Spatial Interpolation Using Conservative Fuzzy Reasoning

Kok Wai Wong, Tamás D. Gedeon
School of Information Technology
Murdoch University
South St, Murdoch
Western Australia 6150
Email: { kwong | tgedeon } @murdoch.edu.au

Patrick M. Wong
School of Petroleum Engineering
University of New South Wales
Sydney, Australia
Email: pm.wong@unsw.edu.au

Abstract

Spatial interpolation is an important feature of a Geographic Information System, which is the procedure used to estimate values at unknown locations within the area covered by existing observations. In this paper, we propose a conservative spatial interpolation technique that incorporates the advantages of local interpolation, Euclidean interpolation, and conservative fuzzy reasoning. The main objective of this paper is to formulate a computationally efficient spatial interpolation technique similar to the IDWA technique that can be used in real time application. The main feature of our spatial interpolation technique is inherited from the concept used in conservative fuzzy interpolation reasoning for interpolating fuzzy rules in sparse fuzzy rule bases. Illustration examples from a rainfall spatial interpolation problem are also used to illustrate the applicability of the proposed technique.

1. Introduction

In a Geographic Information System (GIS), spatial interpolation is an essential feature [1]. Normally, points with known values are used to estimate values at other points. An example is rainfall prediction using spatial interpolation [2, 3, 4]. In this example, interpolation can estimate the amount of rainfall at a location based on the knowledge from the rainfall measurements at nearby locations. All spatial interpolation techniques can be grouped into global and local methods [1]. In a global method, all the information available is used to estimate an unknown value, while local methods only use a sample of the information for estimation. However, in [2, 3, 4], the authors have discussed the functionality of local method and found that they provide better results as compared to global method. Therefore, in this paper, the analysis will be based on local methods.

In a global method, trend surface analysis is normally performed. The equation that can be used to estimate values at other points using a third-order trend surface is:

$$Z = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + b_6X^3 + b_7X^2Y + b_8XY^2 + b_9Y^3 \quad (1)$$

where b coefficients are estimated from the available information points.

Equation (1) can also be written as:

$$Z = f(X,Y) \quad (2)$$

As for a local method, the spatial interpolation of the value Z_i in the *i- th* surface is:

$$Z_i = f(X_i, Y_i) \quad (3)$$

This trend surface analysis could be considered as a subset of polynomial regression [5]. However, in this case, in order to construct an unbiased model, large amount of data is necessary, but this will effectively increase the computational complexity of the system. Another popular spatial interpolation technique used is Kriging [6]. This is normally used to construct a model to describe the rate at which the variance between points changes over space. Kringing involves the construction of a variogram. The estimation of the variogram is complex.

For computational efficiency, the inverse distance weighted averaging (IDWA) interpolation is preferred over the last few spatial interpolation techniques [7]. IDWA is a deterministic estimation technique that uses values at the known locations to estimate the values at the unknown points by using linear combination. The assumption made by IDWA is that the values from the points closer to the unknown points are more representative of the values to be estimated than those points further away. This method is quite similar to the

0-7803-7078-3/01/$10.00 (C)2001 IEEE.

nearest neighbour interpolation method where the unknown points simply take the values of the closest sample point. In [8], the author presented another interpolation technique known as Euclidean interpolation. In this technique, he used attribute data to identify the connections that are independent of the physical distance. The Euclidean distance between the sets of attribute data associated with each unknown point are computed. The interpolation is performed based on the minimum Euclidean distance and the geographic distance.

In this paper, we propose a spatial interpolation technique that incorporates the advantages of local interpolation, Euclidean interpolation, and conservative fuzzy reasoning [10, 11]. The main objective of the analysis is to formulate a computationally efficient spatial interpolation technique similar to IDWA technique that can be used in real time application. Illustration examples extracted from rainfall spatial interpolation [2, 3, 4] problem are also used to illustrate the applicability of the propose technique.

2. Fuzzy Conservative Spatial Interpolation

2.1 *Self-organizing Map (SOM)*

For local spatial interpolation, the first step is to classify the available data into different classes so that the data are split into homogeneous sub-populations. SOM is used to divide the data into sub population and hopefully reduce the complexity of the whole data space to something more homogeneous. The objective in this step is to make use of an unsupervised learning algorithm to sub-divide the whole population. The Self-organizing Map (SOM) is selected for this purpose mainly because it is a fast, easy and reliable unsupervised clustering technique. Beside, SOM has already shown successful application in spatial interpolation as presented in [4].

SOM is designed with the intention to simulate the spatial organizations found in various brain structures and has a close relationship to brain maps [11]. Its main feature is the ability to visualize high dimensional input spaces onto a smaller dimensional display, usually two-dimensional. In this discussion, only two-dimensional arrays will be of interest. Let the input data space $\Re^n$ be mapped by the SOM onto a two-dimensional array with i nodes. Associated with each i node is a parametric reference vector $m_i=[\mu_{i1},\mu_{i2},\ldots,\mu_{i2}]^T \in \Re^n$, where μ_{ij} is the connection weights between node i and input j. Therefore, the input data space $\Re^n$ consisting of input vector $X=[x_1,x_2,\ldots,x_n]^T$, ie $X \in \Re^n$, can be visualized as being connected to all nodes in parallel via a scalar weights μ_{ij}. The aim of the learning is to map all the n input vectors X_n onto m_i by adjusting weights μ_{ij} such that the SOM gives the best match response locations.

SOM can also be said to be a nonlinear projection of the probability density function $p(X)$ of the high dimensional input vector space onto the two-dimensional display map. Normally, to find the best matching node i, the input vector X is compared to all reference vector m_i by searching for the smallest Euclidean distance $\| X - m_i \|$, signified by c.

Therefore,

$$c = \arg\min_i \{\| X - m_i \|\} \qquad (4)$$

or

$$\| X - m_c \| = \min_i \{\| X - m_i \|\} \qquad (5)$$

During the learning process the node that best matches the input vector X is allowed to learn. Those nodes that are close to the node up to a certain distance will also be allowed to learn. The learning process is expressed as:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[X(t) - m_i(t)] \qquad (6)$$

where t is discrete time coordinate
and $h_{ci}(t)$ is the neighbourhood function

After the learning process has converged, the map will display the probability density function $p(X)$ that best describes all the input vectors. At the end of the learning process, an average quantisation error of the map will be generated to indicate how well the map matches the entire input vectors X_n. The average quantisation error is defined as:

$$E = \int \| X - m_c \|^2 \; p(X) dX \qquad (7)$$

After the 2-dimensional map has been trained, the reference vectors that were used in the nodes of the map can be also obtained. In spatial interpolation, the reference vector will be the node center and consists of the input variables (x, y) and the output variable (z). As we like the clusters to be formed to facilitate the concept used in Euclidean interpolation, we propose here to construct the clustering boundaries based on the output reference vector of the nodes. The rule of thumb for deciding on the clustering boundaries is to examine the distance measure between the neighboring reference values. If the distance measure between the present reference node and the neighboring nodes is high, that suggests another cluster.

2.2 *Conservative Fuzzy Reasoning*

After the whole sample space has been subdivided in local homogeneous sub-populations, a spatial interpolation model based on the concept from conservative fuzzy reasoning is constructed for each sub-population. Before formulating the new approach, the characteristics of conservative of fuzziness are shown.

A. Fuzzy Conservative Interpolation Reasoning

Conservative fuzzy rule interpolation is used to interpolate those input instances that cannot find any fuzzy rules to fire. This type of fuzzy rule base is known as a sparse rule base [9, 10, 12]. By referring to Figure 1, the labels r and s indicate the spread of the observation and the rule antecedent, which represent their fuzziness. The labels r' and s' indicate the spread of the conclusion and the rule consequent. By observation, B2 is found to be fuzzier than A2, due to the shallower slope. The A*, A2 and the B*, B2 distances are not normalized, as the values of u and u' are used explicitly as appropriate.

The r' value can be calculated by:

$$r' = r\left(1 + \left|\frac{s'}{u'} - \frac{s}{u}\right|\right) \qquad (8)$$

The interpolated fuzzy rules produced can maintain the local change in fuzziness in the rule base.

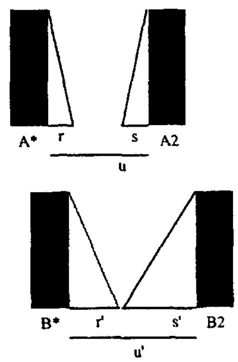

Figure 1: Parameters use in the conservative fuzzy interpolation

B. Conservative Spatial Interpolation Reasoning

The fuzzy sets of the input spaces (x,y) and output value (z) are constructed based on the distance between them. We use the nearest four neighboring points of the unknown locations as the spread of the fuzziness. In this case, we can form a trapezoidal membership with these four location points. For the unknown location, it is transformed as the reference (center) point of the membership function, as (x_0, y_0, and z_0). Those location points on the left and the right of the reference point for x are shown in Figure 2. The notations used for y and z are similar as those used in x. However, the sequence of their orders for y and z may be different.

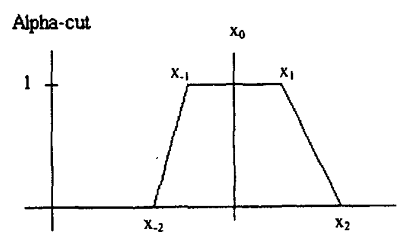

Figure 2: Notations used in the fuzzy sets

For the left side of the reference point:

$$d_{(x_{-2},x_{-1})} = x_{-2} - x_{-1} \qquad (9)$$

$$d_{(x_{-1},x_0)} = x_{-1} - x_0 \qquad (10)$$

$$d_{(x_{-2},x_1)} = x_{-2} - x_1 \qquad (11)$$

$$d_{(y_{-2},y_{-1})} = y_{-2} - y_{-1} \quad (12)$$

$$d_{(y_{-1},y_0)} = y_{-1} - y_0 \quad (13)$$

$$d_{(y_{-2},y_1)} = y_{-2} - y_1 \quad (14)$$

$$d_{inL2} = \sqrt{(d_{(x_{-2},x_{-1})})^2 + (d_{(y_{-2},y_{-1})})^2} \quad (15)$$

$$d_{inL1} = \sqrt{(d_{(x_{-1},x_0)})^2 + (d_{(y_{-1},y_0)})^2} \quad (16)$$

$$d_{inL} = \sqrt{(d_{(x_{-2},x_1)})^2 + (d_{(y_{-2},y_1)})^2} \quad (17)$$

$$d_{(z_{-2},z_{-1})} = z_{-2} - z_{-1} \quad (18)$$

$$d_{(z_{-2},z_1)} = z_{-2} - z_1 \quad (19)$$

The reference point can be calculated as:

$$z_{0L} = z_{-1} + d_{inL1}\left(1 + \left|\frac{d_{(z_{-2},z_{-1})}}{d_{(z_{-2},z_1)}} - \frac{d_{inL2}}{d_{inL}}\right|\right) \quad (20)$$

As for the right side of the reference point, similar calculations are used. The reference point can be calculated as:

$$z_{0R} = z_1 - \left[d_{inR1}\left(1 + \left|\frac{d_{(z_2,z_1)}}{d_{(z_{-1},z_2)}} - \frac{d_{inR2}}{d_{inR}}\right|\right)\right] \quad (21)$$

Therefore the interpolated value at the unknown location can be obtained by taking the aggregation of the two:

$$z_0 = \frac{z_{0L} + z_{0R}}{2} \quad (22)$$

The condition for this spatial interpolation is all the data has to be normalized to be within the same range. For example if x is normalized between 0 and 100, y and z will also need to be normalized between 0 and 100. This spatial interpolation technique will only work well with linear functions, therefore it will only work in local interpolation domain.

3. Illustration Example

In [4], the case study was performed on the data collected on 8th May 1996 in Switzerland. 100 data locations are used as the training data and the other 367 locations data are then used to verify the prediction accuracy. The input variables used are the 2D coordinate position *(x, y)*; and the output is the rainfall measurements *(z)*. The 100 training data points are input into the SOM for unsupervised clustering. In that case study, the authors used a 10 by 10 two-dimensional map. After performing the cluster boundaries determination, the classes are formed. To illustrate this technique, two points from that study are presented as follows.

First unknown point (all values normalized to be between 0 to 100):

$x_{-2} = 29.29$	$y_{-2} = 89.06$	$z_{-2} = 16.22$
$x_{-1} = 33$	$y_{-1} = 78.75$	$z_{-1} = 18.11$
$x_0 = 33.62$	$y_0 = 92.54$	$z_0 = ??$
$x_1 = 41$	$y_1 = 79.97$	$z_1 = 24.59$
$x_2 = 42.58$	$y_2 = 65.05$	$z_2 = 27.3$

The actual z_0 measured at the location that needs to be interpolated is 23.24.

The spatial interpolated z_0 calculated from this conservative spatial interpolation technique is 20.12. The error between the actual and the interpolated z_0 is considered small, as the difference is 3.12.

The second unknown point:

$x_{-2} = 87.44$	$y_{-2} = 74.24$	$z_{-2} = 23.6$
$x_{-1} = 91.6$	$y_{-1} = 87$	$z_{-1} = 24$
$x_0 = 92.63$	$y_0 = 121.7$	$z_0 = ??$
$x_1 = 93.94$	$y_1 = 98.34$	$z_1 = 49.6$
$x_2 = 99.42$	$y_2 = 84.19$	$z_2 = 54.8$

The actual z_0 measured is 37.54, and the spatial interpolated z_0 calculated from this technique is 32.14. The error between the actual and the interpolated z_0 is considered small as well, as the difference is 5.4.

4. Conclusions

This paper has presented a simple and computationally efficient spatial interpolation technique that could be used in real time. The main feature of this spatial interpolation technique is inherited from the concept used in the conservative fuzzy interpolation reasoning for interpolating fuzzy rules in sparse fuzzy rule bases. This paper has also suggested that this technique works well in the local spatial interpolation domain. SOM is used to divide the data into sub population and hopefully reduce the complexity of the whole data space to something more homogeneous. This is the condition for which the conservative spatial interpolation technique could work best. Illustration examples have also shown that this technique could produce reasonable results.

5. References

[1] P. Burrough, and R. McDonnell, *Principles of geographical information systems*, Oxford University Press, New York, 1998.

[2] S. Lee, S. Cho, and P.M. Wong, "Rainfall prediction using artificial neural network", *Journal of Geographic Information and Decision Analysis*, Vol. 2(2), 1998, pp. 254-264.

[3] Y. Huang, T.D. Gedeon, and P.M. Wong, "Spatial interpolation using fuzzy reasoning and genetic algorithms", *Journal of Geographic Information and Decision Analysis*, vol. 2(2), 1998, pp. 223-233.

[4] K.W. Wong, P.M. Wong, T.D. Gedeon, and C.C. Fung, "Rainfall Prediction Using Neural Fuzzy Technique", in G. Dubois, J. Malczewski and M. De Cort (Eds) *Spatial Interpolation Comparison 97*, (in press).

[5] R.H. Myers, *Classical and Modern Regression with Applications*, Boston, PWD-Kent Publishing, 1990.

[6] G. Matheron, "Principles of Geostatistics," *Economic Geology*, 58, 1963, pp. 1246 – 1266.

[7] D. Legates, and C. Willmont, "Mean Seasonal and Spatial Variability in Global Surface Air Temperature," Theor. Appl. Climatol., 41, 1990, pp. 11-21.

[8] C.D. Tomlin, *Geographic Information Systems and Cartographic Modeling*, New Jersey, Prentice Hall, 1990.

[9] T.D. Gedeon, and L.T. Kóczy, "Conservation of relative fuzziness: an additive strategy," *Proceedings IEEE International Conference on Intelligent Engineering Systems (INES'98)*, 1998, pp. 115-120.

[10] T.D. Gedeon, and L.T. Kóczy, "Fuzzy rule interpolation using an additive conservative strategy," *Proceedings 5th European Congress on Intelligent Techniques and Soft Computing (EUFIT '97)*, 1997, pp. 910-913.

[11] T. Kohonen, *Self-Organising Map*, Springer-Verlag, 1995.

[12] L.T. Kóczy, and K. Hirota, "Approximate reasoning by linear rule interpolation and general approximation," *Int. J. Approx. Reason*, Vol. 9, 1993, pp.197-225.