Finding The Maximum Monochromatic Polygon

Sara Khalafi, Alireza Bagheri, Mohammad Mansoor Riahi Kashani

Abstract— Covering the set of points with a geometric shape is an important problem in computational geometry. In this problem, given a set of points in the plane of total size *n* and a geometric shape, covering the points with the geometric shape with minimum perimeter or area is the goal. The points may have different colors or divided into desirable and non-desirable points. On the other hand, in the separability problem, it is required that all of the desirable points lay inside the geometric shape and all of the other lay outside it. There has been a fair amount of work on different kinds of separators such as rectangles, squares, circles, etc. In this paper, a new algorithm based on genetic algorithm is presented for finding the maximum monochromatic polygon, which contains the maximum number of desirable points while avoids non-desirable points. Finding the maximum monochromatic polygon is an important problem in computational geometry which has many applications in different fields. Also, another algorithm is introduced based on triangulation of blue points, which has $O(n^2(\log n+m))$ time, where *n* represents the number of blue points and *m* represents the number of red points. Both algorithms are evaluated and compared to optimal solutions. Both algorithms are near-optimal, i.e. their solutions are close to optimal solutions, but they are not necessarily optimal. Of course, in some cases they yield optimal solutions.

Index Terms— Computational Geometry, Genetic Algorithm, Colored Points Coverage, Separation, Polygon Triangulation

1 Introduction

Computational Geometry is one of the important fields in computer science. Computational geometry is utilized in various fields of computer science including machine learning, computer aided design, computer graphics, etc. Dobkin and Gunopulos [1] studied computing the maximum bichromatic discrepancy, with applications to computer graphics and machine learning. Classification and separation of points within different spaces is one of the important subjects in computational geometry. As we know, one of the most optimal geometric classifiers in these problems is Support Vector Machine (SVM) classifier that possesses power at higher level [2]. The method of classification and separation of points by the aid of computational geometry is similar to SVM and it is also very strong. Like SVM method, this technique of problem solving is purposed for a bi-dimensional space and it may be also generalized to multidimensional spaces with more point sets. There are several techniques for this problem that are solvable by this solution technique. For example, finding the maximum polychromatic polygon or tracking the maximum monochromatic polygon, which absolutely includes certain point, is one question that will be answered by this technique. So far, several methods have been also suggested to find the maximum monochromatic polygon. For instance, Fischer [3] presents a method based on points triangulation and optimal classification that has run time $O(n^3)$.

- *Sara Khalafi, Computer Engineering Department, Islamic Azad University, Tehran North Branch, Tehran, Iran*
  E-mail: sara.khalafi@yahoo.com,Tel: +98-9126471443
- *Alireza Bagheri (Corresponding author), Department of Computer Engineering and IT, Amirkabir University of Technology, Tehran, Iran, Computer Engineering Department, Islamic Azad University, Tehran North Branch, Tehran, Iran*
  E-mail: ar_bagheri@aut.ac.ir,Tel: +98-2164542742
- *Mohammad Mansoor Riahi Kashani, Computer Engineering Department, Islamic Azad University, Tehran North Branch, Tehran, Iran*
  E-mail: m_riahi_kashani@iau-tnb.ac.ir,
  Tel: +98-9123302874

In another essay in [4], a technique is given that reduces the time $O(n^3\log n)$ to linear by means of parallelization and it finds the best solution. In this paper a new method is presented by means of Genetic Algorithm, which is very useful and it substitutes mathematical and computational challenges with genetic intelligent techniques. Also, a simple technique is purposed by means of triangulation, which solves the problem in time $O(n^2\log n)$. Although, both techniques do not have optimal times, a new method is given in computational geometry with reliance on it the execution of these algorithms could be improved. The given technique is very similar to the presented method in [3], of course it is much less complicated than this technique. In the following, we initially describe the general technique presented in both algorithms and then we compare the results of both methods with the optimal solutions.

2 The First Algorithm for Finding The Maximum Monochromatic Polygon

2.1 Review Stage

This algorithm is a simple technique based on triangulation that tries to find the maximum monochromatic polygon [3].

Definition of triangulation vertex: Triangulation vertex denotes a point that all the created triangles intersect commonly at this point. The general technique is implemented in such a way that accordingly each of blue points are superimposed as vertices on points in this triangulation while triangulation with maximum number of adjacent triangles is selected to form a polygon in which there is no red point. The aforesaid polygon is created with composition of these adjacent triangles.

2.2 Method of Forming any Triangulation

Any triangulation is done completely by ordering of points. Namely, when a point is taken as a vertex, other points are ordered proportional to this point based on formed angle. This angle may vary from zero to 360 degrees, as shown in Figure 1.

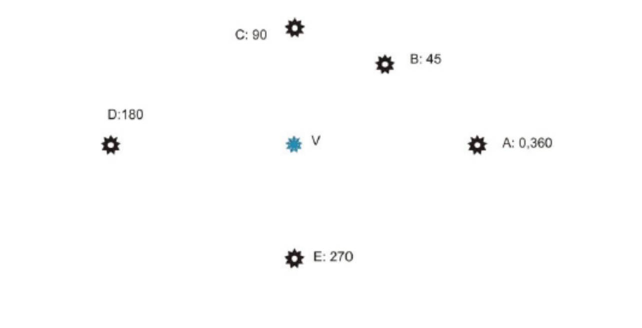

Figure 1. Ordering of points for triangulation.

In this figure, the blue point is vertex V and points of A, B, C, D, E, which has been respectively ordered and their angle has been determined with respect to vertex V. Triangulation starts from V point and by considering two next vertices as other vertices of this triangle in this ordering process so this process passes through a complete round about vertex V to identify all triangles.

2.3 Identifying an Empty Triangle

A blue triangle is assumed as empty provided that no red point exists inside it. In this essay, spiral number is used as the criterion for emptiness of a triangle. Spiral number may be calculated by various techniques. In the simplest way, number of rounds for rotation of a polygon about a point is deemed as spiral number. By this definition, if this number is not zero it means that this point is placed inside the polygon.

2.4 Recognizing Adjacent Triangles

It is an easy and possible task to recognize adjacent triangles with the maximum number of adjacent blue triangles empty from red points.

3 The Secend Algorithm for Finding The Maximum Monochromatic Polygon

3.1 Review Stage

This technique is a genetic algorithm that creates polygon by the same triangulation method of [5].

3.2 Coding of the Algorithm

In this problem, any chromosome represents one polygon that can be an answer for this problem. Also genes are characterized with real numbers each of which displays one number of a blue point. Number of genes also shows the presentation of points in this polygon

11	13	15	8	5	1	12	10	4	2

As it seen here, a chromosome is a vector of real numbers. Any number, which exists in each of genes, is number of a blue point that is involved in formation of this polygon. Regarding this example, blue point no 11 is the first point of the polygon. Blue point no 13 is also the second point in this trend and so on, up to the end of this polygon in point no 2.

3.3 Calculation of Value Function

Value function is used in this algorithm by calculation of number of existing blue points in the polygon and also the red points and their difference. In other words, if a polygon is composed of 20 internal and boundary blue points as well as 5 internal and boundary red points then the amount of the value-function is set as 20 – 5 = 15. Any chromosome with greater value- function is more significant.

3.4 Producing the Initial Population

Here, some chromosomes are produced stochastically with variable length. Any chromosome initially selects a random length with maximum number of blue vertices and at least three vertices. Then, any gene in the chromosome may take a random number ranged from 1 to number of blue points. As a result, one chromosome is produced that represents a polygon. The order of gene placement is determined by an algorithm and through triangulation.

3.5 Normal Choice Operator

This operator selects some chromosomes with respect to their values by means of Roulette Wheel Selection Algorithm so that these chromosomes are involved in this composition and to produce some daughter chromosomes. Any daughter is also a chromosome, which may be better or worse than its parents. Here of course, an ordinary principle may be always used that the probability of reproducing a better daughter from the good parents is higher than producing better daughter from a bad composed parents. Thus, in selection policy, always a higher chance is assumed for the better chromosomes.

3.6 Crossover Operator

This operator is described with an example. Suppose two following chromosomes are selected as parents:

4	14	5	16	1	20	7	8	11	12

12	11	3	4	1	5	10	15	19

We assume that we have totally 20 blue points respectively ranged from one to twenty. Each of two above chromosomes represents a polygon with 10 vertices and another with 9 vertices. A cutoff point is randomly determined for any chromosome at each point so that a chromosome is divided into two segments. Now, the first segment from first chromosome is combined with the second segment of the second chromosome and vice versa so two daughters form. The following figure shows these two daughters:

4	14	5	16	1	15	19

12	11	3	4	1	5	10	20	7	8	1	12

Currently, the repeated genes which represent repetition of vertices are deleted from the chromosomes. The following image is a result of omission of these genes and it displays new chromosomes:

4	14	5	16	1	15	19

12	11	3	4	1	5	10	20	7	8

3.7 Mutation operator

For first mutation operator, the following example indicates the quality of action by this operator:

3	4	5	8	11	12	10

As it seen, the selected genes are numbers 5, 8, and 10. Now, if we suppose that there are 20 blue points the value of any selected gene can be replaced with a random number within the range from one two 20. The following image shows the result of producing the new chromosome:

3	4	13	4	11	12	19

Now, the repeated genes should be omitted so that to produce a sound and healthy chromosome:

3	4	13	11	12	19

At last, this chromosome is ordered by a triangulation algorithm so that it represents an appropriate polygon. For second mutation operator, one chromosome is selected and one gene is added to this chromosome in random position. This gene adds a new vertex to this chromosome. For example, take the following chromosome:

3	4	13	11	12	19

Suppose that the second position is selected randomly. Currently, a gene is added after the second gene that shows a new vertex:

3	4	7	13	11	12	19

3.8 Substitution operator

A few chromosomes from the former generations are directly selected with the higher value among this population and move toward the population in the next generation. Also, among chromosomes in new generation, some chromosomes with the maximum value are directly selected and move toward new generation. The rest of chromosomes are also selected randomly among the chromosomes in the former generation and they form a new generation that includes the same number of former generation of chromosomes. Operation of genetic algorithm is done with certain number of iteration and finally at the last iteration, the best chromosome with higher value is selected as the final chromosome that forms the maximum monochromatic blue polygon.

4 CALCULATION OF RUN TIME FOR ALGORITHMS

In this section running times of the algorithms are computed.

4.1 First method

Triangulation method involves using a linear ordering relative to angle formed between main vertex and other points. This is done for each blue vertex one time. Of we suppose that there are *n* blue vertices, ordering process has *nlogn* time relative to each vertex, so that the time spent for triangulation is $n^2 logn$. No each triangulation has an unknown number of triangles that must be determined. Are there red points in each triangle? Since there are at most *n* triangle in each triangulation and we have supposedly m red points, the time spent for all triangulations equals : $n \times m(n-2)$. A number of neighboring vacant triangles are combined in each triangulation during n^2 period. Finally, in n time, triangulation is selected with the highest number of neighboring triangles for forming polygon. Therefore the spent time equals: $O(n^2 (\log n + m))$.

4.2 The second method

The second method is a genetic algorithm. For production of initial population, a certain number of members are involved. After being developed in a triangulation using the first known vertex in chromosome each member is developed and the time spent equals the length of chromosome. If we suppose that initial population size equals a constant number *p*, and there are *n* number of blue points and *m* number of red points and maximum length of each chromosome is *n*, then the time spent for production of primary population equals: $p \times nlogm$. In order to recognize the presence or absence of red points inside the polygon, at first we must arrange all points around each blue point during $O(n+m)log(n+m)$ time. There would be a space between each two blue points in which the points are either red or blue and these two points are linked through a segment. If these points are red, we must check if they are located at the left side or right side of segment. Each red point could be checked in O(1) time. So, in order to investigate all red points we need O(*m*) time. Given that this must be done for all parts formed by blue points, we need $O(n+m)$ time. Therefore, for recognizing the presence or absence of red points inside the polygon we need $O(n+m)log(n+m)+O(n+m)$ time that is equal to $O(n+m)log(n+m)$. Repetition of genetic algorithm is performed with a certain number. For example, we want to repeat algorithm for *k* times. In each repetition a selection operation is performed in O(*plogp*) time. The time spent for crossover operation equals O(*n*). The time for mutation operation equals O(*n*), too. Substitution operation is performed during O(*p*) time. Overall time of genetic algorithm is equal to $(k \times p)\ (n+m)\ log\ (n+m)$. Since *p* and *k* are both constant numbers and *N* is equal to the sum of red and blue points, the time for this algorithm is: $O(N \log N)$.

TABLE 1 THE VALUES OF VARIABLES AND PROBABILITIES OF GENETIC ALGORITHM

Max_Gen	Pm2	Pm	Pc	Pop_Size
k	0.8	0.85	0.95	P

In This Table: Max_Gen=The number of iterations of genetic algorithm, Pm= The first mutation rate , Pm2= The second mutation rate , Pc=The Cross over rate , Pop_Size= Members of population. k and P are constant numbers.

5 Results and Discussion

The values of variables and probabilities of genetic algorithm are given in Table 1. For testing efficiency of both algorithms, some tests are designed. In these tests, it is supposed that there are 50 points in the plane of which 25 points are blue and 25 points are red. Both algorithms are repeated 10 runs and the results are presented in Table 2. As is evident, for each 10 cases, near optimal answer is found by both algorithms. Of course, in many cases, optimal solution is obtained as well and both algorithms have suitable run time. For testing efficiency of both algorithms, one test is designed. In this test, there are 200 points in the plane of which 110 points are blue and 90 points are red. Both algorithms are run and their results are shown in Figure 2. In this test both algorithms are able to separate the blue and red points, completely.

6 Conclusion

In this article, a new method was introduced by means of genetic algorithm to find the maximum monochromatic polygon. Similarly, a traditional method and technique was designed and implemented for comparison in suitable run time. The second implemented method was a semi-optimum technique that of course results in optimal response in many cases. But genetic algorithm technique is a new semi-optimum method, which is more powerful and stronger than the second technique on many occasions. With reliance on genetic powerful operators, this technique may always find a semi-optimum response and at the same time in many cases it may purpose the optimal answer as well. As a new method, this technique may serve as the start point for calculations and a modern approach in computational geometry and also improving its run time can expose the researchers with a challenge.

TABLE 2 The Results of Both Algorithms

Test no	First algorithm: Number of blue points in polygon	Second algorithm: Number of blue points in polygon	Number of blue points in optimal polygon
1	19	21	23
2	19	20	22
3	13	15	17
4	20	22	22
5	15	16	18
6	16	17	22
7	18	20	20
8	16	18	23
9	17	18	18
10	20	20	20

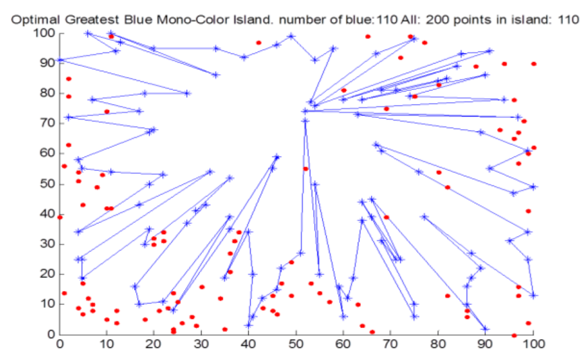

Figure 2. Representation of the used blue and red points in the test

References

[1] D. P. Dobkin, D. Gunopulos, “Computing the maximum bichromatic discrepancy with applications to computer graphics and machine learning,” *Journal of Computer and System Sciences* 52, *pp.* 453–470 , 1996.

[2] J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Kluwer Academic Publishers Hingham,* vol. 2, no. 2, pp. 121-167, June 1998

[3] P. Fischer, "Sequential and parallel algorithms for finding a maximum convex polygon," Elsevier, Computational Geometry: Theory and Applications, vol. 7, no. 3, pp. 187-200, 1997.

[4] C. Bautista-Santiago, J. M. DíAz-BáñEz, D. Lara, P. PéRez-Lantero, P. Urrutia and I. Ventura, "Computing optimal islands," *Elsevier , Operations Research Letters,* vol. 39, no. 4, pp. 246-251, 2011.

[5] D. Goldberg, K. Sastry and G. Kendall, “Genetic Algorithms,” *Search Methodologies*, pp 97-125 , 2001.