ZIBELINE INTERNATIONAL PUBLISHING

Malaysian Sports Journal (MSJ)

DOI : *http://doi.org/10.26480/msj.01.2019.04.07*

ISSN : 2710-5954 (Online)
CODEN: MSJABY

CrossMark

REVIEW ARTICLE

SPORTS COMPETITION STRESSORS BASED ON K-MEANS ALGORITHM

Ranjit Kumar, Zhengwu Liu, Wan Zamri

Institute of Sports Science & Technology, Pune, India
Weinan Normal University, China
Faculty of Applied Sciences, Tunku Abdul Rahman University College, Setapak, Kuala Lumpur, Malaysia
****Corresponding Author Email: zhengwuliu322@126.com***

ARTICLE DETAILS

Article History:

Received 03 August 2019
Accepted 10 September 2019
Available online 08 October 2019

ABSTRACT

In view of the data characteristics of the sports competition stressor, a new improved hierarchical K-Means algorithm is formed by combining the known partition algorithm K-Means algorithm with the hierarchical algorithm. At the same time, the improved algorithm is really applied to the data of sports competition stressor, and the clustering results are obtained and analyzed. The results show that different types of athletes face different pressures due to their own conditions and external environment, and their stressors are quite different. It is concluded that for different types of players, their own circumstances should be considered. It is necessary to conduct emotional guidance to prevent the negative emotion of the athletes.

KEYWORDS

K-Means algorithm, stressors, sports competition, clustering.

1. INTRODUCTION

With the development of competitive sports, the scale of events and the number of viewers are increasing day by day, and the intensity of the competitions on the stadiums is also increasing [1]. These will have a certain pressure impact on the athletes' psychology. Pressure coping has become a problem that cannot be ignored in today's competitive sports [2,3]. Therefore, many scholars and experts have paid more attention to the field of competition pressure. Competition pressure mainly comes from the competition environment. Because of the highly challenging of sports competition itself, long-term stressful conditions are inevitable for athletes. The research shows that the pressure competition will bring many negative effects to athletes and sports organizations, such as causing anxiety, aggressive behavior and low satisfaction, thereby affecting their race performance and physical and mental health [4]. Poor coping skills can significantly reduce the athlete's attention and ability. Therefore, only when athletes fully understand and recognize the pressure factors of their own and adjust their mental state, they will play their competitive level in the arena.

At present, many athletes alleviate this problem by psychological counselling and psychological diagnosis. The traditional psychological counseling is usually asked by professionals, or some questionnaires are provided to evaluate the psychological status of the athletes. However, these methods are inefficient and cannot give advice in time and effectively. Some studies have found that applying data mining related methods to psychological stress coping can solve the situation that the number of athletes is huge and the psychological condition is complex and changeable. It also helps to effectively solve the stress response. According to this, the K-Means algorithm is applied to the analysis of competition stressor. From the aspect of data mining, the data of the athletes' psychological questionnaire evaluation are analyzed and excavated. Valuable information is obtained to assist psychological experts in making effective psychological diagnosis.

2. LITERATURE REVIEW

The K-Means algorithm is the simplest and most common algorithm. It selects the random cluster center as the initial value, and then continuously distributes the similar data into the same cluster until the convergence is finished [5]. The time complexity of K-Means algorithm is very small. However, K-Means algorithm is very sensitive to the initial value. Therefore, it is easy to fall into the local optimal solution. In order to improve the shortcoming of K-Means algorithm, many literatures try to improve K-Means algorithm. Some improvements try to choose a good initial value that makes the algorithm easier to find the global optimal solution. Other improved algorithms attempt to segment and merge the clusters that have been obtained [6,7]. When the variance of a cluster is higher than the specified threshold, the cluster is divided into two clusters. However, when the centroid distance of the two cluster sets is less than the specified threshold, the two cluster sets will be merged. These improvements make the best result of clustering by random initial values. For example, the ISODATA algorithm is a technology that can combine and split the cluster. In ISODATA, if the initial partition is in a "full circle" partition, the best two partitioned sets can be generated. ISODATA first merges cluster {A} and {B, C} as a cluster. The distance between their centroids is the smallest of all cluster centers. Then, for the cluster {D, E, F, G}, it will be cut into two clusters of {D, E} and {F, G}. There are also some improvement algorithms that choose different standard functions. Other improved algorithms, such as the dynamic clustering algorithm, use the maximum likelihood estimation and propose a method to obtain dynamic clustering from the data set. In addition, the regular Mahalanobis distance of the distance measure is improved to obtain the super spherical cluster. The typical clustering K-Means algorithm is generally used for the data set with same direction. On the other hand, based on the time and space complexity, the time and space complexity of classification clustering algorithm is far lower than the hierarchical clustering algorithm. Therefore, combining the advantages of the two, a mixed algorithm which has a good clustering effect is developed.

In fact, there have been many universities abroad that have begun to study the psychological aspects of sports competitions, mainly to study the stressors [8]. Domestic universities are mainly concentrated in physical education institutes or other institutes of physical education, which is a cross discipline which has gradually attracted attention in recent years. At present, researchers generally use cognitive interaction theory to guide the competition pressure [9-11]. Sports psychologists help athletes develop many stress coping strategies, such as competition stress coping strategies, image training, attention training, self-confidence enhancing training and communication skills. As a result, they aim to help athletes enhance team cohesion. In addition, they have made other meaningful studies. Among them, the research on pressure is summarized mainly in the following aspects: the study of stressful events, the cognitive evaluation of stressful events, and the study of coping strategies adopted by athletes. The traditional psychological counseling is usually asked by professionals, or some questionnaires are used to evaluate the psychological status of the athletes [12]. The professionals divide the questionnaire results provided by the athletes into three grades: high, medium and low. Different psychological analysis is made according to different grades [13]. Generally, the athletes' psychology is roughly classified according to their rank. Previous research conclusions often focus on athletes' coping strategies in a specific project, or one element or part of the coping process. Therefore, it cannot effectively analyze the psychological status of the overall personnel. Or, the suggestion is too one-sided and it can't make a comprehensive evaluation for the athletes. When the number of athletes increases, psychological experts cannot effectively judge the athlete according to their personal conditions [14]. Therefore, the efficiency is low, and the coping strategies and results are not ideal. Automatic simulation is carried out from data mining. Through K-Means algorithm and its corresponding technology, it helps psychological experts to analyze and understand athletes' stressors in sports competitions to provide a basis for solving corresponding measures in the later stage [15].

3. METHOD

3.1 Related Formula

Silhouette coefficient: The silhouette coefficient can be represented in a simple image. It shows the advantages and disadvantages of each object in its cluster in the data set. It is assumed that data can form a result using some sort of clustering method, and then the data has been divided into K cluster centers. For any data object i, a(i) represents the average difference between the object i and the other objects in the same cluster. That is, it represents the degree to which the object i is assigned to the current cluster. The more similar i is to other objects in the cluster, the more appropriate it is to assign the object i to the current cluster. The definition formula of a(i) is as follows:

$$\mathrm{a(i)} = \frac{\sum i^n a C_i i^n m i^{dist(i^n, i)}}{|C_i|-1} \tag{1}$$

In the formula, b(i) represents the minimum mean value of the distance between the object i and the other cluster (the cluster is not set by i), that is, the distance between the object i and the nearest neighbor set of its cluster. If the object i is not effectively allocated, the best cluster that it needs to be allocated should be the nearest neighbor to the current assigned cluster set. The definition formula of b(i) is as follows:

$$\mathrm{b(i)} = min_{ci:i \le j \le k, j \ne i} \left\{ \frac{\sum i^n a C_j{}^{dist(i^n, i)}}{|c_i|-1} \right\} \tag{2}$$

Then, the silhouette coefficient can be defined as the following formula:

$$\mathrm{s(i)} = \frac{b(i)-a(i)}{\max\{a(i), b(i)\}} \tag{3}$$

Or, it can be written as follows:

$$\mathrm{s(i)} = \begin{cases} 1 - a(i)/b(i), & a(i) < b(i) \\ 0, & a(i) = b(i) \\ b(i)/a(i) - 1, & a(i) > b(i) \end{cases} \tag{4}$$

According to the above formula, S(i) takes [-1,1]. When a(i)<b(i), because a(i) represents the similarity degree between i and the object in the cluster, the smaller the a(i) is, the better the i is assigned. The larger the b(i) is, the worse the neighbor cluster assigned by I is. Then, at this time, S(i) closing to 1 indicates that the overall allocation of the dataset is better. S(i) closing to 0 indicates that the object i is at the edge of the two clusters. S(i) closing to -1 indicates that the object i is different from the cluster object, and the cluster quality assigned by the object i is poor. The average value of S(i) on the dataset is used to evaluate the quality of the clustering results. For example, the K-Means algorithm is used to cluster. If the selected K value is too large or too small, the value of the obtained S(i) will be significantly different. Thus, the silhouette coefficient can be used as a reference for clustering numbers of the selected data sets.

Similarity function: The average value of the overall cluster is calculated by the overall similarity.

$$\mathrm{similarity} = \frac{\sum x a C_i^{dist(x, c_i)}}{m} \tag{5}$$

In the formula, Ci represents cluster i. ci一 is the center of Ci. x is any object of cluster Ci. dist(x,ci) represents the distance between the object and the cluster Ci center. mi is the number of the objects in the cluster i. The similarity function is used to calculate the tightness in the cluster. The smaller the value is, the closer the cluster is.

Mean variance: It is similar to similarity. The whole cluster has a standard clustering measure function. Mean variance is used to measure the overall situation of clustering. K-Means is to adjust the J to the minimum.

$$\mathrm{J} = \sqrt{\frac{\sum_{i=1}^{k} \sum_{j=i}^{n_i} (x_{ij} - c_i)^2}{n-1}} \tag{6}$$

In the formula, xij represents any data object in the cluster i, and ci represents the cluster center of the cluster set i.

3.2 Specific algorithm

In the analysis of the causes, the effects of small data and more data dimension on the clustering results are considered. According to the data characteristics of sports competition stressor, the improved algorithm is considered. In view of the shortage of traditional hierarchical K-Means algorithm, an improved hierarchical K-Means algorithm is proposed. Setting x is the data of n R-dimensional spaces. After the improvement, the specific algorithm steps are as follows:

Step 1: The original data is processed and the silhouette coefficient is calculated. When the value of K is the maximum, the value is used as the initial value;

Step 2: The agglomerative hierarchical clustering algorithm is used to merge two adjacent clusters to form new cluster;

Step 3: A new cluster center after combination and the average value of the two cluster centers on the previous level are calculated;

Step 4: The step 3 and the step 4 are repeated until the (K-R)(0<=R<=K-2) clusters are obtained. (if K=2, then R=0)

Step 5: The similarity in the cluster of all the partition clusters is calculated respectively;

Step 6: The cluster with smallest similarity is selected, that is, the cluster with the largest class radius. The cluster is decomposed to find the sample xi1 which is away from the class center ci. Then, the sample point xi2 which is away from the class center xi1 is selected.

Step 7: The xi1, xi2 of these two points and other clustering centers are used as new clustering centers, and then the k mean clustering is redone.

Step 8: If the centroid has a change, it returns to step six. Otherwise the algorithm is over and the result is output.

3.3 Clustering of sports competition pressure questionnaire data

The questionnaire is designed, including the competition stressor file, resilience questionnaire, perceived social support questionnaire in sports, athlete and coach relationship questionnaire, athlete burnout scale table, athlete engagement questionnaire. In addition, each set of questionnaires is subdivided into several subdivisions. The cluster data is based on the above five sets of questionnaires. The comprehensive data of nearly 500 athletes are analyzed, and the relevant clustering information is obtained. The original evaluation grade of the questionnaire is generally divided into five subdivisions (fully conformed, more consistent, general, less consistent and completely inconsistent). Therefore, the original score of the data is obtained. The score for each problem is 5 scores, and the cumulative score is the total score of the current subdivision dimension. Because the number of topics corresponding to every dimension is not the same, the normalization process is carried out for all clustering data to prevent weight judgment. After the invalid information is removed, the normalization process is carried out. The normalization formula is as follows:

$$X^{il} = \frac{X - min}{max - min} \quad (7)$$

In the formula, max and min are the maximum and minimum of the sample data, respectively.

According to the improved algorithm, it is necessary to carry out clustering analysis after cleaning and converting the data. First, the approximate number of cluster numbers K is calculated by the silhouette coefficient. Figure 1 shows the curve graph of the value of the silhouette coefficient (re) when the k value is from 2 to 100. According to figure 1, it is known that the number of cluster at 2 is the largest. This indicates that the effect is better when the number of clusters is 2, but this is not the final result. The hierarchical clustering algorithm is used to cluster. According to the improved algorithm, the initial clustering center of the next step of K-means is calculated, which is convenient for the subsequent calculation.

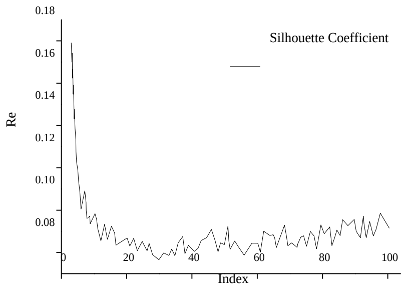

Figure 1: The curve graph of silhouette coefficient.

Then, it is determined that the hierarchical clustering algorithm is used for the experimental data. It is divided into 2 categories to cluster in the later period.

4. RESULTS AND ANALYSIS

The improved hierarchical K-Means algorithm is used to cluster the data sets as well as the 22 dimensions of competition stressor, social support and athlete burnout. The final cluster center is shown in table 1.

Because of too many dimensions, it is more intuitionistic to turn it into a linear graph. Fig.2 is the final cluster center.

As shown in fig.2, the proportion of the first type is the most, the second is followed, and the third is the least. The first type is similar to the second type, and some of the attributes are different. However, the total number of the third type is less, but it is significantly different from the first and the second type. This means that this part of the athletes' mental level is poor. The following is the analysis of different types of athletes:

Table 1: Final cluster center

Number	Final cluster center	Total number of data
1	0.739, 0.125, 0.293, 0.684, 0.800, 0.227, 0.908, 0.624, 0.180, 0.380, 0.317, 0.319, 0.301, 0.085, 0.358, 0.319, 0.219, 0.717, 0.714, 0.733, 0.339	262
2	0.599, 0.110, 0.205, 0.513, 0.570, 0.158, 0.678, 0.412, 0.129, 0.478, 0.435, 0.424, 0.460, 0.108, 0.493, 0.478, 0.418, 0.467, 0.468, 0.523, 0.237	223
3	0.433, 0.874, 0.250, 0.750, 0.667, 0.782, 0.427, 0.760, 0.802, 0.705, 0.091, 0.144, 0.438, 0.882, 0.417, 0.425, 0.617, 0.656, 0.656, 0.573, 0.823	6

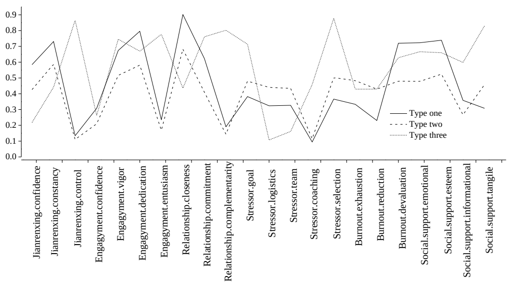

Figure 2: Schematic diagram of the final cluster center

The athletes in first type are better in terms of toughness. But they are better at toughness control. Among them, the score of control sense is low due to the reverse score project. The lower score indicates that the athletes have a better attitude. From the questionnaire, the sense of control mainly refers to the athletes' score on their own sports status and mentality. Then, in the degree of input, their confidence and enthusiasm are low. In addition, in the relationship with the coach, the overall relationship with the coach is better. However, it is poor in complementarity. Complementarity mainly refers to the reaction, psychology and coping situation of the athletes to the coach's guidance. This shows that the athletes in first type have a poor coping situation and the ability needs to be improved in the process of coach guidance. In the stressor score, the overall pressure level of the athletes in first type is low, indicating that the athletes have a better mental quality. In social support, in general, emotional support and information support have a high a score. This shows that the athletes in first type have higher happiness and the highest social support.

Compared with the athletes in first type, the athletes in second type have little difference in terms of toughness and input degree and the relationship with the coach. The score in second type is a little lower than the athletes in first type, and they are of middle level. But in the athletes' burnout, sports exhaustion is the highest of the three types. At the same time, the sense of achievement is also the easiest to reduce. However, in social support, the support degree is the smallest of the three types.

The athletes in third type have the lowest score in the fortitude compared with the first two types. Among them, the control sense of tenacious is the highest, which is the reverse calculation item. These projects mainly indicate that such athletes are easily frustrated and have no confidence in themselves, and thus affect sports. Therefore, their attitude for sport competition needs to be strengthened. In the input meter, the athletes in third type have the highest enthusiasm and can maintain a good enthusiasm. However, in the aspect of competition stressor, it is clear to see that the biggest stressor of athletes in third type comes from the selection and the pressure of the target. Finally, the score of sports negative evaluation is the highest in the athlete burnout scale. This shows that the athletes in third type are incompatible with sports. At the same time, they question the movement and do not concentrate on the sports training, thus producing a conflict psychology.

In a comprehensive view, among the 3 types of athletes, the athletes in first type may belong to the seeded player. Their overall sports quality is high, and they work hard. In addition, their physical fitness is good. However, the input of the athletes is not high enough and needs more encouragement. The athletes in second type may belong to the medium diligent-type player. This kind of athlete may work hard at ordinary times and easily form a burden psychology. However, because they are not qualified athletes, their sense of achievement is also low, and social support is the lowest. It is easy to form an anxiety mentality that falls short of the best but be better than the worst. The athletes in third type are less than the total and belong to the most stressful type. They are in a poor mind state in many ways. Although the enthusiasm of sports is high, it may not be as good as the first two types of athletes. Therefore, athletes are prone to burnout, and even challenge the resistance training. In addition, the athletes in third type may be less likely to join in the large events than the first two. Therefore, they have more pressure. Athletes may be more likely to try to prove themselves, or to completely contradict the extreme emotions of the movement. This kind of account is less than, but the coach needs to pay more attention to it. Coaches need to teach students in accordance with their aptitude to help athletes develop their mental health.

5. CONCLUSION

The improved K-Means algorithm is applied to the stressor data of sports competition. First, the number of clustering in the initial hierarchical clustering is determined according to the silhouette coefficient, and the first hierarchical clustering is carried out. Then, the initial clustering center of the new K-Means algorithm is recalculated according to the improved algorithm. Then, the K-Means algorithm is carried out, and cluster analysis of data sets is used to get the clustering results. According to the results of cluster analysis, the data are analyzed and the conclusion is obtained. The analysis result shows that more encouragement is important for the seeded players to keep them in good state of movement. For the diligent- type players, many aspects of encouragement are needed to improve their sense of achievement. In addition, it is necessary to prevent excessive training and cause unnecessary injury. For the athletes with excessive pressure, the coach should help them to sum up the reasons and set up the sports goals. In addition, the coach should also take more guidance to prevent the negative emotion of the athletes.

REFERENCES

[1] Burgess, N. S., Knight, C. J., Mellalieu, S. D. 2016. Parental stress and coping in elite youth gymnastics: an interpretative phenomenological analysis. Qualitative Research in Sport Exercise & Health, 8(3), 237-256.

[2] Secades, X. G., Molinero, O., Salguero, A. 2016. Relationship Between Resilience and Coping Strategies in Competitive Sport. Percept Mot Skills, 122(1), 336-349.

[3] Nixdorf, I., Frank, R., Beckmann, J. 2015. An explorative study on major stressors and its connection to depression and chronic stress among german elite athletes. Advances in Physical Education, 5(4), 255-262.

[4] Choi, H. S., Johnson, B., Kim, Y. K. 2014. Children's development through sports competition: derivative. Adjustive, generative, and maladaptive approaches. Quest, 66(2), 191-202.

[5] Yuan, Baolan, Zhang. A MAX-MIN CLUSTERING METHOD FOR k-MEANS ALGORITHM OF DATA CLUSTERING. Journal of Industrial & Management Optimization, 8(3), 565-575.

[6] Zhao, L., Hou, X., Hu, J. 2014. Improved k-means algorithm based analysis on massive data of intelligent power utilization. Power System Technology, 38(10), 2715-2720.

[7] Ponnelle, S., Berjot, S. A. 2016. qualitative investigation of organizational, competitive and personal life stressors among amateur athletes. International Journal of Sport Psychology, 47(5), 443-465.

[8] Bebetsos, E., Goulimaris, D. 2015. Examination of "Pre-competition" anxiety levels, of mid-distance runners: A quantitative approach. Polish Psychological Bulletin, 46(3), 498-502.

[9] Thornton, H. R., Delaney, J. A., Duthie, G. M., Scott, B. R., Chivers, W. J., Sanctuary, C. E. (2016. Predicting self-reported illness for professional team-sport athletes. Int J Sports Physiol Perform, 11(4), 543-550.

[10] Dick, K., Alexander, H. D., Moczygemba, J. D. 2016. Use of shelter tubes, grass-specific herbicide, and herbivore exclosures to reduce stressors and improve restoration of semiarid thornscrub forests. Restoration Ecology, 24(6), 785-793.

[11] Arnold, R., Fletcher, D., Daniels, K. 2016. Demographic differences in sport performers' experiences of organizational stressors. Scandinavian Journal of Medicine & Science in Sports, 26(3), 348-358.

[12] Dalmiya, S., Dasgupta, A., Kanti, Datta, S. 2016. Application of wavelet based k-means algorithm in mammogram segmentation. International Journal of Computer Applications, 52(15), 15-19.

[13] Astakhova, N. N., Demidova, L. A., Nikulchev, E. V. 2015. Forecasting method for grouped time series with the use of k-means algorithm. Applied Mathematical Sciences, 9(97), 4813-4830.

[14] Qin, J., Fu, W., Gao, H., Zheng, W. X. 2017. Distributed k-means algorithm and fuzzy c-means algorithm for sensor networks based on multiagent consensus theory. Transactions on Cybernetics, 47(3), 772-783.

[15] Kasuga, H., Yamamoto, H., Okamoto, M. 2016. Color quantization using the fast k-means algorithm. Systems Computers in Japan, 31(8), 33-40.