Asian Journal of Information Technology 5(7): 729-731, 2006

On The Measure of Spatial Centroid in Geography

Duccio Rocchini and [1]Chiara Cateni
Dipartimento di Scienze Ambientali "G. Sarfatti", Università di Siena,
via Mattioli 4, 53100 Siena, Italy
[1]Istituto Comprensivo Chiusi, Piazza Veneto 6, 53043 Chiusi, Italy

Abstract: In this study we deal with the measure of spatial centroid of polygon objects, defined as vectors joining points (e.g. vertices) in a vector space V. In particular, since geographical objects could be viewed as tuples containing both non-spatial (z(i)) and spatial (s(i)) information, we focus on the z(i) and s(i) related measures of centroid and to problems arising from their improper use. In the first case, no significance is given to the real (i.e., statistical) spatial location of the centroid but only on the possibility to directly (e.g. automatically) assign data from polygons to centroids (data storage). On the contrary, in the second case, centroid is a real statistically derived point (i.e., real centre of mass). The mostly used GIS softwares rely the centroid calculation to algorithms only based on a within-polygon assumption, by substantially deriving a z(i) related centroid. The aim of this study is to warn researchers to avoid this type of algorithms when dealing with point pattern analysis, strictly related to the location part s(i) of the object tuple.

Key words: Centroid, centre of mass, point pattern analysis

CENTROID: A SPATIO-STATISTICAL POINT OF VIEW

The centroid is the most useful way to summarise the locations of a set of points[1]. Once a vector space V is defined, a finite set of points, forming a shape, could have more than one centroid, depending on how centroid is defined, but one of these is distinguished, on the strength of a least-squared-distances minimisation criterion[2,3].

Centroid coordinates x,y,…,n are calculated as the weighted mean of a set of coordinates within the X,Y,…,N dimensions.

$$x_c, y_c, \ldots, n_c = \frac{\sum_{i=1}^{m} w_i x_i, w_i y_i, \ldots, w_i n_i}{i} \quad (1)$$

Where $x_c, y_c, \ldots, n_c$ = centroid coordinates in X,Y,…,N dimensions for i points ranging from 1 to m (notice that the minimum number of i is 3 in case of polygon derived centroids), w = weight, in case of simple mean w=1.

Following[3] when averaging a finite set of elements in the vector space V, the mean of the list $\{v_1, \ldots, v_m\}$ can be uniquely characterised as that vector $\bar{v} \in V$ for which

$$\sum_{i=1}^{m} \left(v_i - \bar{v}\right) = 0 \quad (2)$$

As pioneer studies point out the elements of the list $\{v_1, \ldots, v_m\}$ would balance about the centroid, imaging a uniform physical model[4]. This balancing property motivates the alternative term for the average: centre of mass[1,3].

As statisticians and mathematicians well know, as far as the mean of a set of observations is representative, the distribution of observation frequency should follow the Gaussian. The higher the deviation from this curve the higher the deviation from a representative value. Hence, as long as values (e.g. coordinates) distribution is far from the Gaussian, the centroid location derived by an 'average' criterion could be affected by outliers coordinates which may provoke a coordinate shift. In these cases, median criterion appears to be a more efficient method to generate statistically representative centroid. On the other hand, following equation (1), a weighted mean could also overwhelm the problem. As an example, a fascinating and quite simple weighted mean is based on triangular area-weighted calculation (for two dimensions,[5]. Once a point P outside the point cloud (e.g. the set $\{v_1, \ldots, v_m\}$) is defined, triangles joining pairs of original point cloud and the new point are generated. Then the area and the centroid of each triangle are calculated. The resulting centroid (e.g. centroid of the set $\{v_1, \ldots, v_m\}$) is then derived as the mean of each triangle centroid weighted on the triangle area:

Corresponding Author: Duccio Rocchini, Dipartimento di Scienze Ambientali "G. Sarfatti", Università di Siena, via Mattioli 4, 53100 Siena, Italy

$$x_c, y_c = \frac{\sum_{i=1}^{m-1} A(\Delta v_i v_{i+1} p) x, y(\Delta v_i v_{i+1} p)}{\sum_{i=1}^{m-1} A(\Delta v_i v_{i+1} p)} \quad (3)$$

where = centroid coordinates; m = total number of vertices; = area of triangles formed by consecutive pairs of vertices and the point P; = triangle centroid coordinates.

GEOGRAPHICAL OBJECTS WITHIN THE VECTOR SPACE V

Since the development of digital cartography and GIS (Geographical Information Systems), the scientific community has been asked for solving the task of the mathematical representation of spatial entities. In GIScience, spatial entities are represented as objects - e.g. discrete phenomena or entities - or fields - e.g. continuous phenomena[6].

Geographical objects within a vector space V could be viewed as the tuple:

$$\left\{ z_1(i), z_2(i), \ldots, z_k(i) \middle| s(i) \right\}_{i=1,\ldots,n} \quad (4)$$

where z(i) = set of properties for the ith object related to the s(i) spatial location.

Goodchild[7] firstly described this representation of geographical objects, with z(i) and s(i) related to non-spatial and spatial information, respectively (see also[8]). In this study, the location s(i) of spatial objects appears to be the bulk of our reasoning.

The geometric representation of objects relies principally on Euclidean primitives such as points, polylines or polygons, depending on (i) the scale and (ii) the intrinsic nature of the entities under study. Strictly spoken, concerning the (i) point, an object such as a tree could be represented as a point at a 1: 25000 scale or as a polygon at a scale of 1: 2000, depending on the information it conveys; however, a tree could be rarely represented as a line (point ii).

BEWARE OF THE DIFFERENCE: Z(I) AND S(I) RELATED CENTROIDS

Due to byte-related problems in coordinate storage, rather than directly use polygons, most GIS analysts prefer to use centroid on which polygon z information is directly (e.g. automatically) extracted[9]. However, since real (e.g. statistically calculated) centroids could even fall out of the polygons (e.g. consider markedly hollow polygons), in this case a membership condition is to be pursued. This type of centroid (here named z(i) related centroid) has no relations with a statistically derived one (i.e., the real centre of mass, namely the s(i) related centroid), leading only to the possibility to directly (e.g. automatically) assign data from polygons to centroids.

Although a straightforward manner to maintain the non-spatial information z(i) without completely losing the spatial information s(i) is to convert each polygon to a z(i) related centroid, centroid coordinate calculation should follow the previously cited least-square-distances minimisation criterion, in case of point pattern analysis (e.g. spatial analysis on point locations,[10,11]). In this case, researchers are strongly encouraged to avoid algorithms constrained on the membership of centroid within polygons, which unfortunately are the bulk of centroid calculation within the mostly used GIS softwares. As an example, in order to estimate landscape changes, cartographers rely to centroid movements of landscape classes over the time([1] as an example). In these cases, centroid must be accurately defined, since if a z(i) related centroid is used, all spatial information will be lost. Quite a misleading outcome could derive from a study based on non-spatial related centroids to estimate spatial processes!

CONCLUSION

In this study, the measure of spatial centroid was addressed with particular interest to geographic object information. Since objects are defined by tuples combining non-spatial (z) and spatial (s) information, we are claiming that centroid-based analyses must rely to the inherent part of the tuple information to be extracted. That is, if analysis is based on polygon z information storage, centroid must be calculated on the basis of z information so as the derived point fall within the polygons of interest following a membership assumption. On the other hand, if analysis is concerned with spatial information (s) centroid must be calculated as the (simple or weighted) average or the median of point coordinates, without referring to within-polygon assumption.

In this latter case, operator must pay attention to (i) the presence of outlier coordinates and (ii) scale issues. In fact, whereas outliers occur, an average criterion could threat final results. As Haining (2003) point out the smaller the polygon the more representative the centroid will be. However, this could not be a general rule, due to a scale effect. In fact, in contrast to common Euclidean geometry, dimensions and roughness (e.g. a minor or a major number of vertices) of geographic objects (e.g. polygons) are

strictly related to the scale at which the analysis is carried out[13-16].

In general, on the strength of the development of user-friendly tools, analysts are tempted to rely to automatic software-based calculations[17,16,18] however, when performing operations on spatial objects, researchers must keep in mind that centroid derivation is strictly dependent to the final aim of their analysis.

ACKOWLEDGEMENT

We would like to acknowledge Jeff Jennes for his explanations about source code based calculation.

REFERENCES

1. Longley, P.A., M.F. Goodchild, D.J. Maguire and D.W. Rhind, 2003. Geographic Information Systems and Science, John Wiley and Sons, Chichester, New York, Weinheim, Brisbane, Singapore, Toronto.
2. Du, Q. and M. Gunzburger, 2002. andGrid generation and optimization based on centroidal Voronoi tessellations, Applied Mathematics and Computation, 133: 591-607.
3. Groisser, D., 2004. Newton ands method, zeroes of vector fields and the Riemannian center of mass, Advances in Applied Mathematics, 33: 95-135.
4. Kaiser, M.J., 1993. The perimeter centroid of a convex polygon, Applied Mathematics Lett., 6: 17-19.
5. O andRourke, J., 1998. Computational Geometry in C, Cambridge University Press, Cambridge.
6. Burrough, P.A. and R.A. McDonnell, 1998. Principles of Geographic Information Systems, Oxford University Press, Oxford.
7. Goodchild, M.F., M.J. Egenhofer, K.K. Kemp, D.M. Mark and E.S. Sheppard, 1999. Introduction to the Varenius project and, Intl. J. Geographical Information Sci., 13: 731-746.
8. Goodchild, M.F., 2003. The nature and value of geographic information, in Foundations of Geographic Information Science, Ed. by Duckham, M., Goodchild, M. F. and Worboys, M.F., Taylor and Francis, London, New York, p: 19-32.
9. Wise, S., 2002. GIS basics, Taylor and Francis, London.
10. Haining, R., 1990. Spatial data analysis in the social and environmental sciences, Cambridge University Press, Cambridge.
11. Fotheringham, A.S., C. Brunsdon and M. Charlton, 2000. Quantitative Geography: perspectives on spatial data analysis, SAGE, London, Thousand Oaks, New Delhi.
12. Haining, R., 2003. Spatial data analysis: theory and practice, Cambridge University Press, Cambridge.
13. Mandelbrot, B.B., 1985. Self-affine fractals and fractal dimension, Physica scripta, 32: 257-260.
14. Turner, M.G., R.V. O andNeill, R.H. Gardner and B.T. Milne, 1989. Effects of changing spatial scale on the analysis of landscape patterns. Landscape Ecology, 3: 153-162.
15. Turcotte, D.L., 1997. Fractals and chaos in Geology and Geophysics, Cambridge University Press, Cambridge.
16. Rocchini, D., 2005. Resolution problems in calculating landscape metrics and, J. Spatial Science, In press.
17. O andNeill, R.V., J.R. Krummel, R.H. Gardner, G. Sugihara, B. Jackson, D.L. DeAngelis, B.T. Milne, M.G. Turner, B. Zygmunt, S.W. Christensen, V.H. Dale and R.L. Graham, 1988. Indices of landscape pattern and, Landscape Ecology, 1: 153-162.
18. Rocchini, D. and A. Di Rita, 2005. Relief effects on aerial photos geometric correction. Applied Geography, 25: 159-168.