

## CRICKET TEAM SELECTION USING GENETIC ALGORITHM<sup>c</sup>

S.N. OMKAR

*Department of Aerospace Engineering,  
Indian Institute of Science, Bangalore 560 012, India  
E-mail: omkar@aero.iisc.ernet.in*

R. VERMA

*ER&D,  
Ashok Leyland Limited,  
Hosur 635126, India.*

Cricket is a team sport for two teams of eleven players each. Because of its tremendous popularity, it is really difficult to choose right players for a team of 15 or so. Moreover the human judgment cannot be relied upon. A system is formulated which takes into account various available performance data of players and gives an optimized and balanced team, with right number of batsmen, bowler, and wicket-keeper without any human interference, which is limited to entering performance data only.

### 1. Introduction

The encyclopedia defines cricket as "a bat and ball, team game played during the summer in the British Isles and in several countries influenced by the British, such as Australia, New Zealand, India, Pakistan, South Africa, and West Indian nations".

Cricket is played between two teams of 11 players on a grassy field, in the center of which are two wickets - the equivalent of baseball's 'bases'. Although the game play and rules are very different, the basic concept of cricket is similar to that of baseball. Teams bat in successive innings and attempt to score runs, while the opposing team fields and attempts to bring an end to the batting team's innings. After each team has batted an equal number of innings (either one or two, depending on conditions chosen before the game), the team with the most runs wins. This process (an innings) may be repeated once more (a match can last one day or take as many as five).

Cricket is a very popular game and has become an integral part of the

---

<sup>c</sup> INTERNATIONAL CONGRESS ON SPORTS DYNAMICS (ICSD2003), 1-3 September, Melbourne, Australia

culture. The popularity that the game can bring to a player and the financial benefits it can give attracts many aspirants to compete for a place in the team. Because of this the selection committee responsible for choosing players for a particular match, faces a tough job. There are more and more people aspiring to be in the team, and not that they are all bad players, but selection committee has to select the best 15 or so out of them. It is not uncommon to see as many as 300-400 aspirants with varied backgrounds competing for a birth in the team of 11. The competition becomes more severe as the standard of tournament rises from state to national, and from national to international level.

Various factors come into play while selecting a team. A human selection committee will invariably suffer from the shortcomings of unfair or biased judgment, human error, and overlooking of certain important points. A system is thus required which can effectively take into account all factors involved and give the optimal team, with out human interference. This system should take as input various performance characteristics like players history, his average scores if he is a batsmen, wickets taken and runs scored if he is a bowler, whether he is a wicket keeper, his performance as a fielder and so on.

The only science that comes to our rescue is artificial intelligence. In this paper, we have explored genetic algorithm for the selection of team from a list of 'probable players'. A system is generated which can consider various factors and give the optimal team. The results are verified using the case of Indian Cricket Team Selection for world cup 2003.

## **2. Genetic Algorithm**

Many optimization problems are very complex in nature and quiet hard to solve by conventional optimization techniques. Since 1960s, there has been increasing interest in imitating living beings to solve such kind of hard optimization problems. Simulating the natural evolutionary process of human beings results in stochastic optimization techniques called evolutionary algorithms, which can often out perform conventional optimization methods when applied to difficult real world problem [1,2,3,4]. There are currently three main avenues of this research: genetic algorithms (GAs), evolutionary programming (EP), and evolution strategies (ESs). Among them genetic algorithms are perhaps the most widely known type of evolutionary algorithms today.

The useful form of genetic algorithm was described by Goldberg [5]. Genetic algorithms are stochastic search techniques based on the mechanism of natural selection and natural genetics. Genetic algorithms, differing from conventional search techniques, start with an initial set of random solutions called population. Each individual in the population is called a chromosome, representing a solution to the problem in hand. A chromosome is a string of

symbols; it is usually, but not necessarily, a binary bit string. The chromosomes evolve through successive iterations, called generations. During each generation, the chromosomes are evaluated, using some measures of fitness [6]. To create the next generation, new chromosome, called offspring, are formed by either (a) merging two chromosomes from current generation using crossover operator or (b) modifying a chromosome using a mutation operator. A new generation is formed by (a) selecting, according to fitness values, some of the parents and offspring and (b) rejecting others so as to keep population size constant. Fitter chromosomes have higher probability of being selected. After several generations, the algorithm converges to best chromosome, which hopefully represents the optimum or suboptimum solution to the problem.

Usually initialization is assumed to be random. Recombination typically involves crossover and mutation to yield offspring.

1) Crossover: In the two-point crossover approach, two mating parents are selected at random; the random number generator is invoked to identify two sites on the strings, and the strings of 0's and 1's enclosed between the chosen sites are swapped between the mating strings. The crossover rate (denoted by  $P_c$ ) is defined as the ratio of the number of offspring produced in each generation to the population size (usually denoted by *pop\_size*).

2) Mutation: It is a background operator which produces spontaneous random changes in various chromosomes. A simple way to achieve mutation would be to alter one or more genes.

The process of reproduction, crossover and mutation constitute one generation of the GA. After several generations the GA is stopped and the best point among the values taken as the optimal point. Being a probabilistic search method, GA's are very good at finding global maxima. Furthermore, they need only the function values and not gradient information, which makes them easy to use for real systems where accurate gradient information is difficult to obtain.

### 3. Problem Formulation

The fitness of the player is calculated based on these parameters:

- 1) Batting: A player is given some points if he is a batsman. Based on his past performance, he is given more points.
- 2) Bowling: Players fitness as a bowler depends upon the number of wickets he has taken, and average runs he has given.
- 3) Keeping: Players fitness as a keeper depends upon the number of wickets he has taken.
- 4) Fielding: Players fitness as a fielder depends upon the number of catches he has taken.
- 5) Experience: Players fitness depends upon the experience he has playing in

the game. If the player is beyond a certain age, less points are rewarded to him.

- 6) Injuries: Injuries lead to lower performance. Some points are deducted from fitness for injuries.
- 7) Constitution: The team is checked for no less than 4 bowlers, no less than 5 batsmen and no less than 1 keeper. If any of these conditions is not met, fitness is decreased by a certain value called penalty.

The fitness of the team is the addition of fitness of each player comprising it. In addition the system can take care of past performance of the team against a particular team; its performance on a particular pitch, and the recent performance of the team.

### 3.1. Representation

First we need to encode decision variables into strings representing teams; the length of the string depends on the number of members in the team. Each string bit represents a particular player. The required number of bits ( $m_j$ ) for a team is calculated as

$m_j$  = number of players in the team

The mapping from string to the represented team is straight forward. The player corresponding to the position in the table 1 is inducted in the team. Thus the team represented as

1	2	3	4	5
---	---	---	---	---

It consists of five players, from Tendulkar to V.V.S Laxman as in table1.

Table 1. List probable Indian cricket players for world cup

No.	Name	No.	Name	No.	Name
1	Tendulkar	12	Kumble	23	Yohannan
2	Ganguly	13	Harbhajan	24	Ramesh
3	Dravid	14	Srinath	25	Bhandari
4	Sehwag	15	Agarkar	26	Mohanty
5	Laxman	16	Zaheer	27	Sarandeep
6	Yuvraj	17	Nehra	28	Pathan
7	Kaif	18	Dasgupta	29	Rakesh
8	Mongia	19	Kartik	30	Yadav
9	Das	20	Ratra	31	Salvi
10	Bangar	21	Patel		
11	Jaffer	22	Gambhir		

For the present problem, the total length of the chromosome is 15 bits which can be represented as follows

$$v_j = 1 \ 3 \ 2 \ 6 \ 11 \ 5 \ 9 \ 14 \ 7 \ 28 \ 16 \ 20 \ 23 \ 19 \ 22$$

The corresponding value of the variable  $x_1$  is team consisting of players corresponding to the numbers in  $v_j$ .

Initial Population: Initial population is created as 20 individuals of randomly selected 15 numbers between 1 and 30. These individuals are represented as  $v_j, j=1\dots 20$ . The corresponding teams are  $x_j, j = 1\dots 20$ .

### **3.2. Evaluation**

The process of evaluating the fitness of a chromosome consists of following three steps.

Step 1: convert the chromosome's genotype to its phenotype

Step 2: evaluate the objective function.. Here the objective function is the sum of performance parameter of each player and the application of constraints on each team.

Step 3: convert the value of objective function into fitness. For maximization problem, the fitness is simply equal to the value of the objective function. An evaluation function plays the role of the environment and it rates chromosomes in terms of their fitness.

### **3.3. Crossover**

The crossover used here is one cut point method, which randomly cuts one cut point and exchanges the right parts of two parents to generate an offspring. The two parents are selected randomly and the crossover probability is kept at 60% of the population.

### **3.4. Replacement**

Replacement of least fit individuals of the population by the crossover product. In this step the least fit two individuals of the population are replace by the off springs. Then the population is sorted in order of decreasing fitness. This eliminates the need for regular reproduction operation.

### **3.5. Mutation**

This step carries out the mutation operation. The mutation probability is kept very low at 5%.

### **3.6. Reevaluation of fitness**

This step involves reevaluation of fitness of population for selection of best of generation individual, which is reported as result.

## **4. Results and Discussions**

The above system was tested for two different types of situations. Tournament selection, in which, a team is to be selected from 100 or so players. The other case involved international cricket team selection of Indian cricket team for world cup 2003.

### **4.1. Tournament selection**

in this type of selection the selectors have to select 15 players. The list of probable players may be 100 or more. This type of situation is common at district, state or national level cricket tournaments, where pool of candidates is quiet large. The usual method of selection is based on selection trial. A team of observers will observe the players play on a particular day, and decide which player is fit to be chosen. This method may not be the best method to select a player because the performance of a player at a particular day may be good or bad depending on various factors. Thus a deserving player may be left out. To overcome these shortcomings, the decision is made based on the input of past performance data, along with the performance in selection trial, to the program. A set of 100 fictitious players, along with the players of national cricket team was input to the program and the program was able to identify 13 out of 15 players playing in the Indian cricket team. In different runs of program, the fictitious players were changed and the order of players was changed. The final result and the maximum fitness of the chosen team was always same, as shown in figure 1.

### **4.2. International team selection**

in this situation, the list of probable players is considerably reduced. The list of 30 probable players, announced by BCCI, was taken and fed as input to the program along with the performance data. The resulting team of 15 players was compared with the actually selected team and it was found that 13 players were similar. The two non similar players were fresh entrants. The entry of fresh entrants, in order to give encouragement to new and young players, can be taken account of by fixing a few players in the team or by giving extra fitness points to some players.

During the simulation, the fitness of the teams changed as teams became balanced and improved players were included in the teams. These changes occurred in the following stages

- 1) Random team generation. 15 players were chosen randomly and taken as team.
- 2) Teams that did not consist of required number of bowlers, batsmen, and wicket keeper became extinct, whereas a team having required number of these players had a better chance of survival.
- 3) Balanced teams emerged with required number of batsmen, bowler, and wicket keeper. It was no longer needed to get a balanced team, but the performance of players started playing a role.
- 4) Between generation 1-30 (fig 1) and 1-10 (fig 2) the teams were largely unbalanced. Between generations 30-97 (fig 1) and 10-65 (fig 2), the team was balanced but the overall fitness of team depended on individual player fitness.
- 5) From generation 100 onwards (fig 1) and 65 onwards (fig 2) there came a period of stability. All teams nearly always consisted of required number of players and had the best players.

## **5. Conclusions**

A genetic algorithm based method has been developed for selecting optimal team of cricket players. Testing has been done both by considering the selection process at league level and international level. Results are indicative of good potential of the proposed method.

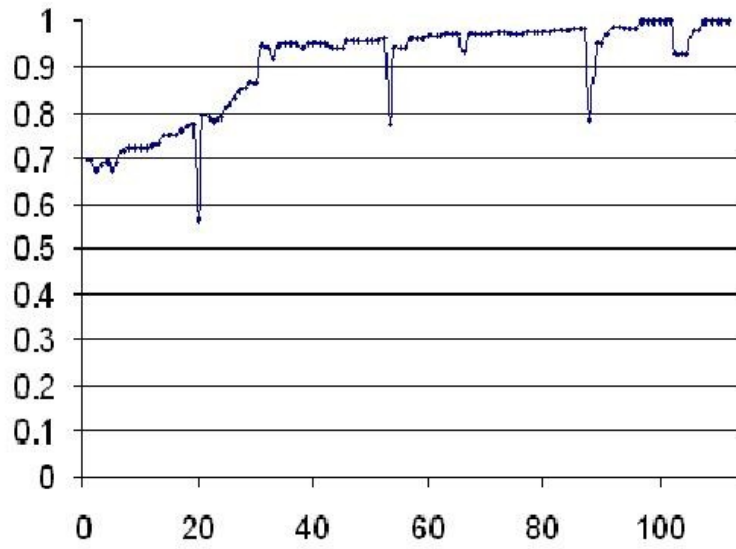


Figure 1. Number of generations versus relative fitness for tournament selection

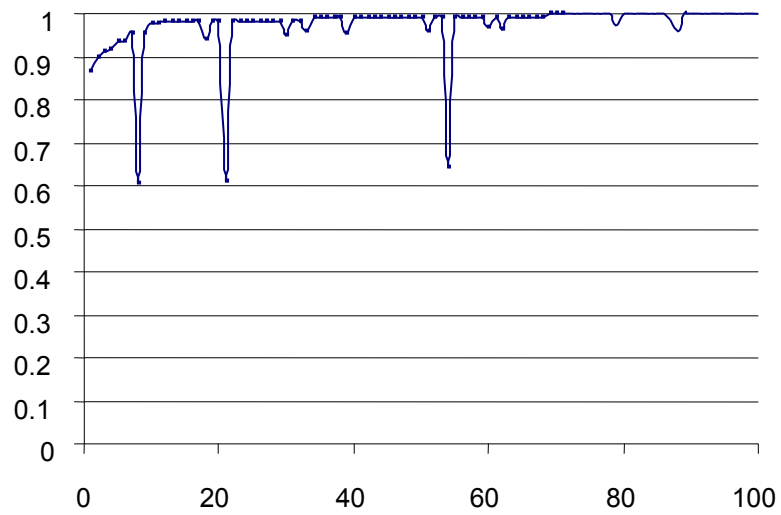


Figure 2. Number of generations versus relative fitness for international team selection



### Acknowledgments

The authors thank the Karnataka State Cricket Association.

### References

1. Back, T., *Evolutionary Algorithms in Theory and Practice*, Oxford University Press, New York, 1996.
2. Schwefel, H., *Numerical Optimization of Computer Models*, John Wiley & Sons, Chichester, 1981.
3. Back, T. and H. Schwefel, *Evolutionary computation: an overview*, pp.20-29.
4. Michalewicz, Z., *Evolutionary computation: practical issues*, pp. 30-39.
5. Goldberg, D., *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA, 1989.
6. Fogel, D. and A. Ghazali, Using fitness distributions to design more efficient evolutionary computations, pp. 11-19.