Journal of Theoretical and Applied Information Technology
31st March 2012. Vol. 37 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

GENETIC ALGORITHM BASED BOWLING COACH FOR CRICKET

HASHAM SHIRAZ ALI, UMAR NAUMAN, FARAZ AHSAN, SAJJAD MOHSIN

Comsats Institute of Information Technology, Islamabad.

E-mail: ha_persia@live.com, umarnauman@comsats.edu.pk, fahsan@comsats.edu.pk, smohsin@comsats.edu.pk

ABSTRACT

Sports are not only means of attaining physical fitness, rather nowadays have become a scientific field where statistical analysis play a vital role to overcome one's weaknesses and exploit that of opponents'. Similarly, in Cricket, bowling in the right areas has always been a headache for the bowlers and the coaches alike, not because bowlers are not accurate enough, but because the 'right areas' vary greatly according to the weather, pitch conditions, field placing, ball condition etc. Above all, the strengths and weaknesses of each batsman varies. With so many variables to consider while bowling, it is hard for a coach to figure out which area is the right one, leaving the bowler with no option other than to stick to the basics. We propose Genetic Algorithms to help the bowler in selecting the right areas to bowl to a particular batsman hence solving this rather difficult coaching problem in a straightforward manner.

Keywords: *Genetic Algorithm, Cricket, Artificial Intelligence, Bowling Strength, Batsman Weakness*

1. INTRODUCTION

Cricket is a very popular game in the Asian subcontinent. It involves a bowler, a batsman and fielders. The bowler bowls to the batsman and tries to get him out while the batsman tries to score runs off the balls bowled by the bowler. A bowler can get a batsman out in a variety of ways like hitting the stumps, getting him out leg before, getting him caught by one of the fielders etc. The ability of a bowler to get a batsman out depends a lot upon the accuracy with which he can hit the right areas on the pitch. If a ball is bowled wide outside off stump there is a high probability that it will be hit for maximum runs by the batsman, hence it's usually considered to be a poor area to bowl to a batsman. Many bowlers have different kind of skills and try to outwit the batsman by using those. Some bowlers vary their pace quite well and as a result induce the batsman in to playing a mistimed shot. Other bowlers can swing the ball in a variety of ways and bamboozle the batsman completely. Some bowlers rely on sheer pace to beat the batsman. Others can spin the ball and try to deceive the batsman[14]. However, the batsmen are also wary of all these tactics and with experience they gain the ability to hit even good deliveries for fours and sixes.

There is no hard and fast rule on how to bowl to a particular batsman because different batsmen have different kinds of strengths and weaknesses. Bowling at fuller length deliveries to a batsman who is weak when dealing with full length deliveries may be a good idea but to a recognized batsman who is strong in that area is a bad idea and may result in a lot of runs. Thus the 'right area' to bowl is different for different batsman. The bowlers and coaches are often faced with this difficult question that which area is the right one to bowl to a particular batsman. Even if they know that a batsman is weak at facing short deliveries the possibility is there that the batsman might be able to dispatch such balls out of the boundary line and will deter the bowler from bowling it in future.

To solve this problem we have focused on strengths and weaknesses of the batsmen on the basis of some initial population. We will then allow the bowler to bowl and the runs scored off each delivery are recorded. After bowling about a hundred deliveries we will get a pretty good idea in terms of which balls are getting hit for fours easily and which are troubling the batsman. Next, we will apply Genetic Algorithms, on the basis of which the bowlers are advised about the right areas to bowl to that particular batsman. Such areas that the genetic algorithm will propose will be based upon the runs that he had scored off the earlier balls. The genetic algorithm

Journal of Theoretical and Applied Information Technology
31st March 2012. Vol. 37 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

will propose the areas where the least runs were scored or where the batsman got out previously.

2. PRELIMINARIES

Genetic Algorithms, inspired by natural evolution, are used to basically solve optimization problems. It comprises of a population and the algorithm selects two individuals based on their fitness value. The higher the fitness value of an individual, the higher are the chances of it getting selected. This is exactly what happens in natural evolution where the fittest genes are selected for procreation. The Genetic Algorithm comprises of three major steps, as shown in figure-1:

- Selection
- Crossover
- Mutation

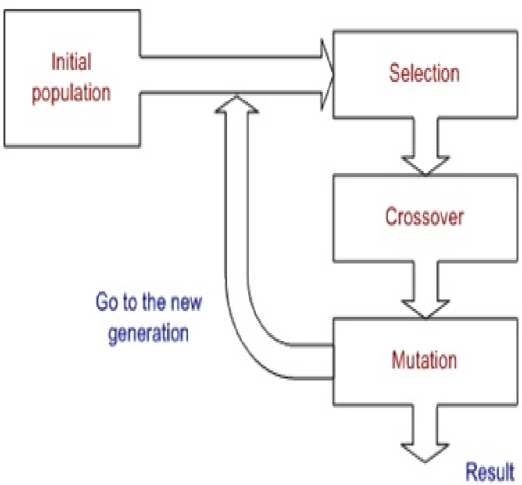

Fig 1. Genetic Algorithm Operation Cycle[16]

The larger the size of the population, the better is the results generated by the Genetic Algorithm. In the first step of selection two parents are selected based on their fitness value. This selection can be performed in many ways but two of the most popular methods are either by the roulette wheel or by using tournament selection. As shown in figure 2, in roulette wheel selection the individuals are distributed along a wheel and selection is performed randomly. The higher the fitness value of an individual the higher is its probability of getting selected. The roulette wheel is spun and an individual is selected as the parent. The roulette wheel is again spun and again an individual is selected as the parent.

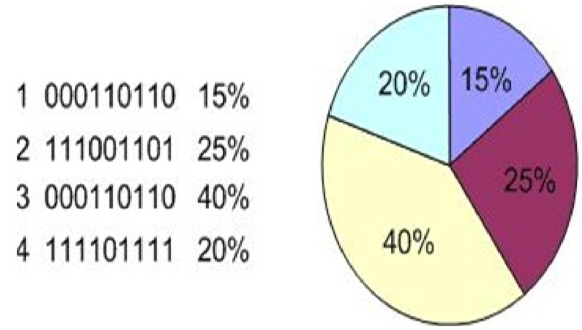

Fig 2. Roulette Wheel Selection[17]

The selection can also be performed using the method of Tournament selection. In tournament selection we simply choose two individuals at random and select the one with the higher fitness value as the parent. This process is repeated so that we have two parents. Now that we have two parents we proceed towards the next step that is crossover. In crossover we simply swap some of the attributes of the two parents to produce two completely unique offspring. The amount of attributes that are exchanged depends upon where we place the cut. We can place one or more cuts as demonstrated by the figure below.

The crossover is the most important step in Genetic Algorithms as it is the first step towards evolution. Two different parents combined to produce two different offsprings that have a better fitness value as compared to their parents. Gradually when parents are selected and crossovers are performed, the offsprings evolve to become individuals that are close to the best among the population. The Genetic Algorithm does not always produce the best or the most optimal results. However, it produces results that are close to optimal results very quickly. A simple example of single point crossover is illustrated in figure 3.

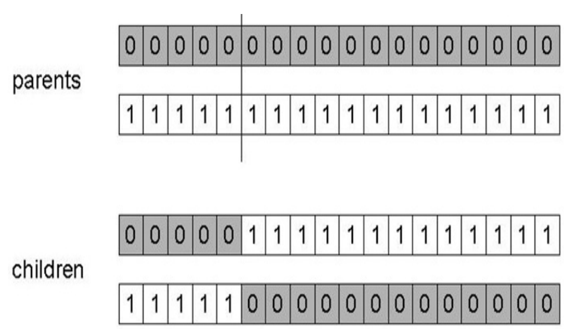

Fig.3 Single Point Crossover[8][19]

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The last step in Genetic Algorithm involves mutation. In nature we have seen how a child is born who possesses an attribute that cannot be related to the genes of either of the parents. For example a child may be born with blond hair while both the parents are dark haired. This is known as mutation. In Genetic Algorithms we perform mutation by simply changing the value of an attribute at random. The probability of the occurrence of mutation is preset. Research has indicated that the probability of mutation should be set at 3% as it is at this value that mutation gives the best results however there is no hard and fast rule to set the probability of mutation. It can be increased or decreased at will. The advantage of performing mutation is that after the Genetic Algorithm has run a certain number of times, it will become static and will stop producing unique offspring. By performing mutation we ensure that the algorithm does not become static and unique offsprings are produced each time[17][20].

3. RELATED LITERATURE

By studying the related literature it was concluded that limited research has been carried out on the game of cricket especially from the perspective of machine learning. Most of the studies revolve around other sports because, barring the Asian subcontinent, cricket is not a very popular game among the rest of the world. And the studies that have been performed are mostly focused on predicting the outcome of a cricket match rather than analyzing the game from the coaching point of view[5]. In other sports various coaching softwares are available. Wenwu Mao and Lijuan Yu[12] used artificial neural networks and genetic algorithms for table tennis coaching. They created skill and tactic diagnostic model for table tennis matches that will help the players in their decision making. Similarly in football, Michael Beetz and Bernhard Kirchlechner[13] proposed high precision microwave technology to track movement of players and ball which would enable coaches to make good tactical decisions and assess a player’s strengths and weaknesses.

The pioneering research in the game of cricket was done by G.H. Wood in 1945 in his research paper ‘Cricket scores and geometrical progression [1]. Wood basically used geometrical distribution to model the total amount of score that could be scored in a match. However his model applied to test matches only as one day internationals were non-existent back then. Bailey and Clarke[2] used statistical models to predict the outcome of a game while it was still in progress. While this was a useful study it failed to provide coaches with any help because the game was already in progress. O.B. Chedzoy[3] in his research paper ‘Issue of the effect of umpiring errors in cricket’ focused on a study of umpiring errors and how an erroneous decision by the umpire can turn the game upside down. However, this issue has been solved recently with the advent of the DRS (Decision Review System) where players can get the decision of the umpire reviewed through TV replays. Moreover, umpiring decisions do not have anything to do with coaching although it might help coaches in predicting the effect on the game in case of a wrong decision given by the umpire. R.Sparks and D. Abrahamson[4] tried to predict the award winners in a match although their study was based on the game of baseball which is similar to cricket. They applied the Naive Bayesian classification to do that but again this raises the question that what is the need to predict the award winners when the match has already started. It might be useful to betters and gamblers but certainly not useful to coaches in any way. Simon Ting and M.V. Chilukuri[21] proposed a time-frequency based pattern recognition technique which helps in decision making and is very useful for umpires. The earlier tools available in this area are Hawk eye, which is used to track the path of the ball and display it graphically, Hot Spot, which is an infrared imaging technique used to identify the location where the ball hit the batsman and Snickometer, which is used to analyze sound and video to identify fine noises such as a snick. K.M Curtis[10] proposed a cricket batting technique analyzer which used fuzzy set theory to analyze the strokes of a batsman and then gives feedback on how well the strokes have been executed, what were the shortcomings and how they can be improved. It relies on the motion of the batsman while he is playing a shot. A similar system was also proposed by M.G Kelly and K.M Curtis[9] which also analyzes batting strokes by capturing the motion of the batsman (through motion sensors) while he is playing the stroke and then comparing it to known strokes and subsequently gives feedback on how well the stroke was executed.

Journal of Theoretical and Applied Information Technology
31st March 2012. Vol. 37 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

We have seen various uses of technology in cricket and it is revealed that little work has been done from the bowler's point of view. The bowler also needs to bowl with accuracy and has to learn various arts like swing, spin and pace variation but technology has made little contribution to the bowler. Hence the need for a system that aids the bowlers and coaches in helping to identify the flaws in bowling is felt very badly. Our system which is explained in the next section helps bowlers in identifying the right areas to bowl to a batsman based solely on the criterion that in which area the batsman scores most and in which area the batsman scores least.

4. PROPOSED METHODOLOGY

The proposed methodology is to help the coaches and the bowlers alike to help them analyze the best areas to bowl to a particular batsman. To go ahead with this approach we first need to generate a large enough sample size. We do that virtually since asking the bowler to go and bowl a hundred deliveries on the pitch seems pretty pointless and will achieve nothing except tire him out. Therefore, we model different batsmen according to their strengths and weaknesses and then bowl to them virtually. Suppose if a batsman X is weak in the good length area but strong with full length deliveries the bowler will spray the ball all over the place and the batsman will be allowed to play his shots. Each ball will have a corresponding score attached to it. So if a batsman is weak in the good length area he won't be able to score much off the balls that pitch in the good length area. Similarly if he is strong in the full length area, it is likely that all the balls that pitch in the full length area will be dispatched for fours or even a six.

In the absence of historic or real data, we have currently generated random data, initially. The data set comprises of for each batsman individually and consists of hundred deliveries that he faced. Once the population of a hundred deliveries has been generated we proceed with the parent selection. We select six parents from the population of hundred using the tournament selection. The score off each delivery acts as the fitness value. In this case the lower the score the higher is the fitness of the individual therefore chances are that the majority of the 6 parents selected will have a score of not more than 2 or 3. After that we perform the crossover. In the crossover we simply swap the y-coordinate value of both the parents while the two parents that we want to mate are selected randomly out of the six selected initially. After the first iteration we have six offspring which are completely unique and with a better fitness as compared to their parents.

Subsequently we run this loop a thousand times and at the end we get a pitch map showing us the appropriate areas to bowl to the batsman where he will find it hard to score runs. Alternatively we can also set the condition that when all the balls start falling in the good length area but that would mean changing the code each time the model of a different batsman is to be tried. This can be complicated and will not go down well with the user. By running the GA 1000 times the results fall within acceptable range and are good enough to be used for real matches.

Table-1: Experimental Parameters

Parameters	Values
Initial Population	100
Selection Type	Tournament Selection
Crossover Type	Single-Point Crossover
Mutation Rate	0-3%
No. of times GA is run	1000
No. of deliveries (output)	30

5. RESULTS

The parameters listed in table-1 are used in our implementation of GA with respect to this problem. Figure 4(a) shows the pitch map of the deliveries that were initially generated (initial population). The dots represent the location where the ball pitched. The dots may not total to a hundred because a few deliveries could have pitched at the exact same location. By looking at this map you can see that the bowler has sprayed the ball all over the place and must have gone for quite a few runs. The model of the batsman in this case is one who is weak at handling good length deliveries, strong at full length deliveries and neutral with short length deliveries, based on the runs scored at each ball.

Journal of Theoretical and Applied Information Technology
31st March 2012. Vol. 37 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

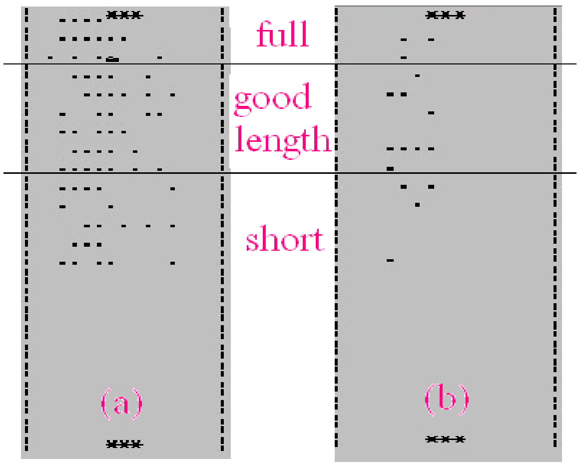

Figure 4: (a) Pitch Map of Initial Population. (b) Pitch Map after Applying GA

Now after applying the GA we get the following results as shown in figure-5. Here, we have not considered the mutation rate so that better precision is achieved. However, in reality it takes quite a practice and hard work before a bowler can pitch the ball exactly where he wants or is being told by the coach. Keeping this physical limitation in mind, error rates are still intact randomly and on average more than 70% deliveries are in the desired slot for that particular batsman.

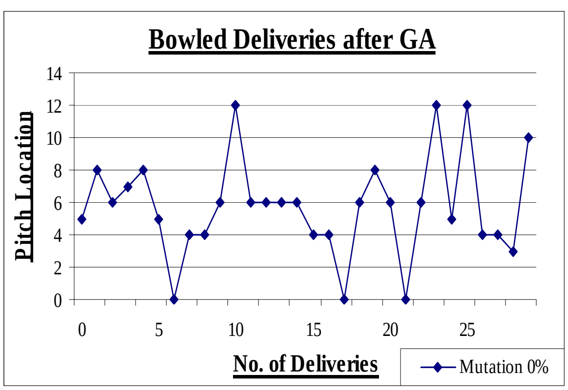

Figure 5: Deliveries pitched after GA being applied.

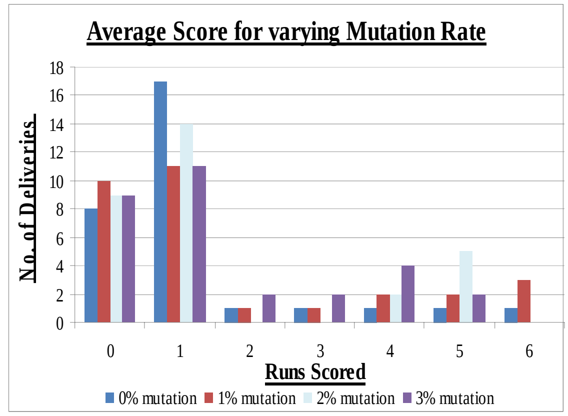

Figure 6: Runs scoring rate after applying GA

Thus, as a next step we compared for various mutation rates ranging from 0-3% as shown in figure 6. Theoretically, the more the mutation rate, the higher the variance of plotting the bowl in undesired area. Similar is the output of the comparison where No Mutation, gets better results. Besides, on average more than 2/3rd of the deliveries are either dot balls (no run) or only a single run was scored.

As it is evident now that the 30 deliveries have pitched in the good length area, more or less. This is an important guide for the coach and the bowler that the batsman is weak in this area and by bowling in this area even if you do not get the batsman out, at least being a bowler you would not be giving away a great deal of runs.

6. CONCLUSION

This work is unique in the sense that it is designed to help coaches in the game of cricket by using an algorithm of machine learning as opposed to the previous works which focused on mostly predicting the outcome of the one day games and would have mostly benefitted the gamblers. However, there are certain shortcomings in this program and future work should revolve around improving this program and bringing in more innovation. For example right now the classification of deliveries is limited to only three areas namely, good length, full length and short length. These areas can be increased to give a more accurate analysis of a batsman's strengths and weaknesses. Also along with the batsman's model other factors that affect the game of cricket like pitch conditions, pressure, weather, humidity etc can also be included. This program is just a demonstration

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

about how GA can be used to create coaching software especially for cricket coaches.

REFERENCES:

[1] G.H. Wood, "Cricket scores and geometrical progression", Journal of the Royal Statistical Society, 1945, Series A, pp.: 12–22.
[2] M. Bailey and S.R. Clarke, "Predicting the match outcome in one day international cricket matches, while the game is in progress", Journal of Sports Science and Medicine, Vol. 5, pp.: 480-487.
[3] O.B. Chedzoy, "Issue of the effect of umpiring errors in cricket Statistician", 1997, Vol. 46, No. 4, pp. 459-527.
[4] R. Sparks and D. Abrahamson, "A mathematical model to predict award winners", Math Horizons, April 2005, pp: 5-13.
[5] A. Kaluarachchi, Aparna S. Varde, "CricAI: A classification based tool to predict ODI matches, 5th International Conference on Information and Automation for Sustainability, 2010.
[6] P. Ross and D. Corne, "Applications of Genetic Algorithms", Transcomputer based parallel processing systems, 1995.
[7] D. E.Goldberg, "Genetic Algorithms in Search, Optimization and Learning", Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA ©1989. ISBN:0201157675
[8] [Online] http://mikilab.doshisha.ac.jp/dia/research/person/jiro/reports/GAparams/GAparams04.html
[9] M G Kelly, K M Curtis, "Fuzzy Recognition of Cricket Batting Strokes Based on Sequences of Body and Bat Postures", SoutheastCon, 2003. Proceedings. IEEE
[10] K M Curtis, "Cricket Batting Technique Analyzer/Trainer: A Proposed Solution to Aid West Indies Cricket", AIKED'10 Proceedings of the 9th WSEAS international conference on Artificial intelligence, knowledge engineering and data bases, 2010
[11] [Online] http://findarticles.com/p/articles/mi_qn4158/is_20000521/ai_n14315223
[12] W. Mao, L. Yu, "Skill and Tactic Diagnosis for Table Tennis Matches Based on Artifical Neural Network and Genetic Algorithm", Sixth International Conference on Natural Computation, 2010
[13] M. Beetz and B. Kirchlechner, "Computerized Real-Time Analysis of Football", Pervasive Computing, IEEE
[14] D. Terry, "The Seventeenth Century Game of Cricket: A Reconstruction of the Game", Journal of Sports History
[15] I. Ahmed, N. R. Prasad, "Quantification of Impact of a Cricket Ball", International Conference on Advanced Computing and Communications, 2006. ADCOM 2006.
[16] [Online] Sugar Signal: Optimize Signal Timings and Intersection Design within ArcGIS. http://www.citilabs.com/modelingtransport/articles/sugar-signal-optimize-signal-timings-and-intersection-design-wit.html
[17] M. Fang, "Layout Optimization for Point-to-Multi-point Wireless Optical Networks via Simulated Annealing & Genetic Algorithm", Senior Project Report, University of Bridgeport, School of Engineering, Fall 2000.
[18] Design research & sustainability: Crossover Operator, Engineering Design Centre, Newcastle University. January 2007. Available Online: http://www.edc.ncl.ac.uk/highlight/rhjanuary2007g03.php/
[19] T. D. Gwiazda, Crossover for single-objective numerical optimization problems, Genetic Algorithms Reference Vol.1, 2006. ISBN 83-923958-3-2.
[20] J. H. Holland, "Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence" ISBN 0262082136, 9780262082136
[21] S. Ting and M. V. Chilukuri, "A Novel Pattern Recognition Technique For An Intelligent Cricket Decision Making System", IEEE Instrumentation and Measurement Technology Conference, I2MTC '09. 2009.