Applied Soft Computing 55 (2017) 319–330

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Fuzzy quality assessment of gridded approximations

Jörg Verstraete

Systems Research Institute, Department of Computer Modelling, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland

ARTICLE INFO

Article history:
Received 12 April 2016
Received in revised form 16 August 2016
Accepted 28 January 2017
Available online 9 February 2017

Keywords:
Spatial data
Gridded data
Quality assessment
Fuzzy set application

ABSTRACT

Spatial data are often represented in a gridded format: a region of interest is overlayed with a raster, and every cell of the raster carries a value that is representative for the cell. Examples for this are e.g. air pollution concentrations or population densities provided on grids. A grid sometimes needs to be remapped onto a different grid, either for representation purposes or to allow operations or interactions with other gridded data, such as provided by the map algebra. Different approaches to remap gridded data exist and more recent approaches use other available data to improve the results. The emerging of these advanced remapping approaches also mandates the need for an objective method for ranking how different mappings conform to a known reference solution, as such a method allows to judge the performance of the different remapping algorithms. In this article, a fuzzy set based method to generate a family of ranking operators to objectively compare how different grids resemble a given reference grid is presented.

1. Introduction

Spatial data comes in different data structures [1,2]; the most common structures are feature based, where basic geometric objects such as points, lines and polygons represent real world features; and grid based, where a discretization of the two dimensional space permits to model a numeric value that is distributed over a region of interest. This latter model is often used for numerical properties obtained through measurements, estimates or simulations. In the gridded model, regular grids are most commonly used; these are grid where all cells have the same shape and size, although irregular grids can be conceived e.g. using administrative borders or other shapes. Examples of the use of regular grids are temperature and concentration of air pollutants; the population density for a country can for example be modelled on an irregular grid defined using city limits and administrative regions. The main issue is that data can be defined on incompatible grids, these are grids that do not have a 1-1 mapping between their cells. In order to deal with such incompatible grids, one grid needs to be remapped on to the other grid. While different techniques exist to achieve this (e.g. [3–5]), much less research has been performed in methods for ranking different grids in order to judge remapping algorithms. Typically general statistical methods are used, but this falls short for the newer remapping methods that aim to detect an underlying spatial distribution in the gridded data: as will be shown, spatial similarity is not well considered in these currently employed statistics. In this contribution, a novel method to rank gridded datasets using fuzzy sets is presented. This ranking is performed with respect to a reference data set, and serves as a quantifiable way to judge the performance for grid remapping algorithms. Key to the presented ranking approach is that it takes into account similarities in underlying spatial distributions. Section 2 elaborates on the problem and explains remapping techniques, Section 3 discusses current approaches for ranking different grids. The novel approach is presented in Section 4, with experiments and conclusion following in respectively Sections 5 and 6.

2. Problem description

2.1. Gridded data

Gridded data models are commonly used to represent numerical properties whose value varies with locations. Generally, a raster which partitions the region of interest is chosen. This raster can be regular or irregular, and can be chosen manually or it can be imposed by the technology used to determine the values that are modelled. In case of a temperature model for example, the grid can be a regular grid with 500 m × 500 m gridcells; whereas in case of population densities, the administrative districts can be used as gridcells. A gridcell is further considered to be the smallest spatial unit: nothing is known about the spatial distribution inside a grid cell. Performing analysis on gridded data requires combining the

E-mail address: jorg.verstraete@ibspan.waw.pl
URL: http://www.ibspan.waw.pl.

http://dx.doi.org/10.1016/j.asoc.2017.01.051
1568-4946/

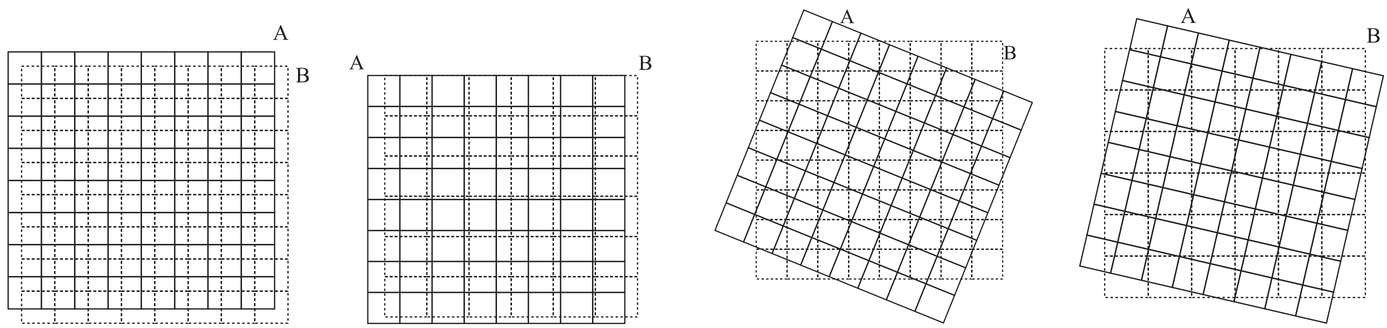

Fig. 1. Different cases that make grids incompatible: rasters are not in the same position, rasters have a different size of cells, rasters are rotated and finally a combination of these.

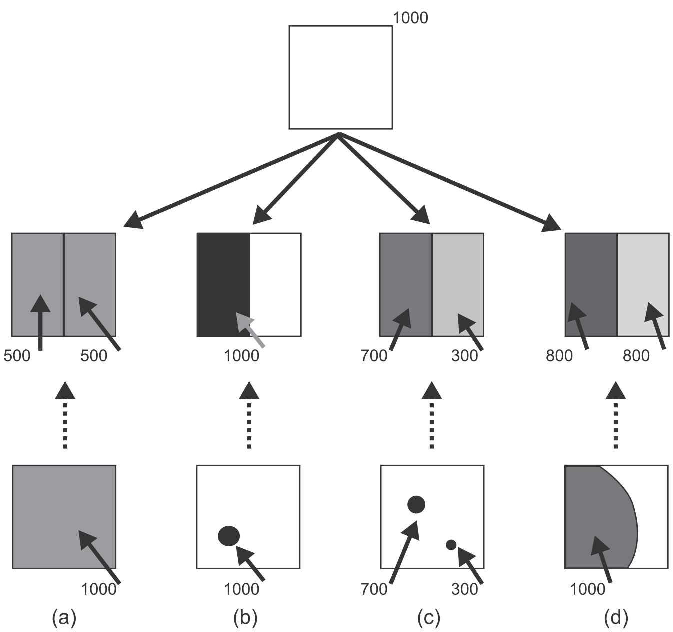

Fig. 2. Different possible underlying distributions: uniform distribution, single point of concentration, two points of concentration and uniform distribution in part of the cell.

grids involved. This is a common overlay operation, for which e.g. Tomlin's map algebra [6] can be used. While the map algebra or other analysis on multiple raster data sets provide for a lot of possibilities, they work on a cell-by-cell basis and cannot be applied when the grids are incompatible, i.e. when there is no 1-1 mapping between the gridcells of the grids involved (e.g. Esri: [7]). Fig. 1 shows several examples of incompatible grids: in Fig. 1a, one grid is slightly shifted compared to the other grid, Fig. 1b shows two grids that have different grid cells, Fig. 1 depicts the case where grids are rotated and the example in Fig. 1d is a combination of aforementioned differences. Resolving incompatible grids is a first step before further analysis can be performed; several methods exists to remap one grid A onto the raster of grid B where the most common one assumes that data are uniformly distributed within each cell and the more advanced methods try to find an underlying spatial distribution. These methods are briefly elaborated on in the next section, as this helps to understand the need for a new assessment of grid approximations.

While the grid itself provides no information regarding the spatial distribution of the data inside a grid cell, the grid is usually an approximation of a real world situation and as such there can be different underlying distributions. Consider the simplified example in Fig. 2: one cell of a grid representing population is shown; the cell has a value of 1000, to indicate that there are 1000 people living in this area. As a simplified case, consider the situation where this gridcell needs to be split into two cells to achieve a remapping. There are many ways for distributing the data over both cells, with four cases shown. The cases in these examples match underlying distributions that are uniform over the region (a), concentrated in one or two villages (b) and (c) or in just part of the region (d), e.g. if there is a nature reserve or a lake. All remapping methods make assumptions regarding the underlying distribution; the more true this assumption the better the remapping will be. The ranking method should however also be made to correctly assess spatial similarity, as will be illustrated in Section 3.

2.2. Remapping incompatible grids

2.2.1. General concept

For this section, assume that there are two grids, A and B. Both grids contain data that needs to be combined, but the grids are incompatible. In order to be able to apply for example the map algebra, one grid needs to be remapped so that there is a one-one mapping between grid cells of A and those of B. In this case, A will be remapped so that it is defined on the same raster as B; the remapped grid of A will be called A'. Consider the grid A, with grid cells c_i^A, $i=1, \ldots, n$ and values for the cells given by $f^A(c_i^A)$, $i=1, \ldots, n$. Using similar notations, the target raster A' onto which A has to be remapped has cells $c_i^{A'}$, $i=1, \ldots, m$ and values $f^{A'}(c_i^{A'})$, $i=1, \ldots, m$ that have to be determined. These values are obtained as the weighted sum:

$$f^{A'}(c_i^{A'}) = \sum_j w_i^j f^A(c_j^A) \tag{1}$$

$$= \sum_{j | c_j^A \cap c^{A'}_i \neq \emptyset} w_i^j f^A(c_j^A) \tag{2}$$

If only the cells of grid A that intersect with the target cell $c_i^{A'}$ will contribute to $f_{A'}(c_i)$, (1) can be modified to only consider the cells of A that have a non-empty intersection (2). This means that the value of the cell $c_i^{A'}$ is only determined by the cells of A that intersect with it. The weights w_i^j need to be determined; for these weights, the following constraints are also assumed:

$$\forall j, \forall i, w_i^j \geq 0 \tag{3}$$

$$\forall j, \sum_i w_i^j = 1 \tag{4}$$

The first constraint indicates that all weights are positive: a cell of A either contributes to a cell in A', or does not; it does not decrease the value of a cell in A'. The second constraint states that all weights sum up to 1: this constraint guarantees that the grid A', the grid after transformation, has the same total value as grid A. A special case of the map overlay for grids is the spatial disaggregation problem [4,5],