Hindawi
Mobile Information Systems
Volume 2022, Article ID 7932684, 12 pages
https://doi.org/10.1155/2022/7932684

Research Article

Study on the Sustainable Development Strategy of School Soccer Based on the Background of Big Data Era

Jun Zhang, Dong Liang, and Zhang Sun

Guangzhou Xinhua University, Guangzhou, Guangdong 510520, China

Correspondence should be addressed to Dong Liang; bcld@xhsysu.edu.cn

Received 11 July 2022; Revised 27 July 2022; Accepted 4 August 2022; Published 20 August 2022

Academic Editor: R. Mo

Big Data is the most popular concept in this era, which is the massive amount of information and related technology generated by the information explosion in the era of "Internet+." Big Data is the most popular concept of our time. With the most advanced technology to collect, analyze, organize, and store data, Big Data can effectively handle all kinds of complex information. Because of this, big data is widely favored by all walks of life. In China's sports industry, the use of big data has become mature and has shown its unique advantages. With the development of campus soccer in China in the past decade, how to use big data to promote the sustainable development of campus soccer in China has become a key issue for sports workers to consider today. Based on the above background, this paper proposes a system combining data mining and personalized data recommendation to collect and analyze the information of campus soccer to promote the sustainable development of campus soccer. First, we propose a data mining method based on deep learning data mining network model combined with migration learning to address the data mining problem. The method uses the knowledge of historical model parameters and applies them to new tasks, thus solving the problem of network training when samples are lacking and improving data utilization and data mining effects. Then, for the data recommendation problem, a new deep learning method is proposed, which performs effective intelligent recommendation by pretraining. In the initial phase, the corresponding low-dimensional embedding vectors are learned, which capture information reflecting the relevance of students to soccer sports. During the prediction phase, a feed-forward neural network is used to model the interaction of student and soccer sport information, where the corresponding pretrained representative vectors are used as inputs to the neural network. Finally, it is experimentally verified that the data mining method proposed in this paper can effectively improve the data mining performance and efficiency, and the proposed data recommendation method possesses better accuracy than the traditional methods. The use of this system can effectively collect and analyze campus soccer information, which helps to develop campus soccer and promote the sustainable development of campus soccer.

1. Introduction

Soccer, as the world's number one ball, has been attracting attention all over the world since its birth. China, as the largest developing country in the world, has also been committed to the development of soccer. After continuous learning and exploration, it embarked on the road of professional development of soccer in the early 1990s. However, more than two decades of learning and efforts have not resulted in rapid progress, and the results of Chinese soccer are still unsatisfactory. To this end, it is particularly important to explore a sustainable path for the development of Chinese soccer, and the government has specifically formulated a strategy for the development of school soccer, placing its hopes on the youth and introducing a series of policies to promote the popularity of school soccer, which has also ushered in unprecedented development opportunities for youth soccer [1]. Of course, opportunities and challenges coexist, and some problems inevitably arise in the process of school soccer development.

Through the survey, it was found that in the institutions visited, there was no long-term development plan for campus soccer, but only regular teaching documents such as training plan, syllabus, and teaching schedule of comprehensive nature. Some institutions have set up school soccer teams, but they are prone to "discontinuity," and there are no incentives for students to participate in the games due to their personal hobbies, so it is difficult to guarantee the training level and game performance. The current problems facing the development of campus soccer are shown in Figure 1, which can be summarized into four main aspects.

(1) Lack of scientific theoretical guidance for sports selection. Sports selection is the beginning of competitive sports, and a good start is the foundation of success. So in a sense, the scientific level of soccer player selection directly affects the development level of soccer in a country [2]. The research of youth selection in China has been lagging behind for a long time, especially some coaches at the grassroots level still uphold the traditional selection concept, relying on subjective experience to select, resulting in many excellent sports talents being buried. The selection of school soccer players also inevitably faces these problems, so scientific theoretical guidance in sports selection is particularly important.

(2) Sports training is difficult to complete with quality and quantity. Athletic training is an essential part of improving athletic performance, and without quality training, it is difficult to achieve the expected results. In the process of training campus soccer players, you can often find a variety of problems, some from the teachers, some from the students' parents, and some from the students themselves, how to balance the relationship between various aspects, and deal with the contradiction between student learning and training, which is also a subject that requires in-depth research. Cultivating an excellent athlete is a long-term process that requires perseverance and persistence, so sports training is the guarantee of athletes' success.

(3) Sports competition fails to form a long-term mechanism. Soccer league is an important part of campus soccer activities and is the main way to evaluate the effect of campus soccer development, and the quality of soccer league development also directly reflects the development of campus soccer [3]. According to the National School Football Competition Program, efforts are made to build a school soccer competition system based on inter-school competitions, interschool competitions and regional competitions, and to improve the four-level league mechanism of school soccer. In recent years, various types of campus soccer competitions have increased, but still no long-term mechanism or scale effect has been formed. Therefore, it is imperative to reform campus soccer, normalize, and scale campus soccer competitions, and establish a campus soccer competition system in line with regional characteristics.

(4) The cultural atmosphere of campus soccer is lacking. Looking at the traditional soccer teams in Europe and South America, it is not difficult to find that they all have a strong soccer atmosphere, and soccer seems to be an indispensable part of their lives, while the soccer culture in China is relatively absent, especially in the campus [4]. Most people look more at your cultural achievements, and under the guidance of such social values and recognition, many young people with athletic talent eventually did not embark on the development of soccer, and the construction of campus soccer culture has become an urgent problem.

In professional soccer, a large amount of data involving tactical behavior analysis is collected, and cooperation with computer science can lead to more rapid development of soccer techniques and tactics because big data presents new ideas on data management and analysis methods commonly used in sports. In recent years, the equipment for data collection on soccer performance has changed day by day and the quality and quantity of data have grown rapidly, resulting in teams having large amounts of data to process on a daily basis [5]. The human form of statistics and analysis has long been unable to match the rapid development of today's soccer game, and the efficiency and speed of big data processing is evident. With the application of big data, technical and tactical performance, which are more abstract and usually evaluated qualitatively, can be quantified and analyzed to provide more scientific suggestions for players' daily training and games.

In tactical performance analysis, passing, as one of the most common and frequent elements, deserves the attention of coaches and players. How to define and distinguish the difference between effective and ineffective passes cannot be determined entirely by subjective judgment, and data-oriented content is needed to support this. Some researchers have assessed the effectiveness or ineffectiveness of passing by developing a model that combines passing effectiveness with consistent offensive performance, rather than relying on the occurrence of infrequent probability events. The model is built to help assess any position, individual tactics, efficiency, player comparisons, and team capabilities, while the method is applied to the game and can also help teams identify important team roles. Tracking data analytics applied to soccer mechanics and tactics demonstrates the efficiency and speed of big data. Whereas previously technical and tactical analysis was only superficially useful for game performance, advances in the data analysis process have now greatly improved the rationalization of athletic performance [6]. Addressing the aggregation regarding performance feature construction, space, and time through multidisciplinary collaboration is critical to unlocking the potential of soccer position tracking data.

Currently, the expected product dissemination of soccer events is more popular. The application of Internet and big data technologies can collect a large amount of data from individual teams and players and display and summarize them through different indicators to predict the performance of teams and players in the game, presenting more intuitive data for fans and optimizing their experience. Many event communication companies will use the information

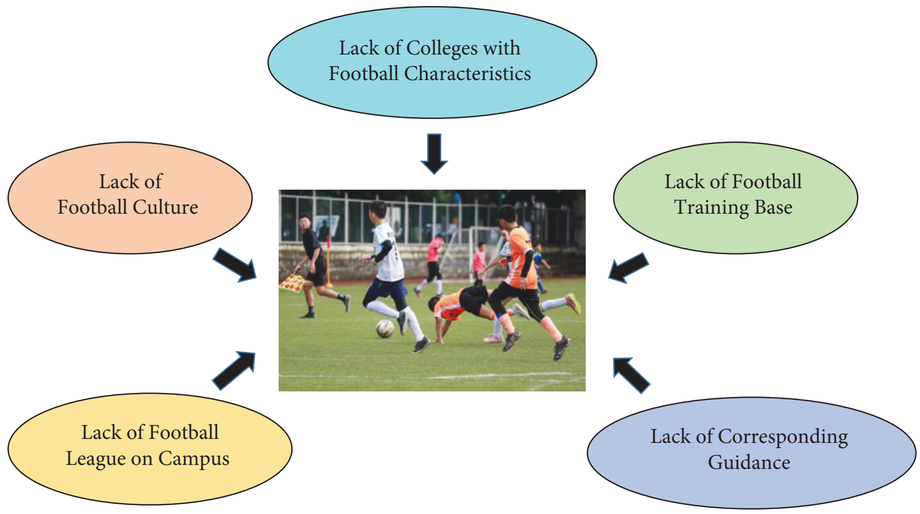

Figure 1: Problems faced by the development of campus football.

technology interaction mode to realize the interaction between commentary and fans, fan-to-fan interaction, etc., to enrich the audience's viewing experience, while the background can make timely adjustments through the collection of user feedback information [7]. The application of big data is also reflected in sports betting. By collecting massive amounts of data, tournament communication companies can provide reference results to fans, who can analyze the results of matches based on the data and attract more people to watch soccer matches. The tournament distribution company can increase the amount of users and collect more data to achieve a virtuous cycle.

Introduce the advanced concept of Internet and big data into the practice of campus soccer and combine big data with student physical fitness test, physical education entrance examination, and four levels of campus soccer league to make technological selection of campus soccer players based on objective data. This can prevent defects such as the static and one-sided nature of the data of the selection index in the past, and also provide certain reference and reference for campus soccer teaching, training, and competition. At the same time, a wide area network for athlete selection should be established nationwide so that the whole country can be played as a single game and resources can be shared.

The main contributions of this paper are as follows: in order to better promote the sustainable development of campus soccer, this paper proposes a combined data mining and data recommendation approach for the collection and analysis of campus soccer sports information. First, a data-mining model based on migration learning, and MMD is proposed for the data-mining task, and MMD is used to measure the distribution differences between source and target data, so that the network model can be adjusted accordingly according to the size of MMD. The model is pretrained using the source domain data and retrained and fine-tuned using the target domain data, thus improving the accuracy and efficiency of data mining. Then, a collaborative filtering recommendation system based on deep neural networks for explicit feedback data is constructed for the data recommendation task, which uses feedforward neural networks to establish the connection between student and soccer information. Finally, the proposed method is experimentally verified to be effective in analyzing campus soccer data, which are conducive to promoting the sustainable development of campus soccer in the era of big data.

2. Related Works

2.1. The Current Situation of Campus Soccer Development. Before exploring school soccer, we must first understand what "school soccer" is and only by understanding its nature and characteristics can the developed intelligent system be close to the actual application needs and serve school soccer to the greatest extent and solve the related problems. According to some researchers, school soccer is a general term for all kinds of soccer activities that focus on cultivating young people's interest in soccer, in the form of competitive matches or games. The purpose of school soccer activities is to improve students' physical fitness, cultivate a spirit of hard work and a sense of cooperation, and its ultimate goal is not just to choose and compete [3]. With the development of large-scale campus soccer activities and the construction and improvement of the four-level soccer league system not only can we popularize soccer knowledge and skills for young students, so that more students understand soccer and are willing to join the sport but also increase the soccer population, discover and cultivate soccer reserve talents from it, and lay a good foundation for the development of soccer in China.

School soccer activities have been carried out for nearly 11 years, and in response to the current situation of school soccer, some researchers point out that certain schools carry out soccer training activities superficially, focusing only on the performance form of efforts, and do not go deep into the organization and promotion of soccer activities. Some schools, due to limited funds, invest most of their funds in cultural learning programs and are unable to provide comprehensive financial support for soccer training programs, resulting in the failure of soccer training programs [8]. The lack of professional soccer teaching staff is also an important reason why school soccer activities cannot be carried out. Most physical education teachers only have basic knowledge of soccer and cannot teach students soccer skills and tactics in depth, which is a great obstacle to the cultivation of soccer talents. In addition, some students' parents only focus on cultural learning and worry that soccer will affect their students' performance in cultural classes, so they do not support students to play soccer.

The aim of school soccer is to make more students participate in soccer, feel the fun of soccer, experience the culture and spirit of soccer, and guide the development of students in all aspects of body and mind. Combining literature and personal experience, the authors conclude that the development of campus soccer faces the following problems in practice: (1) sports safety; (2) a large lack of professional soccer coaches/teachers; (3) inability to objectively and comprehensively reflect players' performance; (4) lack of scientific talent training model; (5) lack of efficient and professional competitive training assistance; (6) lack of "intelligent" management method. The above problems largely restrict the vigorous development of school soccer. Solving the above problems is conducive to the further promotion of school soccer, enhancing the physical fitness of youth and the spirit of teamwork, and promoting China's progress toward a strong sports nation.

Internet+ is a new industry of Internet development under Innovation 2.0, which organically integrates the Internet with various traditional industries to create a new development ecology through cross-border integration, innovation-drive, and structural reshaping. Studying the deep integration of campus soccer and the Internet is conducive to improving the development of intelligent systems for campus soccer [9]. Unlike the inefficient traditional methods, a smart campus soccer system must be combined with the Internet to create a scientific, rational, and efficient campus soccer management and training model.

With the increasing maturity of "Internet+" technology, the emergence of intelligent soccer fields has emerged. Parents only need to connect to the network through 4G or Wi-fi, and they can view students' performance on the smart soccer field in real time with the help of cell phones and other communication devices, using Internet streaming media and other technologies, parents can make real-time scrolling caption comments, interact with other parents, and support one-click forwarding of the game video to the WeChat circle of friends. In addition, through the use of big data and cloud computing technology, ordinary users can have real-time location near the location of the stadium, real-time understanding of the use of the stadium, using cell phones and other intelligent devices to make reservations for the use of the stadium and other operations to enhance the use of soccer fields, improve the rate of reasonable allocation of soccer fields, and greatly solve the problem of soccer field shortage [10]. The combination of Internet technology and campus soccer can realize the function of "intelligent coach" and provide innovative new thinking for campus soccer training.

The combination of Internet and campus soccer not only improves the intelligent teaching level of campus soccer schools but also greatly saves the teaching cost and human cost in the teaching process, and gives full play to the convenience and wisdom of the Internet. For example, in the school soccer work, with the help of Internet technology, the performance data of students at each stage can be recorded, so that a more comprehensive and detailed understanding of the overall level of students [11]. Teachers no longer rely on personal intuition and usual observation to provide targeted training for students, but only need to check the changes in students' performance data to scientifically and reasonably arrange targeted training for students, and in the training process, the intelligent system of campus soccer can also give suggestions on training directions and training programs, greatly reducing teachers' labor intensity. In addition, the intelligent human ball data interaction can also reduce the subjective bias brought by human operation and ensure the objectivity and accuracy of the data.

With the increasing prevalence of big data, soccer, the world's number one sport, needs to progress with the times and keep up with the pace of development. The application of big data in soccer can not only greatly improve the accuracy and speed of information acquisition but also monitor in real time whether the various physiological references in athletes' sports are within a reasonable level range, effectively enhancing the safety and protection of athletes. Only by scientifically using big data technology and adopting an innovative education model can we accelerate the development of Chinese soccer faster.

2.2. Current Status of Data Mining Research. Before data mining, we should clarify our data orientation, determine the direction and scope of data mining, and then implement data mining to avoid data redundancy, data bias and other problems, and avoid blind mining. Data mining is the substantive mining of data, and then select the suitable algorithm for the data research according to the theme, and then implement the data mining work, this link is the core link of data mining work. The main methods of data mining include four categories, as shown in Figure 2.

The decision tree generated in different scenarios will be different, so the decision tree will also be called classification tree, regression tree, etc. The classical algorithms of decision tree data mining methods are the following: (1) Cart algorithm—it is a simple binary tree algorithm, which is often used in simple data to generate a simple binary tree structure. (2) ID3 algorithm [12]—ID3 algorithm is a relatively early algorithm in the decision tree algorithm, and it is based on data information through a series of rules to find the attributes represented by each node in the tree, the entropy of the algorithm as the basis for classification, the data will eventually generate the form of decision trees. (3) C4.5 algorithms [13]—this algorithm uses information gain or entropy to optimize the process of decision tree node classification and improve the decision tree to make it more friendly.

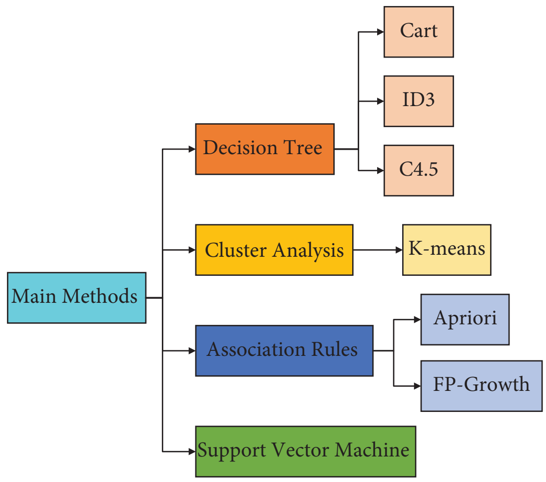

Figure 2: Main methods of data mining.

Clustering analysis is essentially to find out the classification basis of data according to the research topic and according to this basis to classify the data, and refine the data into different types of data sets, and ensure that the data in each set has similarity, and there are differences between different sets, and then use the data visualization technology to represent them, and friendly to the user, that is called cluster analysis. The main algorithm is K-means algorithm, the outstanding advantage of this algorithm is the principle of simple, efficient application, very suitable for the processing of large-scale data.

Association rule analysis is one of the more commonly used methods in data mining work. Association rules refer to the hidden rules of relationship between things, and association rule analysis refers to the process of finding and analyzing the information between things and association rules with set values [14]. The Apriori algorithm can solve the analysis of association rules of corresponding data, but it has some shortcomings, so some researchers proposed the FP-Growth algorithm to make up for the shortcomings of Apriori algorithm in generating candidate item sets.

Support vector machine (SVM), a binary classification model, is defined as a linear classifier with maximum interval in the feature space, which is different from the perceptron when the interval is maximum. The core concept is that the support vector samples play a key role in the recognition problem, and the support vector is the nearest sample point to the classification hyperplane, which is the support vector classifier, and the classification hyperplane is used to divide the sample data into two.

In terms of data mining objects, data mining will mostly favor multimodal data mining in the later stage [15]. At the present stage, most of the data mining is based on the corresponding algorithm, and the algorithm process is not easy to be understood by users, so the data mining visualization research has certain research significance.

In financial investment, it is possible to predict the future trend of stocks through data mining algorithms based on historical stock trading data, thus promoting profitability [16]. In terms of fraud information identification, the identification of fraud information is mainly through data mining methods to obtain certain typical characteristics of fraudulent behavior, and when the relevant business processing or behavior is highly similar to or matches these typical characteristics, it is possible to issue warnings to relevant personnel through some interactive techniques. In addition, data mining has very great potential for application in bank lending, process optimization, oil reserve prediction, drug synthesis and development, chemistry, chemical industry, etc.

2.3. Current Status of Personalized Data Recommendation Research. Recommender system is a new research field combining various disciplines such as data mining, prediction algorithms, and machine learning. In the definition of recommendation system, it is pointed out that in daily life, whether it is an understood event or an unknown event, people always need to make decisions. When facing familiar things, people can often rely on past experiences to make reasonable decisions; however, when facing unknown things, people need others' verbal suggestions, book reviews, movie reviews, recommendations [17]. To make judgments—recommender systems, which match different users with items from a large number of items that match their interest preferences but are not observed by the users, are considered to be becoming an important business with significant economic impact.

The core idea of the content filtering-based recommendation technique is that the user's historical selection record or preference record is used as the reference recommendation, and the items with high correlation with the reference recommendation in other unknown records are mined as the content recommended by the system [18]. Then calculate the similarity between the user's preferences and the recommendation object to be tested in terms of content; finally, rank the similarity between the recommendation object to be tested and the user's preferences, so as to select the recommendation object that matches the user's interest preferences.

The content-based recommendation technology often reduces the timeliness of information due to being time-consuming when dealing with large-scale information content; the collaborative filtering technology is prone to cold start problem when facing new items; and the hybrid recommendation technology is a recommendation method that retains the advantages of different recommendation technologies and avoids their disadvantages by incorporating different algorithms into the recommendation system. The current hybrid recommendation is mainly divided into prefusion, postfusion, and midfusion: (1) Prefusion—refers to the fusion of multiple recommendation algorithms into one model, such as in the process of product recommendation. (2) Midfusion—this hybrid recommendation technology generally [19]. (3) Postfusion—this method places great importance on the recommendation results, mainly by comparing the recommendation effects of different recommendation algorithms to get a more reliable sequence of recommended objects, and finally recommending this sequence to users.

Deep learning algorithms are powerful because they can learn and deal with complex problems such as human beings, analyze and calculate linear or nonlinear feature sequences from multiple dimensions in the face of complex scale data, and automatically learn features that meet users' needs from massive data [20]. Deep learning techniques can not only discover the hidden potential features of user behavior records, but also capture the interaction features of user-user, user-item, and item-item nonlinear relationships, which brings more opportunities for system performance improvement and can overcome some obstacles encountered in traditional recommendation techniques to achieve more accurate recommendations.

3. Algorithm Design

3.1. Data Mining Model Design. Transfer learning is the mainstream machine learning method to address this problem, which transfers the knowledge learned from the previous task to the new task, with the aim of achieving better learning results in the new task. As the driving force of future deep learning, transfer learning can effectively solve the above problems. Therefore, migratory learning is introduced in data mining to make efficient use of big data and explore the commonality of different data sets.

First, the source and target domains are preprocessed accordingly. Then, the original deep neural network model is trained with the source domain data, or the trained network structure and parameters are already available as the source domain classification or prediction model [21]. The difference of data distribution between the source domain and the target domain is analyzed by MMD to obtain the distribution distance. The original network structure is adjusted according to the MMD to obtain a new target domain network, and the parameters obtained from the original network training are selectively migrated. Generally, the source domain and the target domain distribution will have some differences, that is, the MMD is higher than the set threshold, then the network needs to add or replace new hidden layers for the target domain model to learn new knowledge.

It is based on the principle of finding the mean of samples with different distributions by finding a continuous function on the sample space, and then finding the difference between these two means as the difference in means corresponding to these two distributions [22]. The maximum value of this difference is the MMD of the two distributions:

$$MMD[F, p, q] = \sup_{f \in F} \left(E_p[f(x)] - E_q[f(y)] \right). \tag{1}$$

As the size of the observed data set increases, constraints are needed to speed up the convergence of the empirical estimates of MMD. Using the complete inner product space becomes the Hilbert space, and the regenerative kernel Hilbert space can be expressed as the dot product in the space:

$$f(x) = \langle f, \phi(x) \rangle_H. \tag{2}$$

The inner product can be replaced by a kernel function, and for mappings in higher-dimensional spaces, the radial basis kernel function is usually used:

$$k(x, x') = e^{-\|x - x'\|^2 / 2\sigma^2}. \tag{3}$$

In general, MMD can be regarded as the distance between two points in the regenerative kernel Hilbert space, which can be used to measure the distance of two distributions.

According to the above migration steps, the network is not adjusted for the MMD below the set threshold, and the network is directly trained unsupervised with the target domain data based on the original parameters and model, as shown in Figure 3.

The parameter fixing is canceled, and the whole network is fine-tuned using the target domain data to obtain the final target domain network model.

3.2. Personalized Data Recommendation Algorithm Design. The goal of a recommendation system is to recommend content or products that users like or are interested in. In simple terms, a user will give a high rating to content or products that he is interested in or likes. Therefore, it is a simple and effective strategy to recommend a product to a user that he may give a high rating to. In summary, recommendation systems can be broadly viewed as a problem for predicting the user's rating prediction for a product.

A variety of recommendation system models have emerged for the rating prediction problem. In general, most of the current recommendation system models can be classified into two categories: (1) content-based recommendation systems, and (2) collaborative filtering-based recommendation systems. Among them, content-based recommendation systems make recommendations by extracting features from user or product content, such as user profile content, product description content [23]. The collaborative filtering-based model, on the other hand, obtains the available recommendation system by deriving from the user's historical interaction behavior. The user's historical behavior can be the user's click record, purchase record, or rating record of a product on a specific website. Among them, the user's rating records are again the most commonly used. Recommendation systems based on collaborative filtering models are currently gaining more attention because of their higher recommendation accuracy.

Based on the machine-learning algorithm, training a usable recommendation system with the user's historical interaction data is the most commonly used collaborative filtering method. Here, the commonly used machine-learning model is based on matrix decomposition model (MF). However, artificial neural network (ANN)-based models have also started to gain attention in recent years [24]. The MF-based approach can better model simple user-product interactions, but more complex interactions cannot be handled effectively due to the low complexity of the MF model.

Therefore, the model can be divided into two main phases, as shown in Figure 4: (1) feature learning phase: the feature learning model generates the corresponding low-dimensional user and product feature vectors based on the user-user and product-product co-occurrence relationships

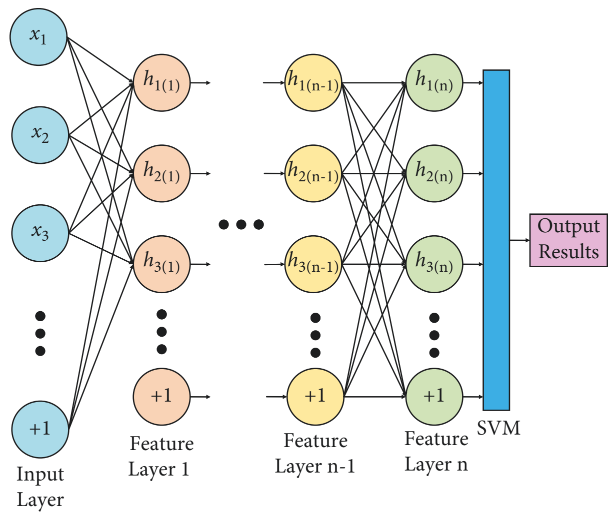

FIGURE 3: Network adjustment process based on transfer learning.

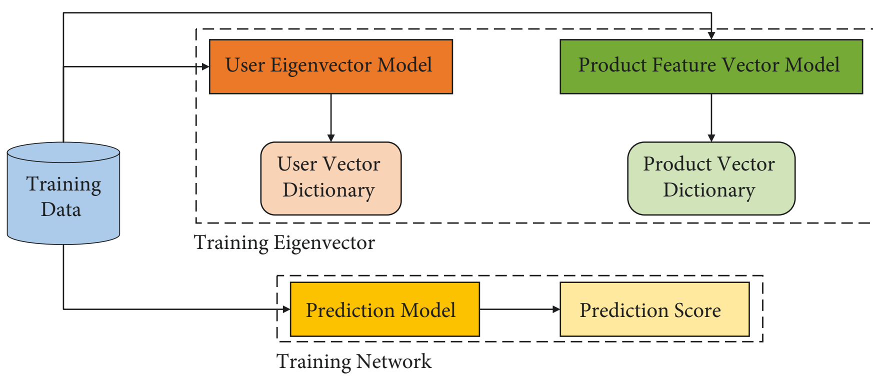

FIGURE 4: Model two-stage learning framework.

through the user's rating matrix. (2) Neural network training phase—the final score is predicted by the rating prediction neural network by using the user's product's CM or RIM feature vector as input, which is calculated by layer-by-layer operations in the network.

Compared to the previous models, the proposed model is able to utilize both co-occurrence and interaction. This is because the pretrained feature vectors can obtain user and product feature vectors that contain co-occurrence features, while the predictive neural network can simulate interaction relationships. Previous models can either handle co-occurrence but not interaction, or the opposite. The pretrained feature vectors can be easily used in different branches of the model. Training the feature vectors and neural networks separately can achieve higher prediction accuracy than joint training.

The product and user feature vectors obtained by the feature learning model reveal the co-occurrence characteristics of users and products, but these feature vectors do not directly yield the final rating prediction results. Therefore, additional components are needed to estimate the user's rating of the product. Since artificial neural networks are able to efficiently extract features and model complex objective functions, as well as fuse input features from multiple perspectives [25]. Therefore, neural networks are an ideal model for rating prediction. In this section, a detailed description of how the neural network is used to generate the final predicted scores from the obtained feature vectors is presented.

The goal of rating prediction is to output a real number of predicted values based on the features of a given user and product as input, and this predicted value represents an estimate of the score given to the product by the user. Thus, the rating prediction problem can be viewed as a regression problem. Therefore, the most straightforward way to use neural networks is to directly input the CM or RIM features of users and products obtained in advance into a feedforward neural network to obtain the prediction values. However, the prediction accuracy of this approach is not high enough [26–29].

This history-based perspective differs from the basic perspective that utilizes only current users and products; it will represent users or products by their historical records. The general architecture of the network is shown in Figure 5. In which, the network takes CM and RIM features as input. The symbols "FC" and "CONV" in the figure indicate the use of fully connected layer and convolutional layer, respectively. In general, the network can be divided into two main parts: (1) multiview feature extraction and (2) integrated prediction.

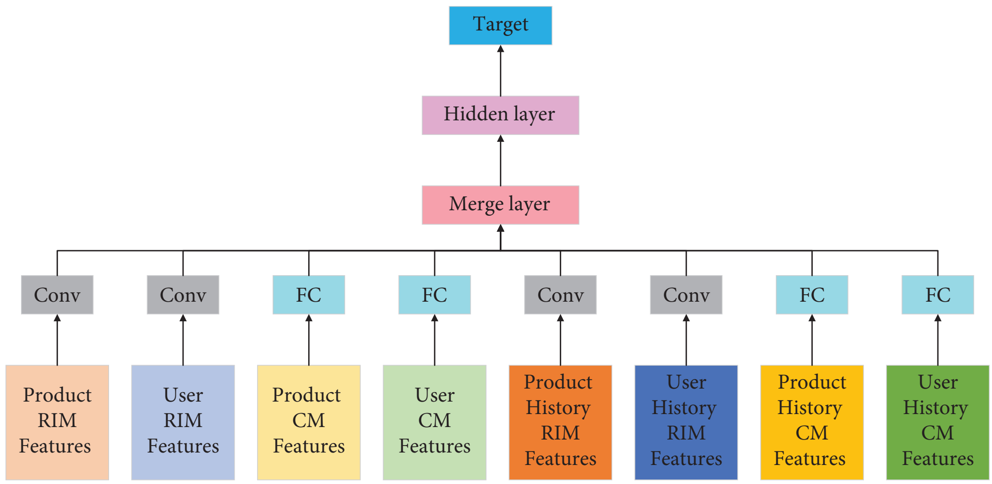

Figure 5: Overall structure of multiview prediction neural network.

As shown in Figure 5, in order to obtain the prediction of the product given by the user, two feature inputs are required in the feature extraction phase under the perspective of current perspective and historical perspective. For the current perspective, the corresponding feature vectors are input. The CM and RIM features of the user from the current and historical perspectives are entered on the left side of the network.

The CM features of users and products exist under different spaces, since they are trained independently of each other. To extract their features, two different fully connected neural networks are introduced to transform the CM features of users and products, respectively. The fully connected neural network is used because the CM feature vectors are holistic and each CM feature vector uniquely represents a user or a product. Then, the corresponding fully connected feature extraction layer is represented as follows:

$$\begin{aligned} \alpha_t(tj) &= g\left(W^e e_j\right), \\ \alpha_u(ui) &= g\left(W^r r_i\right). \end{aligned} \tag{4}$$

In contrast to CM, users and products, given K different rating categories, will each have K RIM feature vectors representing the different ratings. In order to use a unique feature to represent users and products, K different RIM feature vectors connecting users and products are used to represent them:

$$\begin{aligned} v_i(ui) &= \left[r_i^1, r_i^2, \ldots, r_i^k\right], \\ \mu_j(tj) &= \left[e_j^1, e_j^2, \ldots, e_j^k\right]. \end{aligned} \tag{5}$$

In order to efficiently process the historical information effectively, it is first necessary to perform further feature extraction from the set of users' evaluated products or the feature vector corresponding to the set of users who have evaluated products, and then use the extracted features for further prediction by neural networks.

The neural network is trained by the stochastic gradient descent method, and the evaluation results of the network on the validation set are used as a condition to end the training early. If the results do not improve within a few cycles on the validation set, then the training is ended. In addition, if the improvement on the validation set is small in one training cycle, the learning rate of the training will be halved. It is worth noting that any emergent CM or RIM feature vectors are not updated during the network training. This will destroy the semantic meaning of the CM or RIM feature vectors on the corresponding space. In addition, the structure of the multiview network results in the possibility of a feature vector appearing in different branches of the network, which makes training more difficult.

4. Experiments

4.1. Data Mining Experiments. In the experimental data, the ratio of the number of source domain data to target domain data is set to 7 : 3, and several target domains with different distributions are collected. The simulated scenario is a situation where it is difficult to obtain enough target domain data. The variation curves of predicted MAPE with MMD for different models are shown in Figure 6. From the curves, it can be found that the effect of migration learning is correlated with the maximum mean difference MMD between source and target domains, and if the MMD is small enough, it means that the target domain data are highly consistent with the distribution of source domain data, and the migration learning-based model works better than the original model. In this case, the source domain and the target domain can be regarded as in the same domain, then they can be used as a training sample of the network at the same time, if adding new layers may lead to an increase in network parameters, resulting in overfitting, which brings undesirable effects instead. Therefore, we take the measure of fine-tuning the whole original network directly using the target domain data.

The results of representative specific prediction experiments are shown in Table 1. From Table 1, we can see that there is a distance between the source domain and the target domain, that is, the value of MMD is larger, and our model works better. The model not only can effectively migrate the knowledge from the source domain but also allows the network to use the target domain data to learn new

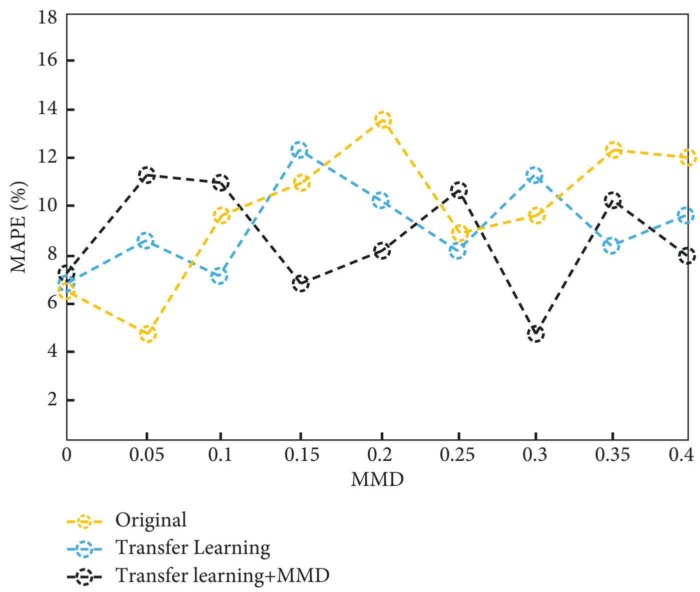

Figure 6: Change curve of MAPE predicted by different models with MMD.

Table 1: Representative prediction experiment results.

MMD	Original (%)	Transfer learning (%)	Transfer learning + MMD
0.045	11.26	9.64	12.92%
0.096	9.74	11.37	10.20%
0.158	19.37	8.96	11.83%
0.213	7.81	12.13	13.56%
0.279	8.35	10.42	15.49%
0.343	12.64	11.58	14.37%

knowledge by adding new layers to achieve the combination of the two domains' knowledge. However, if the MMD is too large, which means that the distribution of the source and target domains is too different, then negative migration will occur. No matter how you adjust the network result, the final migration result is worse than the original network trained with only the target domain data. In this case, the knowledge of the source domain is counterproductive.

A comparison of the prediction curves under different MMD-based models is shown in Figure 7. From Figure 7, it can be seen that our model fits better. In contrast, the original model deviates from the actual curve in several places, and the results are relatively poor. Specifically, the model performs best when the MMD belongs to the interval (0, 0.17), that is, only the target domain data need to be used to fine-tune the trained model for the source domain data. If the MMD is larger and falls in the interval (0.17, 0.35), the model needs to be retrained using the target domain for adding new layers, and finally, the learning rate is reduced and the full network is fine-tuned. When the MMD is larger than 0.35, the migration learning has a negative effect, and it is necessary to find another source domain dataset with stronger relevance.

4.2. Personalized Data Recommendation Experiments. In this section, we test the effectiveness of the proposed model applied to the Top-N recommendation problem. MovieLens-1M is used as the dataset, and the model parameters introduced previously are used. First, the evaluation criteria are introduced, and then the comparison method is presented, and finally, the experimental results and analysis are given. The same consistent test conditions are used. The detailed experimental setup is as follows: with 90% of the data selected as the training set and the rest as the test set. In particular, in the test set, only the samples with a rating of 5 are kept, and the rest are discarded directly. Despite the Top-N problem, the training set is used to train a model for score prediction during the training phase. Therefore, all training and model details follow the previous model. Recall is used as the test criterion to evaluate the performance of Top-N recommendations. The comparison methods are the following: (1) Movie Average—the N products with the highest average scores are recommended to users. (2) Bias Matrix Factorization (BMF)—a MF model that incorporates user and product biases. (3) SVD++—a hybrid model of MF method and neighbor-based method.

Figure 8 shows the recall of different methods at different N. It can be seen from the figure that the proposed method can achieve a recall rate of 0.32 for $N = 10$, and overall, the proposed method is significantly higher than the other models under different N. The experimental results indicate that the proposed method not only has good accuracy in score prediction but also performs well on Top-N. However, this does not mean that a method with high RMSE will necessarily guarantee good Top-N performance. For example, although BMF can achieve an RMSE of 0.85, it has a recall of 0.17 at $N = 10$.

In order to verify the rationality of the multiview structure, the performance under different viewpoint subnetworks and different combinations of subnetworks was tested, and the results are shown in Table 2. In the table, different subnetworks are classified into separate viewpoints and multiviewpoints. Under the current perspective, the performance with CM or RIM feature vectors alone is tested

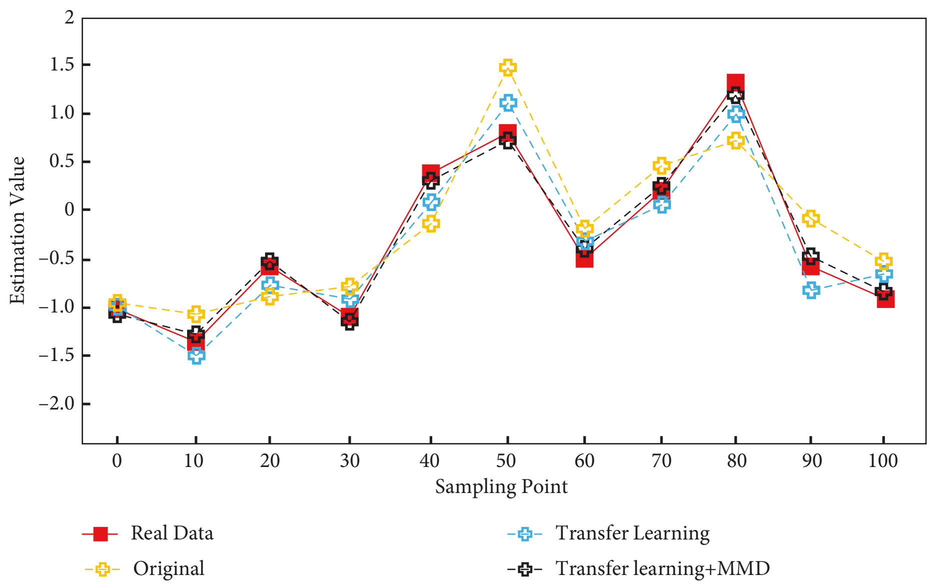

Figure 7: Experimental results of prediction curves of different models.

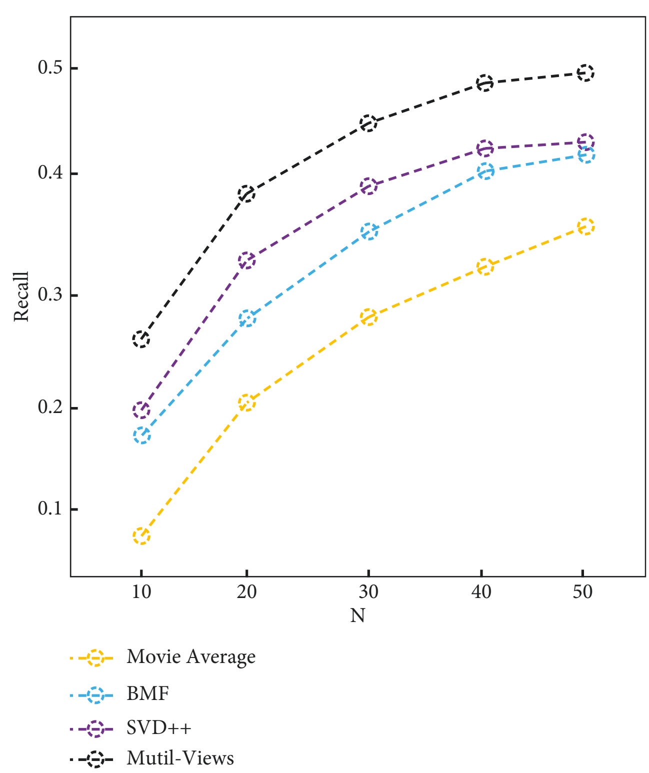

Figure 8: Recall rate of different N values.

separately. Similarly, two separate subnetworks using CM or RIM feature vectors are also tested under the historical view. Note that the use of both CM and RIM features in the current view is also considered as multiview here. Among all the single-view models, using the RIM feature in the historical view is the best performer, reaching 0.834, which is close to the final model's 0.821. This result indicates that the RIM feature in the historical view plays the most critical role in the final prediction of the multiview model. When both RIM and CM are used in the current perspective, the corresponding performance is better than the 0.837 and 0.838 of the CM and RIM single-view models, while the performance of using both RIM and CM in the historical perspective is only equal to the 0.852 of using RIM alone in the historical perspective. However, if all views are used simultaneously, the performance is still further improved to 0.834. Therefore, in general, the use of multiple views is better than the single-view model.

Table 2: RMSE experimental results from different perspectives.

View type		Vector model	RMSE
Single view	Current users and products	CM	0.837
	Current users and products	RIM	0.838
	Historical users and products	CM	0.853
	Historical users and products	RIM	0.852
Multiple view	Current users and products	CM + RIM	0.843
	Historical users and products	CM + RIM	0.834
	Historical and current users and products	CM + RIM	0.821

5. Conclusion

The advent of the era of big data has provided an opportunity for the development of soccer in China, especially for the selection and education of school soccer. It is especially important to deeply integrate big data with school soccer and give full play to the role of "big data" in the selection, teaching, training, and competition of school soccer. Based on the existing big data technology, a method combining data mining and data recommendation is proposed to promote the sustainable development of campus soccer. First, a data mining model based on migration learning and MMD is proposed for the verse mining problem, which improves the data utilization and data mining performance and efficiency. In which, MMD is introduced as a measure of the difference of data distribution between the source domain and the target domain, and the migration learning method and the adjustment network model are designed according to the size of MMD; finally, the data mining model based on migration learning and MMD is established. Then, for the personalized data recommendation problem, a deep collaborative filtering model that does not depend on content features using deep neural networks is proposed. The model learns the feature vectors of users and products, respectively, and the interaction between students and soccer sports information extracted by neural networks using the previously obtained features, and a multiview prediction model with two different feature vectors is applied. Finally, the proposed data-mining model is experimentally verified that the proposed method can not only improve the accuracy of data mining but also improve the efficiency of network learning; the accuracy of the proposed data recommendation algorithm prediction also exceeds the traditional method and the previous deep model. This paper is an attempt for the sustainable development of school soccer in the era of big data, and we hope to provide reference for other researchers. In the future, we plan to conduct research on sustainable development strategies for school soccer based on convolutional neural networks.

Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper is one of the research results of the Higher Education Teaching Reform Project "Academic Achievement Research Based on Learning Participation" in Guangzhou Xinhua University, item number: 2020J016; and The Key Platforms and Major Scientific Research Projects of Education Bureau of Guangdong Province-Young Innovative Talents (Humanities and Social Sciences), China (Grant no. 2015WQNCX187).

References

[1] R. Millington, A. R. Giles, N. van Luijk, and L. M. C. Hayhurst, "Sport for sustainability? The extractives industry, sport, and sustainable development," *Journal of Sport & Social Issues*, vol. 46, no. 3, pp. 293–317, 2022.

[2] S. C. Kong, "Partnership among schools in e-Learning implementation: implications on elements for sustainable development," *Journal of Educational Technology & Society*, vol. 22, no. 1, pp. 28–43, 2019.

[3] B. Lusheng, M. S. Aman, and L. B. Hooi, "Revolutionary of school youth soccer management development in," *Journal of Positive School Psychology*, vol. 6, pp. 4487–4503, 2022.

[4] S. Loland, "Olympic sport and the ideal of sustainable development," *Journal of the Philosophy of Sport*, vol. 33, no. 2, pp. 144–156, 2006.

[5] D. Matarrita-Cascante, M. A. Brennan, and A. E. Luloff, "Community agency and sustainable tourism development: the case of La Fortuna, Costa Rica," *Journal of Sustainable Tourism*, vol. 18, no. 6, pp. 735–756, 2010.

[6] K. Readman, "The soccer finals as an analogy for sustainable futures in assessment," *Primary and Middle Years Educator*, vol. 7, no. 2, pp. 10–15, 2009.

[7] M. O. A. González, J. S. Gonçalves, and R. M. Vasconcelos, "Sustainable development: case study in the implementation of renewable energy in Brazil," *Journal of Cleaner Production*, vol. 142, pp. 461–475, 2017.

[8] B. M. McNoe and D. J. Chalmers, "Injury in community-level soccer: development of an injury surveillance system," *The American Journal of Sports Medicine*, vol. 38, no. 12, pp. 2542–2551, 2010.

[9] C. M. P. Meylan, J. B. Cronin, J. L. Oliver, M. G. Hughes, and S. Manson, "An evidence-based model of power development in youth soccer," *International Journal of Sports Science & Coaching*, vol. 9, no. 5, pp. 1241–1264, 2014.

[10] Z. M. Darus, A. K. A. Rashid, N. A. Hashim, Z. Omar, M. Saruwono, and N. Mohammad, "Development of sustainable campus: universiti Kebangsaan Malaysia planning and strategy," *WSEAS Transactions on Environment and Development*, vol. 5, no. 3, pp. 273–282, 2009.

[11] R. Agricola, J. H. J. M. Bessems, A. Z. Ginai et al., "The development of cam-type deformity in adolescent and young male soccer players," *The American Journal of Sports Medicine*, vol. 40, no. 5, pp. 1099–1106, 2012.

[12] C. Romero and S. Ventura, "Data mining in education," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 12–27, 2013.
[13] R. Baker, "Data mining for education," *International encyclopedia of education*, vol. 7, no. 3, pp. 112–118, 2010.
[14] S. Olafsson, X. Li, and S. Wu, "Operations research and data mining," *European Journal of Operational Research*, vol. 187, no. 3, pp. 1429–1448, 2008.
[15] S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," *Applied Artificial Intelligence*, vol. 17, no. 5-6, pp. 375–381, 2003.
[16] B. Thuraisingham, "A primer for understanding and applying data mining," *It Professional*, vol. 2, no. 1, pp. 28–31, 2000.
[17] C. Wang, Y. Zheng, J. Jiang, and K. Ren, "Toward privacy-preserving personalized recommendation services," *Engineering*, vol. 4, no. 1, pp. 21–28, 2018.
[18] C. H. Chen and C. Toumazou, "Personalized expert recommendation systems for optimized nutrition," *Trends in Personalized Nutrition*, vol. 12, pp. 309–338, 2019.
[19] K. Kesorn, W. Juraphanthong, and A. Salaiwarakul, "Personalized attraction recommendation system for tourists through check-in data," *IEEE Access*, vol. 5, pp. 26703–26721, 2017.
[20] D. Geiger and M. Schader, "Personalized task recommendation in crowdsourcing information systems — current state of the art," *Decision Support Systems*, vol. 65, pp. 3–16, 2014.
[21] K. L. O'Halloran, S. Tan, D. S. Pham, J. Bateman, and A. Vande Moere, "A digital mixed methods research design: integrating multimodal analysis with data mining and information visualization for big data analytics," *Journal of Mixed Methods Research*, vol. 12, no. 1, pp. 11–30, 2018.
[22] C. H. Wu, "Data mining applied to material acquisition budget allocation for libraries: design and development," *Expert Systems with Applications*, vol. 25, no. 3, pp. 401–411, 2003.
[23] Z. Huang, D. Zeng, and H. Chen, "A comparison of collaborative-filtering recommendation algorithms for E-commerce," *IEEE Intelligent Systems*, vol. 22, no. 5, pp. 68–78, 2007.
[24] L. Xie, W. Zhou, and Y. Li, "Application of improved recommendation system based on spark platform in big data analysis," *Cybernetics and Information Technologies*, vol. 16, no. 6, pp. 245–255, 2016.
[25] Z. Liu, Y. Ma, H. Zheng, D. Liu, and J. Liu, "Human resource recommendation algorithm based on improved frequent itemset mining," *Future Generation Computer Systems*, vol. 126, pp. 284–288, 2022.
[26] W. Gang, A. Gunasekaran, E. Ngai, and T. Papadopoulos, "Big data analytics in logistics and supply chain management: certain Investigations for research and applications," *International Journal of Production Economics*, vol. 176, pp. 98–110, 2016.
[27] D. Opresnik and M. Taisch, "The value of Big Data in servitization," *International Journal of Production Economics*, vol. 165, pp. 174–184, 2015.
[28] Y. Sun, H. Song, A. J. Jara, and R Bie, "Internet of things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, 2016.
[29] M. Hong, M. Razaviyayn, Z. Q. Luo, and J. S. Pang, "A unified algorithmic framework for block-structured optimization involving big data: with applications in machine learning and signal processing," *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 57–77, 2016.