Correlation-Aware Sport Training Evaluation for Players With Trust Based on Mahalanobis Distance

TENGFEI FAN,1 SHENGLI TIAN,2 ZHICHEN HU,3 XINTONG FAN,4 AND SIFENG WANG4

1Department of Physical Education, Qufu Normal University, Rizhao 276826, China
21st Middle School of Juye County, Heze 274999, China
3School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China
4School of Computer Science, Qufu Normal University, Rizhao 276826, China

Corresponding author: Tengfei Fan (18678703694@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 42050102.

ABSTRACT With the widely-adopted idea of health and longevity, sports have become one of the most popular entertainment ways of the public. For the majority of sports, players need to know about their concrete physical conditions in a real time manner so as to pursue a good sport score or ranking in a competition or a race. Generally, we can achieve the above goal through analyzing and evaluating the daily training scores of each player. However, there are often multiple physical trainings for players and various correlations exist among them, which significantly decrease the fairness and trust of player training score evaluation and ranking since traditional multi-dimensional data integration solutions are often based on a strong hypothesis, i.e., the involved multiple dimensions are independent with each other. In view of this shortcoming, we introduce the Mahalanobis Distance into the multi-dimensional player training score evaluation and further propose a correlation-aware player training score evaluation method with trust (abbreviated as CPE MD) based on Mahalanobis Distance. As Mahalanobis Distance can eliminate the hidden linear correlations among the involved multiple dimensions, we can guarantee the fairness and trust of Mahalanobis Distance-based player training score evaluation and ranking results. At last, we use a case study to show the feasibility of CPE MD in this paper.

INDEX TERMS Sports, trust, player training score evaluation, multiple dimensions, dimension correlation, Mahalanobis distance.

I. INTRODUCTION

With the continuous progress of social reform and economic developments, people’s living conditions have gained considerable improvements [1]–[3]. In this situation, more people are paying their attentions to the improvement of life health quality. With the widely-adopted idea of health and longevity, sports have become one of the most popular entertainment ways of the public so as to ensure the high-quality physical conditions in daily life [5]–[7]. As a result, more and more people are engaging in various sport activities and have gained another identity or role, i.e., the so-called players.

For the majority of sports, players need to know about their concrete physical conditions in a real time manner so as to pursue a good sport score or ranking in a competition or a race or an exercise. Generally, we can achieve the above goal through analyzing and evaluating the daily training scores of each player. However, there are often multiple physical trainings for players, such as football, running, swimming, volleyball, tennis ball and so on. To get a comprehensive sport score of a player, we often need to calculate the sum of the multiple scores corresponding to the multiple sport items. Thus, through the derived comprehensive score of each player, we can evaluate the practical physical conditions of players and rank them accordingly if necessary.

However, the above comprehensive score calculation solutions are often confronted with a big challenge, i.e., they often assume that the involved multiple sport training scores are independent with each other. While actually, for different sport items, their training scores are often correlated with each other [8]–[11]. For example, a player who swims well...
often has high lung vital capacity; and therefore, he or she is prone to get a higher training score when taking the running sport. Likewise, a taller man often player better than a shorter man in terms of volleyball and basketball due to the body height advantages of the former man. Therefore, in this situation, it is inappropriate to evaluate the comprehensive score of a player by calculating the sum of his or her engaged multiple sport items directly, since the hidden correlations among different sport training scores may decrease the fairness and trust.

Considering this challenging issue, we introduce Mahalanobis Distance into the multi-dimensional player training score evaluation problem and further propose a correlation-aware player training score evaluation method with trust (abbreviated as CPE\(_{MD}\)) based on Mahalanobis Distance. As Mahalanobis Distance can eliminate the hidden linear correlations among the multiple dimensions involved in the player training score evaluation process, we can guarantee the fairness and trust of Mahalanobis Distance-based player training score evaluation and ranking results output by CPE\(_{MD}\) method.

In summary, the contribution of this research paper is three-aspect.

1. We study the hidden correlations among the multiple training scores of players and regard it as a key factor that influences the objectiveness, fairness and trust of the player sport physique evaluation.
2. We introduce the Mahalanobis Distance into the multi-dimensional player training score evaluation problem and further propose a correlation-aware player training score evaluation method with trust, i.e., CPE\(_{MD}\).
3. A case study and a set of experiments are provided to demonstrate the concrete processes of CPE\(_{MD}\) method, through which we show the feasibility of our proposal in eliminating the linear correlation relationships among the multiple sport training scores of players.

We organize the rest structure of our paper as follows. Related work is presented in Section 2 to introduce the current research status of the field. In Section 3, a motivating example is shown to better describe the research significance and importance of our paper. Concrete process or procedure of our proposed CPE\(_{MD}\) method is clarified in Section 4 with more details. A case study constructed from real-world applications is presented in Section 5 and experiment evaluation is made in Section 6. At last, in Section 7, we simply conclude the research paper and point out the possible investigation directions or topics in the upcoming research work.

II. RELATED WORK

Sport training score evaluation is an essential aspect to objectively observe the health conditions or sport physique of players. A big volume of researchers has paid their attentions on the topics and resolutions associated with player evaluation. In [12], the authors analyze the multiple factors or aspects that could be recruited to objectively evaluate a player in football, no matter grassroot players or high-level players. The mentioned factors do not include the traditional factors at club levels, such as player transfer, squad formation and strategic planning. Therefore, it is more probable that excellent grassroot players could be evaluated and mined from massive candidates. This evaluation strategy is beneficial to both candidate grassroot players and team manager. In [13], the authors introduce psychophysiology into the player quality or level evaluation. In concrete, the authors assess the sport level of each player through collecting the answers of various physiological questions designed by the authors beforehand. Thus, through the obtained “question-answer” pairs, we can approximately evaluate the category of a player especially when he or she is confronted with different challenges or difficulties. Loyalty is an important aspect for evaluating a player in various sport items. Motivated by this hypothesis, the authors in [14] investigate the loyalty status of golf players in current Vietnam. In their provided study and statistical analysis, a player’s perceived value and satisfaction have played key roles in keeping the loyalty of the player. As a contrast, the service quality of golf course is not as important as other influencing factors, which is an interesting phenomenon that is discovered in the novel research work.

In [15], different guidance or strategies are provisioned by the authors for different roles involved in a basketball team or club. On one hand, they offer quantified models or strategies for guaranteeing the healthy running or operations of clubs or teams. On the other hand, they design a set of effective metrics or evaluation tools tailored to the characteristics of basketball item, through which the player conditions or types can be depicted and profiled conveniently. Likewise, in [16], the authors develop an algorithm to evaluate and pick out the most appropriate players for a match or a competition. In concrete, they propose a sport level evaluation framework for players, through which the real-time competitive conditions of a player can be monitored and evaluated in a quantified way. The provided framework includes a variety of evaluation algorithms that are tailored to the sport characteristics or features of cricket item. In addition, from the perspective of clubs or teams, the high trust between players, clubs and spectators are crucial to maintain the healthy collaborations among them. Inspired by this observation, the authors in [17] study the trust relationship among different stakeholders in a sport team or club. To achieve this goal, blockchain technology is recruited to guarantee the dependability of released player transfer information by the club, because fake player transfer news or information probably decreases the trust or satisfactions of spectators towards involved clubs. In this way, sport clubs can advertise their player transfer fee information that is sensitive enough to their spectators without revealing partial sensitive club privacy.

Recently, various machine learning techniques are brought forth and have been widely applied in various business domains. A typical application domain is sport business or industry. In [18], machine learning techniques are employed to predict a considerable number of issues associated with sport clubs, such as player sport conditions, player injury
conditions, club economy running conditions, club competition status, and so on. In addition, the paper also studies the long-term prediction of players or clubs. For example, whether the club competition status this year is better than last year? what is the concrete ranking of a club in the next season? who will win the next season in a football league? Such predictions have provided a good basis for the healthy running of a football club because it can enact appropriate running strategy according to the predicted results. Similar work is done in [19] where the authors also use machine learning techniques to predict the various scores or performances associated with football clubs. A featured point in [19] is that it offers a set of evaluation metrics that are especially suitable for the football players and football clubs as well as their involved football matches. In addition, the authors also investigate the features or patterns that discriminate the top players from the other players who are not qualified enough. Traditional player evaluation mechanisms are often based on statistical techniques in mathematics, which is prone to produce delayed statistical results and overlook other valuable information hidden in non-statistical data records. In view of this shortcoming, a novel player evaluation method is proposed in [20] which is mainly based on the deep learning technology. In concrete, the authors use the statistical data in matches to quantify the player performances while use the news articles after matches to evaluate the player performances qualitatively. Thus, through combining these quantified and qualitative evaluation results, we can objectively evaluate the overall performances of a player comprehensively. Similar work is done in [21] where the Convolutional Neural Network (CNN) technique is used to achieve the same player evaluation goal as in [20].

With the above analyses, we can find that although many researchers have paid their attentions on player performance or score evaluation, they seldom consider the possible inner correlations among the involved multiple dimensions in evaluation. As a consequence, the finally derived player evaluation results are probably inaccurate and not trusted enough, which do harm to the health conditions or performance evaluation of players. Considering this drawback, we introduce the Mahalanobis Distance into the multi-dimensional player training score evaluation problem and further propose a correlation-aware player training score evaluation method with trust, i.e., CPE$_{MD}$ based on Mahalanobis Distance. As Mahalanobis Distance can eliminate the hidden linear correlations among the multiple dimensions involved in the player training score evaluation process, we can guarantee the fairness and trust of Mahalanobis Distance-based player training score evaluation and ranking results outputted by CPE$_{MD}$ method. Details of CPE$_{MD}$ method will be described briefly in the following sections.

### III. MOTIVATION

A motivating example is presented in Fig.1 to further describe the research significance of our paper. As the example shows, a player Tom engaged in four sports: running, biking, weight lifting and boating. As the sports of running and biking both require players to own high leg strength and endurance, Tom’s score on running and biking is often correlated with each other. Likewise, the sports of weight lifting and boating both require players to own high arm strength and endurance, Tom’s scores on weight lifting and boating are also correlated with each other. In summary, different correlations are existed in the four sports engaged by Tom. In this situation, simple addition of the four sport scores will lead to an inaccurate and untrusted evaluation result for Tom, which challenges the objective and fair player evaluation.

In view of this challenge, we introduce Mahalanobis Distance into the multi-dimensional player training score evaluation problem and further propose a correlation-aware player training score evaluation method with trust, i.e., CPE$_{MD}$ based on Mahalanobis Distance. As Mahalanobis Distance can eliminate the hidden linear correlations among the multiple dimensions involved in the player training score evaluation process, we can guarantee the fairness and trust of Mahalanobis Distance-based player training score evaluation results outputted by CPE$_{MD}$ method. Concrete steps of CPE$_{MD}$ will be clarified in the following sections.

### IV. SOLUTION: CPE$_{MD}$

In this section, the novel player training score evaluation method CPE$_{MD}$ will be introduced in detail, to minimize or reduce the possibly existed linear correlations among multiple training scores of players and thereby enhance the trust of the final player evaluation results. The proposed idea is mainly based on the Mahalanobis Distance which can introduce the covariance information among different training scores into the evaluation process while covariance depicts
the correlations among different dimensions well. In general, CPE\textsubscript{MD} consists of the five steps in Fig.2.

**Step 1: Calculate average value vector \( \mu \) of sport items.** According to \( n \) players’ training scores over \( m \) sport items, we calculate the average value vector \( \mu \) corresponding to the \( m \) sport items.

**Step 2: Calculate covariance matrix \( \text{Cov} \) of sport items.** According to the players-sport score matrix and the average value vector \( \mu \), we calculate the covariance matrix \( \text{Cov} \) corresponding to the \( m \) sport items.

**Step 3: Calculate Mahalanobis Distance of each player.** According to average value vector \( \mu \), covariance matrix \( \text{Cov} \) and each player’s training score, we calculate the Mahalanobis Distance of each player.

**Step 4: Player evaluation and ranking.** According to the Mahalanobis Distance of each player, we evaluate and rank the players in descending order.

**FIGURE 2.** Procedure of CPE\textsubscript{MD}.

**Step 1: Calculate average value vector \( \mu \) of sport items.**

We assume that there are \( n \) players, i.e., \( p_1, \ldots, p_n \) and \( m \) sport items, i.e., \( sp_1, \ldots, sp_m \). According to the \( n \) players’ training scores over \( m \) sport items, we can get a player-sport training score matrix \( M \) which is an \( n \times m \) matrix as presented in (1). In matrix \( M \), \( s_{i,j} \) denotes the training score of player \( p_i \) over sport item \( sp_j \).

\[
M = \begin{bmatrix}
sp_1 & \cdots & sp_m \\
s_{1,1} & \cdots & s_{1,m} \\
\vdots & \ddots & \vdots \\
p_n & \cdots & s_{n,m}
\end{bmatrix} \tag{1}
\]

Next, we calculate the average value vector \( \mu \) corresponding to the \( m \) sport items. Concretely, for each column \( sp_j \) (\( 1 \leq j \leq m \)), we calculate its average value \( av_j \) based on the \( n \) training scores of \( sp_j \), i.e., \( s_{1,j}, \ldots, s_{n,j} \). In other words, equation (2) holds. Then with \( av_j \) (\( 1 \leq j \leq m \)), we can get the average value vector \( \mu \) as in (3).

\[
av_j = \frac{\sum_{i=1}^{n} s_{i,j}}{n} \tag{2}
\]

\[
\mu = (av_1, av_2, \ldots, av_m) \tag{3}
\]

**Step 2: Calculate covariance matrix \( \text{Cov} \) of sport items.**

In this step, we calculate the covariance matrix \( \text{Cov} \) of the \( m \) sport items: \( sp_1, \ldots, sp_m \). In concrete, according to the player-sport training score matrix \( M \) and the average value vector \( \mu \), we can obtain covariance matrix \( \text{Cov} \) as follows.

First, for matrix \( M \), we use the average value vector \( \mu \) \((av_1, av_2, \ldots, av_m)\) to realize the data regularization. Concretely, for each entry \( s_{i,j} \) of matrix \( M \), we execute the operation in equation (4) and then get a new entry value \( s_{i,j}^{\#} \). Afterwards, we get a new matrix \( M^{\#} \) as in (5).

\[
s_{i,j}^{\#} = s_{i,j} - av_j \tag{4}
\]

\[
M^{\#} = \begin{bmatrix}
sp_1 & \cdots & sp_m \\
\vdots & \ddots & \vdots \\
p_n & \cdots & s_{n,m}
\end{bmatrix} \tag{5}
\]

Then we can calculate the covariance matrix \( \text{Cov} \) of the \( m \) sport items as in (6) where \( M^{\#T} \) is the transposed matrix of \( M^{\#} \). After that, we can get a \( m \times m \) matrix \( \text{Cov} \). Since \( M^{\#} \) is an \( n \times m \) matrix, \( \text{Cov} \) is an \( n \times m \) matrix. According to Mahalanobis Distance theory, matrix \( \text{Cov} \) reflects the linear correlation information among the \( m \) sport training scores. For example, in Fig.1, as the sports of running and biking both require players to own high arm strength and endurance, Tom’s scores on the running and biking are often correlated with each other. Likewise, the sports of weight lifting and boating both require players to own high leg strength and endurance, Tom’s scores on weight lifting and boating are also correlated with each other. Thus, we can consider to reduce the linear correlation information among different dimensions by using the covariance matrix \( \text{Cov} \). Here, if no correlations are present among different dimensions, the entries in covariance matrix \( \text{Cov} \) would be equal to zero.

\[
\text{Cov} = \frac{1}{m-1} * M^{\#T} * M^{\#} \tag{6}
\]

**Step 3: Calculate Mahalanobis Distance of each player.**

According to matrix \( M \) in (1), we can get a player training score vector \( \tilde{p}_i \) for each player \( p_i \) (\( 1 \leq i \leq n \)). For example, \( \tilde{p}_1 = (s_{1,1}, s_{1,2}, \ldots, s_{1,m}) \). Then through the covariance matrix \( \text{Cov} \), we can calculate the Mahalanobis Distance of each player \( p_i \) (\( 1 \leq i \leq n \)) based on the equation in (7). Here, \( D_M(p_i) \) denotes the Mahalanobis Distance of player \( p_i \), \( \tilde{p}_i^T \) is the transposed vector of \( \tilde{p}_i \), covariance matrix \( \text{Cov} \) is available from (6) and \( \text{Cov}^{-1} \) represents the inverse matrix of matrix \( \text{Cov} \).

\[
D_M(p_i) = \sqrt{\tilde{p}_i^T * \text{Cov}^{-1} * \tilde{p}_i} \tag{7}
\]

Thus, through the calculation operation in equation (7), we can obtain the Mahalanobis Distance \( D_M(p_i) \) of player \( p_i \) (\( 1 \leq i \leq n \)). Actually, \( D_M(p_i) \) measure the Mahalanobis Distance between each player’s training score vector \( \tilde{p}_i \) and the original point 0. As covariance matrix \( \text{Cov} \) is considered when calculating the Mahalanobis Distance \( D_M(p_i) \), the linear correlations among the training scores of \( m \) sport items (i.e., \( sp_1, \ldots, sp_m \)) of \( p_i \) could be alleviated or minimized as much as possible. As a consequence, the obtained Mahalanobis Distance \( D_M(p_i) \) for player \( p_i \) could be regarded as an objective measurement or evaluation metric of \( p_i \)’s comprehensive training score even if multiple training scores are available for \( p_i \).
Step 4: Player evaluation and ranking.
In Step 3, we have obtained the Mahalanobis Distance $D_M(p_i)$ for player $p_i$ ($1 \leq i \leq n$). According to the Mahalanobis Distance theory, the larger $D_M(p_i)$ is, the farther that $p_i$ is away from the original point 0. As original point 0 is an ideal reference point for evaluating all the players, we hope that the Mahalanobis Distance $D_M(p_i)$ “the larger, the better”.

With the above analyses, we can evaluate and rank the training scores of all $n$ players $p_1, \ldots, p_n$ based on the Mahalanobis Distance $D_M(p_i)$ (“the larger, the better”).

Algorithm 1 CPE$_{MD}$

```
Inputs:
(1) $p_1, \ldots, p_n$: players
(2) $sp_1, \ldots, sp_m$: sport items
(3) $M$: player-sport score matrix

Output:
(1) $\text{Ranked}\_\text{list}(p_1, \ldots, p_n)$

1 For $j = 1$ to $m$
2 \hspace{1em} sum$_j = 0$
3 \hspace{1em} For $i = 1$ to $n$
4 \hspace{2em} sum$_j = sum_j + s_{i,j}$
5 \hspace{1em} End For
6 \hspace{1em} av$_j = sum_j/n$
7 End For
8 $\mu = (av_1, av_2, \ldots, av_m)$
9 For $j = 1$ to $m$
10 \hspace{1em} For $i = 1$ to $n$
11 \hspace{2em} $s_{i,j} = s_{i,j} - av_j$
12 \hspace{1em} End For
13 End For
14 Calculate $M^\#$ by (5)
15 Calculate Cov by (6)
16 For $i = 1$ to $n$
17 \hspace{1em} $\tilde{p}_i = (s_{i,1}, s_{i,2}, \ldots, s_{i,m})$
18 \hspace{1em} Calculate $D_M(p_i)$ by (7)
19 End For
20 For $i = 1$ to $n$
21 \hspace{1em} Rank $p_i$ by $D_M(p_i)$ in descending order
22 \hspace{1em} Put $p_i$ in $\text{Ranked}\_\text{list}(p_1, \ldots, p_n)$
23 End For
24 Return $\text{Ranked}\_\text{list}(p_1, \ldots, p_n)$
```

Formally, the pseudo code of the proposed CPE$_{MD}$ method is described in Algorithm 1.

V. CASE STUDY
To further introduce the main idea of our proposed CPE$_{MD}$ method, a case study is presented in this section to demonstrate the concrete steps introduced in Section IV. In the case study, there are four players (i.e., $p_1$, $p_2$, $p_3$, $p_4$) and two sport items (i.e., $sp_1$, $sp_2$) as well as the resulted $4 \times 2$ matrix $M$ as follows.

\[
M = \begin{bmatrix}
    sp_1 & sp_2 \\
    p_1 & 3 \\ p_2 & 5 \\ p_3 & 2 \\ p_4 & 8 \\
\end{bmatrix}
\]

Step 1: Calculate average value vector $\mu$ of sport items.
With the player-sport matrix $M$, for the two sport items (i.e., $sp_1$, $sp_2$), we calculate the average values of $sp_1$ and $sp_2$, respectively. The results are: $av_1 = 4.5$, $av_2 = 4$. Therefore, according to the equation (3), the average value vector $\mu = (4.5, 4)$.

Step 2: Calculate covariance matrix Cov of sport items.
According to the player-sport matrix $M$ and the average value vector $\mu$ (derived in Step 1, i.e., $\mu = (4.5, 4)$), we can calculate the new matrix $M^\#$ based on equations in (4)-(5), as follows.

\[
M^\# = \begin{bmatrix}
    sp_1 & sp_2 \\
    p_1 & -1.5 & 0 \\
    p_2 & 0.5 & 2 \\
    p_3 & -2.5 & -2 \\
    p_4 & 4 & 0 \\
\end{bmatrix}
\]

Next, according to derived $M^\#$ and equation (6), we can calculate the covariance matrix Cov as follows. As there are only two sport items $sp_1$ and $sp_2$, the covariance matrix Cov is a $2 \times 2$ matrix. Furthermore, we can calculate the inverse matrix of Cov, i.e., $Cov^{-1}$ as below.

\[
Cov = \begin{bmatrix}
    7 & 2 \\
    2 & 2.667 \\
\end{bmatrix}
\]

\[
Cov^{-1} = \begin{bmatrix}
    0.18 & -0.13 \\
    -0.13 & 0.48 \\
\end{bmatrix}
\]

Step 3: Calculate Mahalanobis Distance of each player.
In this step, according to the covariance matrix Cov as well as its inverse matrix $Cov^{-1}$ (derived in Step 2) and the equation in (7), we can calculate the Mahalanobis Distance of each of the four players, i.e., $p_1$, $p_2$, $p_3$ and $p_4$. In concrete, the four players’ Mahalanobis Distances from the original point 0 are listed as below.

\[
D_M(p_1) = 0.6364 \\
D_M(p_2) = 1.3058 \\
D_M(p_3) = 1.3209 \\
D_M(p_4) = 1.6971
\]

Step 4: Player evaluation and ranking.
According to the Mahalanobis Distances the four players derived in Step 3, i.e., $D_M(p_1)$, $D_M(p_2)$, $D_M(p_3)$ and $D_M(p_4)$, we can rank the four players (i.e., $p_1$, $p_2$, $p_3$, $p_4$) in descending order as follows. In other words, the training score of player $p_4$ is the best while the training score of player $p_1$ is the worst.

\[
p_4 > p_3 > p_2 > p_1
\]
VI. EVALUATIONS

To further show the innovation of the proposed CPE$_{MD}$ method in this paper, we test a set of experiments as follows. In concrete, the experiment dataset used for evaluations is WS-DREAM. Two methods are used for comparison purpose are: Euc-D (Euclidean Distance based evaluation) and Cos-D (Cosine Distance based evaluation). Experiments are all run in a desk computer with 8 GB memory and 2.4 GHz processor and the software configurations include Win 7 OS and Python 3.6. Experiments are executed 100 times and we report their final average values for evaluation. Concrete comparison results are shown as follows.

Test 1: Covariance comparison of three methods.

The advantage of CPE$_{MD}$ method is that it considers dimension correlation. Therefore, the covariance of CPE$_{MD}$ should be small. To validate this hypothesis, we compare the covariances of three methods with respect to the size of used records for each user in WS-DREAM. The results are reported in Fig.3 where Fig.3(a) records the covariances of response time in WS-DREAM and Fig.3(b) records the covariances of throughput in WS-DREAM.

It can be seen from Fig.3(a) that the covariances of three methods all increase with the growth of the number of records, which is due to the fact that: more records used for evaluation will bring larger accumulated covariance value. However, our CPE$_{MD}$ outperforms Ecu-D and Cos-D methods because CPE$_{MD}$ considers the correlation among different dimensions. Similar results could also be observed in Fig.3(b) whose results will not be analyzed repeatedly.

Test 2: Time cost of three methods.

In CPE$_{MD}$ method, the covariance matrix among different dimensions is used for objective evaluation, which will bring additional time cost compared to other methods. To validate this hypothesis, we compare the time costs of three methods with respect to the size of used records for each user in WS-DREAM and the size of users for each dimension in WS-DREAM. The results are reported in Fig.4 where Fig.4(a) records the time costs of three methods with respect to the size of records and Fig.4(b) records the time costs of three methods with respect to the size of users. Here, we only consider the response time dimension of WS-DREAM dataset.
It is seen from Fig. 4(a) that the time costs of three methods all increase with the growth of the number of records, which is due to the fact that more records used for evaluation will bring additional processing time. Moreover, our CPE\textsubscript{MD} does not perform better than Ecu-D and Cos-D methods because CPE\textsubscript{MD} considers the correlation among different dimensions. However, the time cost of CPE\textsubscript{MD} is close to those of Ecu-D and Cos-D. Similar results could also be observed in Fig. 4(b).

VII. CONCLUSION
Training score evaluation is a promising way to know about the physical conditions of a player. However, there are often multiple physical trainings for players and various correlations are existent among them, which significantly decreases the fairness and trust of player training score evaluation and ranking because traditional multi-dimensional data integration solutions are often based on a strong hypothesis, i.e., the involved multiple dimensions are independent with each other. In view of this shortcoming, we introduce Mahalanobis Distance into the multi-dimensional player training score evaluation and further propose a correlation-aware player training score evaluation method based on Mahalanobis Distance, i.e., CPE\textsubscript{MD}. Through Mahalanobis Distance, we can minimize the hidden linear correlations among multiple training scores of players, so as to guarantee the fairness and trust of player evaluation and ranking. Finally, a case study is presented to show the concrete procedure of CPE\textsubscript{MD}.

However, player evaluation is often a multi-dimensional decision-making problem while multi-dimensional decision-making problems often involves multiple-type data fusion, weight assignment as well as the optimization issue [22]–[29]. Therefore, we will further study the multi-source data integration problem with weighting. In addition, Mahalanobis Distance requires additional time cost to compute the covariance matrix of different dimensions; therefore, its time complexity is not very low. While time cost is critical for real world applications especially for the big data scenario [23], [30]–[35]. Therefore, we would continuously refine our algorithm to further reduce its time costs so as to meet the quick response requirements from users. At last, in our CPE\textsubscript{MD} method, user privacy is not considered since we more focus on correlation elimination. While in most big data applications, privacy is always a critical concern especially when users are sensitive to their personal data [36]–[41]. Considering this drawback, we will further investigate the current research work associated with privacy protection and refine CPE\textsubscript{MD} by introducing user privacy.

REFERENCES
[1] M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour, and E. M. Aggoune, “Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk,” IEEE Access, vol. 7, pp. 129551–129583, 2019.
[2] C.-H. Chen, “A cell probe-based method for vehicle speed estimation,” IEICE Trans. Fundam. Electron., Commun. Comput. Sci., vol. E103-A, no. 1, pp. 265–267, 2020.
[3] C.-H. Chen, F.-J. Hwang, and H.-Y. Kung, “Travel time prediction system based on data clustering for waste collection vehicles,” IEICE Trans. Inf. Syst., vol. E102-D, no. 7, pp. 1374–1383, 2019.
[4] V. Chamola, V. Hassija, V. Gupta, and M. Guizani, “A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact,” IEEE Access, vol. 8, PP 90225–90265, 2020.
[5] K. Xu, F. Wang, H. Wang, and B. Yang, “Detecting fake news over online social media via domain reputations and content understanding,” Tsinghua Sci. Technol., vol. 25, no. 1, pp. 20–27, Feb. 2020.
[6] Y. Liu, A. Pei, F. Wang, Y. Yang, X. Zhang, H. Dai, L. Qi, and R. Ma, “An attention-based category-aware GRU model for the next POI recommendation,” Int. J. Intell. Syst., vol. 36, no. 7, pp. 3174–3189, Jul. 2021, doi: 10.1002/int.22412.
[7] Z. Sun, Y. Wang, Z. Cai, T. Liu, X. Tong, and N. Jiang, “A two-stage privacy protection mechanism based on blockchain in mobile crowdsourcing,” Int. J. Intell. Syst., vol. 36, no. 5, pp. 2058–2080, May 2021, doi: 10.1002/int.22371.
[8] L. Qi, C. Hu, X. Zhang, M. R. Khosravi, S. Sharma, S. Pang, and T. Wang, “Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment,” IEEE Trans. Ind. Informat., vol. 17, no. 6, pp. 4159–4167, Jun. 2021, doi: 10.1109/TII.2020.3012157.
[9] R. Kumar, S. Kumar, R. C. Poonia, V. Singh, L. Raja, V. Bhatnagar, and P. Agrawal, “Analysis and predictions of spread, recovery, and death caused by COVID-19 in India,” Big Data Mining Anal., vol. 4, no. 2, pp. 65–75, Jun. 2021.
[10] Z. Cai and Z. He, “Trading private range counting over big IoT data,” in Proc. IEEE 39th Int. Conf. Distribut. Computat. Syst. (ICDCS), Jul. 2019, pp. 144–153, doi: 10.1109/ICDCS.2019.00023.
[11] G. Wu, X. Guo, and B. Xu, “BAM: A block-based Bayesian method for detecting genome-wide associations with multiple diseases,” Tsinghua Sci. Technol., vol. 25, no. 5, pp. 678–689, Oct. 2020.
[12] R. Pariath, S. Shah, A. Surve, and J. Mittal, “Player performance prediction in football game,” in Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA), Mar. 2018, pp. 1148–1153, doi: 10.1109/ICECA.2018.8474750.
[13] M. Klarkowski, “The psychophysiological evaluation of the player experience,” Ph.D. dissertation, Queensland Univ. Technol., Brisbane, QLD, Australia, 2017.
[14] N. T. K. Thanh, L.-A. Tuan, and N. D. Nguyen, “Evaluation of factors impacting the player loyalty in golf course business,” Manage. Sci. Lett., vol. 11, no. 1, pp. 155–164, 2021.
[15] Z. Terner and A. Franks, “Modeling player and team performance in basketball,” Annu. Rev. Statist. Appl., vol. 8, no. 1, pp. 1–23, Mar. 2021, doi: 10.1146/annurev-statistics-040720-015536.
[16] M. A. Uddin, M. H. Haldar, S. Ahmed, and U. K. Acharjee, “CRICRATE: A cricket match conduction and player evaluation framework,” in Emerging Technologies in Data Mining and Information Security (Advances in Intelligent Systems and Computing), vol. 755, 2018, pp. 491–500.
[17] C. Liu, Z. Li, S. Liu, J. Xie, C. Yan, and W. Huang, “Trusted player transfer evaluation for sport markets based on blockchain and locality-sensitive hashing,” IEEE Access, vol. 9, pp. 87332–87339, 2021.
[18] V. C. Pantzalis and C. Tjortjis, “Sports analytics for football league table and player performance prediction,” in Proc. 11th Int. Conf. Inf., Intell., Syst. Appl. (HISA), Jul. 2020, pp. 1–8, doi: 10.1109/HISA50023.2020.9284352.
[19] L. Pappalardo, P. Cintia, P. Ferragina, E. Massucco, D. Pedreschi, and F. Giannotti, “PlayeRank: Data-driven performance evaluation and player ranking in soccer via a machine learning approach,” ACM Trans. Intell. Syst. Technol., vol. 10, no. 5, pp. 1–27, Nov. 2019.
[20] Y. J. Park, H. S. Kim, D. Kim, H. Lee, S. B. Kim, and P. Kang, “A deep learning-based sports player evaluation model based on game statistics and news articles,” Knowl.-Based Syst., vol. 138, pp. 15–26, Dec. 2017.
[21] Y. Ban, “Software-defined information-centric networking based exercise intensity evaluation of volleyball player: An efficient convolutional neural network method,” Internet Technol. Lett., vol. 4, no. 5, p. e246, Sep. 2021, doi: 10.1002/itl2.246.
[22] L. Qi, Q. He, F. Chen, X. Zhang, W. Dou, and Q. Ni, “Data-driven web APIs recommendation for building web applications,” IEEE Trans. Big Data, early access, Feb. 24, 2020, doi: 10.1109/TBDATA.2020.2975587.
[23] Y. N. Malek, M. Najib, M. Bakhoyua, and M. Essaaidi, “Multivariate deep learning approach for electric vehicle speed forecasting,” Big Data Mining Anal., vol. 4, no. 1, pp. 56–64, Mar. 2021.
Y. Xu, J. Ren, Y. Zhang, C. Zhang, B. Shen, and Y. Zhang, “Blockchain for the COVID-19 pandemic: A federated learning solution,” Tsinghua Sci. Technol., vol. 26, no. 5, pp. 759–771, Oct. 2021.

Z. Lu, Y. Wang, Y. Li, X. Tong, C. Mu, and C. Yu, “Data-driven many-objective crowd worker selection for mobile crowdsourcing in industrial IoT,” IEEE Trans. Ind. Informat., early access, Apr. 30, 2021, doi: 10.1109/TII.2021.3076811.

Y. Zhang, X. Lei, Z. Fang, and Y. Pan, “CircRNA-disease associations prediction based on metapath2vec++ and matrix factorization,” Big Data Mining Anal., vol. 3, no. 4, pp. 280–291, Dec. 2020.

F. Wang, H. Zhu, G. Srivastava, S. Li, M. R. Khoshavi, and L. Qi, “Robust collaborative filtering recommendation with user-item-trust records,” IEEE Trans. Computat. Social Syst., early access, Mar. 17, 2021, doi: 10.1109/TCSST.2021.3064213.

J. Luo, J. Li, L. Jiao, and J. Cai, “On the effective parallelization and near-optimal deployment of service function chains,” IEEE Trans. Parallel Distrib. Syst., vol. 32, no. 5, pp. 1238–1255, May 2021.

J. Cai, Z. Huang, L. Liao, J. Luo, and W.-X. Liu, “APPM: Adaptive parallel processing mechanism for service function chains,” IEEE Trans. Netw. Service Manage., vol. 18, no. 2, pp. 1540–1555, Jun. 2021.

X. Yang, X. Jia, M. Yuan, and D.-M. Yan, “Real-time facial pose estimation and tracking by coarse-to-fine iterative optimization,” Tsinghua Sci. Technol., vol. 25, no. 5, pp. 690–700, Oct. 2020.

Y. Liu, D. Li, S. Wang, F. Wang, W. Dou, X. Xu, S. Li, R. Ma, and L. Qi, “A long short-term memory-based model for greenhouse climate prediction,” Int. J. Intell. Syst., Aug. 2021, doi: 10.1002/int.22620.

Y. Jin, W. Guo, and Y. Zhang, “A time-aware dynamic service quality prediction approach for services,” Tsinghua Sci. Technol., vol. 25, no. 2, pp. 227–238, Apr. 2020.

L. Kong, L. Wang, W. Gong, C. Yan, Y. Duan, and L. Qi, “LSH-aware multitype health data prediction with privacy preservation in edge environment,” World Wide Web, vol. 3, pp. 1–6, Sep. 2021, doi: 10.1007/s11280-021-00941-z.

J. Li, H. Jiao, J. Wang, Z. Liu, and J. Wu, “Online real-time trajectory analysis based on adaptive time interval clustering algorithm,” Big Data Mining Anal., vol. 3, no. 2, pp. 131–142, 2020.

H. Liu, H. Kou, C. Yan, and L. Qi, “Link prediction in paper citation network to construct paper correlation graph,” EURASIP J. Wireless Commun. Netw., vol. 2019, no. 1, pp. 1–12, Dec. 2019.

Y. Xu, C. Zhang, G. Wang, Z. Qin, and Q. Zeng, “A blockchain-enabled deduplicatable data auditing mechanism for network storage services,” IEEE Trans. Emerg. Topics Computat., vol. 9, no. 3, pp. 1421–1432, Jul. 2021, doi: 10.1109/TETC.2020.3005610.

Y. Wang, Y. Gao, Y. Li, and X. Tong, “A worker-selection incentive mechanism for optimizing platform-centric mobile crowdsourcing systems,” Computat. Netw., vol. 171, Apr. 2020, Art. no. 107144.

Y. Xu, C. Zhang, Q. Zeng, G. Wang, J. Ren, and Y. Zhang, “Blockchain-enabled accountability mechanism against information leakage in vertical industry services,” IEEE Trans. Netw. Sci. Eng., vol. 8, no. 2, pp. 1202–1213, Apr. 2021, doi: 10.1109/TNSE.2020.2976607.

Z. Cai, Z. Xiong, H. Xu, P. Wang, W. Li, and Y. Pan, “Generative adversarial networks: A survey toward private and secure applications,” ACM Computat. Surveys, vol. 54, no. 6, pp. 1–38, Jul. 2021.

Y. Xu, L. Qi, W. Dou, and J. Yu, “Privacy-preserving and scalable service recommendation based on SimHash in a distributed cloud environment,” Complexity, vol. 2017, May 2017, Art. no. 3437854.

Y. Xu, J. Ren, Y. Zhang, C. Zhang, B. Shen, and Y. Zhang, “Blockchain empowered arbitrable data auditing scheme for network storage as a service,” IEEE Trans. Computat. Sci., vol. 13, no. 2, pp. 289–300, Mar. 2020.

SIFENG WANG received the bachelor’s and master’s degrees from Qufu Normal University, China, in 2011 and 2014, respectively. He is currently a Teacher with the 1st Middle School of Juye County, China. His research interests include school sports and humane sociology of sports.

ZHICHEN HU received the bachelor’s degree from Nanjing University of Information Science and Technology, China, in 2020, where he is currently pursuing the master’s degree in computer science and technology. His research interests include knowledge graph, cloud computing, and big data.

XINTONG FAN received the bachelor’s degree from Qufu Normal University, China, in 2020, where she is currently pursuing the master’s degree in computer science and technology. Her research interests include big data and recommender systems.

TENGFEI FAN received the bachelor’s degree in sport education from Beijing Sport University, China, in 2011. He is currently a Lecturer with the Department of Physical Education, Qufu Normal University. He has published several research papers in international journals and conferences. His research interest includes sport education.