Towards Memetic Algorithms in GIScience: An Adaptive Multi-Objective Algorithm for Optimized Delineation of Neighborhood Boundaries

G. Fraley, M. Jankowska, P. Jankowski

Department of Geography, San Diego State University, 5500 Campanile Dr, San Diego, CA 92182-4493
Email: fraley@rohan.sdsu.edu

1. Introduction

Practical geographic problems often involve multiple objectives used by analysts to describe the quality of a solution. Multiple-objective genetic algorithms (MOGA) are capable of generating a population of solutions where each solution represents a trade-off in optimality between the multiple objectives. In the nomenclature of MOGA, the algorithm's decisions towards optimality are driven by natural selection, which prunes solutions towards specific objectives. This can be thought of as a global search heuristic that applies to all problems. Recent research in computer science has explored the design of memetic algorithms, which add cultural learning components into genetic algorithms (GAs) to exploit problem-specific knowledge (Knowles and Corne 2005). The notion of culture, or local social rules and norms, can adapt global search heuristics to produce specific solutions that are pertinent for the problem at hand. It is suggested that incorporating problem-specific knowledge can improve the performance of a MOGA (Corne and Knowles 2003). Specifically, geographic problems can be characterized by spatial patterns and topology, which are rarely included in MOGAs, but could be leveraged as part of a memetic algorithm for generating efficient solutions that incorporate spatial characteristics. Yet, little research has been presented which explicitly includes geographic constructs in genetic algorithms. Memetic algorithms are an extension of GAs. In a memetic algorithm, after the genetic operators are applied to the population, each solution in the population undergoes a refinement step that seeks to improve objective performance.

This study focuses on the design of a memetic algorithm, which employs NSGA-II (Deb et al 2002) as the method of global search for solutions, and the AMOEBA spatial autocorrelation clustering technique (Aldstadt and Getis 2006) as the problem-specific local search heuristic to improve individual solution quality. The local heuristic is specifically selected for the problem of neighborhood boundary delineation for health modeling in Accra, Ghana, Africa. An optimization problem is formulated with two objective functions. By applying spatial characteristics to the MOGA search heuristic, optimized neighborhoods are explored within the function space.

2. Background

Spatial optimization methods belong to a class of techniques that aim at generating solutions representing a location or geographic arrangement of locations that perform well on the problem objectives. Upon formulating an optimization problem an analyst must decide how to generate a solution or even multiple alternative solutions to the problem. It is proven by "No Free Lunch" (NFL) theorems that all search and optimization algorithms are equal in average performance over all problems (Corne and Knowles 2003, Wolpert and Macready 1997). What follows from the NFL theorems is the supposition that incorporating problem-specific knowledge to guide the search for optimal solutions can make one algorithm perform better than another (Wolpert and Macready 2005, Corne and Knowles 2003).

Genetic algorithms are heuristics inspired by Darwinian ideas of evolution suggesting that fit members will emerge in a population through natural selection. A genetic algorithm begins with an initial randomly generated population of individuals, or potential solutions. Decision variables (locations, for example) of the optimization problem are termed chromosomes in the GA, which are often a string of bits or an array of values, where each value is called a gene. Typically, individual solutions with high fitness (parents) are selected for reproduction using genetic operators such as selection, crossover, and mutation, which will hopefully result in higher performing children.

MOGA heuristics may require a large number of computations because many solutions are evaluated at each generation. The geographic information science literature has devoted little attention to the design of MOGAs that incorporate problem specific knowledge to improve convergence towards a Pareto set, and also reduce computation time. Xiao (2008) presents a framework to exploit spatial structure in GAs, and strives to avoid approaches that are too specific to individual problems. Tong et al. (2009) introduce in their single objective genetic algorithm a crossover operation that is specific to facility location problems and incorporates the geographic arrangement of facilities to promote dispersion. In this research we introduce a memetic algorithm that leverages geographic knowledge of the diversity in average neighborhood size and the diversity of the average neighborhood local spatial autocorrelation within the MOGA population.

3. Methodology

The memetic algorithm in this research draws on geographic knowledge by (1) adapting genetic operator probabilities based on the diversity in spatial pattern among the population of solutions, and (2) introducing a spatially driven refinement step into a MOGA based on the AMOEBA spatial clustering heuristic. Fuzzy adaptation is a technique that examines the population throughout execution of the algorithm and adjusts operator probabilities to attain better solutions (Tarokh 2008). We use a geographic metric to examine the population and provide feedback by changing operator probabilities.

The AMOEBA algorithm for spatial clustering (Aldstadt and Getis 2006) is used as a local search heuristic for the memetic algorithm. AMOEBA is a greedy algorithm that builds contiguous areas (clusters) from polygons by using the local spatial autocorrelation of a polygon attribute. In AMOEBA, a seed (starting polygon) is chosen, and adjacent polygons are added to form a cluster if they increase the spatial autocorrelation statistic for the cluster of polygons. This research uses the local Gi* statistic of spatial autocorrelation (Ord and Getis 1995) to guide the AMOEBA algorithm.

3.1 Model Formulation

One task within a larger project examining health in Accra, Ghana is the optimal delineation of neighborhood (cluster) boundaries using enumeration areas (EAs). Data concerning health outcomes is only present for a subset of EAs, but variables that impact health are measured for all EAs and utilized as inputs for boundary delineation with the AMOEBA algorithm. An ideal neighborhood solution would demonstrate high spatial autocorrelation for a known factor that influences health, as well as high R^2 values for the regression model of a health outcome. The individual EAs in a given neighborhood are used to estimate the health outcome at the neighborhood level in the regression model. Two objective functions are formulated to determine optimality of a neighborhood solution. The first is to maximize the R^2 of a linear regression that estimates average body mass index (BMI) of a neighborhood, and the second is to maximize the average Gi* value of neighborhoods in the solution, thus maximizing homogeneity of the neighborhood characteristic of interest, socio-economic status. The goal is not necessarily to find a neighborhood set that best predicts health outcomes, but rather to explore and understand how small scale geographical changes can impact both spatial autocorrelation of neighborhoods, and global statistical models.

3.2 A Spatial Memetic Algorithm

The non-dominated sorting genetic algorithm II (NSGA-II) (Deb et al. 2002) is the MOGA that we use as the basis for our memetic algorithm. First AMOEBA is used to generate a neighborhood cluster of socio-economic status from each possible seed enumeration area. For initial MOGA solutions these neighborhoods are randomly selected until each enumeration area belongs to a cluster, and overlapping clusters are merged to create a spatially contiguous neighborhood map. The genetic crossover operator is applied by spatially overlaying and intersecting two solutions, then repairing any neighborhoods that do not meet the contiguity constraint (Xiao, 2008). Swapping the neighborhood membership of EAs that lie on the boundary of two neighborhoods carries out the genetic mutation operator in the algorithm.

NSGA-II was augmented to include a fuzzy adaptation (Tarokh 2008) of the probabilities assigned to each genetic operator (mutation and crossover) based on two metrics, one describing the diversity of the population of solutions in terms of objective function values, or spread (Deb 2001), and another describing the spatial diversity of the population using the variance in the average area of clusters within the population. This methodology promotes geographically different solutions in terms of average neighborhood size and also strives to generate solutions that represent a wide range of objective function trade-offs. Fuzzy rules were defined with the notion that in a diverse solution set crossover should be applied with higher probability to create more high-performing individuals in the population, but in a low diversity solution set mutation should be applied with higher probability to increase diversity of the population by introducing solutions with potentially dissimilar chromosomes.

After applying genetic operators and adaptation in each iteration, the AMOEBA algorithm is used to modify clusters by applying the mutation operator to the solution, but only accepting mutations of the genes if they increase the Gi* of socio-economic status in the solution's neighborhood clusters (Figure 1).

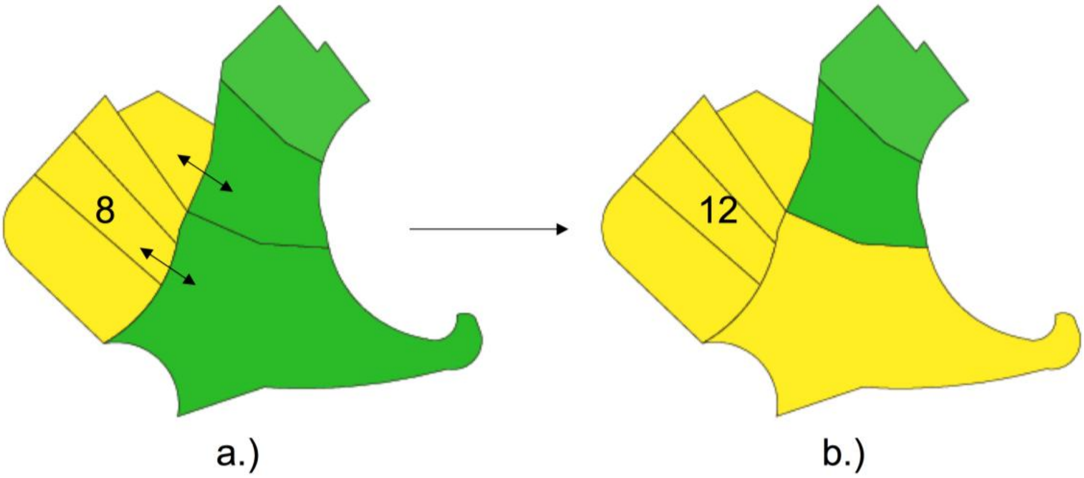

Figure 1. AMOEBA is used as a local improvement step in the memetic algorithm. A small subset of two neighborhoods delineated in Accra, Ghana is shown in (a.). The seed neighborhood (yellow) starts with a Gi* value of 8. Each boundary EA is swapped with the adjacent neighborhood (green) and if the Gi* value increases it is kept as part of the seed neighborhood. In this example, the southernmost adjacent EA produces a neighborhood with a greater Gi* value of 12, while the northernmost adjacent EA does not increase the Gi* value and so is not included in the resulting neighborhood (b.).

4. Conclusion

A spatial memetic algorithm was designed with the ability to provide a set of decision options that are diverse in objective values and geographically unique. In comparison to solving the optimization model with a standard implementation of NSGA-II, spatial fuzzy adaptation preserved objective diversity effectively. Our memetic algorithm leverages knowledge of the local spatial autocorrelation of socio-economic status in neighborhoods to guide local improvements to solutions. Improving individuals by introducing the notion of cultural, problem-specific, knowledge using the AMOEBA algorithm increased the average objective performances of decision alternatives. Further research will probe into the implications of using the delineated neighborhoods of the memetic algorithm solutions as the spatial units for analysis related to the BMI health outcome. It is expected that the form of the memetic algorithm utilized in this research is applicable to a wide variety of geographic representations of decision problems. Future work will recommend a framework for implementing adaptation and local spatial search heuristics in the memetic stage of the algorithm for problems with the generalized geographic structures of points, networks (lines), and areas (polygons).

References

Aldstadt J and Getis A., 2006. Using AMOEBA to create a spatial weights matrix and identify spatial clusters. *Geographical Analysis*, 38(4): 327-343.

Corne D and Knowles J, 2003. No Free Lunch and Free Leftovers Theorems for Multiobjective Optimisation Problems. *Evolutionary Multi-Criterion Optimization*, 66, Springer.

Deb K, 2001. Multi-Objective Optimization Using Evolutionary Algorithms. Wiley.

Deb K, Pratap A, Agarwal S and Meyarivan T, 2002. A Fast and Elitist Multi-Objective Genetic Algorithm: NSGA-II. *IEEE Transaction on Evolutionary Computation*, 6(2): 181-197.

Knowles J and Corne D, 2005. Memetic Algorithms for Multiobjective Optimization: Issues, Methods and Prospects. *Recent Advances in Memetic Algorithms.* Springer Berlin / Heidelberg.

Ord J and Getis A, 1995. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, 27: 286-306.

Tarokh M, 2008. Hybrid Intelligent Path Planning for Articulated Rovers in Rough Terrain. *Fuzzy Sets and Systems*, 159 (21): 2927-2937.

Tong D, Murray A and Xiao N, 2009. Heuristics in Spatial Analysis: A Genetic Algorithm for Coverage Maximization. *Annals of the Association of American Geographers*, 99(4): 698 - 711.

Wolpert D and Macready W, 1997. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1:67–82.

Wolpert D and Macready W, 2005. Coevolutionary free lunches. *IEEE Transactions on Evolutionary Computation*, 9(6):721-735.

Xiao N, 2008. A unified conceptual framework for geographical optimization using evolutionary algorithms. *Annals of the Association of American Geographers* 98(4): 795–817.