



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

Extracting the roles of different players in soccer using an automatic clustering algorithm

Iman Behravan^{1,*}, Seyed Hamid Zahiri², Seyed Mohammad Razavi³, Roberto Trasarti⁴

¹Department of Electrical engineering, University of Birjand, Birjand, Iran

²Department of Electrical engineering, University of Birjand, Birjand, Iran

³Department of Electrical engineering, University of Birjand, Birjand, Iran

⁴KDD Lab, ISTI-CNR, Pisa, Italy

*Corresponding Author: (Phone: + 98 915 163 2809; Email: i.behravan@birjand.ac.ir)

Abstract

Nowadays analyzing team sports using artificial intelligence algorithms has become one of the most interesting topics for data scientists. On the other hand, it is really important for the coaches and managers to analyze their own and opponents' performance. So team sport companies support data scientists to get good results based on the valuable knowledge extracted by the data scientists from massive and unstructured datasets. Clustering, which is the process of grouping data points in a dataset into several clusters based on their similarities, is one of the most important data mining and big data mining tools which has been widely used in different fields. In this research a new automatic big data clustering algorithm, based on a swarm intelligence method, is used to cluster a big dataset containing centers of players' performance in different soccer matches to extract the role of different players in a soccer match. The proposed method consists of two phases. In the first phase the algorithm tries to find proper number of clusters and in the second phase it tries to find the position of the centroids. The dataset which is used in this research contains the center of several players' performance in about 93000 soccer matches which means that the dataset has 93000 objects.

Key words: artificial intelligence , Big data, Big data clustering, football analysis, swarm intelligence.

1. INTRODUCTION

Nowadays analyzing the performance of the players in the team sports such as soccer, basketball and volleyball has become very important for the managers and the coaches to make critical decisions and also to design suitable tactics for different matches. Recently artificial intelligence algorithms have been used in this field to extract valuable knowledge about the performance of the players and also the performance of the team. So different clubs and sport companies invest their resources to



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

analyze their own team performance as well as the performance of their opponents in a tournament. Analyzing the performance of the soccer players and also soccer teams using artificial intelligence algorithms and data mining techniques has become one of the most interesting topics in these years among data scientists. Data scientists try to discover valuable knowledge from big datasets generated by advanced technologies such as soccer logs [1,2] to analyze the performance of the players as well as the performance of the teams. Also these huge data sets can be used to solve the problem of rating the players in different leagues or international tournaments. Clustering, which is the process of grouping samples in a dataset into different groups (or clusters) based on their similarities, is used, as an important data mining method, in this paper to extract different roles of players in soccer based on their performance in different parts of the soccer field. Up to now different clustering methods have been introduced such as *K-means* [3] and *fuzzy C-means* [4] but in this paper a new automatic big data clustering method based on a swarm intelligence algorithm called particle swarm optimization (PSO) [5], is introduced and used to cluster a dataset containing 93000 objects which are the center of players' performance in 93000 matches in 5 prominent European leagues.

2. RELATED WORKS

In [6] Dutch et al. introduced a metric called flow centrality based on the passes result in shot, to evaluate the performance of the players. They used this metric to rank the players in Euro cup 2008 and reported that 8 players in the list of top 20 players selected by UEFA were also selected by their method. In another research [7] Brooks et al. introduced another metric called PSV (Pass Shots Value) to rank the players playing in La Liga. As well as flow centrality, PSV is based on the important passes result in shots. So these two metrics are not suitable to evaluate the performance of the defenders. Because they try to stop opponent's attacks rather than sending important passes. In a more comprehensive research, Pappalardo et al. introduced a new framework for ranking players [8]. In this method, which is called PlayRank, they not only consider important passes, but also other events like tackles, crosses, shots, fouls and etc are considered which make their method more accurate and reliable. In fact, they rank the players based on their roles. For this purpose, in one step they extracted different roles of players using *K-means*. They used *K-means* several times with different number of clusters to cluster a dataset containing the center of players' performance in several matches. Each time they calculated the quality of clustering using *Silhouette* index.. Also in another kind of research Stanojevic and Gyarmati tried to extract the relationship between the performance of the players and their market value [9].

3. PARTICLE SWARM OPTIMIZATION ALGORITHM

Particle swarm optimization algorithm first suggested by Kennedy and Eberhart in 1995. This algorithm searches the solution space of an optimization problem to find the optimum solution, using a population of individuals. Each individual, called particle, is a potential solution of the optimization problem. These particles search the solution space by simulating the behaviour of different species in group such as birds' flocking. Each particle has a vector which shows its position in the solution



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

space and also a velocity vector which shows the direction of its movement in the next iteration. The following equations show how these individuals move in the solution space and search different areas to find the optimum point.

$$v_{id}^{t+1} = w \cdot v_{id}^t + c_1 \cdot rand \cdot (p_{best}^d - x_{id}^t) + c_2 \cdot rand \cdot (p_{gbest}^d - x_{id}^t) \quad (1)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^t \quad (2)$$

In these equations, v_{id} is the d th dimension of the velocity of the i th particle, x denotes the position of the particle, t is the number of iteration, c_1 and c_2 are learning factors, $rand$ is a positive random number between 0 and 1 under normal distribution, w is the inertia weight coefficient, p_{best} is the best position of the particle from the beginning to current iteration and p_{gbest} shows the position of the best particle in each iteration which has the best fitness amount in the population. According to these equations c_1 and c_2 are two controlling factors which determine the direction of the movement. If $c_1 > c_2$ then the particles tend to move toward p_{best} and search the area around their best position observed from the first iteration to the current iteration. If $c_1 < c_2$ then they tend to go near the position of the best particle in the population. So c_1 and c_2 are two controlling parameters that their amount should be determined optimally. In this paper c_1 and c_2 are exponentially changed during the search process. In the beginning iterations c_1 has a high amount while c_2 is low and in the last iterations vice versa. Using this method the algorithm can escape from local optima.

4. PROPOSED METHOD

4.1 NON-AUTOMATIC CLUSTERING USING PSO

For solving any optimization problem using swarm intelligence algorithms, first, the encoding schema of the search agents should be determined. In this research, since the algorithm searches the solution space to find the best centroids, each particle should contain the centroids. So the position of each particle is a vector containing $k \times p$ cells where k is the number of centroids and p is the number of features. Figure 1 shows a particle containing n centroids of a 2 dimensional dataset. In this figure C_{ij} is j th dimension of the i th centroid.

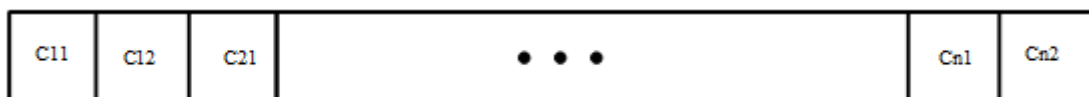


Figure 1. A particle containing n centroids for a 2-D dataset

After encoding the particles, a suitable fitness function should be selected or designed for measuring the quality of the particles. Up to now several indexes are introduced for measuring the quality of a solution to the clustering problem such as *Silhouette index* [10], *Davies-Bouldin index* [11], *Dunn index* [12] and etc. In this research *Calinski-Harabasz index* [13] is used to calculate the fitness



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

function due to its low complexity in compare to *Silhouette* and also its capability in finding good solutions. To calculate the fitness amount (quality) of each particle, first, all of the samples in the dataset are assigned to their nearest cluster center, after that fitness amount is calculated using the following equations:

$$VRC = \frac{SS_B}{SS_W} \times \frac{(N-k)}{k-1} \quad (3)$$

$$SS_B = \sum_{i=1}^k n_i \|m_i - m\|^2 \quad (4)$$

$$SS_W = \sum_{i=1}^k \sum_{x \in c_i} \|x - m_i\|^2 \quad (5)$$

In these equations, SS_B is the overall between-cluster variance, SS_W is the overall within-cluster variance, k is the number of centroids, N is the number of samples, m_i is the i th centroid, m is the overall mean of the data points, x is a data point, c_i is the i th cluster and $\|m_i - m\|$ is the Euclidean distance between two vectors. Good solutions have high VRC amount. In this research clustering problem is converted to a minimization problem so the following equation is used as fitness function:

$$f = \frac{1}{VRC} \quad (6)$$

In the first step of the algorithm a random population should be generated. In fact, each individual (particle) should be positioned randomly in the solution space. So for each particle k data points are selected randomly from the dataset as the centroids. In the second step, the quality of each particle is measured using equations 3 to 6. After evaluating each particle they start to move in the solution space through equations 1 and 2. For each particle, after movement a set of samples is produced by selecting j samples randomly from the dataset and each centroid is replaced with the position of the nearest sample. This procedure continues until the last iteration. Figure 2 shows the pseudo code of the proposed algorithm.

4.2 AUTOMATIC CLUSTERING ALGORITHM (APSO-CLUSTERING)

One of the main challenges in clustering especially big data clustering is finding the number of clusters, which is beyond the ability of conventional clustering algorithms such as *K-means* and *fuzzy C-means*. Although a new version of *K-means*, called *X-means* [14], has been introduced which can find the number of clusters, but according to our experiments its performance is not acceptable especially for clustering big datasets. To solve this problem and improve our algorithm, another phase is added to our method to search the solution space for proper number of clusters. In this phase a sequence of PSO-Clustering algorithm is run with different number of clusters to find the best possible solution. In this structure, in the beginning a random integer number is generated for k and PSO-Clustering is run to find k clusters. For the next step, number of clusters is changed using the following equation:



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

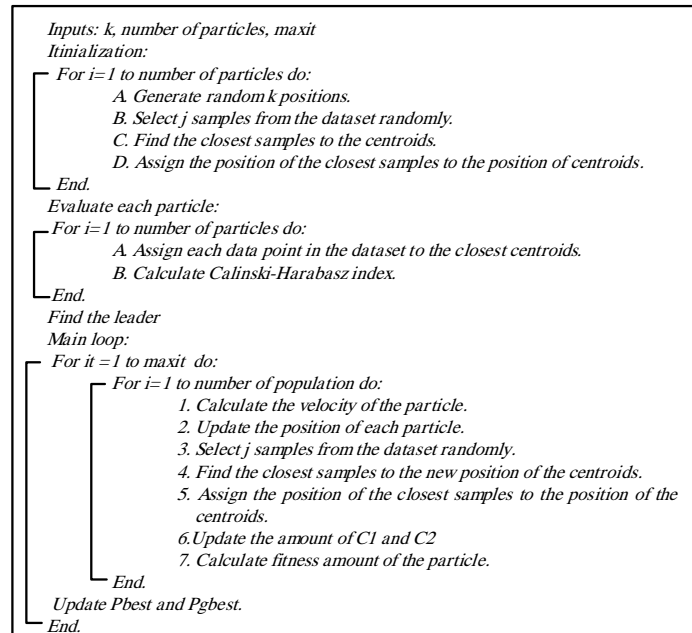


Figure 2- Pseudo code of the PSO-Clustering algorithm

$$k_{new} = k \pm \varepsilon \quad (7)$$

In this equation, ε is a random integer number. After that again PSO-Clustering will be run with k_{new} cluster centers. If the quality of the final result is better than solution found in the previous step, then k will be updated using the following equation:

$$k = k_{new} \quad (8)$$

This procedure is repeated until the last step of the sequence and finally k shows the best solution for the number of clusters. In our experiments three sequences are run and the best solution found by these three structures is selected as the proper number of clusters. After that in the second phase, PSO-Clustering algorithm will be run to find k centroids. In fact, in the first phase the algorithm searches the solution space to find the number of clusters and in the second phase it tries to find the position of the centroids. Figure 3 shows the pseudo code of the APSO-Clustering algorithm.

5. EXPERIMENTAL RESULTS ON SYNTHETIC DATASETS

To evaluate the effectiveness of the proposed method its accuracy is calculated on 6 synthetic datasets [15] using *Rand* [16] and *NMI* [17] indexes and its performance is compared with *K-means* and *X-means*. These datasets are briefly introduced in table 1. Tables 2 to 4 indicate the average results for *S*, *A* and *high dimensional* datasets respectively, after several experiments.



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

```

Stage 1:
Inputs: number of sequences, length of sequence, number of population
For i=1 to number of sequences do:
  1. k = Generate a random integer number
  2. beginning population = Generate a random population
  3. Current state.fitness = inf
  4. Current state.k=∅
  for j=1 to length of sequence do:
    if (j==1) do:
      current state.k=k
    end
    best fitness = PSO-Clustering(k, beginning population)
    if best fitness < current state.fitness
      current state.fitness=best fitness
      current state.k=k
    end
    k = current state.k ± ε
  end
  Bestk[i] = current state.k
End
Output: find the best solution which has the best fitness amount and its corresponding k
and its corresponding beginning population
Stage 2:
Final output = PSO-Clustering(best k, beginning population)

```

Figure 3- Pseudo code of the automatic PSO-Clustering algorithm (APSO-Clustering algorithm)

These tables demonstrate the effectiveness and high accuracy of the *APSO-Clustering* algorithm in both *Rand* index and *NMI* index. *NMI* index shows the similarity of the shapes of the clusters detected by the clustering algorithm to the shapes of real clusters. In fact, just finding the number of clusters accurately doesn't mean that the quality of the clustering is high and clusters are found correctly. High amount of *NMI* shows that the detected clusters are similar to the real clusters and the algorithm has clustered the data precisely. Also according to these tables, the accuracy of the proposed algorithm in finding the number of clusters is remarkable. So higher *NMI* and *rand* index besides higher accuracy in finding the number of clusters, indicate the superiority of *APSO-Clustering* algorithm over the other two conventional clustering algorithms. Actually, these tables demonstrate the power of the proposed method in clustering datasets containing several clusters and also high dimensional datasets. Figures 4 to 6 show the centroids found by the *APSO-Clustering* algorithm for one experiment on *S1*, *A1* and *dim064* respectively. In these figures the blue circles are the samples of the dataset and the grey stars are the centroids. In figure 5 the extra centroid detected by the automatic algorithm is shown by a red star. Figures 4 and 5 show that all of the clusters are accurately detected by the proposed automatic clustering algorithm. According to the tables and figures, *APSO-Clustering*'s performance on high-dimensional datasets and datasets with high number of clusters is remarkable. In the next section the experimental results on a real big dataset is reported.

5. REAL DATA

The dataset which is processed in this research, is generated by collecting data from whoscored.com website. These data are related to the areas that different players touched the ball in almost 93000 soccer matches. In a soccer match each player, based on his role, touches the ball in different areas of the field several times and each data point in this dataset is the center of the area in which a



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

specified player touched the ball in a specified match. The dataset contains the centers of the areas in which 2005 different players touched the ball in 93000 matches.

Table 1- Overview of the synthetic datasets

	<i>Dataset</i>	<i>Number of data points</i>	<i>Number of features</i>	<i>Number of clusters</i>
<i>S datasets</i>	<i>S1</i>	5000	2	15
	<i>S2</i>			
<i>A datasets</i>	<i>A1</i>	3000	2	20
	<i>A2</i>	5250	2	35
<i>High dimensional datasets</i>	<i>Dim064</i>	1024	64	16
	<i>Dim1024</i>	1024	1024	16

Table 2- Average results for S datasets

<i>Method</i>	<i>Rand index</i>		<i>NMI</i>		<i>Number of detected clusters</i>	
	<i>S1</i>	<i>S2</i>	<i>S1</i>	<i>S2</i>	<i>S1</i>	<i>S2</i>
<i>APSO-Clustering</i>	0.9959	0.9911	0.9710	0.9413	16.25	15.75
<i>X-means</i>	0.9225	0.9353	0.8341	0.8091	8	9
<i>K-means</i>	0.9901	0.9777	0.9489	0.8878	-	-



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

Table 3- Average results for A datasets

<i>Method</i>	<i>Rand index</i>		<i>NMI</i>		<i>Number of detected clusters</i>	
	<i>A1</i>	<i>A2</i>	<i>A1</i>	<i>A2</i>	<i>A1</i>	<i>A2</i>
<i>APSO-Clustering</i>	0.9981	0.9975	0.9897	0.9752	20.25	38.33
<i>X-means</i>	0.862	0.9104	0.7208	0.7482	6	9
<i>K-means</i>	0.9877	0.9924	0.9574	0.9586	-	-

Table 4- Average results for high dimensional datasets

<i>Method</i>	<i>Rand index</i>		<i>NMI</i>		<i>Number of detected clusters</i>	
	<i>Dim064</i>	<i>Dim1024</i>	<i>Dim064</i>	<i>Dim1024</i>	<i>Dim064</i>	<i>Dim1024</i>
<i>APSO-Clustering</i>	0.999	0.999	0.9987	0.9988	17	17
<i>X-means</i>	0.9844	0.9922	0.9682	0.9843	14	15
<i>K-means</i>	0.9984	1	0.9966	1	-	-

6. FINDING DIFFERENT ROLES IN SOCCER

Types of events created by different players in a soccer match are correlated to their roles. In fact, different players are supposed to do different tasks based on their roles in soccer. For example, a defender should stop the opponent's attacks and prevent them to score while a forward should try to score and use the opportunities created for him/her. Usually a defender makes more tackles and interceptions than a forward, while a forward makes more shots and passes in a game. This means that comparing players without considering their roles is meaningless. Therefore, using an effective method to extract different roles of Players, is necessary. *APSO-Clustering* algorithm found 8 different clusters which indicate 8 different roles in a soccer match. Figure 7 shows the output of the algorithm including data points (blue circles) and detected centroids (grey stars). Also figure 8 shows the equivalent role for each cluster and table 5 contains the name of each role. These names are selected from Pappalardo and his colleagues' paper [8].



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

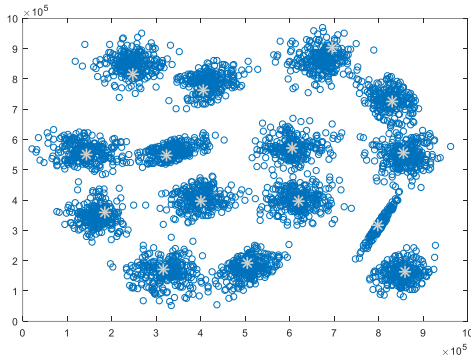


Figure 4- Centroids of S1 dataset found by APSO-Clustering algorithm

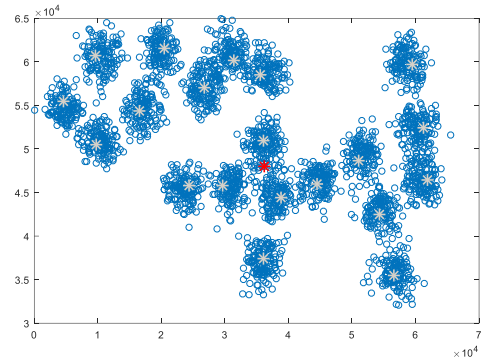


Figure 5- Centroids of A1 dataset found by APSO-Clustering algorithm

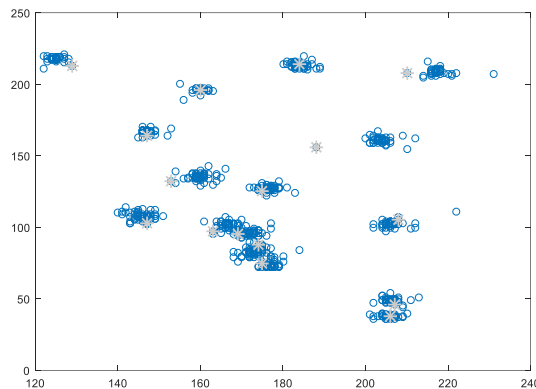


Figure 6- Centroids of dim064 dataset found by APSO-Clustering algorithm

Although each player, depending on his skills, is asked to play in a specified area, but no one expects not to see him in the other parts of the field. This really depends on the situation of the match. For example, for a team losing the match and needing just one goal to equalize, sometimes also defenders participate in attacks to score. Also some players have the ability to play in different areas and more than one role in different matches or even in one match. Extracting frequency of touching the ball, or creating different events (pass, shots, fouls, cross and etc.) in different areas is important for the coaches to find movement patterns of the players and the team, design effective strategies for each match, select the best players based on the selected strategies and preparing the team for every situation. To analyze better, we found the normalized frequency of touching the ball in 8 different areas for each player. So a new dataset including 2005 objects (players) and 8 attributes, indicating the frequency of touching the ball by different players in different areas, is created.



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

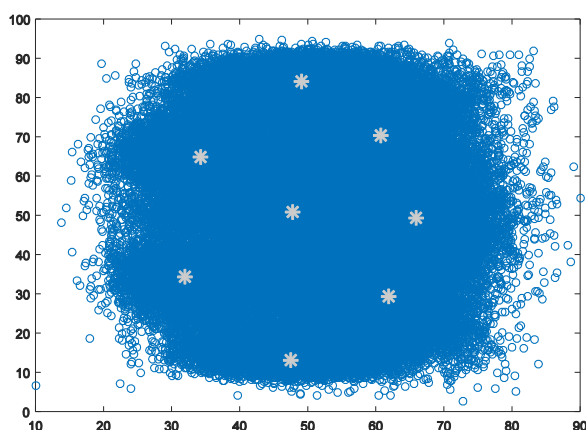


Figure 7- 8 different clusters found by the *APSO-Clustering* algorithm

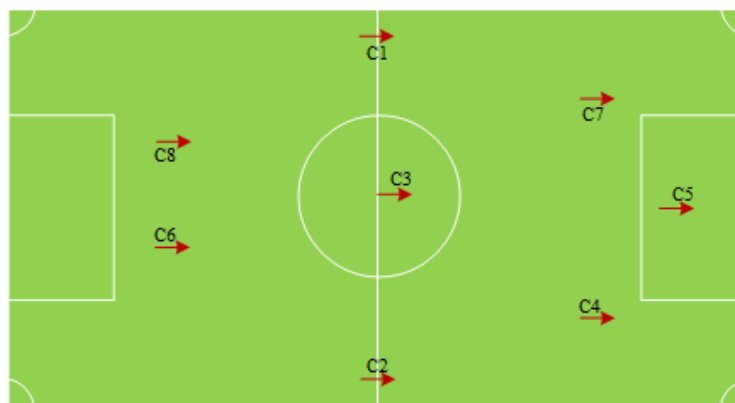


Figure 8- 8 different roles of players

Clustering this dataset, we found 8 centroids which are indicated in table 6. Each centroid contains 8 elements showing the average touching frequency of the ball by different players in different areas of the field. The biggest element in each centroid, shows that most of the players belong to the cluster play in that specified area. For example, in the first cluster, players touched the ball (on average) 78 percent in the left side of the field (C1). This shows that they play as left fielder, left wing or left back (according to table 5). Also they created 15 percent of their events in C8 area where they had to play as a left defender. In other words, these players' main role is left fielder or left wing and depending on the situation of the match or depending on the strategies, they also played the role of a left defender (or left back). This concept is shown in figure 9. Also the box plot of the players belong to the first cluster is shown in figure 10. This figure emphasizes that although some of the players of this cluster have played in C2, C5 and C8 areas but these players often played in C1 area



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

as left fielder. Table 7 shows the performance of three players of the first cluster. This table indicates how many times each of these players touched the ball in different areas. According to this table, the performance center of the first player was C1 in 69 matches, C2 and C8 in 13 matches and C5 in 2 matches. second centroid, according to table 6, contains players who have touched the ball 58 percent of the times in C8 area while 17 percent of the events, they created, were in C1 area. This shows that these players often play as defenders. Their roles are shown in figure 11 and the corresponding box plot is shown in figure 12. Figures 13 and 14 show the same information for the third cluster.

Table 5- Name of different roles

<i>Role</i>	<i>name</i>
<i>C1</i>	Left fielder, left wing, left back
<i>C2</i>	Right fielder, right wing, right back
<i>C3</i>	Central midfielder, internal midfielder
<i>C4</i>	Right wing, right forward
<i>C5</i>	Central forward
<i>C6</i>	Right back
<i>C7</i>	Left wing, left forward
<i>C8</i>	Left back, left central back

These players usually try to make good opportunities for forwards and wingers and also they should participate in defense and help the defenders. Their play is focused on the central part of the field. Figure 14 clearly shows the stability of the central midfielders. In compare to the other roles they are the most stable players. In other words, this analysis shows the importance of the role of central fielders. Usually they don't change their position and their role is really important both in defending and attacking.



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

Table 6- Eight centroids extracted from ball touching frequency dataset

roles centroids	<i>C1</i>	<i>C2</i>	<i>C3</i>	<i>C4</i>	<i>C5</i>	<i>C6</i>	<i>C7</i>	<i>C8</i>
<i>1</i>	0.788	0.022	0	0.011	0.022	0	0	0.155
<i>2</i>	0.176	0	0	0	0.117	0.117	0	0.588
<i>3</i>	0	0	0.914	0	0.085	0	0	0
<i>4</i>	0.058	0	0.176	0	0.588	0	0	0.176
<i>5</i>	0	0.040	0	0.163	0	0	0.795	0
<i>6</i>	0	0.782	0	0.102	0	0	0.115	0
<i>7</i>	0	0	0	0.698	0	0	0.301	0
<i>8</i>	0	0	0.047	0.023	0	0.88	0.023	0.023

Table 7- touching frequency of the ball in 8 different areas by three players of the first cluster

<i>Player</i>	<i>C1</i>	<i>C2</i>	<i>C3</i>	<i>C4</i>	<i>C5</i>	<i>C6</i>	<i>C7</i>	<i>C8</i>
<i>1</i>	69	13	0	0	2	0	0	13
<i>2</i>	29	2	0	0	1	0	0	1
<i>3</i>	16	5	1	0	2	3	0	10



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

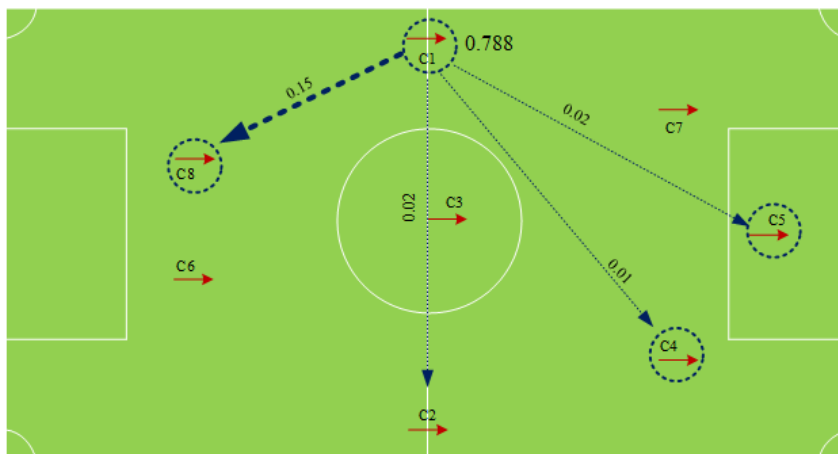


Figure 9- Different areas of touching the ball by left fielders, left wingers or left backs

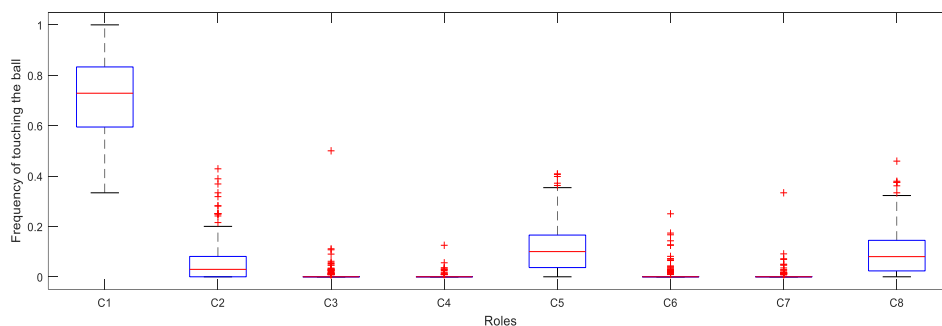


Figure 10- Box plot of the players belong to the first cluster

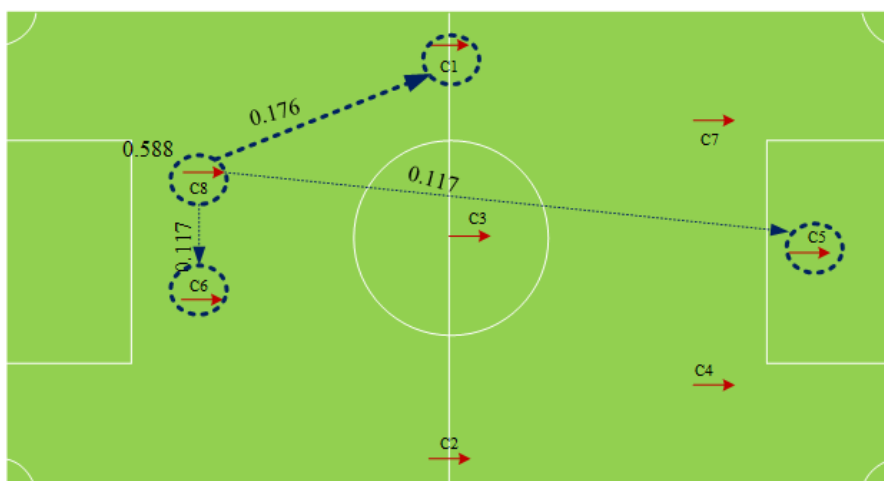


Figure 11- The areas of touching the ball by left back players.



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

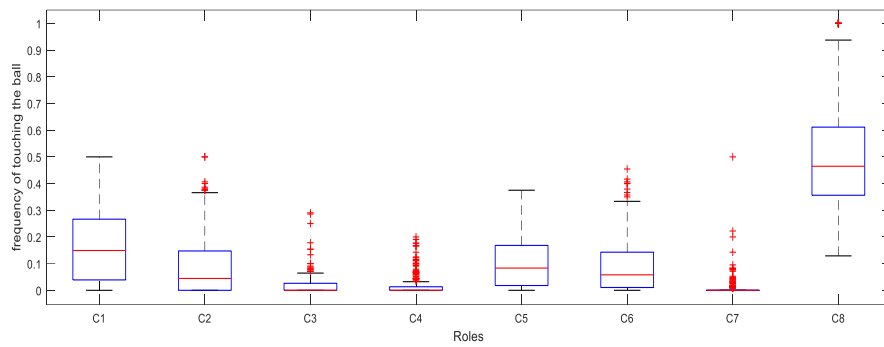


Figure 12- Box plot of the players belong to the second cluster

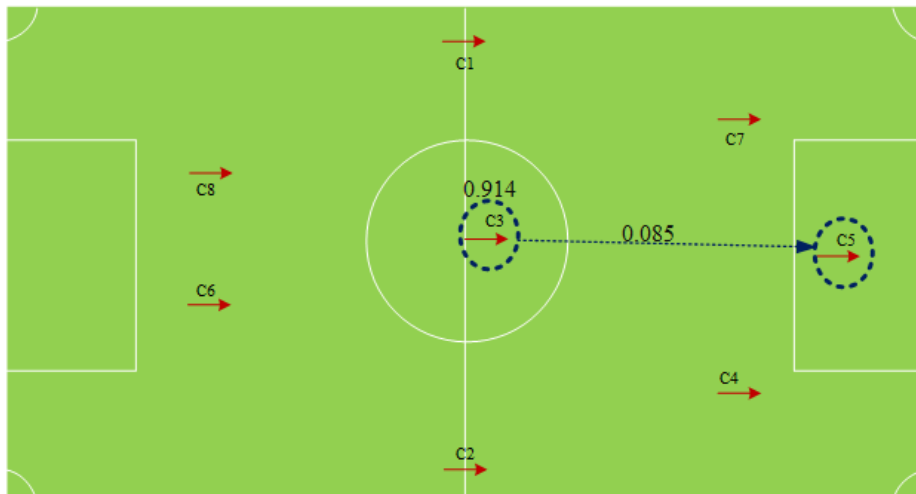


Figure 13- Central fielders touches the ball in C3 area 91 percent of the times.

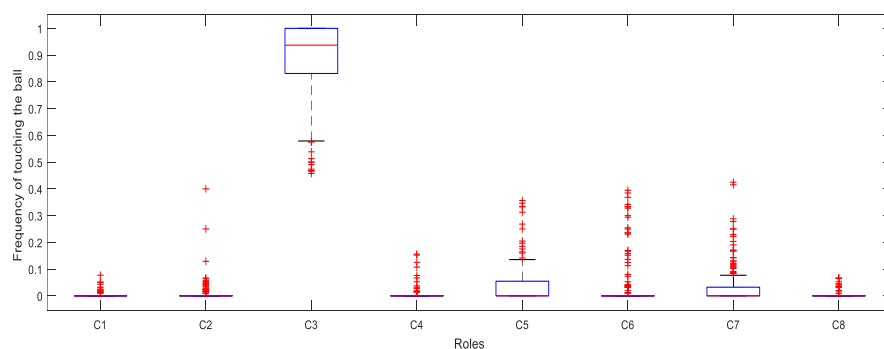


Figure 14- Box plot of the players belong to the third cluster



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

The last row of table 6 (cluster 8) shows the importance of players playing in C6 area. Most of the players in cluster 8, are right back defenders. Usually they don't play in the other areas, on the other hand just other kind of defenders, who have mostly played in C8 area, have played their role in some matches. This means that these players have a key role in defending. Also according to 5th and 7th rows of table 6, we can infer that wingers are the most important players in attack. They change their position rarely. Right wingers have played the role of left wingers in 30 percent of the matches and on the other hand, left wingers have played the role of right wingers in 16 percent of the matches. These statistics show that wingers (left or right) can change the side of their play, but it is not common to see them playing another role. So these results show that, right back, central midfielder and wingers are the most important and effective roles. Which means that having top players in these roles is necessary for a team trying to win the title.

7. CONCLUSION AND FUTURE WORKS

In this paper a new automatic clustering method called, *APSO-Clustering*, is used to cluster a soccer dataset for finding different roles of soccer players. The proposed method detected eight distinct clusters indicating eight roles in a team. In the next step the players who have changed their positions in different matches, have been found. In fact, for each player centers of the areas he played in each match are found and a new dataset including 2005 objects (players) and eight features showing the frequency of playing in each role, has been created. Interesting results which are achieved by clustering this dataset, indicate three key roles in each team: right back, central midfielder and wingers. In addition to that, generally the results demonstrate the power and effectiveness of *APSO-Clustering* algorithm in clustering massive datasets. The proposed method not only finds the position of the centroids accurately, but also it finds the number of clusters precisely which is very important specially for clustering big datasets. For future works, it is good to use this method to analyze the performance of each team in a season and make a good comparison between the players based on their roles. Also using another swarm intelligence algorithm to cluster these kind of datasets and compare the results with PSO is interesting.

References

- [1] Gudmundsson J, Horton M. Spatio-Temporal Analysis of Team Sports. *ACM Comput Surv.* 2017;50(2):1-34.
- [2] Rein R, Memmert D. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. SpringerPlus. 2016;5(1):1410.
- [3] Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society Series C (Applied Statistics).* 1979;28(1):100-8.
- [4] Rezaee MJ, Jozmaleki M, Valipour M. Integrating dynamic fuzzy C-means, data envelopment analysis and artificial neural network to online prediction performance of companies in stock exchange. *Physica A: Statistical Mechanics and its Applications.* 2018;489:78-93.
- [5] Shi Y, Eberhart RC, editors. Empirical study of particle swarm optimization. *Evolutionary computation, 1999 CEC 99 Proceedings of the 1999 congress on; 1999: IEEE.*
- [6] Duch J, Waitzman JS, Amaral LAN. Quantifying the performance of individual players in a team activity. *PLoS one.* 2010;5(6):e10937.



International Congress of Sciences and Innovative Technologies

■ Sept. 5-6, 2018 Babol Noshirvani University of Technology- Babol, Iran

- [7] Brooks J, Kerr M, Guttag J, editors. Developing a data-driven player ranking in soccer using predictive model weights. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016: ACM.
- [8] Pappalardo L, Cintia P, Ferragina P, Pedreschi D, Giannotti F. PlayeRank: Multi-dimensional and role-aware rating of soccer player performance. arXiv preprint arXiv:180204987. 2018.
- [9] Stanojevic R, Gyarmati L, editors. Towards data-driven football player assessment. Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on; 2016: IEEE.
- [10] Starczewski A, Krzyżak A, editors. Performance evaluation of the silhouette index. International Conference on Artificial Intelligence and Soft Computing; 2015: Springer.
- [11] Maulik U, Bandyopadhyay S. Performance evaluation of some clustering algorithms and validity indices. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002;24(12):1650-4.
- [12] Pakhira MK, Bandyopadhyay S, Maulik U. Validity index for crisp and fuzzy clusters. Pattern recognition. 2004;37(3):487-501.
- [13] Caliński T, Harabasz J. A dendrite method for cluster analysis. Communications in Statistics-theory and Methods. 1974;3(1):1-27.
- [14] Pelleg D, Moore AW, editors. X-means: Extending k-means with efficient estimation of the number of clusters. Icml; 2000.
- [15] School of Computing University of Eastern Finland. clustering basic benchmarks 2015 [Available from: <https://cs.joensuu.fi/sipu/datasets/>].
- [16] Rand WM. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association. 1971;66(336):846-50.
- [17] McDaid AF, Greene D, Hurley N. Normalized mutual information to evaluate overlapping community finding algorithms. arXiv preprint arXiv:11102515. 2011.