An Empirical Study of Stressors on Athletes in Sports Competition Based on Data Mining Technology

Qiyi Du
Wuhan Institute of Physical Education, Wuhan 430079, China

Abstract

In recent years, the development of sports undertakings has become increasingly high, and the scale and number of sports events are increasing. With the continuous development of information technology, the training methods of athletes show a scientific and systematic trend, which makes the rivalry between athletes more and more intense in sports events. In sports competition, athletes inevitably will have some psychological or physiological stress, which leads to a decline in the motor skills of athletes and makes it difficult for them to give full play to their strength, causing a significant negative impact on competition results. Therefore, more and more relevant experts and scholars are engaged in the study of stressors on athletes in sports competition. In sports competition, athletes tend to be under stress from many aspects. How to find out the stressors, put forward solutions to the problems and reduce athletes' psychological and physical stress in sports competitions so as to improve athletes' competition results has become a key issue facing scientific research in sports. Therefore, based on the data mining technology, the author, in this paper, analyzes the stressors on athletes in sports competition, aiming to provide a reference for improving athletes' performance in competition.

Keywords: Data Mining Technology, Sports Competition, Stressor.

1. SUMMARY OF RESEARCH

1.1 Background

With the continuous development of sports undertakings, the increasing scale and number of sports events and formation of a relatively well-developed sports value chain have spurred China's economic development. At the same time, in sports competition, the stress on athletes is also increasing, which has a great impact on athletes' performance. How to avoid the stress caused by sports competition has become a key issue facing research in the field of sports. In general, athletes will be subject to both physical and psychological stress, of which the physical stress is mainly reflected in the insufficient pre-exercise warm-up or physical damage due to long-term over-training, etc., and the psychological stress mainly refers to athlete's anxiety, aggressive behavior, low satisfaction and even pressure from public opinion or interpersonal pressure. Relevant surveys show that due to the extremely challenging and competitive sports competition itself, the vast majority of athletes will be in a state of high pressure for a long time in the course of athletic competition or preparing for an athletic competition. This high pressure condition can result in a decrease in an athlete's motor skills, a significant increase in muscle tension, and a significant reduction in attention and motor skills. Therefore, before the athletic competition, in order to be able to get better results in the sports arena, we should learn about the psychological stress of athletes and take targeted solutions to actively adjust their mental state and reduce their stress.

1.2 Literature review

Psychological conditions are characterized by ambiguity and uncertainty. Nowadays, there are professional counselors who provide psychological counseling and diagnosis to athletes before every big game. However, most of the time, facing the psychological stress of athletes, they can only adopt a simple method of isolation and psychological counseling, which is hard to solve the real psychological problems of athletes. In super large games, there are more athletes, and it is more difficult for psychological counselors to meet the needs of each athlete, which makes it hard to provide personalized psychological counseling services to athletes (Lou and Liu, 2016). The traditional psychological counseling for athletes is mainly carried out through communication between athletes and professional counselors and the measurement of psychological stress of athletes by means of
questionnaires and other. This approach can play a certain role but is relatively low in efficiency, and it is difficult to play a constructive role. Athletes still have a lot of psychological stress after consultation and it is hard to get rid of the stress. At present, in China, the General Administration of Sport has developed a system for management of competition stress on athletes. By using data mining technology, we can better find out the psychological stress existing in athletes and put forward solutions for them. It can effectively relieve the psychological stress of athletes and improve the performance of athletic competition (Wu and Wang, 2016). In the measurement of stressors of athletes in sports competition, the K-Means clustering algorithm has a better performance compared with other algorithms. By using the K-Means clustering algorithm to carry out a clustering analysis of sports competition stressor data, combining with the main features of the content mining algorithm, a user preference model can be established. This model can be matched with the sports competition stress coping strategy text library to get the final matching results. And the appropriate strategy recommendations given to athletes according to the matching results, together with the psychological counseling given by mental health staff, allows athletes to get better psychological counseling and suggestions, so as to enhance their mental health (Hu and Xu, 2013).

2. OVERVIEW OF THEORY OF ALGORITHMS

2.1 Overview of clustering analysis algorithm

Data mining is a process of extracting valuable or hidden information from massive data. There are many algorithms for data mining, and clustering analysis algorithm is one of the most important among them (Meng et al., 2013). The core concept of clustering analysis algorithm is to classify a large number of information data into different categories to obtain more information. The clustering analysis algorithm mainly includes the processes of raw data preparation, feature selection and extraction, approximation measurement, clustering or grouping, and clustering results assessment as shown in Figure 1.

![Figure 1. The Main Flow of Clustering Analysis Algorithm](image)

The first process is raw data preparation. The data mainly come from data of stressors on athletes in sports competition. All the data needs to be processed effectively and then standardized and dimensioned according to their characteristics (Yang and Zhang, 2014). In addition, the uncontrollable factors that have nothing to do with the stress of athletes should be removed.

Followed by the feature extraction, which is to extract a number of the most effective clustering feature subsets from the feature set, so as to form a new data set. Feature extraction is essentially a method of converting one or several kinds of data in massive data.

The second process is approximation measurement, which is to use distance functions to test the similarity of different data and datasets (Li, 2014). Those with high similarity are the data having similar features;

The third process is clustering or grouping. In this process, the clustering algorithm is mainly used, and the clustering algorithm has diverse methods - one is hard clustering, that is, to give a clear partitioning result, and another is fuzzy clustering, that is, for each data, given its membership of the cluster;

And the last process is clustering results assessment - in order to ensure the scientific and effective clustering results, the results obtained need to be assessed. There are many ways in which clustering results can be assessed. Some methods mainly measure the goodness of fit of clustering data, while others determine the degree of matching between clusters and benchmarks (Jiang and Qiu, 2014). The main methods are object matching and related test assessment methods.
In clustering analysis algorithms, it is difficult to classify the large number of algorithms involved, mainly because different classification methods can be generated according to different classification ways, each method of which has one or more categories of features (Li et al., 2013). Based on the basic idea of clustering analysis algorithm, clustering analysis algorithm can be divided into hierarchical clustering, partition-based clustering, clustering based on density and grid and other clustering, as shown in Figure 2:

![Clustering Algorithm Classification Sketch Map](image)

**Figure 2. Clustering Algorithm Classification Sketch Map**

### 2.2 Hierarchical clustering algorithm

Hierarchical clustering algorithm is a kind of classification algorithm to present the contents in the form of dendrogram, which has many fields of application, mainly in biosystematics and evolutionary theory, etc. As for the hierarchical clustering algorithm, there are two kinds of clustering process: one is to combine several small clusters into a large one, which is called agglomerative algorithm, and the other is to split a large cluster into small clusters, which is called partitioning algorithm (Li et al., 2013). No matter what kind of method is used, the core objective of hierarchical clustering algorithm is to change the obtained information data into a dendrogram, so as to obtain the clustering of relevant hierarchies. The steps of hierarchical clustering algorithm are as follows:

**Step 1**
Take each data object as a separate category, calculate the distance between different data objects and get them arranged according to the distance from small to large;

**Step 2**
Merge the results obtained after the ranking according to different distances. The closer the distance, the higher the similarity between the two categories of data and vice versa;

**Step 3**
While sorting out the distance between the existing data, calculate the distance between the new categories generated according to Step 2, which still forms a list of distances between categories; repeat the above steps until all data object connections become a category;

**Step 4**
Through the above steps, the results can be constructed into a hierarchical nesting nested dendrogram, and each hierarchy can form an abstract view of data.

In the clustering process, hierarchical clustering algorithm mainly describes the similarity between two categories of data and the connection rules between clusters. The calculation method of similarity between different data mainly adopts the Euclidean distance, and the link rules mainly include the following points:

**First**, the single linkage aggregation rule, that is, the shortest distance. The formula is as follows:

$$d(o_i, o_k) = \min_{x \in o, y \in o} \| x - y \|$$ (1)

**Second**, the complete linkage rules, that is, the longest distance. The formula is as follows:

$$d(o_i, o_k) = \max_{x \in o, y \in o} \| x - y \|$$ (2)
Third, the average linkage aggregation between clusters, that is, the average linkage distance between clusters. The formula is as follows:

\[ d(o_i, o_k) = \left( \frac{1}{n_i n_k} \sum_{x \in o_i} \sum_{y \in o_k} \|x - y\| \right) \]

(3)

Fourth, the average linkage aggregation within a cluster, that is, the average linkage distance between a cluster. The formula is as follows:

\[ d(o_i, o_k) = \left( \frac{1}{n_i n_k} \sum_{x \in o_i} \sum_{y \in o_k} \|x - y\| \right) \]

(4)

Fifth, Ward's criterion, that is, the sum of squares of deviations. The formula is as follows:

\[ d(o_i, o_k) = (1/C(n_i + n_k, 2)) \sum_{x,y \in (o_i, o_k)} \|x - y\| \]

(5)

3. A STUDY OF STRESSORS ON ATHLETES IN SPORTS COMPETITION BASED ON DATA MINING TECHNOLOGY

3.1 Questionnaire and object selection

In order to get a better understanding of the stressors on athletes in sports competition and to segment the stressors and put forward solutions in a targeted manner, in this study, the author mainly investigates various indicators of athletes in the course of athletic competition by means of questionnaire. As for the form of questionnaire, the author adopts the Organizational Stressor Indicator for Sport Performers (OSI-SP) put forward by Richard Arnold and his team from the University of Bath, UK. The questionnaire consists of 23 assessment indicators and provides a comprehensive assessment of stressors on athletes (Zhang, 2013).

And the objects of study are 200 students from an institute of physical education. The grade distribution of students is shown in Figure 3:

![Figure 3. Grade Distribution of Students](image)

Among them, 72 are freshmen, accounting for 36% of the total, 66 are sophomores, accounting for 33%, 42 are juniors, accounting for 21%, and 20 are seniors, accounting for 10%. The questionnaire is divided into five assessment levels, which are completely disagree, partially disagree, basically agree, mostly agree and completely agree, and the corresponding scores are 1-5 points (Yu, 2015).

3.2 Questionnaire on competition stressors

The main content of the questionnaire is divided into five parts, the first part is the goal and development of athletes, which mainly reflects the personal career development plan of athletes; the second part is athletes' rear
services and practice, which mainly refers to the level of team management and equipment, training ground and the rear services arrangements before and after sports competition; the third part is the team of athletes, which means the team's atmosphere, interpersonal relationships, etc.; the fourth part is the relationship between athletes and coaches, which mainly reflects the ability of coaches and relationships between the two; the fifth part is the selection of athletes, namely the influence of the team's personnel selection on athletes (Li and Liu, 2010). The details are shown in Table 1.

Table 1 Competition Stresor Questionnaire

<table>
<thead>
<tr>
<th></th>
<th>Responsibility for the team</th>
<th>6</th>
<th>Teammate attitude</th>
<th>11</th>
<th>Organization arrangement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Relationship with coach</td>
<td>7</td>
<td>Age</td>
<td>12</td>
<td>Injured</td>
</tr>
<tr>
<td>2</td>
<td>Rules and regulations</td>
<td>8</td>
<td>Interpersonal relationship</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Coach character</td>
<td>9</td>
<td>Selection of competition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Accommodation</td>
<td>10</td>
<td>Training arrangement</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.3 Questionnaire on psychological resilience

In sports competition, the questionnaire on athletes' psychological resilience mainly includes the following dimensions:

The first dimension is self-confidence, which means whether an athlete has the confidence to play and win the competition. The second dimension is appraisal, which means that whether an athlete can face the difficulties in special conditions or when undertaking the corresponding tasks and whether there will be a problem of inability to concentrate. The third dimension is the sense of control, that is, in the face of losing the game, whether an athlete can make an objective analysis of the problems of his own, and whether there will be a bad mood, such as irritability, etc. (Yao et al., 2013). The details are shown in Table 2:

Table 2 Resilience Inventory

<table>
<thead>
<tr>
<th></th>
<th>Self-recovery</th>
<th>6</th>
<th>Compression</th>
<th>11</th>
<th>Have particularity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Worry</td>
<td>7</td>
<td>Prone to depression</td>
<td>12</td>
<td>Self-challenge</td>
</tr>
<tr>
<td>2</td>
<td>Accomplish tasks</td>
<td>8</td>
<td>Choose to give up</td>
<td>13</td>
<td>Competitive spirit</td>
</tr>
<tr>
<td>3</td>
<td>Self-doubt</td>
<td>9</td>
<td>Prone to anxiety</td>
<td>14</td>
<td>Confident enough</td>
</tr>
<tr>
<td>4</td>
<td>Lack of confidence</td>
<td>10</td>
<td>Unable to concentrate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.4 Questionnaire on athletes' input

In this questionnaire, it mainly investigates athletes' input in the sports competition. It mainly contains the following dimensions:

The first dimension is self-confidence, which means that whether an athlete has enough confidence in reaching his sports goals, and whether he believes that the effort is proportional to the reward. The second dimension is vitality, which is mainly reflected in whether an athlete has enough passion during training and competition. The third dimension is risk, which means that whether an athlete is willing to dedicate himself to training and constantly outdo himself. The fourth dimension is enthusiasm, which means that whether an athlete really loves sports and feels happy during training and competition (Wang and Qi, 2011). The details are shown in Table 3:

Table 3 Athlete Input Questionnaire

<table>
<thead>
<tr>
<th></th>
<th>Confident enough</th>
<th>7</th>
<th>Achieve goals</th>
<th>13</th>
<th>Sufficient capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Passionate</td>
<td>8</td>
<td>Passion surging</td>
<td>14</td>
<td>wide awake</td>
</tr>
<tr>
<td>2</td>
<td>Achieve goals</td>
<td>9</td>
<td>Technical qualification</td>
<td>15</td>
<td>Willing to work hard</td>
</tr>
<tr>
<td>3</td>
<td>Excitement</td>
<td>10</td>
<td>Sufficient vitality</td>
<td>16</td>
<td>Find fun</td>
</tr>
<tr>
<td>4</td>
<td>Can succeed</td>
<td>11</td>
<td>Devote oneself wholeheartedly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Be full of go</td>
<td>12</td>
<td>Enjoy sports</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.5 Questionnaire on burnout of athletes

Occupational burnout is a problem prone to occur in all occupations. The questionnaire on burnout of athletes is mainly used to investigate the existence of occupational burnout in athletes, which mainly includes the following dimensions:

The first dimension is emotional or physical exhaustion, which is mainly reflected whether there is depression or physical exhaustion in sports competition. The second dimension is decline in achievement, which is mainly reflected in whether an athlete is dissatisfied with the results or has a sense of accomplishment or not. The third dimension is sports depreciation, which means that whether an athlete feels no value for his efforts (Chen, 2013). The details are shown in Table 4:

<table>
<thead>
<tr>
<th></th>
<th>Do good</th>
<th>6</th>
<th>Don't care about grades</th>
<th>11</th>
<th>Not paying attention to winning or losing</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Strong sense of tiredness</td>
<td>7</td>
<td>Can not play the ability</td>
<td>12</td>
<td>Exhaustion of body and mind</td>
</tr>
<tr>
<td>3</td>
<td>Want to do something else?</td>
<td>8</td>
<td>It's going to crash</td>
<td>13</td>
<td>Conflicting emotions</td>
</tr>
<tr>
<td>4</td>
<td>Extreme fatigue</td>
<td>9</td>
<td>Unable to concentrate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>No good grades</td>
<td>10</td>
<td>Tired</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. CONCLUSIONS

In this article, the author studies stressors on athletes in sports competition mainly by using data mining technology. It shows that the traditional method for stressor assessment has some defects, and the psychological counseling used is characterized by subjectivity and blindness. It relies too much on the experience of psychological counseling experts and does not use in-depth information technology. Clustering and recommendation algorithms based on data mining technology can effectively solve the drawbacks of the traditional model and provide an important reference for psychological counseling experts to develop counseling plans. Besides, the use of OSI-SP put forward by Richard Arnold and his team is of great significance for assessing the stress of athletes during sports competition.

REFERENCES

Li J. (2014). College basketball players' different stressors, coping efficacy and coping styles, Journal of Wuhan Sports Institute, 48 (10), 82-86.
Wang L.D., Qi L. (2011). Comparison of EMG and plantar pressure between jump players in different levels of basketball jump players, Sports science, 32 (05), 67-70.