A Hybrid System Approach to Determine the Ranking of a Debutant Country in Eurovision

Alberto Ochoa
Universidad Autónoma de Ciudad Juárez, Chi, México
Email: cbr_lad7@yahoo.com.mx

Angel E. Muñoz-Zavala and Arturo Hernández-Aguirre
Centro de Investigación en Matemáticas, Guanajuato, México
Email: {aemz, artha}@cimat.mx

***Abstract*— The goal of the present work is to apply the computational properties of cultural technology, such as data mining and, to propose the solution of a real problem about society modeling: the Eurovision Song Contest. We analyze the voting behavior and ratings of judges using data mining techniques. The dataset makes it possible to analyze the determinants of success, and gives a rare opportunity to run a direct test of vote trading from logrolling. We show that they are rather driven by linguistic and cultural proximities between singers and voting countries. With this information it is possible to predict the rank of a new country, distributing a number of votes of all the participants. A computer model is proposed and solved using a technique based on Particle Swarm Optimization.**

***Index Terms*— PESO, particle swarm optimization, data mining, social modeling**

I. INTRODUCTION

The Eurovision Song Contest (ESC) was born in 1955 and held for the first time in Lugano, Switzerland, in 1956, with seven countries competing. The number of participants increased to 16 in 1961. Non-European countries can also take part: Israel, Morocco, Turkey, Armenia and Georgia are now regular participants. In 2008, Azerbaijan and San Marino will participate for the first time. Since 2002, there are 24 slots for finalists, four of which are reserved for the Big Four (France, Germany, Spain and the United Kingdom). Other countries are guaranteed a slot every other year. Every year the ESC is broadcasted by television, and since 1985, this happens via satellite. In 2001, the contest is broadcast live all around the world. Nowadays, it is watched by several hundred millions of people [9].

The scoring system has changed several times. Since 1975, the first year in our dataset, the 11 (16 between 1988 and 1997) jury members in each country (often a popular jury, not consisting of experts), can rate on a scale from 1 to 10. Tele-voting was introduced in 1998, so that every citizen can participate, and according to Haan et. all [9], "in many countries, the number of people calling in to register their vote is in the hundreds of thousands".

The ratings are normalized so that the favorite song gets 12 points, the next one 10, and then 8, 7, 6, 5, 4, 3, 2 and 1. This allows each voting country to give positive ratings to ten other countries. Participating countries cannot vote for their nationals.

The order in which candidates perform is randomly drawn before the competition starts. After the performance ends, countries are asked to cast their votes. Results are announced country by country, in the same order in which participants perform. Participants are ranked according to their aggregate score.

Eurovision have been studied with different perspectives, for example the compatibility between countries [4] and the political and cultural structures of Europe [18], the persistent structure of hegemony in the Eurovision Song Contest [19], cultural voting [6] and the analysis about Grand Prix which evaluate many countries participating in different years and with different many of countries competing [12], among others.

This research is novelty because it analyzes the behavior of all countries when a new country joins the new ESC. The objective is to estimate the final ranking of Azerbaijan and San Marino, the new contenders in Eurovision Song Contest 2008.

The organization of this article is the following. In Section 2, the 52 past editions of ESC are analyzed and, a priori knowledge about the voting patterns and relationships between neighbor countries is extracted. Next, the problem statement is given in Section 3. The COPSO algorithm is thoroughly explained in Section 4. In Section 5, our approach is tested in the ESC 2007. The experiments and the analysis applied to estimate the final ranking of Azerbaijan and San Marino in ESC 2008 are explained in Section 6. Conclusions are provided in Section 7.

II. EUROVISION RANKING USING DATA MINING

Data mining is the search of global patterns and the existent relationships among the data of immense databases, but that are hidden in them inside the vast quantity of information [3]. These relationships represent knowledge of value about the objects that are in the database. This information is not necessarily a faithful copy of the information stored in the databases. Rather, is the information that one can deduce from the database. One of the main problems in data mining is that the number of possible extracted relationships is exponential [17]. Therefore, there are a great variety of machine's learning heuristics that have been proposed for the discovery of knowledge in databases [16].

One of the most popular approaches to represent the results of data mining is to use decision trees. A decision tree provides a procedure to recognize a given case for a concept. It is a “divide and conquer” strategy for the acquisition of the concept (instance). The decision trees have been useful in a great variety of practical cases in science and engineering; in our case we use data mining to characterize the historical voting behavior for each country. Thus, we selected societies that have participated and characterized its behavior based on its votes previously emitted, which allowed to describe so much to the society as to the individual.

The purpose is to explain v_{ij}, the vote (that is, the number of points) cast by the people of country $i \ni L$ in evaluating the performer of country $j \ni L$ ($i \neq j$, since country i can not vote for its own candidate), where L is the total number of participating countries.

If countries i and j ($i \neq j$) exchanged their votes, without taking into account any other feature, the voting equation could simply be written

$$v_{ji} = \alpha_{ij} v_{ij} + u_{ij} \quad (1)$$

where α_{ij} is a commitment parameter, and u_{ij} a random disturbance. If exchanges of votes were “perfect”, and both countries kept their commitment, α_{ij} would be equal to 1.

More generally, such an equation should contain variables x_{ik}, $k=\{1,...,K\}$ representing the characteristics (language in which sing -English, French, Italian -, lyrics, music, genre and others) of a performer (singer or band) from country i, and variables z_{it}, $t = \{1,..., T\}$ representing the performances of the country i along its T_i participations in the ESC

$$v_{ji} = \alpha_{ij} v_{ij} + \beta \sum_{k=1}^{K} x_{ik} + \gamma \sum_{t=1}^{T_I} z_{it} + u_{ij} \quad (2)$$

where β and γ are parameters to be estimated. The information associated with beta parameter is related with the attributes of performance of a song (music, lyrics, and language) and her/his/their interpreter(s). The information associated with gamma parameter is related with the performance of a country during the ESC participations (example: Armenia has participated in 2006 and 2007).

A problem is concerned with the fact that v_{ij} will appear on the other side of the equation for the observation concerning the vote of country i for the singer representing country j. This can be dealt with in several ways. First, and this is the easiest way, instead of using v_{ij} in the right-hand side, one can use the vote cast in the previous competition, say v^{-1}_{ij}, though one could think that countries would not necessarily keep their commitment over time. An alternative is to use only half of the observations along all ECS editions; thus, every v_{ij} that appears in the right-hand side of the equation is not used in the left-hand side.

The voting equation is estimated by linear methods. The influence of the order in which musicians appear in a competition has often been outlined. In [5,7,8] we observe that in one of the top-ranked international piano competitions, the Queen Elizabeth competitions, those who perform first are less likely to receive high ratings. Similar observations are made by [9] for the contest that we are dealing with. The exogenous order in which candidates perform is thus included as determinant. Other variables include (a) a dummy for host country, determined by the citizenship of the previous year’s winner-the variable takes the value 1 for the performer whose citizenship is the same as that of the host country, (b) the language, in which the artist sings (English, French, Spanish, Italian, etc.), (c) gender of the artist, and (d) whether the artist signs alone, in a duet or in a group. The last group of variables will include linguistic and cultural distances between voters and performers, and may dispense us from using variables that characterize voters.

National culture differences are represented by the four dimensions studied in [14, 2]. The authors claim that these ideas started with a research project across subsidiaries of the multinational corporation IBM in 64 countries. Subsequent studies by others covered students in 23 countries, elites in 19 countries, commercial airline pilots in 23 countries, up market consumers in 15 countries, and civil service managers in 14 countries. These studies identified and scored the four following dimensions that make for “cultural distances”:

(a) Power Distance: It measures the extent to which the less powerful members of a society accept that power is distributed unequally; it focuses in the degree of equality between individuals;
(b) Individualism: It measures the degree to which individuals in a society are integrated into groups; it focuses on the degree a society reinforces individual or collective achievement and interpersonal relationships;
(c) Masculinity: It refers to the distribution of roles between genders in a society; it focuses on the degree a society reinforces the traditional masculine work role of male achievement, control, and power;
(d) Uncertainty Avoidance: It deals with a society’s tolerance for uncertainty or ambiguity, and refers to man’s search for truth.

Table 1 illustrates the correlations between the cultural distances and native languages for the countries that are present in our sample. Uncertainty avoidance is correlated with three other variables, but otherwise, distances seem to pick very different dimensions of people’s behavior.

One of the most interesting characteristics observed in this experiment were the diversity of the cultural patterns established by each community. The structured scenes associated with the agents can not be reproduced in general, so that the time and space belong to a given moment. They represent a unique form, needs and innovator of adaptive behavior which solves a followed computational problem of a complex change of relations. The generated configurations can metaphorically be related to the knowledge of the behavior of the community with respect to an optimization problem (to make alliances, to obtain a better ranking). Columns (a) to (d) of Table 2 contain the results of an estimation of equation 2 (by ordinary least squares, OLS).

We first observe that quality always plays a very significant role, which should of course not be surprising. Logrolling is significant only in (a), in which no account is taken of linguistic and cultural distances. It ceases to be so in all the other equations once linguistic and/or cultural distances are also accounted for. Note that even when the coefficient is significantly different from zero, its value is very small. Order of appearance plays no role, while among the other variables, the only one which has some influence is “sung in French”. Though not all distance coefficients are significantly different from 0 at the 5 percent probability level, they all are negative (the larger the distance, the lower the rating).

The Table 3 Column 2 presents the expected performance rates for 2008. The performance rate tries to predict the country rank through environment variables observed along the 52 ECS editions.

The Table 3 shows the performance rate of the last ECS where Ukraine had the highest rate. In the ECS 2007 the winner was Serbia which had a performance rate of 0.55, below the top-10. The performance rates were estimated based on the characteristics listed in Table 4 and the country performance along previously participations in every ESC editions. For example, in ESC 2007, 42 countries participated; hence it was more complex to obtain a second place, than in opposition, a country that obtained second place in ESC 1981 when only 20 countries participated.

Obviously, for the new contenders, Azerbaijan and San Marino, there is not historical information available. The information obtained through data mining, denotes a similar behavior of countries into the same neighborhood and with similar characteristics (language, territorial extension, religion, etc.). Thus, the historical performance for Azerbaijan was calculated from Armenia, Georgia, Bosnia & Herzegovina and Turkey; and for San Marino was calculated from Italy, Switzerland, Andorra, Monaco, Malta and Luxembourg.

TABLE II.
CORRELATIONS BETWEEN CULTURAL DISTANCES AND LINGUISTIC

	Language	Power	Indiv.	Masc.	U. A.
Language	1				
Power	0.205	1			
Indiv.	0.254	0.111	1		
Masc.	-0.092	0.031	-0.128	1	
U. A.	0.319	0.567	0.404	0.083	1

TABLE I.
CULTURAL DISTANCES VS CONTENDER CHARACTERISTICS

	(a)	(b)	(c)	(d)
Quality	0.911 (0.03)	0.914 (0.03)	0.901 (0.03)	0.905 (0.03)
Logrolling	0.028 (0.01)	0.022 (0.01)	0.018 (0.01)	0.016 (0.01)
Order of perf	0.003 (0.01)	0.002 (0.01)	0.004 (0.01)	0.003 (0.01)
Host Country	0.177 (0.24)	0.191 (0.24)	0.155 (0.24)	0.171 (0.24)
Sung in English	0.14 (0.14)	0.193 (0.14)	0.101 (0.14)	0.135 (0.14)
Sung in French	0.353 (0.17)	0.354 (0.17)	0.343 (0.18)	0.347 (0.18)
Male Singer	0.139 (0.13)	0.148 (0.13)	0.147 (0.13)	0.154 (0.13)
Duet	0.223 (0.20)	0.147 (0.20)	0.203 (0.20)	0.174 (0.20)
Group	0.1 (0.13)	0.08 (0.13)	0.087 (0.13)	0.079 (0.13)
Language	-	-1.142 (0.22)	-	-0.634 (0.24)

TABLE III.
PERFORMANCE RATES

Country	2008	2007
Armenia	0.87	0.64
Ukraine	0.81	0.77
Georgia	0.79	0.61
Serbia	0.78	0.55
Azerbaijan	0.77	-
Ireland	0.68	0.69
Belarus	0.66	0.61
Sweden	0.65	0.64
Turkey	0.63	0.6
Finland	0.62	0.51
Malta	0.61	0.58
Russia	0.6	0.59
Albania	0.59	0.58
Greece	0.58	0.55
Israel	0.57	0.53
Slovenia	0.56	0.54
Bosnia & Herzegovina	0.55	0.51
Hungary	0.54	0.51
Poland	0.53	0.52
Croatia	0.52	0.51
Latvia	0.51	0.49
Belgium	0.49	0.47
France	0.48	0.46
Romania	0.46	0.43
Germany	0.45	0.42
Spain	0.44	0.37
FYR Macedonia	0.43	0.42
United Kingdom	0.42	0.43
Bulgaria	0.4	0.41
Norway	0.39	0.38
The Netherlands	0.37	0.39
Iceland	0.36	0.35
Estonia	0.35	0.34
Portugal	0.34	0.37
Lithuania	0.33	0.34
Moldova	0.32	0.36
Denmark	0.31	0.33
Cyprus	0.3	0.28
Montenegro	0.29	0.21
Switzerland	0.25	0.26
Czech Republic	0.22	0.21
San Marino	0.14	-
Andorra	0.11	0.08

The parameters used by the model to calculate the performance rate are: β = 0.4 and γ = 0.6. The model used to calculate the values of Table 3 is the following:

$$r_i = 0.4\sum_{k=1}^{7} x_{ik} + 0.6\sum_{t=1}^{T_I} z_{it} \quad (3)$$

TABLE IV.
CONTENDER CHARACTERISTICS

Characteristic	Quality Factor
Language	0.30
Lyric and Topic	0.25
Musical Arrangement	0.20
Musical Genre	0.15
International Fame	0.10
Sex of Singer	-0.10
Number of Singers	-0.15

Equation (3) is a synthesis of the voting model presented in (2). The missed term $\alpha_{ij}v_{ij}$ represents the voting behavior expected between countries *i* and *j*. A robust model was developed adding probability terms that reflect the voting history between judge country *i* and contender country *j* (v_{ij}). The complete model and its implicit problem are explained in the next section.

III. PROBLEM STATEMENT

The objective of this paper is to estimate the position rank of the new contenders, Azerbaijan and San Marino. This implies to estimate the final voting matrix, where every cell *j*, *i* represents the score gives to contender *i* by country *j*; that is v_{ji}.

For attaining a good prediction, the model should controls the voting behavior between judges and contenders taking into account the historical performance that reflects the cultural empathy, the commonality of regions, the returning voting patterns, etc. The estimated performance rate could guide the model towards an optimal voting configuration according to the current expectations of the experts.

The next objective function posses these two important features of the ESC, the voting behavior and the performance rate explained in the previous Section. Notice that (3) is part of (4).

Maximize

$$f = \sum_{i=1}^{C}\sum_{j=1}^{N} c_{ij} + 4\sum_{i=1}^{C}\sum_{k=1}^{S} p_{ik} + \frac{2}{\max_S}\sum_{i=1}^{C} s_i * r_i \quad (4)$$

Subject to:

- Country *j* can not vote for itself.
- Country *j* just can vote one time for contender *i*.
- Country *j* just can give a score *k* to only one contender *i*.

Where *N* is the number of voting countries, *C* is the number of contenders, *S* is the number available scores S={12, 10, 8, 7, 6, 5, 4, 3, 2, 1}, and max_S =12 is the maximum score. The first two terms represents the voting process and the last term represents the performance of the final ranking.

In the first term of (4), c_{ij} is the probability that a score *k* was given by country *j* for a contender country *i*. Table 5 shows an example of the probabilities c_{ij} of Finland for score *s*=12. Along 52 ECS editions, Finland has received 19 times a score of 12 points from 11 different countries. Sweden and Iceland are the countries which have voted more times for Finland, both with 3 editions. Therefore, they are the countries with highest probabilities c_{ij}.

In the second term of (4), p_{ik} is the probability that country *i* receives a score *k* from country *j*. Table 6 shows an example of the probabilities p_{ik} of Finland with Germany. Along 52 ECS editions, Finland has received 16 votes from Germany. In 4 times, Germany has given a score of 1 point to Finland; thus, it is the score with highest probability p_{ik}.

For the last term of (4), s_i represents the scores sum got by a contender country *i* from every country $i \neq j$; and r_i represents the expected performance rate of the country *i* in the competition (see Table 3).

The probabilities c_{ij} and p_{ik} were calculated based on the previous Eurovision editions. The probabilities for Azerbaijan and San Marino were calculated observing the behavior of the voting along 52 ECS editions between a mature country and a new contender.

The model explained in this section solves the combinatorial problem implicit in Equation (4). The best solution is an estimation of the final voting table of the ECS 2008 (for predicting the position rank of Azerbaijan and San Marino). The constrained optimization problem has two parts. In the first part, the problem is to find the optimal combination that maximizes the sum of probabilities, the first two terms of (4). This implies 42 voting countries, subject to the mentioned constraints, which must assign 10 different scores (S) to 25 contender countries, resulting in 2.99×10^{14} possible combinations.

In the second part, the total sum of the votes obtained by every contender country is calculated. The vote sums (s_i) are used to calculate the weighted sum presented in the third term of (4). This implies again to find the optimal combination out of 2.99×10^{14} possible solutions.

The maximization of both parts of the problem generates a tradeoff between the voting behavior and the performance rate. For solving the optimization problem, we use a simple and innovative PSO-based algorithm, which is thoroughly explained in the next section.

TABLE V.
EXAMPLE OF COUNTRY VOTING PROBABILITY

Contender (*i*)	Country (*j*)	Score (*k*)	Frequence (f_i)	c_{ij} f_i/TF_k
Finland	Denmark	12	1	0.0526
Finland	Estonia	12	2	0.1053
Finland	Germany	12	2	0.1053
Finland	Greece	12	1	0.0526
Finland	Iceland	12	3	0.1579
Finland	Ireland	12	1	0.0526
Finland	Norway	12	1	0.0526
Finland	Poland	12	2	0.1053
Finland	Sweden	12	3	0.1579
Finland	Switzerland	12	1	0.0526
Finland	U. Kingdom	12	2	0.1053
			TF_k=19	1.0000

TABLE VI.
EXAMPLE OF SCORE VOTING PROBABILITY

Contender (*i*)	**Country** (*j*)	**Score** (*k*)	**Frequence** (f_i)	p_{ik} f_i/TF_k
Finland	Germany	12	2	0.1250
Finland	Germany	10	1	0.0625
Finland	Germany	7	1	0.0625
Finland	Germany	6	1	0.0625
Finland	Germany	5	3	0.1875
Finland	Germany	4	1	0.0625
Finland	Germany	3	1	0.0625
Finland	Germany	2	2	0.1250
Finland	Germany	1	4	0.2500
			TF_k=16	1.0000

IV. CONSTRAINED OPTIMIZATION VIA PSO

Particle Swarm Optimization (PSO) [10] algorithm is inspired by the motion of a bird flock. A member of the flock is called “particle”. In PSO, the source of diversity, called variation, comes from two sources. One is the difference between the particle’s current position x_t and the global best G_{Best} (best solution found by the flock), and the other is the difference between the particle’s current position x_t and its best historical value P_{Best} (best solution found by the particle). Although variation provides diversity, it can only be sustained by for a limited number of generations because convergence of the flock to the best is necessary to refine the solution. The velocity equation combines the local information of the particle with global information of the flock, in the following way.

$$v_{t+1} = wv_t + \varphi_1(P_{Best} - x_t) + \varphi_2(G_{Best} - x_t)$$
$$x_{t+1} = x_t + v_{t+1} \quad (5)$$

where vt is the velocity vector, w is the inertia factor, φ_1 and φ_2 are the acceleration coefficients. The second term is called the cognitive component, while the last term is called the social component.

A leader can be global to all the flock, or local to a flock’s neighborhood. Flock neighborhoods have a structure that defines the way information is concentrated and then distributed among its members. The most common flock organizations are shown in Fig. 1. The organization of the flock affects search capacity and convergence.

The particle swarm algorithm is simple to implement and has low computational cost. PSO has a disadvantage, its quick convergence reduces exploration. In this paper we use a modifed PSO to avoid this problem, the Constrained Optimization via PSO (COPSO) algorithm [13]. The COPSO algorithm has two important contributions: the singly-linked neighborhood structure, and the perturbation of the particle’s best.

The original ring structure is implemented by a doubly-linked list, as shown in Fig. 2-a. COPSO uses an alternative ring implementation, the singly-linked list, shown in Fig. 2-b. This structure improved the success of experimental results by a very important factor. Although more details are not provided, the advantages of the new structure can be explained as follows (see Fig. 2).

First, we describe a PSO based on the original ring structure (doubly-linked). Assume particle *k* is the best of the flock. Since *k*−*1* and *k*+*1* have particle *k* in their own neighborhood, but *k* is the best, then particles *k*−*1* and *k*+*1* are directly pulled by *k*. Simultaneously, particle *k* has particles *k*−*1* and *k*+*1* as neighbors. Therefore, *k* attracts *k*−*1* and *k*−*1* attracts *k*. After some generations, particles *k*−*1* and *k* converge to the same point due to the double link. Now, we analyses a PSO based on a singly-linked ring, and assume particle *k* is again the best of flock. But particle *k* now has particles *k*−*2* and *k*+*1* as neighbors (not *k*−*1* and *k*+*1* as in the double link). Since particle *k*+*1* has particles *k*−*1* and *k*+*2* as neighbors, and *k*−*1* has particles *k*−*3* and *k* as neighbors. Then, *k* attracts *k*−*1* but *k*−*1* only attracts *k* through particle *k*+*1*. Therefore, the particle in between cancels the mutual attraction, and in consequence reduces the convergence of the flock.

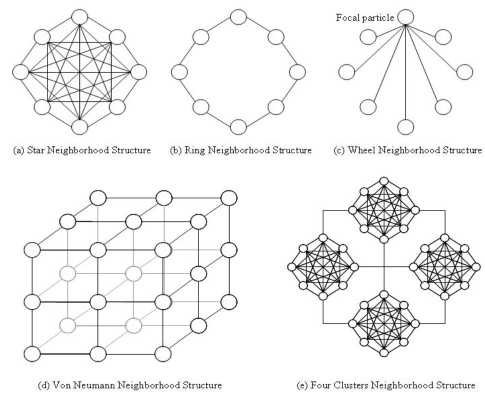

Figure 1. Neighborhood structures for PSO.

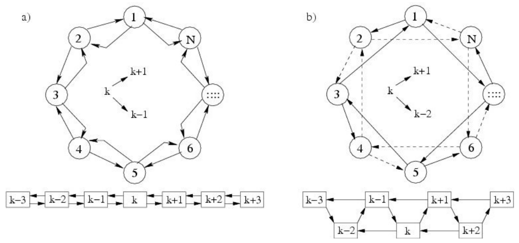

Figure 2. Neighborhood structures for PSO.

For each particle *i*, the members of a neighborhood of size *n* are selected by the next algorithm.

(a) Set *step* = 1
(b) Set *switch* = 1 (pick from left or right side)
(c) Include in the neighborhood the particle
(d) *i*+*switch***step*
(e) Increment *step* = *step* + 1
(f) Calculate *switch* = −*switch*
(g) Repeat step 3 until *neighborhood_size* = *n*.

COPSO improves the local best PSO algorithm with external perturbation operators applied to the best visited location P_{Best}. This approach keeps diversity and guide the flock towards good spots without destroying its self organization capacity. Flying the particles remains the main task of PSO. A view of COPSO algorithm with its three components is shown in Fig. 3.

In the first stage the standard local PSO based on singly-linked structure runs one iteration [11]. Then the perturbations are applied to P_{Best} in the next two stages.

```
X0 = Rand(LL, UL)
F0 = Fitness ( X0 )
PBest= X0
FBest= F0
Function COPSO
For i = 1 To maxgenerations

                    Stage 1
    LBest = LocalBest ( FBest)
    Vi+1 = Velocity ( Vi, Xi, PBest, LBest)
    Xi+1 = Xi + Vi+1
    Fi+1 = Fitness ( Xi+1 )
    ParticleBest ( PBest, Xi+1, FBest, Fi+1 )

                    Stage 2
    If(U(0, 1) < p )
        PTemp = C-Perturbation (PBest )
        FTemp = Fitness ( PTemp )
        ParticleBest ( PBest, PTemp, FBest, FTemp )
    End If

                    Stage 3
    If(U(0, 1) < p )
        PTemp = M-Perturbation (PBest )
        FTemp = Fitness ( PTemp )
        ParticleBest ( PBest, PTemp, FBest, FTemp )
    End If

End For
```

Figure 3. Pseudo-code of COPSO algorithm.

The goal of the second stage is to add a perturbation generated from the linear combination of three random vectors. This perturbation is preferred over other operators because it preserves the distribution of the population (also used for reproduction by the differential evolution algorithm [15]). In COPSO this perturbation is called "C-Perturbation". It is applied to the members of P_{Best} to yield a set of temporal particles P_{Temp}. Then each member of P_{Temp} is compared with its corresponding father and P_{Best} is updated with the child if it wins the tournament. Fig. 4 shows the pseudo-code of the C-Perturbation operator.

In the third stage every vector is perturbed again so a particle could be deviated from its current direction as responding to external, maybe more promissory, stimuli. This perturbation is implemented by adding small random numbers (from a uniform distribution) to every design variable. The perturbation, called "M-Perturbation", is applied to every member of P_{Best} to yield a set of temporal particles P_{Temp}. Then each member of P_{Temp} is compared with its corresponding father and P_{Best} is updated with the child if it wins the tournament. Fig. 5 shows the pseudo-code of the M-Perturbation operator. The perturbation is added to every dimension of the decision vector with probability p=1/d (d is the dimension of the decision variable vector).

The perturbation operators have the additional advantage of keeping the self-organization potential of the flock since they only work on the P_{Best} memory. In Fig. 3 the main algorithm of COPSO is listed. p is a linearly decreasing probability from 1.0 to 0 (according to the function evaluations), LL and UL are the lower and upper limits of the search space.

In the next sections COPSO is applied to maximize (4) for estimating the final ranking of the new contender countries.

V. EXPERIMENTS ESC 2007: MODEL VALIDATION

For assessing the performance of the proposed model,

```
For k = 0 To n
    For j = 0 To d
        r = U(0, 1)
        p1 = Random(n)
        p2 = Random(n)
        PTemp[k,j] = P[k,j] + r (P[p1,j] - P[p2,j])
    End For
End For
```

Figure 5. Pseudo-code of C-Perturbation.

```
For k = 0 To n
    For j = 0 To d
        r = U(0, 1)
        If r ≤ 1/d Then
            PTemp[k,j] = Rand(LL, UL)
        Else
            PTemp[k, j] = P[k, j]
    End For
End For
```

Figure 4. Pseudo-code of M-Perturbation.

we used it to estimate the final ranking of the countries which competed for the first time on the ESC 2007: Czech Republic, Georgia, Montenegro and Serbia.

In the ESC 2007, the mentioned 4 countries competed in the unique Semi-Final stage against other 24 countries for obtaining just 10 places to the Final stage. In the Final stage, 14 countries were waiting for competing against the first 10 places of the Semi-Final. The 14 finalists were composed by the "Big Four" (Germany, Spain, France and United Kingdom) and the first 10 places of the ESC 2006. The list of contenders for every stage is available in the web host of the ESC 2008 [1].

For estimating the voting matrix, 30 runs of the each experiment, Semi-Final and Final, were performed to obtain a better estimation of the final ranking. In every run, 350000 function evaluations were performed. For the Semi-Final stage, the result of every run is a voting matrix and a rating list from 1 to 28. The average along the 30 runs was calculated for every contender. Next, the average ranking was obtained to determine the 10 countries which are going to contend in the Final stage. Table 7 shows the results of the 4 new contenders in the Semi-Final ESC 2007.

The results placed in the Final stage only 2 out of 4 new contenders: Georgia and Serbia. At this point, the model predictions are according the reality, since in ESC 2007, Czech Republic and Montenegro did not attain to reach the Final stage.

For the Final stage, the result of every run is a voting matrix and a rating list from 1 to 24. Three measures are calculated from the 30 runs: average, median and interquartile range. The interquartile range has a comprehensiveness of 50% around the median value (second quartile Q_2), which is calculated through the lower quartile Q_1 (first quartile) and upper quartile Q_3 (third quartile). In descriptive statistics, a quartile is any of the three values which divide the sorted data set into four equal parts, so that each part represents 1/4*th* of the sampled population. The difference between the upper and lower quartiles is called the interquartile range. The results of the Final stage for Georgia and Serbia are showed in Table 8.

In ESC 2007, Georgia obtained the 12*th* position in the Final stage, which is a value into the estimated interquartile range and very close to the average value predicted along 30 runs. Nevertheless, the predictions for Serbia are far away from the reality. Serbia was the winner of the ESC 2007. Maybe, anyone of the ESC experts would have not guessed that a new contender would win the contest. The ESC 2007 results showed how as hard can be to estimate the ESC's behavior. In the next section, the estimation of our approach for ESC 2008 is presented.

VI. EXPERIMENTS ESC 2008

This year, the ESC consist of three stages: 2 Semi-Finals stages and a Final stage. 43 countries will be represented in the Eurovision Song Contest - Belgrade 2008. Five of them are automatically qualified for the Final: The "Big Four" (France, Germany, Spain and United Kingdom) and the winner of the ESC 2007, Serbia (host country). On January 28th was determined which 19 countries are represented in the First Semi-Final, and which 19 in the Second Semi-Final. The complete list of countries which will contend in the every Semi-Final is available in the web host of the ESC 2008 [1].

France, Germany, Spain, United Kingdom and Serbia will be voting in one of the two Semi-Finals. Germany and Spain will vote in the First Semi-Final, and France, United Kingdom and Serbia in the Second Semi-Final. Who will be voting in each group was determined by draw, also on January 28th [1].

The objective of this experiment is to predict the final ranking for Azerbaijan and San Marino. Azerbaijan and San Marino will compete in the First Semi-Final for winning a place in the Final stage. For this experiment 30 runs were performed with 350000 function evaluations. The top-10 of the 30 runs is presented in Table 9.

The same experiment was applied to obtain the top-10 of the Second Semi-Final. Table 9 shows the countries which complete the contenders in the Final stage.

The results of the First Semi-Final indicate that just Azerbaijan could attain a place in the Final stage. The results for Azerbaijan and San Marino contrast to each other because Azerbaijan obtained the first place and San Marino the last place, at the First Semi-Final (see Table 10).

For estimating the final ranking of Azerbaijan in the ESC 2008, 30 runs were performed with 350000 function evaluations. The average, median and interquartile range for the 30 runs were calculated. Table 11 presents the results of the Final stage.

The experiments predict a 7*th* place for Azerbaijan in the ESC 2008. Also, the results estimate a interquartile range equal to 4, that is a final ranking from 3*rd* to 11*th* in the final Eurovision 2008, which will be on May 24th, 2008.

VII. CONCLUSIONS

The prediction of future events is a hard task, also, impossible in several topics. There are several methods that have been used as an auxiliary tool for building estimation models (Data Mining, Regression models, Support Vector Machines, Neural Networks).

In this work, data mining and evolutionary computation are combined for predicting the behavior of the European society in a music contest. A huge set of models to predict the final ranking of the Eurovision Song Contest can be found in its web host [1]. These works have not focused their attention in model the behavior of the European society when a new country competes for the first time in the ESC.

Our approach propose a model that includes two main features: voting behavior and cultural characteristics. The model incorporates historical information about the vote assignation, that European society has performed along previously ESC editions. Besides, the model includes information about intrinsic characteristics of the contender that represents a country (language, lyric,

TABLE VII.
SEMI-FINAL ESC 2007

Contender	Average (30 runs)	Average Ranking
Georgia	9.87	7
Serbia	10.70	8
Czech Republic	24.33	27
Montenegro	26.13	28

TABLE IX.
FINAL ESC 2007

(30 runs)	Georgia	Serbia
Average	11.3	15.97
Median	10.5	17.5
Interquartile Range	7 - 16	12 - 22

TABLE VIII.
SEMI-FINAL TOP-10

First Semi-Final	Second Semi-Final
Azerbaijan	Ukraine
Armenia	Georgia
Bosnia & Herzegovina	Albania
Finland	Sweden
Russia	Belarus
Ireland	Malta
Israel	Turkey
Poland	Latvia
Romania	Croatia
Slovenia	Hungary

TABLE X.
FIRST SEMI-FINAL ESC 2008

Contender	Average (30 runs)	Average Ranking
Azerbaijan	5.7	1
San Marino	18.6	19

genre, in others). The prediction model implies to solve a constrained optimization problem. COPSO is a simple optimization algorithm that has been tested in several benchmark problems of the state-of-the-art in constrained optimization. The results obtained by COPSO have been competitive. The prediction performance could be judged or rated when the Eurovision Song Contest - Belgrade 2008 will develop, on May 24th, 2008. A future research is obtain votes from a Mexican society (Chihuahua) and to describe the similarity with another society in Europe, for example Gibraltar (A society mixture with English people, but influenced with Spanish Culture and simulated a similar society that likely music in English but lives in Mediterranean), and analyze the voting of this society, we simulated an artificial society using Cultural

TABLE XI.
AZERBAIJAN IN THE ESC 2008

Average	Median	Interquartile Range
7.47	**6. 5**	**3 - 11**

Algorithms to predict the behavior of a sample of 100 people in Mexico, with their information is possible understand the possible inclusion in the future of Gibraltar, Kosovo or Sahrawi Arab Democratic Republic, maybe in Eurovision 2017.

ACKNOWLEDGMENT

The second author acknowledges support from National Council of Science and Technology, CONACYT, to pursue graduate studies at the Center for Research in Mathematics.

REFERENCES

[1] http://www.eurovision.tv/page/home.
[2] Z. Dennis. Culture and Organizations. McGraw-Hill, London, 1991.
[3] T. H. et al. Towards robust indexing for ranked queries. In Proceedings of VLDB'06, Seoul, Korea, 2006.
[4] D. Fenn, O. Suleman, J. Efstathiou, and N. Johnson. Eurovision make its mind up? Connections, cliques, and compability between countries in the Eurovision song contest. Physic A: Statistical Mechanics and its Applications, arXiv:physics/0505071v1, 2005.
[5] J. R. Flôres and V. Ginsburgh. The queen Elizabeth musical competition: How fair is the final ranking? Journal of the Royal Statistical Society, Series D, (45):97–104, 1996.
[6] V. Ginsburgh and A. Noury. Cultural voting: The Eurovision song contest. http://ssrn.com/abstract=884379, 2005.
[7] V. Ginsburgh and J. Ours. Expert opinion and compensation: Evidence from a musical competition. American Economic Review, (93):289–298, 2003.
[8] H. Glejser and B. Heyndels. The ranking of finalists in the queen Elisabeth international music competition. Journal of Cultural Economics, (25):109–129, 2001.
[9] M. Haan, D. Gerhard, and D. Peter. Expect judgment versus public opinion. Evidence from the Eurovision Song Contest. 2003.
[10] J. Kennedy and R. Eberhart. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, pages 1942–1948. IEEE Press, November 1995.
[11] J. Kennedy and R. Eberhart. The Particle Swarm: Social Adaptation in Information-Processing Systems. McGraw-Hill, London, 1999.
[12] H. Moser. Twelve points grand prix Eurovision, analyse einer fankultur. Verlag Pesstalozzianum, 1999.
[13] A. Muñoz, A. Hernández, and E. Villa: Constrained optimization via particle evolutionary swarm optimization algorithm (PESO). In: Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2005, pp.209-216. Association for Computing Machinery, (2005)
[14] A. Ochoa and P. Meneguzzi. Italianitá: Discovering a Pygmalion effect on Italian communities using data mining. In Proceedings of CORE'2006.
[15] K. Price, R. Storn, and J. Lampinen. Differential Evolution: A practical approach to global optimization. Springer Verlag, Berlin, Germany, 2005.
[16] M. Rauhlen. Culture's Consequences. Beverly Hills, California: Sage, 1997.
[17] T. Suaremi. Understand social groups. In Proceedings of NDAM'2003, Reykjavik, Iceland, 2006.
[18] G. Yair. Unite Europe: The political and cultural structures of Europe as reflected in the Eurovision song contest. Social Networks, 17(2):147–161, 1995.
[19] G. Yair and D. Maman. The persistent structure of hegemony in the Eurovision song contest. Acta Sociologica, 39:309–325, 1996.

Alberto Ochoa Ortiz-Zezzatti (Bs'94–Eng.Master'00–PhD'04-Postdoctoral Researcher'06 & Industrial Postdoctoral Research'09). He joined the Juarez City University in 2008. He has 1 book, and 7 chapters in books related with AI. He has supervised 8 PhD theses, 11 M.Sc. theses and 27 undergraduate theses. He participated in the organization of COMCEV'07, COMCEV'08, HAIS'07, HAIS'08, HAIS'09, ENC'06, ENC'07, ENC'08 and MICAI'08. His research interests include evolutionary computation, natural processing language, anthropometrics characterization and Social Data Mining.

Angel E. Muñoz-Zavala received a B.Sc. degree in Industrial Engineering from Universidad Autónoma de Guadalajara (UAG) in Guadalajara, México in 2002, and a M.Sc. in Mathematics and Computation from de Center of Research Mathematics (CIMAT) in Guanajuato, México in 2004. He is currently working towards the Ph.D. degree in the Department of Computer Sciences at CIMAT. He has authored and co-authored one journal paper, over 10 conference papers and two book chapters. His current research interests are vehicle routing problem, particle swarm optimization and system reliability. He was awarded a scholarship from National Council of Science and Technology (CONACYT), to pursue graduate studies at CIMAT, in Guanajuato, México.

Arturo Hernández Aguirre received a BSc in Electronics from Universidad Autónoma Metropolitana in 1982, and an MSc and a PhD in Computer Science from Tulane University in 1998 and 2001, respectively. Hernández joined the Center for Research in Mathematic in 2001, where he works as a researcher at the Computer Science Department. He has authored over 16 journal papers, 100 conference papers and six book chapters. His research interests include evolutionary computation and bio inspired algorithms for global and constrained optimization.