National Conference on Engineering Innovations and Solutions (NCEIS – 2018)
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 4 | Issue 6 | ISSN : 2456-3307

Match Winner Prediction Using Tennis Players Data Analytics

Rakshith M S, S Preetham, Maria Simon, Sahana S P, Mrs. Rajini S
Final Year Students, Department of ISE, Vidyavardhaka College of Engineering, Mysuru, Karnataka, India

ABSTRACT

Tennis is one of the popular sports in the world. Many Analysts have find out who will be the Winner of the Match by Using Players Statistics and Players Past Results. This paper proposes analysis of Tennis Players and a technique to predict the Winner of the Tennis Match. The proposed technique combines Players Statistics, Players Past Results, Players Performance on Different Type of Court and The Present Year Performance based on Multi-Layer Perceptron with Back-Propagation learning algorithm.

Keywords: Tennis, Neural Network, Multi-Layer Perceptron, Back-Propagation learning algorithm.

I. INTRODUCTION

Nowadays, tennis is one of the popular sports in the world. In every year, there are four major Grand Slam Tennis Tournaments that is Australian Open, French Open, US Open and Wimbledon. The four majors have different court surfaces; Australian and US Open is played on hard court, French Open is played on clay and Wimbledon is played on grass. Each court surface has its own characteristics and makes difference in speed and bounce of the ball. Clay court has a slower paced ball and a fairly true bounce with more spin. Hard court has a faster paced ball and very true bounce. Grass court has a faster paced ball and more erratic bounce. The scoring system for both men's and women's matches are same, that is the player who wins two-sets is the winner of the match. The scoring system for men's Grand Slam Tournaments is different, that is the player who wins three-sets is the winner of the match.

Due to the growth of sport competition, analysis and prediction are widely used in many kinds of sports, especially tennis. The tennis prediction model is created to evaluate the chance of winning. Most people believe that the first serve person in the set has more advantage than another because most of the games often go like that so the first serve affect to the games score. Similarly, lots of players always make fault in the first serve and do better in the second serve so second serve might affect to the games score too. Nevertheless, the first serve and the second serve affect to the games score but there is another criteria that might be refuting an advantage of serves, it is the returns of serve. Moreover, the surface characteristics also affect to the players, e.g., some players perform better on grass but they may get worse on clay.

The first tennis model was proposed by Kemeny and Snell which has only one parameter; probability of each player winning a point. Furthermore, Barnett and Clarke proposed the prediction of a match played at the Australian Open by using Markov chain model set up in Microsoft Excel which has the probability of player A winning a point if player A is serving and the probability of player B winning a point if player B is serving as inputs.

Many papers calculate winning percentage of players considering serving and receiving for statistics. Some of the problems associated with statistics are the data is not updated, the data contains many features and environmental data that is the surface of the court.

The purpose of this paper is to combine player statistics, past results, environmental data and present year performance based on Multi-Layer Perceptron. MLP is the basic of Artificial Neural Networks (ANN). ANNs are powerful technique to solve real world classification problems and have the learning ability from experience to improve their performance. ANNs are particularly effective for predicting outcome when the networks have large database of examples to draw on and able to deal with incomplete information or noisy data.

II. PROPOSED METHOD

A. Analysis

In our paper the analysis of tennis players is done based on winning percentage, first serve winning percentage, second serve winning percentage, break points winning percentage, return of serve winning percentage and the injuries of a players in different type of court. Here Injuries play important role because if a player had major injury he or she cannot play tennis and takes many days to recover.

B. Multi-Layer Perceptron (MLP)

An Artificial Neural Network (ANN) is a mathematical model or computational model based on biological neural network. The network consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. ANN is an adaptive system for which its structure can be changed using external and internal information flowing through the network during the learning phase. For the learning models, there are three major types of learning: supervised learning, unsupervised learning, and reinforcement learning.

The Multi-Layer Perceptron (MLP) is a supervised learning neural network with the input layer, hidden layer, and output layer. One input fed to one node of the network on the input layer corresponds to one input feature. In the case, N neurons are used to represent the N features of the input vector. The input layer gives out the corresponding input vector to each neuron in the hidden layer.

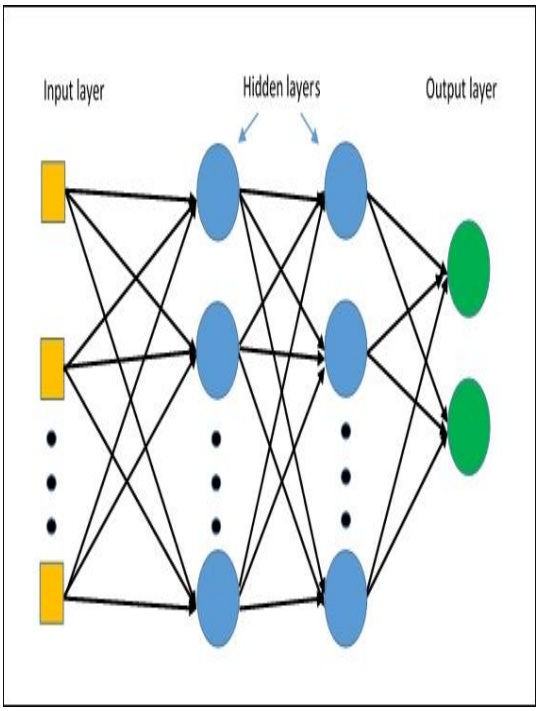

Figure 1. MLP

Some issues for training the MLP network; defining how many number of the hidden layers used in the network, deciding how many number of neurons to use in each hidden layer, finding a technique to avoid local minima, converging to an optimal solution in a reasonable period of time, and validating the neural network to avoid over fitting problem.

a. Input Features

The Input features are

1. Winning percentage on the first serve, this feature represents a chance of the player to get point on the first serve.
2. Winning percentage on the second serve, this feature represents a chance of the player to get point on the second serve.

3. Winning percentage on break point, this feature represents a chance of the player to get point when he faces the break point game.
4. Winning percentage on return of serve, this feature represents a chance of the player to gets point on receiving from opponent's serve.
5. Winning percentage of played match, this feature represents a chance of the player to win the overall matches played.
6. Type of Court: Hard Court, this feature represents the match that play on hard court. Clay Court, this feature represents the match that play on clay court. Grass Court, this feature represents the match that play on grass court.

In Tennis Match all the above inputs are considered for two players because it is played between two players.

b. Workflow

All the required data is fetched from the Database, it includes data of two players, some data are used as training set and some as test set, for testing real time data can be used. Then training and test data are together fed to MLP to get the results.

III. EXPERIMENTS AND RESULTS

A. Data Managing

The proposed method gets the collected data of the tournaments and the data is used as the input of MLP. The Input Features are statistical data, by using these player statistics and present year performance can be found. The Head to Head Match results of two players can be calculated for the past results. The environmental data that is performance based on type of court can be calculated.

B. MLP Modeling

The model consists of results based on player statistics and past results, results based on type of court and present year performance. The above are input nodes which contain input features mentioned in 2.3 if required.

To find the suitable MLP model, the learning parameters are adjusted until the error is reduced into acceptable value.

C. Training Data

The statistical data and environmental data of match played is obtained from ATP repository and www.atpworldtour.com (ATP World Tour).All the tournaments data is considered.

D. Results

By using the above model the accuracy of predictions is around 70%.

IV. CONCLUSION

In this paper an approach to create the tennis prediction model is shown. MLP is used to predict the winner of the Tennis Matches. In this model we concentrate on some important input features and consider the present year performance of the players. Coaches and Players can use this to devise strategies, Fact-based decision making, Building better teams, Improving players performance and Spotting top performers and Retaining and Engaging fans.

V. REFERENCES

1. Amornchai Somboonphokkaphan, Suphakant Phimoltares and Chidchanok Lursinsap, "Tennis Winner Prediction based on Time-Series History with Neural Modeling". Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong.
2. Designing for Insight: A Case Study from Tennis Player Analysis Kim Albrecht and Burcu Yucesoy Center for Complex Network Research. Published by the IEEE Computer Society, IEEE Computer Graphics and Applications, July/August 2016.
3. JG.Kemeny and J.L.Snell, "Finite Markov chains," Princeton, New Jersey D. Van Nostrand, 1960.

4. Barnett T. and Clarke S.R., "Combining player statistics to predict outcomes of tennis matches," IMA Journal of Management Mathematics, 16(2), 113-120, 2005.
5. Barnett T. and Clarke S.R., "Using Microsoft Excel to model a tennis match," In Proceedings of the 6M&CS, G. Cohen and T. Langtry eds., 63-68, 2002.