A Hybrid System Using PSO and Data Mining for Determining the Ranking of a New Participant in Eurovision

Alberto Ochoa Ortíz
Instituto de Ingeniería y Tecnología, Universidad Autónoma de Ciudad Juárez
Av. del Charro 610 Norte
Ciudad Juárez, Chihuahua, México
cbr_lad7@yahoo.com.mx

Angel E. Muñoz Zavala and Arturo Hernández Aguirre
Centro de Investigación en Matemáticas
Jalisco S/N, Valenciana
Guanajuato, Gto, México
aemz@cimat.mx, artha@cimat.mx

ABSTRACT

The intention of the present work is to apply data mining and PSO to propose the solution of a specific problem about society modelling. We analyze the voting behavior and ratings of judges in a popular song contest held every year in Europe. 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 final rank of a new country in the contest.

Categories and Subject Descriptors: I.2[Artificial Intelligence]Miscellaneous

General Terms: Algorithms, Experimentation.

Keywords: PSO, Data Mining, Social Modeling.

1. INTRODUCTION

The Eurovision Song Contest (ESC) was born in 1955. Eurovision have been studied with different perspectives, for example the compatibility between countries [1] and the political and cultural structures of Europe [2], cultural voting [3] and the analysis about Grand Prix which evaluate many countries participating in different years and with different countries competing [4], among others.

This paper is novely because analyze the behavior of all countries when arrived a new country in a new ESC. The objective is estimated the final ranking of Azerbaijan and San Marino, the new contenders in Eurovision Song Contest 2008.

2. EUROVISION RANKING MODEL

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

The Table 1 presents the expected performance rates for 2008. The performance rate tries to predict the country rank through enviroment variables observed along 52 ESC editions. The performance rates were estimated based on the characteristics listed in Table 2 and the country performance along previously participations in every ESC editions.

Table 1: 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	-
⋮	⋮	⋮
San Marino	0.14	-
Andorra	0.11	0.08

Table 2: 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

Obviuosly, for the new contenders, Azerbaijan and San Marino, not historical information is 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 and Luxembourg.

The model used to calculate the values of Table 1 is the following:

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

where x_{ik}, $k = \{1, ..., K\}$ represents the characteristics (lan-

GECCO'08, July 12–16, 2008, Atlanta, Georgia, USA.
ACM 978-1-60558-130-9/08/07.

guage, lyric, music, genre and others) of a performer (singer or band) from country i, and z_{it}, $t = \{1, ..., T\}$ represents the performances of the country i along its T_i participations in the ESC.

A robust model was developed adding probability terms that reflect the voting history between a judge country i and a contender country j (v_{ij}). The complete model and its implicit problem are explained in the next section.

3. 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}.

The next function posses two important features of the ESC: the voting behavior and the performance rate explained in the previous section. Notice that Equation (1) is part of Equation (2).

Maximize

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

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 Equation (2), c_{ij} is the probability that a score k was given by country j for a contender country i. For example, along 52 ESC editions, Finland have 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 Equation (2), p_{ik} is the probability that country i receives a score k from country j. For example, along 52 ESC editions, Finland have received 16 votes from Germany. In 4 times, Germany have given a score of 1 point to Finland; thus, it is the score with highest probability p_{ik}.

For the last term of Equation (2), s_i represents the scores sum got by a contender country i from every country $j \neq i$; and r_i represents the expected performance rate of the country i in the competition.

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 ESC editions between a mature country and a new contender.

4. EXPERIMENTS ESC 2008

This year, the ESC consist of three stages: 2 Semi-Finals and a Final. 43 countries will be represented in the ESC - 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, it 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 every Semi-Final is available in the web host of the ESC 2008.

Table 3: First Semi-Final ESC 2008

Contender	Average (30 runs)	Final Position
Azerbaijan	5.17	1
San Marino	18.6	19

Table 4: Azerbaijan in the ESC 2008.

Average	Median	Interquartile Range
7.47	6.5	3 - 11

Azerbaijan and San Marino will compete in the First Semi-Final for winning a place in the Final stage. For estimating the rank a constrained optimization problem is solved using the *COPSO* algorithm [5]. For this experiment 30 runs were performed with $350,000$ function evaluations. The results for both Azerbaijan and San Marino are presented in Table 3.

The experiments predicted that only Azerbaijan will pass to the Final Stage. For estimating the final ranking of Azerbaijan in the ESC 2008, 30 runs were performed with $350,000$ function evaluations. The average, median and interquartile range for the 30 runs were calculated. Table 4 presents the results of the Final stage. The experiments predicted a $7th$ place for Azerbaijan in the ESC 2008. Also, the results estimated an interquartile range equal to 4, that is a final ranking from $3rd$ to $11th$ in the Eurovision 2008.

5. CONCLUSIONS

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 have performed along previuosly ESC editions. Besides, the model includes information about intrinsic characteristics of the contender that represents a country (lenguage, lyric, genre, etc.).

The prediction performance could be judged or rated when the Eurovision Song Contest - Belgrade 2008 will develop, on May 24th, 2008.

6. REFERENCES

[1] D. Fenn, O. Suleman, J. Efstathiou, N.F. Johnson. Eurovision Make Its Mind Up? Connections, cliques, and compability between countries in The Eurovision Song Contest. Physica A: Statistical Mechanics and its Applications, 2005.

[2] G. Yair. Unite Unite Europe: The political and cultural structures of Europe as reflected in the Eurovision Song Contest. Social Networks, Vol. 17 (2), pp. 147-161, 1995.

[3] V. Ginsburgh, A. Noury. Cultural Voting: The Eurovision Song Contest. http://ssrn.com/abstract=884379, 2005.

[4] H. Moser. Twelve points Grand Prix Eurovision, Analyse einer Fankultur. Verlag Pesstalozzianum, 1999.

[5] A.E. Munoz, A. Hernández, E.R. Villa. Constrained optimization via particle evolutionary swarm optimization algorithm (PESO). Proceedings of the 2005 Conference on Genetic and Evolutionary Computation, pp. 209-216, Washington DC, USA, 2005.