	[1.89 3.824 5.724]

As a result of the training, 27 rules of fuzzy products are formed, which are described in Table 4.

Step 4. Calculation of the attractiveness matrix of the studied area and its fuzzy inference. The resulting fuzzy inference systems form the basis for constructing spatial distributions of the territory membership to the urbanized. For this purpose, the method of construction of recreational potentials maps was used (Ontiveros-Robles, Melin, & Castillo, 2019). Map of the territory *T* is covered by a rectangle *Π* = [*a*, *b*] · [*c*, *d*]. Obviously, the rectangle *Π* contains the set (territory) *T* (*T* ⊂ *Π*).

Table 4
A fuzzy rule base for fuzzy inference systems

Distance to the city center	Distance to the railway station	Distance to the nearest road	Degree of the TMU (Gaussian)	Degree of the TMU (Bell)	Degree of the TMU (Trapezoidal)	Degree of the TMU (Triangular)
PS	PS	PS	1	0.7946	0.9678	1
PS	PS	PM	0.01009	0	0.4993	0.5752
PS	PS	PB	1	1	0	1
PS	PM	PS	1	1	0.8382	1
PS	PM	PM	0	0.45552	0.2961	0
PS	PM	PB	0	0	0	0
PS	PB	PS	0	0.8318	0.1988	0.07919
PS	PB	PM	1	1	1	1
PS	PB	PB	0	0	0	0
PM	PS	PS	0.8182	0.7723	0.5157	0.7555
PM	PS	PM	0	0.1472	0.1156	0.08501
PM	PS	PB	0	0	0.07971	0.07224
PM	PM	PS	0	0.4819	0	0
PM	PM	PM	0.04031	0.0551	0.07995	0
PM	PM	PB	0.4775	1	0.01572	0.03573
PM	PB	PS	0	1	0.04309	0
PM	PB	PM	1	0.8374	0.3348	0.71
PM	PB	PB	1	1	1	1
PB	PS	PS	0	0	0.6864	0
PB	PS	PM	1	1	0.2733	1
PB	PS	PB	0	0	0.03523	0
PB	PM	PS	1	1	0.1674	0.695
PB	PM	PM	0	0	0	0
PB	PM	PB	0.1522	0.04285	0.01209	0.799
PB	PB	PS	0	0	0	0
PB	PB	PM	0	0	0	0
PB	PB	PB	0	0.3151	0	0.7014

The rectangle *П* is broken by a grid $\Delta = \Delta_x \cdot \Delta_y$, where (*a*, *c*), (*b*, *d*) are the coordinates of opposite vertices of the rectangle; *П*, Δ_x, Δ_y are the sizes of the edges of the grid, $\{x_k\}$, $\{y_l\}$ are the coordinates of grid nodes, *N*, *M* are the number of parts to which the bound of the rectangle *П* is broken [*a*, *b*] and [c, *d*], accordingly.

$$\Delta_x = \bigcup_{k=0}^{N} \{x_k\} \tag{25}$$

$$\Delta_y = \bigcup_{l=0}^{M} \{y_l\} \tag{26}$$

For each grid node, the values of the input parameters of the model of spatial distribution of the territory membership to the urbanized (1) are determined. The result of the calculation is a matrix of size 1,000,000 × 3. The resulting matrices were used as arguments of the evalfis function of the FuzzyLogic Toolbox package, with which fuzzy inference was performed in the built fuzzy inference systems. At the output, matrices were obtained showing the degree of the territory membership to the urbanized for each of the nodes, and corresponding spatial distribution graphs were built (Witten & Sander, 1983). A schematic representation of the method is shown in Figures 4 and 5.

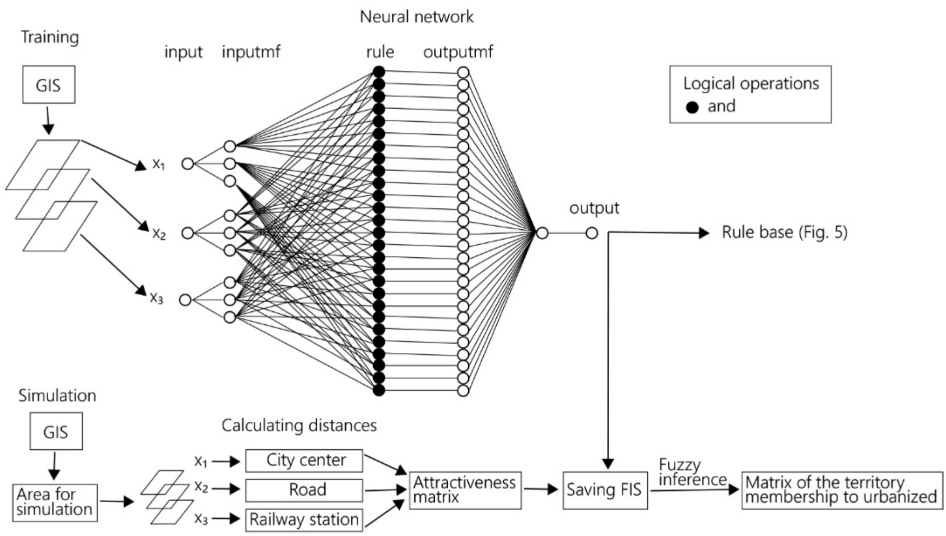

Figure 4. Schematic representation of the method of determining the degree of the territory membership to the urbanized.

The Figure 4 shows a schematic implementation of the method of determining the degree of the territory membership to the urbanized, which consists of two stages: training a fuzzy hybrid neural network and calculating the matrix of the territory membership to the urbanized. The training stage contains the first three steps of the described method, the calculation stage—steps 3 and 4.

The algorithm for the selection and conversion of geospatial data from GIS was developed, namely, the coordinates of the attractors (vectors of the roads, city center, railway station) and arbitrary points of the study area on the basis of an exchange file, which transmits the calculated distances between the elements (the path to the attractor, “Manhattan” distance). The conversion of the calculated distances to ASCII format made it possible to unify the input parameters of many models of forecasting the development of tourist infrastructure. The algorithm is a sequence of steps:

Step 1. Selection in the MapInfo GIS the prediction area, that is, the border of the left upper and right lower areas, the coordinates of which are written as rectangular coordinates in an intermediate text file. The generated text file contains only rectangular coordinates of the prediction area and is used for intermediate calculations. With each new area selection, the file entries change accordingly.

Step 2. Determination of the roads’ coordinates. Since the map is a set of layers with different semantic content, the layers “Roads” and “Highways” were used to determine the coordinates of the roads. The determined coordinates of the study area are recorded in separate GIS MapInfo tables.

Step 3. Determination of the attractors’ coordinates. In this step, the coordinates of the railway station of the study area (the layer “Station”) and the coordinates of the city center (layers “Cities”, “Villages”) are determined. The coordinates of the attractors are written to the MapInfo intermediate tables.

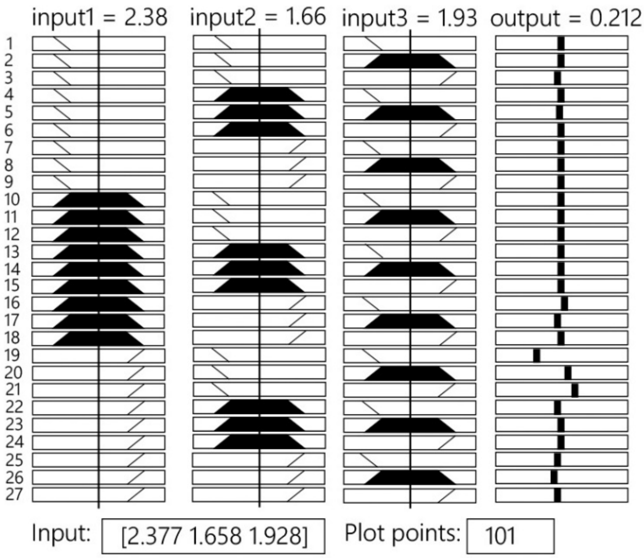

Figure 5. Rule base.

Step 4. Calculating the distances to the attractors and vectors of the roads. The selection of the forecasting area is accompanied by the partitioning of it by the grid (29–30). For each node of the grid, a distance calculation is performed, with subsequent recording into one separate MapInfo GIS table.

Step 5. Forming the exchange file. After performing the distance calculation, SQL selects values from the MapInfo result table and the table is exported to an ASCII Matlab data exchange file that is used to pass the input parameters in the tourism infrastructure development forecasting model.

Conclusion

The paper identifies major factors that influence the development of infrastructure for cities and small tourist settlements. Taking into account the specificity of the Carpathian region, methods for calculation and generating the spatial distribution of the territory membership to the urbanized is presented. The developed algorithm of picking and conversion of geospatial data from GIS for the knowledge base generation and the method for determining the spatial distribution of the territory membership to the urbanized are based on the hybrid neural network.

The proposed methods give possibilities to obtain spatial forms of the main tourist settlements of the Carpathian region in the form of matrices, which is the main parameter of a number of models developed by urban planners to determine the harmony of urban development. Methods can be applied to settlements of different sizes and functions, which will help the relevant private and government organizations to determine the feasibility and effectiveness of territories and infrastructure development. The spatial distributions of the territory membership to urbanized allowed to solve the problem of determining the most attractive places for the development of tourist infrastructure. This, in turn, allows the decision-makers to draw conclusions about investing in new facilities, developing related facilities and building the existing infrastructure.

References

Batty, M., & Longley, P. (1994). *Fractal Cities*: *A Geometry of Form and Function.* London, UK: Academic Press.

Chen, W., Panahi, M., Khosravi, K., Pourghasemi, H. R., Rezaie, F., & Parvinnezhad, D. (2019). Spatial Prediction of Groundwater Potentiality Using ANFIS Ensembled With Teaching-Learning-Based and Biogeography-Based Optimization. *Journal of Hydrology*, *572*, 435–448. https://doi.org/10.1016/j.jhydrol.2019.03.013

Crols, T., Vanderhaegen, S., Canters, F., Engelen, G., Poelmans, L., Uljee, I., & White, R. (2017). Downdating High-Resolution Population Density Maps Using Sealed Surface Cover Time Series. *Landscape and Urban Planning*, *160*, 96–106. https://doi.org/10.1016/j.landurbplan.2016.12.009

Deng, Z., Zhu, X., He, Q., & Tang, L. (2019). Land Use/Land Cover Classification Using Time Series Landsat 8 Images in a Heavily Urbanized Area. *Advances in Space Research*, *63*(7), 2144–2154. https://doi.org/10.1016/j.asr.2018.12.005

El-Hallaq, M. A., & Habboub, M. O. (2015). Using Cellular Automata-Markov Analysis and Multi Criteria Evaluation for Predicting the Shape of the Dead Sea. *Advances in Remote Sensing*, *4*(1), 83–95. https://doi.org/10.4236/ars.2015.41008

Herold, M., Goldstein, N. C., & Clarke, K. C. (2003). The Spatiotemporal Form of Urban Growth: Measurement, Analysis and Modeling. *Remote Sensing of Environment*, *86*(3), 286–302. https://doi.org/10.1016/S0034-4257(03)00075-0

Koohpayma, J., Tahooni, A., Jelokhani-Niaraki, M., & Jokar Arsanjani, J. (2019). Spatial Analysis of Curb-Park Violations and Their Relationship with Points of Interest: A Case Study of Tehran, Iran. *Sustainability*, *11*(22), 6336. https://doi.org/10.3390/su11226336

Lu, D., Li, L., Li, G., Fan, P., Ouyang, Z., & Moran, E. (2018). Examining Spatial Patterns of Urban Distribution and Impacts of Physical Conditions on Urbanization in Coastal and Inland Metropoles. *Remote Sensing*, *10*(7), 1101. https://doi.org/10.3390/rs10071101

Munteanu, C., Radeloff, V., Griffiths, P., Halada, L., Kaim, D., Knorn, J., . . . Stych, P. (2017). Land Change in the Carpathian Region Before and After Major Institutional Changes. In G. Gutman & V. Radeloff (Eds.), Land-Cover and Land-Use Changes in Eastern Europe after the Collapse of the Soviet Union in 1991 (pp. 57–90). https://doi.org/10.1007/978-3-319-42638-9_4

Ontiveros-Robles, E., Melin, P., & Castillo, O. (2019). Comparative Analysis of Type-1 Fuzzy Inference Systems with Different Sugeno Polynomial Orders Applied to Diagnosis Problems. In R. B. Kearfott, I. Batyrshin, M. Reformat, M. Ceberio, & V. Kreinovich (Eds.), *Fuzzy Techniques: Theory and Applications* (pp. 453–465). https://doi.org/10.1007/978-3-030-21920-8_41

Ragagnin Pimentel, M., & Pereira, M. de L. (2020). Tourism Spatial Distribution in Brazil: A Recent Assessment Using Thematic Cartography. *Journal of Hospitality and Tourism Insights*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/JHTI-03-2019-0051

Romão, J., Kourtit, K., Neuts, B., & Nijkamp, P. (2018). The Smart City as a Common Place for Tourists and Residents: A Structural Analysis of the Determinants of Urban Attractiveness. *Cities*, *78*, 67–75. https://doi.org/10.1016/j.cities.2017.11.007

Serasinghe Pathiranage, I. S., Kantakumar, L. N., & Sundaramoorthy, S. (2018). Remote Sensing Data and SLEUTH Urban Growth Model: As Decision Support Tools for Urban Planning. *Chinese Geographical Science*, *28*(2), 274–286. https://doi.org/10.1007/s11769-018-0946-6

Shafizadeh-Moghadam, H., Tayyebi, A., & Helbich, M. (2017). Transition Index Maps for Urban Growth Simulation: Application of Artificial Neural Networks, Weight of Evidence and Fuzzy Multi-Criteria Evaluation. *Environmental Monitoring and Assessment*, *189*(6), 300. https://doi.org/10.1007/s10661-017-5986-3

Vyklyuk, Y. (2009). Crystal's fractal growth patterns in the fuzzy potential field for prognostication of socio-economic processes. *Information Technology for Economics & Management*, *5(1)*, 1. Retrieved from http://www.item.woiz.polsl.pl/issue5.1/pdf/Crystalsfractalgrowthpatterninthefuzzypotentialfieldforprognosticationofsocioeconomicprocesses.pdf

White, R., & Engelen, G. (1993). Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land-Use Patterns. *Environment and Planning A*, *25(8)*, 1175–1199. https://doi.org/10.1068/a251175

Witten, T. A., & Sander, L. M. (1983). Diffusion-limited Aggregation. *Physical Review B*, *27*(9), 5686–5697. https://doi.org/10.1103/PhysRevB.27.5686

Wu, F. (2002). Calibration of Stochastic Cellular Automata: The Application to Rural-Urban Land Conversions. *International Journal of Geographical Information Science*, *16*(8), 795–818. https://doi.org/10.1080/13658810210157769

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., & Mai, K. (2017). Sensing Spatial Distribution of Urban Land Use by Integrating Points-Of-Interest and Google Word2Vec Model. *International Journal of Geographical Information Science*, *31*(4), 825–848. https://doi.org/10.1080/13658816.2016.1244608

Yu, W., Ai, T., He, Y., & Shao, S. (2017). Spatial Co-Location Pattern Mining of Facility Points-Of-Interest Improved by Network Neighborhood and Distance Decay Effects. *International Journal of Geographical Information Science*, *31*(2), 280–296. https://doi.org/10.1080/13658816.2016.1194423