SPATIAL ANALYSIS OF SCHIZOPHRENIA PREVALENCE USING A MULTIOBJECTIVE EVOLUTIONARY ALGORITHM

Carlos R. García-Alonso
Department of Management and Quantitative Methods
ETEA, Business Administration Faculty. University of Córdoba
Córdoba, Córdoba, SPAIN
cgarcia@etea.com

José A. Salinas-Pérez
Department of Management and Quantitative Methods
ETEA, Business Administration Faculty. University of Córdoba
Córdoba, Córdoba, SPAIN

Luis Salvador-Carulla
PSICOST Scientific Association
Jerez de la Frontera, Cádiz, SPAIN

Abstract
A group of geographically close spatial units that shows a similar prevalence pattern – significantly high- is called a hotspot and, when it is low or drain-like –outlet effect, this area is called a coldspot. Both areas can be identified through the evaluation of the degree of agreement between Local Indicators of Spatial Aggregation (LISA) and Bayesian Conditional Autorregresive (CAR) scores. A Multi Objective Evolutionary Algorithm (MOEA) has been designed and tested to evaluate this degree of agreement and QQ-Plots are used to identify hot and cold-spots. The MOEA uses four different strategies to evaluate the fitness function including a fuzzy approach and the standard SPEA2. Using this methodology the spatial distribution throughout the time span (years 2004, 2006, 2007 and 2008) of schizophrenia prevalence has been analysed in 770 municipalities in Andalusia (southern region in Spain). Our procedure is a robust method to identify both hot and cold-spots in the space and can be useful to organize the spatial distribution of health-care services.

Keywords: Spatial distribution of illnesses, Autocorrelation analysis; Conditional autorregresive model; Multiobjective evolutionary algorithms; Geographical information systems.
Introduction

The study of the spatial –geographical- distribution of relevant phenomena is especially important in health care research [1-2] and in health economics [3-4]. This scientific interest is mainly justified by the potential existence of spatial dependence of the prevalence in specific illnesses. There exists spatial prevalence when the value of a variable in a geographical location –i.e. prevalence- depends on the values that correspond to its spatial neighborhoods. The existence of spatial autocorrelation can be due to reaction and/or interaction effects [5]. The former means that there is a spatial risk in common, the latter appears when there is a transmission process.

Techniques for spatial analysis of geographical areas with a statistically significant high or low autocorrelated prevalence of any given disorder are especially interesting in health resources planning and management [6] because they can be used to both highlight relevant concentrations of illnesses as well as to help an efficient location of health-care units.

There are many spatial analysis techniques based on econometrics from Poisson analysis [7] to artificial intelligence [8-9]; in between well-known techniques as Local Indicators of Spatial Aggregation (LISA) [10-11] and Bayesian methods [1, 4]. Unfortunately, results obtained using different methods differ [6, 12]. This disparity hampers the validity and the usability of spatial analysis in health care planning because there is not any clear scientific evidence on which is the best methodological approach [13-15].

The objectives of this research are: i) to design and develop a computer-based tool to identify and geographically locate highly or lowly autocorrelated zones –‘hotspots’ and ‘coldspots’- which merges different spatial analysis methods using a Multiobjective Evolutionary Algorithm (MOEA), and ii) to carry out a demonstration study in a geographical area (Andalusia, Spain). This study takes into account specific LISA and Bayesian methods but our tool can be generalized including other methods.

Spatial autocorrelation: hotspots and coldspots

In spatial analysis, a hotspot is a group of close spatial units (usually polygons) where the phenomenon under study shows a significantly high value of autocorrelation/Bayesian risk scores. When this value is statistically significant low then we have a coldspot. The analysis of identifying and locating both hot and coldspots can be then included in the generalized clustering techniques.

Taking into account the previous definitions, the identification of hotspots when the autocorrelated prevalence of a illness is studied could evidence the existence of a geographical zone where the illness is significantly high due to unknown reasons. On the contrary, identified coldspots evidence a spatial outlet effect, that is: a spatial unit with very high prevalence surrounded by units with very low prevalence scores or viceversa. Both spatial phenomena are interesting for planning purposes.
Hybridising methods to identify and locate hot and coldspots

Three standard LISA methods were selected to evaluate local autocorrelation: Moran’s $I$, Geary’s $C$ and Getis and Ord’s $G$. In addition to the previous methods, Bayesian risks $B$, Bayesian Conditional Autoregressive (CAR) model, were also calculated for all the spatial units (SU). For the MOEA, a solution was a set of groups of SU (1), each group was evaluated by the means and the standard deviations (SD) of the corresponding LISA and CAR scores ($\bar{I}, \bar{SD}_I, \bar{C}, \bar{SD}_C, \bar{G}, \bar{SD}_G$) as well as the minimum distance between them $D$.

$\{(SU_1^1, SU_1^2, ..., SU_1^3), (SU_2^2, SU_2^3, ..., SU_2^3), ..., (SU_n^1, SU_n^2, ..., SU_n^3)\}$ \hspace{1cm} (1)

In order to identify hotspots, objectives were: $Max\bar{I}, Min\bar{SD}_I, Min\bar{C}, Min\bar{SD}_C, Max\bar{G}, Min\bar{SD}_G, Max\bar{B}, Min\bar{SD}_B, Min\bar{D}$. On the contrary, to identify coldspots: $Min\bar{I}, Min\bar{SD}_I, Max\bar{C}, Min\bar{SD}_C, Min\bar{G}, Min\bar{SD}_G, Max\bar{B}, Min\bar{SD}_B, Min\bar{D}$. In both cases the algorithm tried to find high-low autocorrelated (optimizing LISA and CAR scores), uniform (minimizing SD) and close (minimizing D) geographical areas. The MOEA used four fitness functions (SPEA2, weighted sum of objectives, ranking selection and fuzzy evaluation of weighted objectives) for evaluating solutions once the standard genetic operators (selection, mutation and crossover) determined a new one. Elitism was also taken into consideration. The iterative process of improving solutions stopped when a predefined maximum number of generations $g$ was reached or when the fitness of the solution obtained did not improve previous ones (mean squared error). In order to avoid unexpected local optima, the MOEA generated $s$ seeds at random. The structure of the MOEA is described in Table 1.

SU in the hot and cold spots were identified using Q-Q Plots taking into consideration an exponential fit. Finally, they were located in the space using a standard geographical information system.

Results of the pilot testing

The MOEA was checked analysing the schizophrenia prevalence in Andalusia, the greater region in Spain. 770 municipalities, the greatest spatial precision, were included in the analysis for the years 2004, 2006, 2007 and 2008.

Results showed relevant discrepancies between the spatial distribution of LISA scores and Bayesian risks (Figure 1) in the analysed years. The existence of high-low autocorrelated prevalence areas was evident but it was impossible to identify their specific SU and geographical limits by simply analysing maps. Using our proposed MOEA, hot and colds were identified and located in the space (Figure 2). When all the fitness functions agreed, a SU was really considered as a member in the hot-coldspot (level 4 in Figure 2), when only one fitness function identified a SU as a potential member of a spot, a level 1 was assigned to locate it in the space.

The spatial location of hot and colspots for the schizophrenia prevalence in Andalusia remains constant throughout the time span. This area is located in a rural and mountainous area and surprisingly without specialized mental health-care units.
Table 1: Structure of the MOEA to identify hot and coldspots in the space

<table>
<thead>
<tr>
<th></th>
<th>Create the external Non-dominated Solution File (NDF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td><strong>For</strong> ( i = 1 ) to ( s^{(1)} ) <strong>do</strong></td>
</tr>
<tr>
<td>3</td>
<td>Initial population design, ( S )</td>
</tr>
<tr>
<td>4</td>
<td><strong>For</strong> ( i = 1 ) to ( g^{(2)} ) <strong>do</strong></td>
</tr>
<tr>
<td>5</td>
<td>Compute the fitness of each individual in ( S ) and</td>
</tr>
<tr>
<td></td>
<td>( NDF )</td>
</tr>
<tr>
<td>6</td>
<td>Preserve all non-dominated solutions in ( NDF )</td>
</tr>
<tr>
<td>7</td>
<td>If ( NDF ) is too big then the truncated operator</td>
</tr>
<tr>
<td></td>
<td>removes unnecessary solutions</td>
</tr>
<tr>
<td>8</td>
<td>Empty registers in ( NDF ) are filled out using</td>
</tr>
<tr>
<td></td>
<td>dominated solutions</td>
</tr>
<tr>
<td>9</td>
<td>Binary tournament selection with replacement</td>
</tr>
<tr>
<td>10</td>
<td>Crossover (Elitism is optional)</td>
</tr>
<tr>
<td>11</td>
<td>Mutation (Elitism is optional)</td>
</tr>
<tr>
<td>12</td>
<td>Repair process</td>
</tr>
<tr>
<td></td>
<td>Elitism is optional. Two stopping criteria (statistical</td>
</tr>
<tr>
<td></td>
<td>error) are analysed.</td>
</tr>
<tr>
<td></td>
<td>Simple or double.</td>
</tr>
<tr>
<td></td>
<td>At random, distance-based or fitness-based.</td>
</tr>
<tr>
<td></td>
<td>Structural and technical infeasibilities.</td>
</tr>
<tr>
<td>13</td>
<td><strong>end for</strong></td>
</tr>
<tr>
<td>14</td>
<td><strong>end for</strong></td>
</tr>
</tbody>
</table>

(1) Number of different initial solutions \( s \). (2) Number of generations \( g \).

Main conclusions

Our MOEA has demonstrated to be a useful and robust method to identify and locate hot and coldspots for selected illnesses. It can be adjusted to incorporate new spatial analysis procedures as objectives, so it can be easily generalized. The identification of hot and coldspots can be formulated as a non-supervised clustering where each spatial econometric model may be considered an objective and where unknown algebraic and spatial trade-offs must be handled. In a specific pilot test, our MOEA identified a hot-spot in a small isolated health area of Andalusia which had been previously traced by other methods [16] using a more accurate health database (case-register) [17].

As the use of spatial analysis increases, there is mounting demand for international standards on the principles and methods of spatial analysis. Merging tools for spatial analysis are needed in health technology assessment and spatial analysis. The spatial distribution of schizophrenia in Andalusia at municipality level analysed by the MOEA shows that this tool is strategically relevant for health care management.

Acknowledgements

This study was supported by the Ministry of Health of Spain [PI08/90752]. We thank the mental health program in Andalusia (Andalusian Health Service) their ability to access the System of Mental Health Information System of Andalusia.
Figure 1: Spatial distribution of autocorrelation scores (Moran’s I, Geary’s C and Getis and Ord’s G) and Bayesian Conditional Autorregresive (CAR) risks: Schizophrenia 2006.

Figure 2: Evolution throughout the time span of the spatial location of schizophrenia hotspots
References