

## K Nearest Neighbor Algorithm for Finding Soccer Talent

Mohammad Bazmara<sup>1</sup>, Shahram Jafari<sup>2</sup>

School of Electrical and Computer Engineering, Shiraz University

---

### ABSTRACT

In this paper, the nearest neighbor algorithm is used to evaluate soccer talents for suitable positions. In other words, the nearest neighbor algorithm is used as a suitable tool to investigate which position is suitable for a player considering his skills and characteristics. The combination of the information provided by some papers was used to specify the characteristics of each soccer player. Soccer experts have found out that each soccer player must own some characteristics to succeed in his position. The proposed method has some advantages. Firstly, the best player for a special position can be chosen among many candidates. Secondly, this pattern overcomes the selection done by soccer coaches based on their sense and experience, and the need to professional experts and coaches decreases too. Furthermore, this method is so simple to understand, implement and use. The selection process using the proposed method is done using real data and the results show the efficiency and reliability of this method.

**KEYWORDS:** soccer players, nearest neighbor algorithm, indexes, player position, nearest neighbor classifier.

---

### 1. INTRODUCTION

Most parents are interested to see their children successful in a sport major, for instance, as a national or international champion. This interest is mostly towards soccer, the world's most popular sport. The process of discovering to attend organized practice programs is one of the most important things which have appeared in soccer for years. Everyone can learn to sing, draw, and etc. a few people, however, reach to high levels of skill and capability. So, discovering talented people and assigning them to special positions is very important [1].

In this paper, first of all, players' position in soccer will be discussed and then the scientific and experimental indexes and criteria of each player's special position in soccer are evaluated and their application and effectiveness in practice will be surveyed. At last, it's relation to the nearest neighbor classifier algorithm and the use of this algorithm in finding the suitable position for each player is offered. Talent finding programs have long existed in many countries and have been followed with different methods and models. The most sophisticated applications of sport talent finding have come into existence in the Eastern Bloc countries, particularly in the former East Germany and the Soviet Union [8, 9]. The method relies on the ability of fully-experienced experts. As we know, artificial intelligent (AI) methods have not been widely used for finding soccer talents. Vladan Papić et al [10] in his work entitled as "Identification of sport talents using a web-oriented expert system with a fuzzy module" attempted to predict and plan the most suitable sports for the person who has taken the test. It is important to note that only two parameters were used in this study. Peter F. Lamb et al [11] used artificial neural networks to analyze the sport of golf. Mezyk and Unold [12] used machine learning models to learn about the sport of swimming exercise. In this method combining fuzzy immune algorithm was used. In general, there is no scientific and applicable equation for recognition of each player's position in soccer. This procedure is done by coaches using their experiences and observations of players [1].

In our proposed method, nearest neighbor classifier is used to produce a pattern investigating each player's quality and goodness for all possible soccer positions. Against the current methods, we have specifically focused on soccer which has not been researched upon so far. The process initializes gathering information about soccer players, and then a sport expert assigns values to every index and criterion. An expert is a person whose knowledge in a special field is gained gradually and by learning and experiencing. The aim in this paper is to design the nearest neighbor algorithm (classifier) and to determine the distance on the basis of expert knowledge in order to eliminate wrong recognitions based on experience in selecting the proper position and increasing the efficiency in this process. As you know, this subject has never been surveyed in literature.

The scientific contribution of this paper are: 1) The need for experts in finding soccer talent will not be necessary 2) enormous costs for clubs and schools will be avoided 3) the decrease in different decisions by different experts for talent finding 4) talented players with high accuracy will be found.

The remainder of the paper was organized as follows: In section II we want to explain the application of the nearest neighbor classifier in finding soccer talents, in section III, we discuss the research method and represent the important features gathered from different works related to our study for finding soccer talents, in section VI, we implement and evaluate the results. Finally, in section VII, the result is analyzed briefly.

---

\*Corresponding Author: Mohammad Bazmara, School of Electrical and Computer Engineering, Shiraz University  
E-mail: Mohamad.bazmara@gmail.com

### I. The application of nearest neighbor classifier in finding soccer talents

Soccer specialists believe that each soccer player involves particular characteristics for discovering the soccer talents, which discovery of these soccer talents only relies on the personal experiences and sensation by which the real behavioral phenomenon shows the differences in implementing the skills in soccer activities such as playing soccer, so that advancement in playing soccer involves diverse speed. The player talent while used in the personal differences means the capacity of the person to play soccer and use his skills and abilities to play it as well. The player talent success level in performing the movement skills relies on their ability in using the skills. Through this the researchers believe that movement ability is the infrastructure for other abilities in the soccer match, and the individuals in different levels involve the movement ability, so that these abilities are interrelated with each other among the individuals. Along this, one of the main factors in using the movement abilities in the individuals in soccer game is the forecasting the advancement in their skills, and the relation between implementation of the skill in the early moment of learning is realized within the implementation in upcoming stages. Through this, there are points about the abilities of individuals – for instance this matter could be mentioned that children before going to the 8 years old, have not to be participated in the competitive soccer which the adults have a role in these games. This is due to the point that if the children do not raise up and get ready for competitive soccer, in this case mental pressure would be realized for them, so that the failure in the competitions would put the worst effect on their mind, this may lead to the leaving the soccer forever. Surely, parents are interested in seeing the soccer success in their children, so that they are also interested to see their children as the future champion. For this, the discovery process of these abilities for participating in the organized plans are the very important factors in soccer domain, so that no practical approach for identifying the success in the soccer players in soccer field has been observed. Finally, it has to be mentioned that soccer specialists could discover the talent and skills of the individuals, which some of the talents of the individuals may be latent. In fact, the coaches in the soccer could use their personal opinion and outlooks in order to choose the best position, for this not approach or formula has been recommended to choose the sufficient player.

People usually refer to similar pre-tested solutions when faced with new problems. Nearest neighbor classifier is a classifying technique which uses an edition of this method. In this method, the fact that, in which class a new case is to get included, is decided using  $k$  similar cases or neighbors. The number of cases for each class is counted and the new case is included in the classes owning more neighbors [2].

For instance, there are three classes of data:  $\omega_1$ ,  $\omega_2$ ,  $\omega_3$  in figure 1.

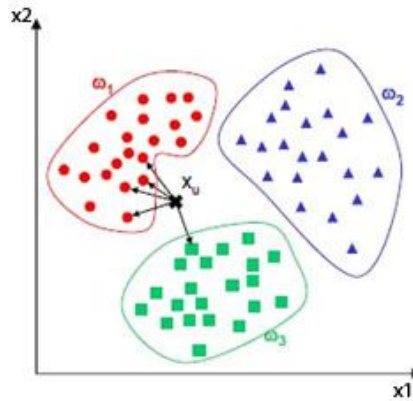


Fig. 1: three classes of data  $\omega_1$ ,  $\omega_2$ ,  $\omega_3$

Suppose each data to be the representative of each player and the squares bunch ( $\omega_3$ ) represents defenders, circles bunch ( $\omega_1$ ) represents attackers, and the triangles bunch ( $\omega_2$ ) represent halfbacks. If  $k$  is chosen to be 5, and the talents and suitable position of player  $X$  is to be found, we consider the 5 nearest player to this player. As it can be seen in figure 1, 4 near data belong to  $\omega_1$  (the attackers) and just one data belongs to  $\omega_3$  (the defenders). So we conclude that the player is better to be used as attacker. To better understand, the flowchart of  $K$  nearest neighbor algorithm is shown in Figure 2.

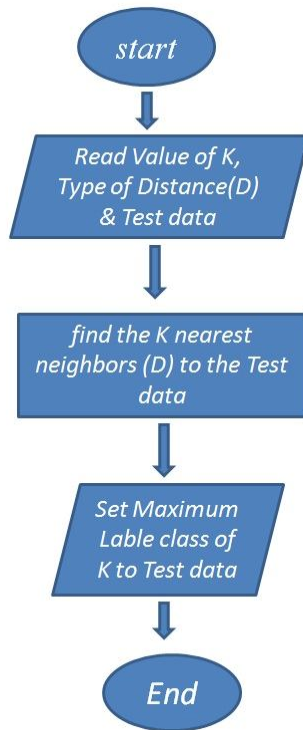


Fig. 2: The flowchart of K nearest neighbor classifier procedure

After defining k, the type of distance should be specified. In this research different types of distances are used and the results are reported. The most known type of these distances is Euclidian distance.

## II. RESEARCH METHOD

Firstly, different positions in soccer should be known so that the players can be classified on the basis of them. In every soccer team, the positions or the place of players in field are as follows [3]:

- a) Goalkeeper
- b) Sweeper
- c) Wing backs
- d) Right back and Left back
- e) Defensive midfielder
- f) Wingers
- g) Wide midfielder
- h) Center midfielder
- i) Attacking midfielder
- j) Secondary striker
- k) Forward

On the other hand, the attributes of each soccer player are as follows [4, 5, 7]:

- A) Physical qualities:
  - i) Speed
  - ii) Agility
  - iii) Jumping
  - iv) Acceleration
  - v) Height
  - vi) physical Strength
  - vii) Vigor
- B) Mental qualities[4,6]:
  - i) the ability to read game
  - ii) Calmness
  - iii) Creativity
  - iv) boldness

- v) Confidence and decision making
- vi) Leadership
- C) Technical qualities[4]:
  - i) Finishing
  - ii) Passing
  - iii) Shooting
  - iv) heading ability
  - v) tackling
  - vi) carrying the ball
  - vii) Dribbling
  - viii) single play
  - ix) crossing
  - x) backing
  - xi) Attacking play

It is worth mentioning that goalkeeper is a special position which differs from other positions in some characteristics like "Overhead exit" and "person to person battles". So we ignore this position as a separate position.

In order to select talented players for positions, the above characteristics must be taken into consideration.

For positioning players the following properties should be considered:

Properties of a sweeper:

the ability to read game, calmness, leadership, agility, speed[3,5,6].

Properties of Wing backs:

Backing, boldness, height, physical strength, heading ability [4,7].

Properties of Right and left backs:

Speed, acceleration, tackling, Crossing, agility [4,5].

Properties of Defensive midfielders:

Attacking play, physical strength, shooting, backing, the ability to read game, vigor[5,6]

Properties of center midfielders:

Vigor, passing, shooting, creativity, confidence and decision making, leadership[3,4]

Properties of wingers:

Vigor, crossing, passing, tackling, speed, carrying the ball[4,7].

Properties of attacking midfielders:

Creative, carrying the ball, passing, calmness, dribbling, shooting[4,6,7]

Properties of wide midfielders:

Crossing, carrying the ball, passing, dribbling, speed, acceleration, agility, vigor[4,5]

Properties of secondary strikers:

height, the ability of single play, passing, agility, physical strength, heading ability[3,4]

Properties of forward:

Finishing, shooting, heading ability, speed, calmness, dribbling, Creativity[3,4,6]

### III. IMPLEMENTATION AND EVALUATION OF RESULTS

The nearest neighbor algorithm for finding soccer talents was performed as an independent test in 3 soccer institutes and analyzed by experts and university professors. An experimental group composed of 150 soccer player aged between 18 and 31 was selected and each player played in one of 10 positions decided by the coach.

Now, the nearest neighbor classifier should be implemented. In order to do this, two factors should be considered, the value of k and the type of distance. Since there is no equation to specify these two factors for nearest neighbor classifier, we considered different situations to find out which type of distance with what k value works better.

K was selected as 1 to 10 and 4 type of distance studied for each value.

- i) Euclidian distance
- ii) Cityblock distance
- iii) Cosine distance
- iv) Correlation distance

That is defined as follows:

- **Euclidean distance:** The *Euclidean distance* between two points, a and b, with k dimensions is calculated as:

$$\sqrt{\sum_{j=1}^k (a_j - b_j)^2} \quad (1)$$

- **Cityblock distance:** The *City block distance* between two points,  $a$  and  $b$ , with  $k$  dimensions is calculated as:

$$\sum_{j=1}^k |a_j - b_j| \tag{2}$$

- **Cosine distance:** The *Cosine distance* between two points,  $a$  and  $b$ , with  $k$  dimensions is calculated as:

$$\frac{\sum_{j=1}^k a_j \times b_j}{norm(a) \times norm(b)} \tag{3}$$

Where

$$norm(a) = \sqrt{\sum_{j=1}^k a_j^2} \quad \& \quad norm(b) = \sqrt{\sum_{j=1}^k b_j^2}$$

- **Correlation distance:** The *Correlation distance* between two points,  $a$  and  $b$ , with  $k$  dimensions is calculated as:

$$\frac{cov(a,b)}{std(a) \times std(b)} \tag{4}$$

where

$$cov(a,b) = \frac{1}{k} \sum_{j=1}^k (a_j - \bar{a}) \times (b_j - \bar{b})$$

$$\bar{a} = \frac{1}{k} \sum_{j=1}^k a_j \quad \& \quad \bar{b} = \frac{1}{k} \sum_{j=1}^k b_j$$

In order to determine the algorithm accuracy, the data in data set was selected as test data one at a time, and classified considering  $k$  and distance. If the algorithm assigns a player to his real position, the process of finding talent has worked, and vice versa. The results are reported in table 1. For better understanding of the accuracy of data, the table is presented visually in figure 3.

Table 1: the nearest neighbor classifier accuracy

Distance K	K=1	K=3	K=5	K=10	AVG
Euclidean	82.00	87.50	87.50	87.50	86.13
Cityblock	72.66	72.66	87.50	87.50	80.08
Cosine	96.00	98.00	96.66	97.33	96.99
Correlation	93.00	97.00	97.33	96.00	95.83
AVG	85.91	88.79	92.25	92.08	

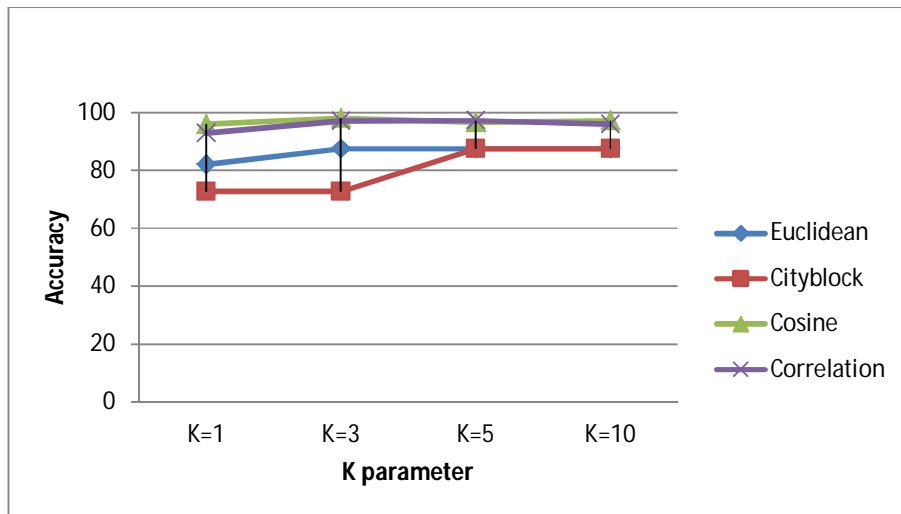


Fig. 3: the nearest neighbor classifier accuracy according to different values of K parameter and distance type.

It can be concluded from the above results that:

- K is better to be chosen as 5 or 10.
- It is better to use Cosine and correlation type distances for this classifier

At last, 42 experts were asked to describe the results of nearest neighbor classifier, which led us to following 5 possible descriptions:

– Completely Agree – Mainly Agree – Uncertain – Mainly Disagree – Strongly Disagree.

Table 2: An expert description of Nearest Neighbor Algorithm

comment	Number
Completely Agree	32
Mainly Agree	2
Uncertain	6
Mainly Disagree	2
Strongly Disagree	0

#### IV. CONCLUSION

Soccer coaches act according to their experiences and personal attitudes for selecting the position of players, Also there is no special formula for such a selection and different coaches may decide different player position. In our proposed method, we used K nearest neighbor classifier to choose the proper position for the soccer player. The accuracy achieved by the experimental result was 97.33%. This was achieved by selection of the best K parameter and suitable type of distance. 42 experts who were asked to describe the results of nearest neighbor classifier verified the exactness of the K nearest neighbor result.

#### REFERENCES

- 1- Ali Nejad, Hamid, 2005. Finding Aptitudes in Sport, Physical Education and Sport Science Research Institute.
- 2- David Hand, Heikki Mannila , Padhraic Smyth, 2001. Principles of Data Mining. The MIT Press .
- 3- Huntsville Soccer Club, 2011, SOCCER POSITIONS, Box 5187, www.huntsvillesoccer.ca/pdf/HSC%20-%20Soccer%20Positions.pdf.
- 4- Duraskovic Ratomir, Joksimovic Aleksandar, Joksimovic Stanimir, 2004. The Most important Attributes of soccer players, Physical Education and Sport Vol. 2, No 1, , pp. 13 – 24.
- 5- Ostojic, Sergej , Physical & physiological characteristics of elite Serbian soccer Players , UDC 796.332:1 59.9(497..11) , Institute of sports Medicine.
- 6- Raver R, Gettman L, Pollock M, cooper k , 1976. A physiological & mental evaluation of professional soccer players, British journal of Sports Medicine ,Br J Sports Med;10:209-216 doi:10.1136/bjism.10.4.209.
- 7- McIntyre, Hall M, Physiological profile in relation to playing position of elite college Gaelic footballers , Institute of Technology Tralee, Clash, Tralee, County Kerry, Republic of Ireland.
- 8- William A.M, Perceptual skill in soccer implication for talent identification and development, journal of sport science, 2000, 18, 737-75.
- 9- Reilly. T., anthropometric and physiological predisposition for elite soccer. Journal of sport science, 2000, swp 18: 669-683.
- 10- Vladan Papic, Nenad Rogulj et al., 2008, Identification of sport talents using a web-oriented expert system with a fuzzy modul, Expert Systems with Applications.
- 11- Peter F. Lamb, Roger M. Bartlett, Anthony Robins, 2011, Artificial neural networks for analyzing inter-limb coordination: The golf chip shot, Human Movement Science.
- 12- Edward Mezyk, Olgierd Unold, 2011, Machine learning approach to model sport training, Computers in Human Behavior.