CO-227

FACILITY SITTING USING GIS AND GENETIC ALGORITHMS

MEHDIPOUR F.(1), MESGARI M.S.(2), GOLCHIN B.(3), AKBARI M.R.(1)

(1) National cartographic center, TABRIZ, IRAN, ISLAMIC REPUBLIC OF ; (2) Faculty of Geodesy and Geomatics Eng. K. N. Toosi University of Technology, TEHRAN, IRAN, ISLAMIC REPUBLIC OF ; (3) Islamic Azad University of Ahar, AHAR, IRAN, ISLAMIC REPUBLIC OF

ABSTRACT

One of the important questions, in the establishment of a facility in an area, is about its optimum location. Geo-spatial information systems provide a variety of analytical tools for the integration of different spatial data, related to the parameters affecting the suitability of a location.

In this regard, proper models, algorithms and ideas are required to be developed, in order to find the proper site. This article elaborates the usability of genetic algorithms in GIS environment for site selection. In addition, it describes the difficulties related to GA's implementation in comparison with other methods.

In the literature, the suitability of this algorithm for complex and non-structured problems is not properly reported. Probably because of its complex logic, the algorithm is rarely used for spatial data integration in site selection process.

In this article, the application of GA, for the selection of proper sites for building service areas beside roads, is reported. First, the data and analysis requirements of the application were studied. On the basis of this, the components of GA were designed. Then, the required data were collected, processed and entered into the algorithm. As a result, the proper areas for the establishment of service areas were selected. As a final point, a comparison is made between the selected areas and the results of other approaches and methods.

KEYWORDS

facility siting, Genetic Algorithms, GIS, geospatial analysis, Service areas.

1. INTRODUCTION

Today Genetic Algorithms as common search and optimization techniques are widely applied for solving real and complex problems. By scanning all over the search space and evaluation of all possible solutions, these algorithms can lead to optimum or pseudo optimum solution. In recent years, GAs are developed for solving spatial problems and implanting spatial analysis. One of these analyses is optimal location search that is frequently required in many urban applications for locating one or more facilities. However, the search may become very complex when it involves multiple sites, various constraints and multiple objectives.

In this paper, we will focus in GA usage in finding best location for constructing some service areas along with roads between cities.

2. WHAT IS GENETIC ALGORITHM?

Genetic algorithms are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover (Goldberg 1989).

GAs are implemented as a computer simulation in which a population of abstract representations (called chromosomes or the genotype) of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem evolves toward better solutions. The evolution usually starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified (mutated or recombined) to form a new population. The new population is then used in the next iteration of the algorithm.

3. GA PROCEDURE

A typical genetic algorithm requires two things to be defined:

1. A genetic representation of the solution domain
2. A fitness function to evaluate the solution domain

Once we have the genetic representation and the fitness function defined, GA proceeds to initialize a population of solutions randomly, and then improve it through repetitive application of mutation, crossover, and selection operators (Hyma et. al, 2010).

3.1 Initialization

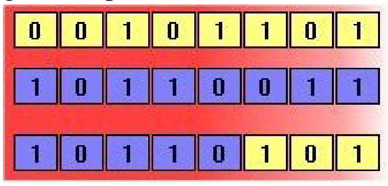
Initially many individual solutions are randomly generated to form an initial population. The population size depends on the nature of the problem, but typically contains several hundreds or thousands of possible solutions. Occasionally, the solutions may be "seeded" in areas where optimal solutions are likely to be found.

3.2 Selection

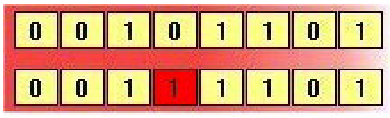
During each successive epoch, a proportion of the existing population is selected to breed a new generation. Individual solutions are selected through a fitness-based process, where fitter solutions (as measured by a fitness function) are typically more likely to be selected. Certain selection methods rate the fitness of each solution and preferentially select the best solutions. Other methods rate only a random sample of the population, as this process may be very time-consuming. Popular and well-studied selection methods include roullet wheel selection and tournament selection.

3.3 Reproduction

The next step is to generate a second generation population of solutions from those selected through genetic operators: crossover (also called recombination), and/or mutation (Figure 1).



a) Operation of Crossover



b) Operation of mutation

Figure 1. GA operations

For each new solution to be produced, a pair of "parent" solutions is selected for breeding from the pool selected previously. By producing a "child" solution using the above methods of crossover and mutation, a new solution is created which typically shares many of the characteristics of its "parents". New parents are selected for each child, and the process continues until a new population of solutions of appropriate size is generated.

3.4 Termination

This generational process is repeated until a termination condition has been reached. Common terminating conditions are

-A solution is found that satisfies minimum criteria

-Fixed number of generations reached

-Allocated budget (computation time/money) reached

-The highest ranking solution's fitness is reaching or has reached a plateau such that successive iterations no longer produce better results

-Manual inspection

-Combinations of the above

Pseudo-code algorithm is shown in figure 2.

```
Begin
    t ← 0;
    Initialize P(t);
    Evaluate P(t);
    While (not terminate condition) do
        Begin
            Recombine P(t) to yield C(t);
            Evaluate C(t);
            Select P(t+1) from C(t);
            t ← t + 1;
        End
    End
End
```

figure 2. Pseudo-code algorithm

4. GA CAPABILITIES FOR FINDING BEST LOCATION

An often encountered spatial decision problem is to search for the best site or sites to accommodate one or more facilities to generate the best utility values (e.g. the maximum population coverage and minimum transport cost). Traditional location-allocation methods before GIS data were available only use relatively small datasets (Church 1999). The general facility location problem and its variants, including most location-allocation and p-median problems, are known to be NP-hard combinatorial optimization problems. Most of these traditional methods cannot easily handle thousands of demand points and sites in GIS datasets (Church 1999). This is especially a problem when raster data with many cells are used. Some types of data aggregation have been used in dealing with large data sets (X. Li et al., 2005). Although agglomerative clustering can be used to find good solutions, this problem can be complicated by including the items that need to be stored at each location. In this case, clustering will not yield an optimal solution. Goodchild (1979) has pointed out that data aggregation can have a great effect in the absolute location of a specific facility. Consequently, there is now substantial literature on heuristic algorithms for a variety of location problems, among which can be found the well-known simulated annealing algorithm (Simha et al., 2001).

Location search usually requires the use of optimization tools. There are two categories of optimization methods. The first is the local optimization method, such as simplex, Gauss-Newton, and the Levenberg-Marquart (Zhan et al. 2003). These local optimization algorithms have limitations because the search may be trapped in local minima or maxima and their success is heavily dependent on the choice of initial values. The second is the global optimization method which can avoid such problems. These algorithms include simulated annealing and genetic algorithms (GAs). Studies have indicated that GAs are attractive global search tools suitable for the multimodal objective functions (Zhan et al. 2003). GAs have advantages for global optimization without using complicated calculations, and they are effective especially when the number of parameters is very large (Jin and Wang 2001). GAs are stochastic search algorithms for searching optimal solutions in large and complex non-linear spaces.

The objective of this paper is to explore the capability of GAs in solving high-dimensional optimization problems in raster data model. Although GAs have been widely used for searching optimal parameter values, there are limited studies on the integration of GAs and GIS for solving optimization problems in resource and environmental management. There are many mathematical methods which can find optimization solutions very quickly for fairly ‘‘well-behaved’’ problems. However, these traditional methods tend to break down when the problem is not so ‘‘wellbehaved’’. The use of evolutionary algorithms should be very efficient in solving a lot of spatial decision problems. The integration of GAs and GIS can help to find optimal solutions for a variety of geographical problems. This study will demonstrate that complex spatial search problems involving multiple-objectives and constraints can be conveniently tackled by using GAs and GIS.

5. CASE STUDY

In this paper we want to use GA for finding best location of n service areas along whit a part of major road between cities of Tabriz and Bazargan located in east and west Azerbaijan states of Iran. This road is one of important gates of country's international transportation and its imports and exports (Figure3).

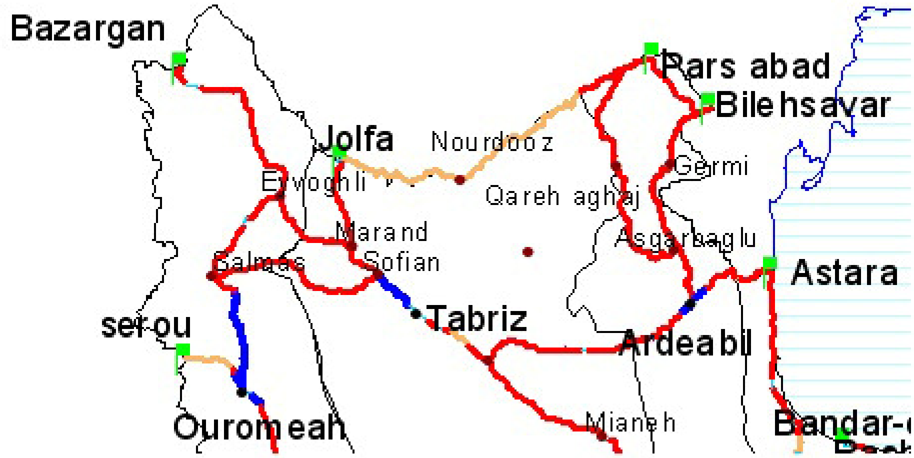


Figure 3. Case study map

6. DESIGN AND IMPLEMENTATION A GENETIC ALGORITHM FOR SITTING SERVICE AREAS

6.1 Required data layers

According to our studies and interviews a set of effective factor in service area locating process was produced and the useful data layers were determined. These data layers have been classified into three main class as follow:

i) Land use: cultivation, forest, building block and industry complex

ii) Underlying equipments: water resources, power line, oil and gas tank, communication tower

iii) Safety factors: distance from road,turn, cross, village, bridge and slope.

The final places shouldn't be located on listed land use layers, have access to underlying equipments and be far from dangerous features.

Required data layers gathered from Road Maintenance & Transportation Organization and National Cartographic Center in SHP format and the scale of 1:250000. Afterwards, some GIS ready methods including determination of projection system and datum, conversion vector layers to raster and producing distance map carried out on shp files that input layers got ready to enter into our designed GA's fitness function.

6.2 Designed model

In best locating problems, the suitable regions have been extracted from a geographical zone which is the search space of problem. Since service areas can be constructed in at least 800m2 areas, raster data model seems to be the suitable search space and we can do our search on cells of matrices that have almost 0.5 hectares of area and derived from rasterized data layers. Therefore, the model's individuals are the cells of all raster data layers. According to this, all of gathered data layers were converted to raster datasets with 75m*75m cell size and were exported from ArcGIS to 1103*1465 matrices. Now, there are two components including row number and column number for each cell which are considered as genes of individuals.

Thus our genes or decision variables are the numbers of each cell's row and column. Evaluation (fitness) function is the index overlay formula for data integration that causes the problem to be solved in single objective manner (figure4):

$$w_{out} = \frac{\sum_{i=1}^{n} w_i \times s_{ij}}{\sum_{i=1}^{n} w_i}$$

figure 4. The fitness function

Where wi is weight of each factor and sij is weight of every suitability class in factors. Finally, we have a constraint of optimization that is locating service areas on safe regions according to safety factors.

After designing our search framework, the site selection model was implemented using Matlab7.0 calculation software and genetic parameters defined as below:

Number of generations: 100

Population size: 100

Crossover rate: 0.8

Mutation rate: 0.05

Selection methods: roulette wheel and tournament

As an extra part, we integrated data layers using map algebra and same weight values in order to be able to compare the output map from GA model with the common linear Index Overlay method.

7. RESULTS

The results of 10 execution of algorithm for finding 10 sites along the road with two common selection methods are shown in tables1 and 2 and the process of each selection method is shown in figures 5 and 6. The figures show that although both of methods lead to same places neighborhood, but the convergence rate and individuals' distribution around optima in tournament selection is too less than roulette wheel method. So tournament selection finds the optima very faster than roulette wheel selection. The main reason could be provided about this result is the execution type of each method. Because tournament method chooses the high ranked parents to produce next generations while roulette wheel method gives equal chance to all parents to have child.

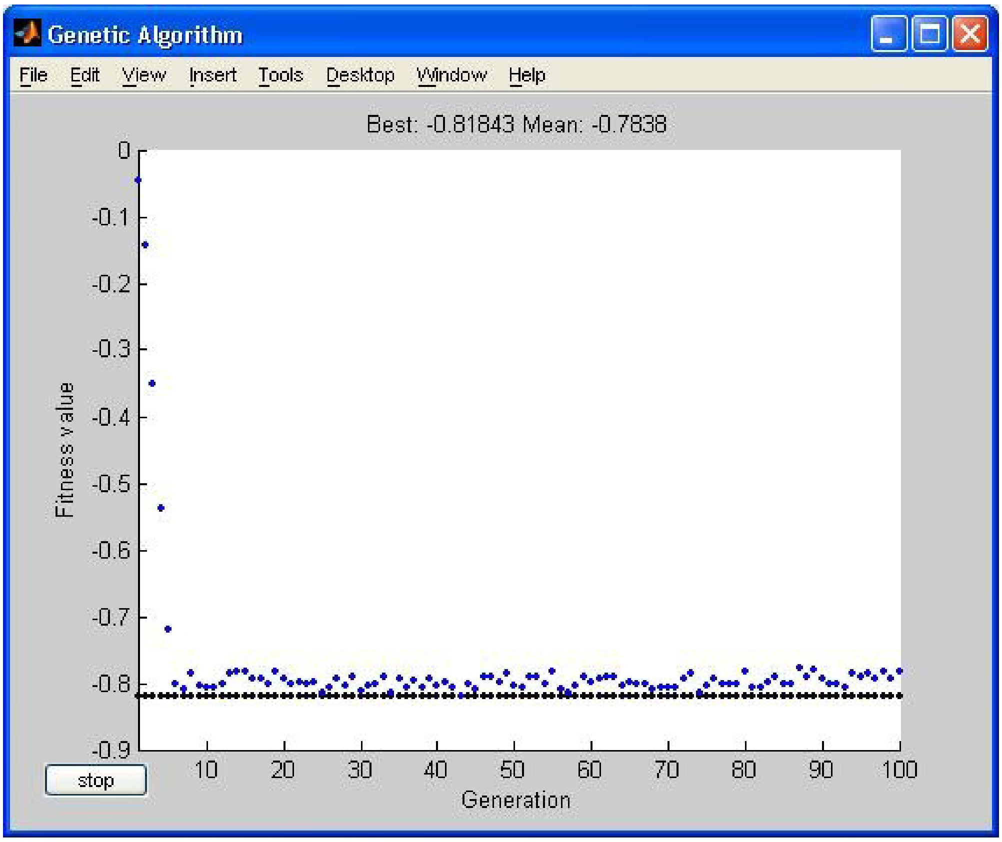


Figure 5. Convergence rate of tournament method

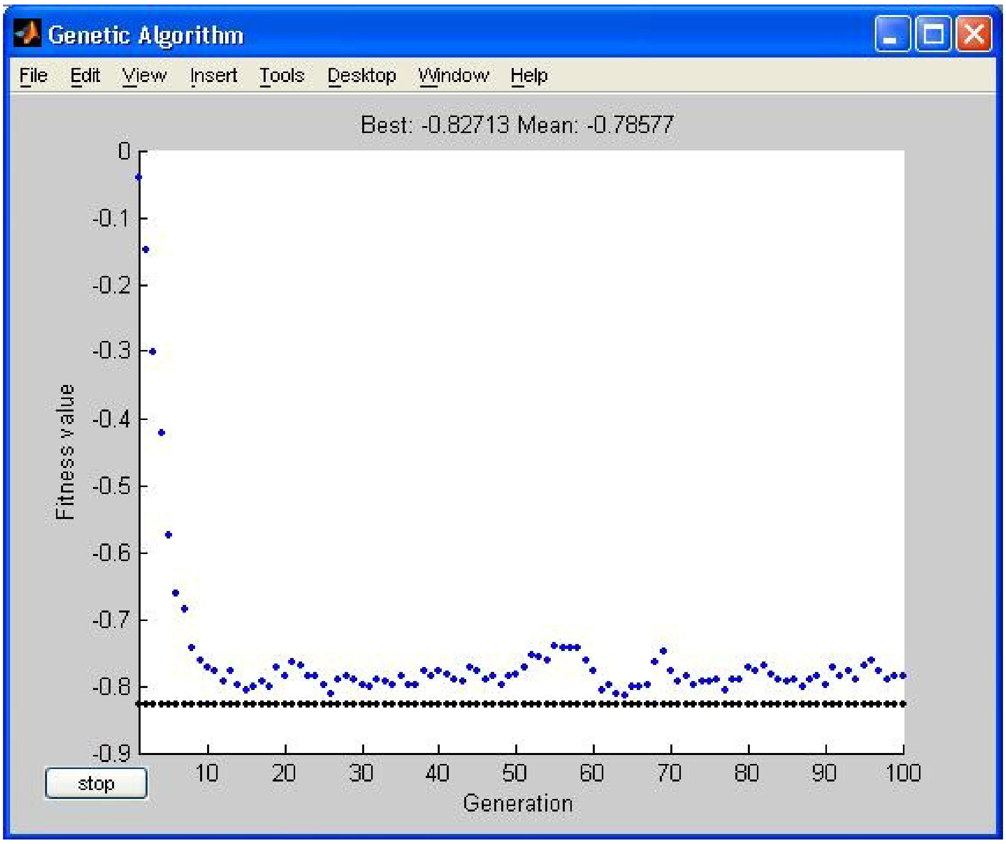


Figure 6. Convergence rate of roulette wheel method

No.	1	2	3	4	5	6	7	8	9	10
Row	99	114	822	279	95	794	133	601	404	422
Column	96	120	1198	219	90	879	122	849	356	368

Table 1.Location of optimum sites using roulette wheel method

No.	1	2	3	4	5	6	7	8	9	10
Row	408	113	97	865	90	248	841	414	781	94
Column	349	118	98	1257	139	183	1196	377	1141	134

Table 2. Location of optimum sites using tournament method

As another result, a comparison of suitable places with areas produced by index overlay method shows that however GA's places are located on the index overlay output areas (figure7), GAs never receives to index overlay's best point and highest weight (0.8314), though it uses same function to evaluate the solutions. This is because of GA's nature and property in finding approximate solutions instead of exact ones. On the other hand, GAs will be the most useful approach, when we can design a mathematically complex fitness function which models all decision makers' preferences according to each criterion.

Another point to consider is the time consumption of executing GA. With the above parameters, GA runs in about 4 seconds that in great values of parameters it runs slow and slower. This is a disadvantage of GA that can be obviated by designing algorithms without unnecessary searches.

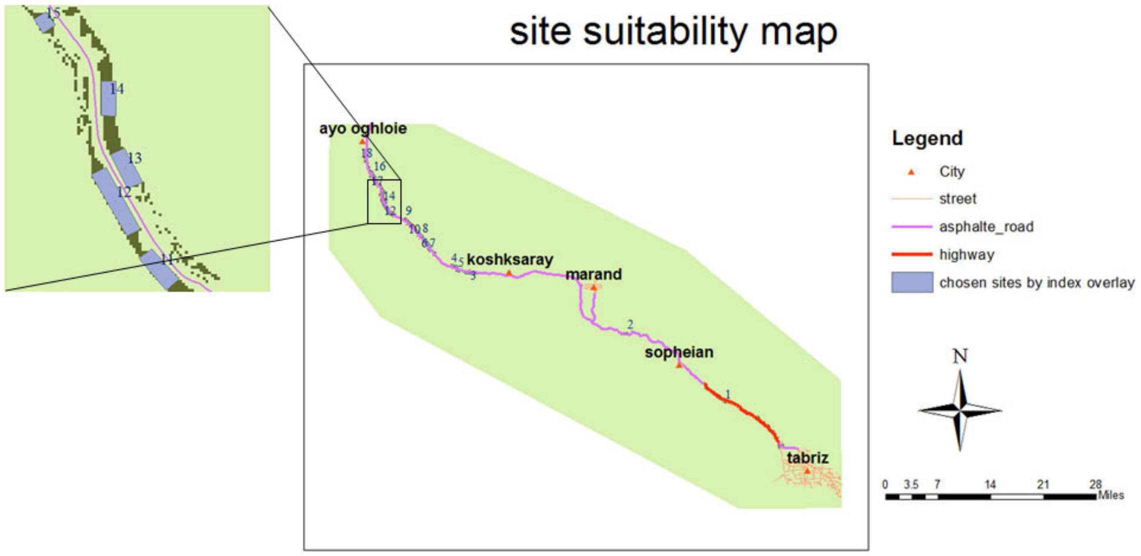


Figure 7. Results of data integration using index overlay method

8. CONCLUSION AND FURTHER PLANS

This study demonstrates that genetic algorithms are capable of producing satisfactory results for optimal location search under complex situations. This method has been tested by solving a spatial search problem which is to allocate a facility according to the land use, underlying equipments and safety constraints derived from a GIS. The GA algorithm becomes very effective through the use of the mechanism of natural selection in biology. The proposed method can be used as a planning tool to solve location search problems under multiple-objectives.

This method has been tested in Tabriz-Bazargan road, a densely transported road. The study indicates that the proposed GA method can be conveniently integrated with GIS to retrieve spatial data and it finds better solutions if it uses stronger evaluation function. It's recommended that to test this algorithm using vector data set for getting more precise result of points or polygon, although it will be more complex than our study. Another point is using Arcobjects toolkit for development of such application in ArcGIS environment. This can do effective help for achieving much more flexibility on favorite parameters of GA.

REFERENCES

1- CHURCH, R.L., 1999, Location modeling and GIS. In Geographical information systems: Volume 1, edited by P.A. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind (New York: John Wiley & Sons, Inc.), pp. 293–303.

2- GOLDBERG, D.E., 1989, Genetic Algorithms in Search, Optimisation and Machine Learning (Reading, MA: Addison-Wesley).

3- GOODCHILD, M.F., 1979, Aggregation problem in location-allocation. Geographical Analysis, 11(3), pp. 240–255.

4- HOLLAND, J., 1992, Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence (Cambridge, Mass: MIT Press).

5- Hyma, J., Jhansi, Y. and Anuradha, S., 2010, A new hybridized approach of PSO & GA for document clustering, International Journal of Engineering Science and Technology, Vol. 2(5), pp.1221-1226.

6- JIN, Y.Q. and WANG, Y., 2001, A genetic algorithm to simultaneously retrieve land surface roughness and soil wetness, International Journal of remote sensing, 22(16), pp. 3093–3099.

7- Li, X. and Yeh, A.G., 2005, Integration of genetic algorithms and GIS for optimal location search, International Journal of Geographical Information Science, Vol. 19, No. 5, May 2005, 581– 601.

8- SIMHA, R., CAI, W.D. and SPITKOVSKY, V., 2001, Simulated N-body: new particle physics-based heuristics for a Euclidean location-allocation problem. Journal of Heuristics, 7(1), pp. 23–36.

9- ZHAN, H.G., LEE, Z.P., SHI, P., CHEN, C.Q. and CARDER, K.L., 2003, Retrieval of water optical properties for optically deep waters using genetic algorithms, IEEE Transactions on Geo-science and Remote Sensing, 41(5), pp. 1123–1128.