World Academy of Science, Engineering and Technology
International Journal of Humanities and Social Sciences
Vol:6, No:11, 2012

Spatial Analysis and Statistics for Zoning of Urban Areas

Benedetto Manganelli, and Beniamino Murgante

***Abstract*—**The use of statistical data and of the neural networks, capable of elaborate a series of data and territorial info, have allowed the making of a model useful in the subdivision of urban places into homogeneous zone under the profile of a social, real estate, environmental and urbanist background of a city. The development of homogeneous zone has fiscal and urbanist advantages. The tools in the model proposed, able to be adapted to the dynamic changes of the city, allow the application of the zoning fast and dynamic.

***Keywords*—**Homogeneous Urban Areas, Multidimensional Scaling, Neural Network, Real Estate Market, Urban Planning.

I. Introduction

THE model, built on a general draft, is made to respond, in an efficient way to the need of identify zones in the territory where there is a common ground to identify urban and environmental qualities.

The characteristic that the model refers to are the ones of the Real Estate. These characteristics that are in the proposed model are grouped into two different categories:

1. Characteristics morphological-environmental and normative;
2. Characteristics economical-structural.

The following characteristics belong to the first category:

- Land orography;
- The morphology of the building (typology and planimetric density);
- The presence of natural or artificial barriers(for example roads, rail roads, rivers and so forth);
- The presence of relevant point, entropic or natural features, (for example a park or a stadium and so on);
- The purpose of the land utilization.
- The economic and structural characteristics are the following:
- The infrastructures and facility equipments;
- Service equipments;
- The existing building typologies and their consistency;
- The condition of the buildings and the equipment;
- The condition of the private buildings;
- The level of the values and revenues.

This distinction is because of the way the model is, that provides a sequence of operative phases: the first phase proceeds from arguments deductible from the exam of variables that describe the morphologic-environmental and normative characteristics. Then there are two other phases of inductive analysis based on the elaboration of the data built upon the economic and structural characteristics.

The two different approaches- argumentative for the first phase, numeric for the others, require that the knowledge of the information has different levels of deepening. While the information concerning the morphological- environmental and normative characteristics do not need, in the development of the model, of a representation in “measure”, for the economic-structural characters it is imperative the construction of the ordinal scale measurement.

II. The Model Phases

A. Phase 1

The first phase there is a subdivision of the urban city, considered here, in a not regular grid that generates the so called “attitudinal areas”. The grid is built starting from a photogram survey reported on maps of appropriate scale (possibly no more than 1/2000), following points that usually are the same of anthropic elements of the urban building (railroads, roads and so forth) and the ones that belong to the natural territory (rivers, slopes, and so on.); these are elements that could create discontinuity. On the base of the arguments done on the morphologic-environmental- normative characters, the grid separates urban areas for which are recognized organized forms of the building for type or planimetric density, or they even circumscribe urban buildings around points (anthropic or natural) big enough to be considered and with possible relevance on assets and revenues. It is important that the areas found in the grid are homogeneous so that they can establish a preliminary subdivision of the territory that describes in details physical, economical and environmental characteristics of the land. For this reason, it is useful that the grid remark, when there are not enough informations from the morphological-environmental characters, the norms that establish the use of the land. Therefore, for this model phase, we can use more complex classification methods, such as combining GIS with multicriteria methods [1], [2].

B. Phase 2

Once we have created a map for the space in question, then we study the economic and structural characters. The phases following the construction of the map require the need of informations to be used with statistic-mathematics models. For

Benedetto Manganelli, School of Engineering, University of Basilicata, Viale dell’Ateneo Lucano 10, 85100, Potenza (phone: 39-0971-205171; fax: 39-0971-205185; e-mail: benedetto.manganelli@unibas.it).
Beniamino Murgante, School of Engineering, University of Basilicata, Viale dell’Ateneo Lucano 10, 85100, Potenza (e-mail beniamino.murgante@unibas.it).

this reason for every grid there is a map in which the variables, economic-structural characteristics, are measured on apposite scales, mostly ordinals.

The mesaurements of (p) economic- structural characters defines the qualitative-quantitative profile of every area in question. The parameters found constitute the elements of a numeric matrix A with demensions n x p, the rows describe the qualitative profiles that identify the single areas in question. This matrix is used to explore some statistic data to identify homogeneous groups among the units. Among the techniques of multivariate analysis of the data, has been used in this phase the Multidimensional Scaling (MDS), which has as its specific use just what exploratory, and in addition offers the advantage of being able to analyze data expressed on an ordinal scale model, not a metric one.

The MDS is able to treat any set of data that defines measures of similarity or dissimilarity between the units, and generate as output a geometric pattern that represents that information in a limited number of sizes.

The use of MDS therefore requires, if it is not already known, the determination of an index of dissimilarity (or similarity) able to approximate a measure of the mutual distance between the pairs of units (Id). The generic index Id_{ij}, between the units i and j the unit, can be constructed, starting from the initial matrix of data, as a function of the respective row vectors which, remember, define how ordinal - translated into appropriate numerical codes - the variables used to characterize the attitudinal areas. When the variables are expressed on an ordinal scale, and if at the same time one wants to take into account the correlation existing between the same variables (in this case it is likely the correlation), is usable as an index of dissimilarity the "Mahalanobis distance". The Mahalanobis distance, compared to other measures such as those Euclidean distance and the City Block distance, is a function of the entire set of sample units, therefore changes as this also remain unchanged when the carriers on which is calculated, while it is invariant for transformations of scale of the variables and for modifications of the weights of the same. This measure is obtained as the solution to the following expression:

$$Md_{ij} = c \cdot \left[(x_i - x_j)' \cdot S^{-1} \cdot (x_i - x_j) \right]^{1/2}, \quad (1)$$

where x_i and xj are vectors which define the quality profile of units, S is the covariance matrix and c is a correction factor which takes into account the number n of sample observations, calculated with the following:

$$c = \sqrt{\frac{n}{n-1}}. \quad (2)$$

The index of dissimilarity, defined by the calculation of the Mahalanobis distance, must however corrected in function of a latent variable, not detectable and therefore not explicitly included in those that constitute the qualitative profile of the unit, but implied in the binding of spatial contiguity between the different attitudinal areas. “Everything is related to everything else, but near things are more related than distant things” [3].

We want so that the exploratory statistical analysis performed on the matrix of "profiles" (database) is "bound", for example that the configuration of the MDS output is generated by a measure of the distance "modified" by a bond of spatial contiguity. This constraint defines an additional indicator of similarity or dissimilarity that joins those obtained explicitly by p economic structural variables [10].

The distance "modified" (d *) is a function of the difference between the assumed mode of descriptive variables p is the effective mutual physical distance between the different units. The function takes the following form:

$$d^*_{ij} = f\left(Md_{ij}, \psi_{ij}\right), \quad (3)$$

where Md_{ij} is the Mahalanobis distance, ψ_{ij} is a measure of the geographic distance between the units i and j, and f is an increasing function of Md_{ij} and decreasing ψ_{ij}. In particular ψ_{ij} is zero if units i and j are not contiguous, mail is equal to the distance between the barycentres territorial units i and j when the latter are contiguous. For f takes the following algebraic expression:

$$f = ZMd_{ij} - \pi \cdot Z\psi_{ij}, \quad (4)$$

where Z means that the variables are standardized in the interval (0, 1), and π is the weight of the bond connected to the spatial contiguity. The weighting of the bond must be properly measured by comparing the results for different values imposed on the coefficient π.

The standardization is necessary to eliminate the distortions arising from the use of measurement units and orders of magnitude different for the two variables (Md and ψ); having also the function f to generate a measure of distance (logically positive) the standardization choice is the following:

$$ZX_i = \frac{X_i - \min}{\max - \min}. \quad (5)$$

For a comparison of possible approaches to be used to take into account the constraint of contiguity of statistical units in a two-dimensional [4], [5] and [6], [18].

After determining the distances "changed" between the territorial units of the same analysis performed with the technique of MDS - in the "non-metric model" - returns the configuration of the distribution of units in a limited number of sizes. The choice of the number of dimensions depends on the purpose for which the technique is employed, which is generally to obtain a simple representation of the units to search for any links between the same and / or give an interpretation to the size. The lower the number of dimensions, the lower the variance of the observed dissimilarity (distances modified) explained by the configuration obtained, up to affect the validity of the solution. The quality of the solution is measured with indices *stress*, *s-stress* and *RQS*. With reference to the *stress* you consider inadequate values higher than 20%, excellent values less than 5% [7], [17]. The two-dimensional representation is logically more simply interpretable.

C. Phase 3

The observation of two-dimensional configuration of the MDS output allows the identification of homogenous groups, that is to say groups of units in the graphic representation that are very close and therefore visually distinguishable.

These units are portions of micro-climates. Depending on the total number of units and the expected number of micro-climates, you can define the minimum number for the formation of a group.

Likewise in the configuration are identifiable very isolated from other single units, which by their strong dissimilarities define portions of micro zones other than those already identified. All other statistical units, for which it is possible to recognize the configuration belonging to a group or the other, are called "margin areas".

D. Phase 4

In this phase the model proposed, in order to confer on the margin areas of the class to which they belong, or in order to assign these areas to one of the groups already formed, employs a "neural network". Neural networks are, therefore, systems of nonlinear dynamic analysis and multidimensional capable of providing results comparable or superior to those for approximation of the more complex statistical methods, but at the same time that their use is extremely easy and fast because it does not require knowledge of an explicit model [8]. Neural networks learn it "on their own" to solve a problem, usually with the help of a series of examples (learning with supervisor) or self-organizing independently (unsupervised learning).

Another advantage offered by the use of neural networks is to be able to treat data "noisy" that is affected by measurement errors, incomplete or even incorrect.

The "noise" of the input data is damped by the sum of the partial processing that occur in individual neurons which have a degree of activation variable between a minimum and a maximum. For input slightly different, the network provides the same solution as the minimum differences in the level of activation of neurons affected in input are canceled in the propagation of the signal to the output unit.

This makes them particularly suitable application when the request is not a solution algebraically precise, but rather when both pursued the certainty (or "robustness") of the result. The best performances are obtained in the solution of problems of classification or prediction of the evolution of a phenomenon. In the training phase the structure of the network is defined, in other words, the estimated weights of the connections. In this phase are employed iterative processes that through successive adjustments of the weights come to define the network that best fits the data.. The training algorithms most commonly used are: back-propagation algorithm, genetic algorithms, competitive learning. It is instead the researcher choosing architecture and transfer functions.

There are types of standard networks widely experimented in the solution of specific problems and in particular in proceedings of classification of units described by a large set of parameters [9],[11], [14], [15]. That used in this case is the Multi-Layers Perception (MLP).

The multi-layers perception is a neural network of the kind fully connected multilayer (each unit is connected with all those of the next level) and feed forward (the signal propagation is unidirectional). The transfer function is, for simplicity, common to all the neurons, of non-linear type. The sigmoid function is chosen. Remains the choice of who conducts the analysis determining the number of hidden layers and the number of neurons which composes each layer. This choice does not come from any assumptions inherent in the analyzed phenomenon; also there are no strict criteria to define the optimal number of layers and hidden neurons.

It is derived instead from the specific experience gained in similar applications and from the general one, for which a greater number of neurons increases the ability to approximate functions very complex, but in turn increases the time for training and reduces the ability of the network to generalize. A large number of neurons produces a network too specialized on the data used in the training phase (overfitting). The network learns too much from the examples of the training set up to be interpreted as structural relationships of the phenomenon even those due to a component rather erratic.

The training in this type of network can only be supervised. The training algorithm is the back-propagation, which provides the iterative adjustment of the weights with the rule of the gradient. This rule is equivalent to a drop-down, along the surface of a parabola defined by the function of mean square error, in the direction of the maximum slope.

The MLP network is trained to recognize and store the ideal characteristics of the objects that make up the training set, that is, the prototypes of which are known to be the values of the explanatory variables and those variables explained (target). In the present case the training set is constituted by the attitudinal areas already classified in the second phase of the proposed model, the explanatory variables are those that constitute the profile qualitative-quantitative (economic and structural characteristics), the variable explained (nominal variable) is instead the class that is assigned to.

It should be noted that in this third phase is excluded from the indicator of similarity between the areas connected to the spatial contiguity constraint. This indicator, used for the definition of the prototypes, has not logically more utility in the classification of areas defined margin.

The training procedure involves two steps that are repeated in an iterative manner (forming later periods) at the end of the first (forward pass) associated with the network drives on the prototype class ideal estimated from the current weights on the second step (backward pass) the weights are adjusted so as to reduce the minimum error between the actual target of the training set and the estimated.

The calculated error identifies the direction and magnitude of the correction to be made to the weights of the units (neurons) output, which transmit the information of the error committed in all neurons of the previous layer, and so on, up to correct all the weights of the network.

Concluded training, the network is able to recognize for the single attitudinal area initially between those of margin - that for the network constitutes a combination of unknown parameters - which, between the classes ideal (or homogeneous zone) assigned to the prototypes, closest thing to it. This completes the classification. The result must be represented graphically with a map where each territorial unit (attitudinal area) is highlighted with a texture or a color corresponding to the class. The territorial partition is finally completed by assigning a different area to all attitudinal areas belonging to the same class and physically contiguous, as well as to those areas which are instead isolated. We can now complete the process of zoning.

III. Conclusion

The establishment of micro-zones homogeneous physical characteristics, economic and environmental administrations represents a fundamental moment in the government of the territory. Zoning makes it possible to formulate functional models for the management of tax revenues or planning, more consistent with the real transformation of the territory. The articulation of territory in homogeneous areas, facilitating for example the necessary procedures to Class Transfer of Property and defines, together with a parametric model for the calculation and updating of income, a flexible and dynamic mechanism capable of ensuring greater transparency and fairness of the tax system [16]. Zoning built upon the economic and structural characteristics is also critical step in the process of urban equalization [12].

The research, based on the statistical analysis of the data (Multidimensional Scaling) and the use of artificial intelligence (Neural Networks) has defined a model whose application - beyond the apparent complexity of the structure of the model itself - it is easy to and fast enough. The data on which the model processes the zoning concern only the extrinsic characteristics of real estate. This makes the procedure as well as rapid obviously more transparent. The experiments carried out on some city "sample" highlighted the benefits achievable with the use of the model, namely the speed of the procedure, and the consistency of the results to the real conditions of the area. The use of neural networks also allows for perfect adaptability of the model to changes in the area. Neural networks do not apply rules known a priori, but are capable of changing their structure because they are able to learn "on their own" to recognize existing relationships from the observation data.

Appendix

The application of the model in the below is intended to clarify the operational phases of the procedure and to provide a verification of the consistency of results. The urban used as a reference for testing is the city of Avellino.

The territorial division (phase 1) carried out a mapping scale 1/2000 and on the basis of arguments put forward by reading the information on the same map, as well as on that of the urban zoning, produced (n) 73 attitudinal areas. The graphical representation is in Fig. 1.

Fig. 1 Subdivision of urban places into attitudinal areas

B ($n \times n$) is the matrix of Mahalanobis distances between the attitudinal areas. Then a matrix C of the same size ($n \times n$) is constructed, whose elements are expressed dichotomous scale, such that 1 (0) indicates the (non-) spatial contiguity between the attitudinal areas.

After determining the coordinates of the barycentres of the individual attitudinal areas, geographic distances between them are calculated and reported these in a matrix D ($n \times n$) symmetric.

The elements of the matrices B and D are normalized in the interval (0, 1). It is thereby defined a matrix of indices of dissimilarity whose elements are given by the following:

$$\left[a_{ij} - \pi \cdot b_{ij} \cdot c_{ij}\right], \qquad (6)$$

with bij, cij and dij the generic elements respectively of the matrices B, C and D.

The weight of the indicator relative to the constraint of spatial contiguity between the attitudinal areas (π) is set equal to 0.20; in percentage terms is determined that in 10% of that resulting from the complex of factors that define the index of dissimilarity (0.20 * 0.50 = 0.10).

On the basis of the measure "correct" of distance has been generated, by the application of Multidimensional Scaling (model not metric), the dimensional configuration represented in Fig. 2.

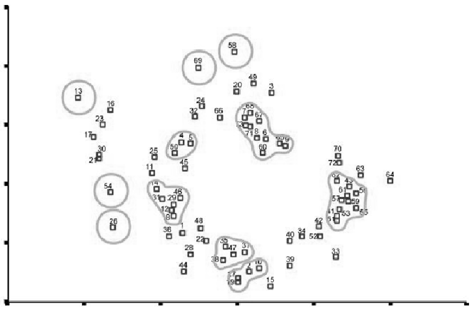

Fig. 2 Configuration of the MDS output

The phase 3 of the model is implemented by applying the following criteria:

a) the representation of the units is transformed from punctual in a circle with a radius, which is identical for all, to be defined in such a way that the group is composed of elements with overlapping portions.

b) The dissimilarity is measured by increasing the radius of the circles to a size for which all units, except for some - those dissimilar -, have overlapping portions of the circle.

From this graphics processing 11 shares of homogeneous areas result for a total of 42 units classified as Fig. 2 shows.

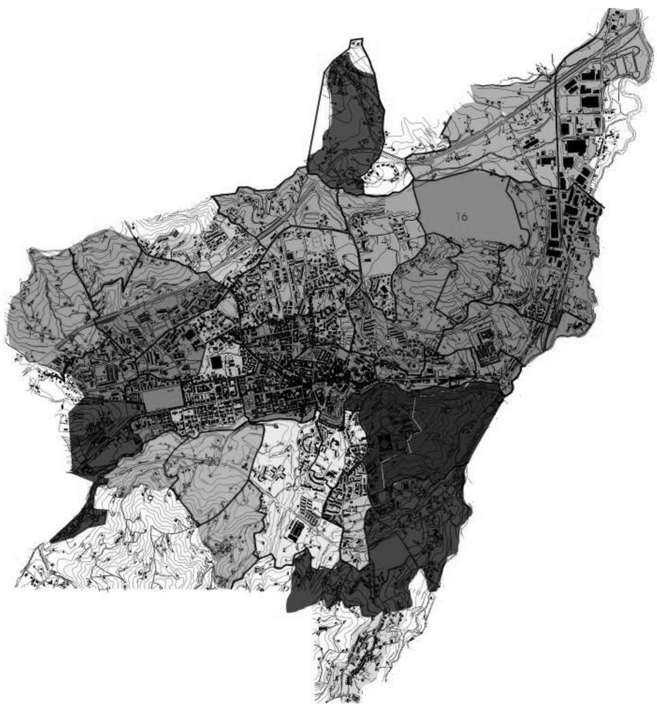

Fig. 3 Mapping of homogeneous zones

These units form the prototypes (training set) on which train the neural network. The multi-layer perception network used in the application is defined by a single layer of neurons and by a non-linear transfer function (sigmoid). The processing performed on the matrix of the profiles and with the selection of a target nominal (symbol) for the training set, assigns to the margin areas of the class to which they belong. The graphical representation of the output of the network (some units of the training set undergo a correction of the target) is shown in Fig. 3. Each homogeneous zone - represented by a different color - is identified by spatially contiguous areas of the same class and therefore defines areas of urban land within which there is homogeneity of physical and economic factors

ACKNOWLEDGMENT

This paper is to be attributed in equal parts to the two authors.

REFERENCES

[1] B. Manganelli, B. Murgante, “Analyzing Periuban Fringe with Rough Set”, in World Academy of Science, Engineering and Technology, ser. ICUPRDIS 2012, 2012. issn: p 2010-376X e 2010-3778 [Online]. Available: http://www.waset.org/journals/waset/v71.php.

[2] G. De Mare, P. Morano, and A. Nestico’, “Multi-criteria spatial analysis for the localization of production structures. Analytic hierarchy process and geographical information systems in the case of expanding an industrial area”, in World Academy of Science, Engineering and Technology, ser. ICUPRD 2012, 2012, issn: p 2010-376X e 2010-3778. [Online]. Available: http://www.waset.org/journals/waset/v71.php.

[3] W.R. Tobler, “A Computer Model Simulating Urban Growth in the Detroit Region” in *Economic Geography*, Vol. 46, Clark University, 1970, pp. 234-240.

[4] A.D. Gordon. “A survey of constrained classification”, in *Computational Statistics & Data Analysis*, n. 21, Elsevier Science, 1996, pp. 17-29.

[5] S. Zani, “Classificazione di unità territoriali e spaziali”, in S. Zani (eds.) *Metodi statistici per le analisi territoriali*, Milano: Franco Angeli, 1993, ch. 2.

[6] Z. Pan, J.P. LeSage, “Using spatial contiguity as prior information in vector autoregressive models”, in *Economics Letters*, Vol. 47, Elsevier Science, 1995, pp. 137-142.

[7] J.B. Kruskal, “Non metric multidimensional scaling: a numerical method”, in Psychometrika, Vol. 29, New York: Springer-Verlag, 1964 pp. 115-129.

[8] G. Fabbri, R. Orsini, *Reti neurali per le scienze economiche*. Franco Muzzio editore, Padova, 1993.

[9] H. Hruschka, M. Natter, “Comparing performance of feedforward neural nets and K-means for cluster-based marked segmentation”, in *European Journal of Operational Research*, Vol. 114, Elsevier Science, 1999, pp. 346-353.

[10] F. Murtagh, “A survey of algorithms for continuity-constrained clustering and related problems”, in *The Computer Journal*, Vol. 28, Oxford University Press, 1985, pp. 82-88.

[11] C. Bishop, *Neural networks for pattern recognition*, Oxford: Oxford University Press, 1995.

[12] B. Manganelli, “Stime e valutazioni economiche nella perequazione urbanistica”, (Published Conference Proceedings) in *INPUT 2010*, *6th Conferenza Nazionale di Informatica e Pianificazione* Territoriale, Vol. 3, Potenza, 2010, pp. 173-187.

[13] M.M. Fischer, “Computational neural networks - Tools for spatial data analysis”, in M. M. Fischer and Y. Leung (eds.): *Geo Computational Modelling: Techniques and Applications*, Springer, 2001, pp. 15-34.

[14] Li X. Yeh AGO, “Urban Simulation Using Neural Networks and Cellular Automata for Land Use Planning”, in: D. Richardson, P.V. Oosterom (eds.) *Advances in Spatial Data Handling*, Springer, 2002, pp 452-464.

[15] A. Bayes, A. Raquib, “Modeling Urban Land Cover Growth Dynamics Using Multi-Temporal Satellite Images: A Case Study of Dhaka, Bangladesh” in *ISPRS International Journal of Geo-Information*, Vol. 1, Open Access, doi:10.3390/ijgi1010003, 2012, pp. 3-31.

[16] B. Manganelli, G. Basile, “Appreciation Model of a Public Entity Set of Assets in the Accounting System”, submitted and accepted at 2nd Global Conference for Academic Research on Management and Economics which is being held 2012-12-7 at Kuala Lumpur, Malaysia, 2012.

[17] T.F. Cox, M.A.A. Cox, *Multidimensional scaling,* (2nd ed.) Chapman & Hall/CRC, 2001.

[18] G. De Mare, T. Lenza, and R. Conte, “Economic evaluations using genetic algorithms to determine the territorial impact caused by high speed railways”, in World Academy of Science, Engineering and Technology, ser. ICUPRD 2012, 2012. [Online]. Available: http://www.waset.org/journals/waset/v71.php.