Computer Fraud and Security
ISSN (online): 1873-7056

Construction of Information Security Model for Talent Technology Training Based on Improved Machine Learning Algorithms

Weixue Han [1,2,*], Jingjing Duan [1,2]
[1]*School of Economics,Wuhan Business University , Wuhan,Hubei,430058, China*
[2]*Technical University of Ostrava, Ostrava-Poruba, 70833,Czech Republic*
**Email:Hmomo1134@163.com*

Abstract:

In order to improve the training quality of technical and skilled talents, this paper combines big data technology and machine learning technology to study the training of skilled talents, proposes corresponding data mining algorithms based on the characteristics of technical talents data, and combines machine learning algorithms to improve the data processing process. Moreover, this paper combines the improved algorithm to construct technical and skilled talents training and evaluation model, analyzes the system structure framework and basic logical structure, and obtains the system function structure. After building an intelligent system, this paper explores the actual effects of the training model of technical and skilled talents based on big data and machine learning. Finally, this paper combines simulation experiments to verify the system separately. Through data analysis, it can be known that the method and system proposed in this paper have a certain feasibility.

Keywords: big data; machine learning; skills; talent training;information security

1 INTRODUCTION

The rapid development of science and technology has led to a drastic change in the technological basis of modern production. At this time, a new group of industries is growing rapidly, and its proportion in the gross national product continues to rise. At the same time, due to its penetration and influence, many traditional industries have been reborn and developed. Moreover, changes in the industrial structure have brought about a series of chain reactions and major changes in the labor employment structure, regional economic structure, production organization and social structure, and even international economic relations [1]. This kind of basic technology that has caused a sharp change in contemporary production is high and new technology. The collection of enterprises based on high and new technology and engaged in the research, development, production and technical services of one or more high and new technologies and their products is a high-tech industry. The key technologies possessed by this industry are often difficult to develop, but once they are successfully developed, they have higher economic and social benefits [2]. With the advent of international economic globalization and the knowledge economy, high-tech industries are developing vigorously all over the world, and their development has become the focus of attention and competition among countries all over the world. The economic competitiveness of a country depends to a large extent on the ability and level of high technology and its industrialization. The competition of high and new technology seems to be the competition of products, labor services, and technology in form, but in essence it is the competition of human capital, especially the competition of high-tech talents with high technology [3].

This article combines big data technology and machine learning technology to study the training of skilled talents, and proposes a scientific talent training system to provide references for subsequent training of skilled talents.

2 RELATED WORK

The shortage of skilled personnel is an issue of great concern to all sectors of society. The academic community has also put forward many constructive methods on how to increase the training of highly skilled personnel. The literature [4] proposed that the society’s misplaced evaluation of the value of high-skilled talents is the root cause of their shortage. The literature [5] put forward suggestions on how to reduce the supply and demand of high-skilled talents to develop high-skilled human resources, such as establishing modern enterprise education and giving full play to the role of the main body of the enterprise. The literature [6] pointed out that compared with the cost of "one-off" academic education, the cost of developing high-skilled talents is higher, which is also one of the main reasons why most companies are unwilling to conduct employee training. The literature [8] pointed out that one of the main reasons leading to the shortage of high-skilled talents is people's outdated concepts and the high risk of dedicated human capital investment. The literature [9] believed that people's emphasis on academic qualifications and negligence of skills, poor remuneration for technical talents and other reasons are the fuse that caused the shortage of high-skilled talents. In summary, the shortage of high-skilled talents in China is mainly due to people's outdated ideas and the relatively high investment cost of high-skilled talents.

Literature [10] proposes that high-skilled talents are technicians and high-skilled workers who serve the production and service industries. The tasks they have to complete are highly complex and highly technical. These highly skilled personnel have mastered a relatively high theoretical basis and possess proficient operating skills. Literature [11] believes that only human resources who have mastered advanced operating skills and technology, can perform creative work, and make positive contributions to spiritual civilization and material civilization are highly skilled talents. Literature [12] believes that high-skilled talents are high-quality laborers who can complete key and highly difficult actions that are difficult for intermediate-skilled talents, and also possess theoretical knowledge and practical ability, and have innovative capabilities and can operate modern equipment. Literature [13] believes that the concept of high-skilled talents is broad and comprehensive, and that high-skilled talents are creative talents with comprehensive qualities. Literature [14] refers to those who have obtained the professional certification and corresponding ranks of senior workers, technicians and senior technicians in the front line of production and service as high-skilled talents. These personnel can solve key operation techniques and master specialized operation skills and knowledge theories.

The literature [15] shows that "accidents" can cause talents to gather. The core idea of the theory is that the centripetal force generated by the interaction between transportation costs, increasing returns, and the flow of factors is a model that causes the accumulation of talents and industries. The key to this model is to maintain a high degree of flexibility and sensitivity in the level of income in the flow of talents. Literature [16] believes that enterprises or companies are clustered in connected or unconnected departments. Literature [17] analyzes the externality of knowledge external economy, production cost, and transaction cost, and elaborates the horizontal talent gathering mode of external economy and knowledge sharing. This is the knowledge sharing and knowledge sharing under the premise of talent gathering. The core of knowledge externalization. The literature [18] shows that the important explanatory variable of vertical talent accumulation is transaction cost. He proposed that under the premise of the interaction of information cost, transfer cost, transaction cost, etc., the theoretical basis of vertical talent gathering is information cost and transaction cost. Literature [19] has done research on the vertical talent gathering mode, and believes that high-tech will form a strong radiation force for talent gathering, and the vertical talent gathering mode can be derived from the structure of transaction activities in the production process. Literature [20] believes that the power of talent gathering is affected by five factors, namely, the number of opportunities, the imitation of employers, the ability of employers to recognize and select talents, the status of intermediaries and economic agents, and the space for future improvement. . The literature [21] pointed out that the main centripetal force of talent gathering includes various factors such as the #l-department scale economy, knowledge spillover effect and so on.

3 TALENT TRAINING QUALITY ANALYSIS ALGORITHM BASED ON BIG DATA AND MACHINE LEARNING

First, a constructive covering algorithm is proposed. Based on the M-P neuron model, a rule based on domain coverage is derived. This method has the advantages of high recognition rate and fast calculation speed.

The main process of the coverage algorithm is: The algorithm is given an input skill talent sample set $K = \{x^1, x^2, \ldots, x^k, x^l \in R^n\}$ (K is a point set of n-dimensional Euclidean space), suppose K is divided into s subsets $K^1 = \{x^1, x^2, \ldots, x^{m(1)}\}$,...,$K^s = \{x^{m(s-1)+2}, x^{m(s-1)+2}, \ldots, x^k\}$, and the output of the skill talent sample points belonging to K^i are all y^i. Among them, $y^i = (0, \ldots, 1, 0, \ldots, 0)$ (that is, the vector in which the i-th component is 1, and the remaining components are O), i=1, 2,..., s, that is, there are s cases in which the output of each skilled talent sample. According to the following process, the spherical domain coverage of the i-th skill talent sample K' is constructed: the skill talent sample point $x^i \in K^i$ that has not been covered in the skill talent sample set is randomly taken out.

$$d_1 = max\{< x, x^i >\}, x \notin X^i \quad (1)$$

$$d_2 = min\{< x, x^i > | < x, x^i >> d_1\}, x \notin X^i \quad (2)$$

$$r = (d_1 + d_2)/2 \quad (3)$$

Among them, d_1 represents the distance between the center of the current class x^i and the nearest sample point of skilled talents in different classes, and d_2 represents the distance between the center of the current class and the farthest sample point of skilled talents in the same class when the distance is small d_1. $< x, x^i >$ represents the inner product of the vectors x and x^j, and r is used as the radius to construct the spherical domain coverage. Then, the center of gravity of all skilled talent samples in the coverage is solved. After that, we take this center of gravity as the class center and reconstruct the coverage area according to the operations in operations (1), (2), and (3) until the new coverage cannot cover more skilled talent sample points. Then, we perform a translation operation on the coverage center, and continue to repeat the operations in (1) (2) (3) to reconstruct the coverage area. Similarly, until the new coverage cannot cover more skilled talent sample points, the spherical domain coverage of K' is obtained. Among them, the specific translation algorithm refers to the literature. The above method is repeated to obtain the coverage set of all fields of the sample set of skilled talents.

It is known that the user set is $U = \{u_1, u_2, \dots, u_m\}$, the service set is $S = \{s_1, s_2, \dots, s_n \quad\}$, the target user is u, and the target service is s.

(1) The specific steps of the user-based recommendation method are as follows:

Step 1. User v and target user u have selected the same service, and the similarity Sim(u,v) between target user u and user v is calculated by PCC. The calculation formula is as follows:

$$Sim(u, v) = \frac{\sum_{s \in S_{u,v}} (R_{u,s} - \overline{R_u})(R_{v,s} - \overline{R_v})}{\sqrt{\sum_{s \in S_{u,v}} (R_{u,s} - \overline{R_u})^2} \sqrt{\sum_{s \in S_{u,v}} (R_{v,s} - \overline{R_v})^2}} \quad (4)$$

Among them, $S_{u,v}$ represents the set of services jointly called by user u and user v, $R_{u,s}/R_{v,s}$ respectively represents the evaluation value of user u/v on service s, and $\overline{R_u}/\overline{R_v}$ represents the average value of user u/v's evaluation value of all services called respectively.

Step 2. Based on the similarity information between users calculated in Step 1, the algorithm selects the first k users to form the target user's similar user set S(u);

Step 3. After obtaining the similar user set Su) of the target user through Step 2, the collaborative filtering recommendation method based on similar users predicts the QoS value of the target user missing in the user-service matrix. The prediction formula is as follows:

$$P_u(u, s) = \overline{R_u} + \frac{\sum_{v \in S(u)} Sim(u,v)(R_{v,s} - \overline{R_v})}{\sum_{v \in S(u)} Sim(u,v)} \quad (5)$$

Step 4. After predicting the evaluation value in Step 3, the algorithm selects the service with the best evaluation value to recommend to the target user, and finally completes the high-quality service recommendation that meets the user's needs.

(2) The specific steps of the service-based recommendation method are as follows:

Step 1. Service i and service j have been called by the same user. The similarity Sim(i,j) between the service owner and service j is calculated by PCC. The calculation formula is as follows:

$$Sim(i, j) = \frac{\sum_{u \in U_{i,j}}^{\Sigma} \left(R\overline{R}i_{u,i}()(R_{u,j} - \overline{R_j}) \right)}{\sqrt{\sum_{u \in U_{i,j}} (R_{u,j} - \overline{R_j})^2} \sqrt{\sum_{u \in U_{i,j}}^{\Sigma} \left(R\overline{R}i_{u,j}()^2 \right)}} \quad (6)$$

Among them, $U_{i,j}$ represents the set of users who have called service i and service j at the same time, and $\overline{R_i}/\overline{R_j}$ respectively represents the average value of the evaluation values of all users who have called service visit i/j.

Step 2. Based on the similarity information between services calculated in Step 1, the algorithm selects the first k services to form the similar service set si) of the target service:

Step 3. After obtaining the similar service set Si of the target service through Step 2, the collaborative filtering recommendation method based on similar services predicts the evaluation value of the target service missing in the user-service matrix. The prediction formula is as follows:

$$P_i\left(u, i()\overline{R_i}\frac{\sum_{j\in S(i)} Sim(i,j)(R_{u,j}-\overline{R_j})}{\sum_{j\in S(i)} Sim(i,j)}\right) \quad (7)$$

Step 4: After the evaluation value prediction in Step 3, the service with the best evaluation value is selected and recommended to the target user, and the high-quality service recommendation that finally meets the user's needs is completed.

(3) The specific steps of the recommended method of mixing are as follows:

In order to make full use of the information of similar users and similar services, and alleviate the data sparsity problem of the user-service matrix, researchers have proposed a hybrid recommendation method. It uses the mixing factor a to integrate the user-based recommendation method and the service-based recommendation method, and the prediction formula is as follows:

$$P(u,i) = \lambda \cdot P_u(u,i) + (1-\lambda) \cdot P_i(u,i) \quad (8)$$

Among them, the value range of the mixing factor λ is [0,1], and the mixing factor λ can make the prediction adapt to different environments.

In this paper, we mainly choose 6 standard indicators to measure the performance of the clustering algorithm: (1) Davies-bouldins index (DBI); (2) Dunn validity index (DVI); (3) Normalized mutual information (NM); (4) Clustering loss function; (5) Silhouette index (SI); (6) SD index (SDI). 6 kinds of index calculation formulas are as follows:

$$DBI = \frac{1}{k}\sum_{i=1}^{k\sum} \underset{j=1}{max}\left(\frac{\overline{C_i}+\overline{C_j}}{\left\|\omega_i-\omega_j\right\|_2}\right) \quad (9)$$

$$DVI = \frac{\underset{0<m\neq n\geq k}{min}\left\{\underset{\forall x_i\in\Omega_m, x_j\in\Omega_n}{min}\{\|x_i-x_j\|\}\right\}}{\underset{0<m\leq k}{min}\left\{\underset{\forall x_i,x_j\in\Omega_m}{min}\{\|x_i-x_j\|\}\right\}} \quad (10)$$

$$NMI = I\left(X,Y()\sqrt{H(X)H(Y)}\right) \quad (11)$$

$$\phi_Y(C) = \sum_{x\in D} d^2(x,C) = \sum_{x\in D} \underset{i=1,2,\ldots,k}{min}\|x-c_i\| \quad (12)$$

$$SI = \sum_{0<i\leq k}\frac{b(i)-a(i)}{max\{a(i),b(i)\}} \quad (13)$$

$$SDI(k) = a.Scatt(k) + Dos(k) \quad (14)$$

When,

$$Scatt(k) = \frac{1}{k}\sum_{0<i\leq k}\frac{\|\sigma(v_i)\|}{\|\sigma(D)\|}, Dis(k) = \frac{D_{max}}{D_{min}\sum_{0<j\leq k}\left[\sum_{0<j\leq k}\|v_i-v_j\|\right]^{-1}} \quad (15)$$

Among them, k represents the number of categories. In DBI, $\overline{C_i}$ represents the average distance from each point in the i-th cluster to the cluster center, and $\left\|\omega_i-\omega_j\right\|$ represents the distance between the two cluster centers of cluster i and cluster j. The value interval of the SI indicator is [-1,1]. SDI is based on the average scattering of the clusters and the total separation of the clusters

(1) Test skill talent data set

In order to fully verify the effectiveness of the C-K-means algorithm, this paper selects 7 skilled talent data sets for data simulation experiment verification, as shown in Table 1.

Table 1 Description of 7 Skilled Talent Data Sets

The name of the skill talent dataset	Number of attributes	Number of skilled talent samples	Number of categories (k)
Iris	4	150	3
Wine	13	178	3
Abalone	8	4,177	29
GaussS	3	10,000	Unknown category
PAM	57	4,601	Unknown category
Cloud	10	1,024	Unknown category
Individual household	7	2,049,280	Unknown category

This article uses the maximum minimum standardization method, and the normalization formula is as follows:

$$x_i^j = \begin{cases} \dfrac{x_i^{\max} - x_i^j(ori)}{x_i^{\max} - x_j^{\min}}, if\ x_j^{\max} \neq x_j^{\min} \\ 1, otherwise \end{cases} \tag{16}$$

(2) Comparison algorithm

The recommended framework proposed in this paper is mainly divided into 4 stages.

1. Based on the historical data of user set U's evaluation of service S, the C-K-means algorithm proposed in this paper clusters users or services with similar QoS values into a cluster. After covering the clustering, the algorithm further performs granularity analysis on the clustering results to obtain the final clustering results, and obtains the user association matrix M_u and the service association matrix M_s by covering the clustering information.

2. The algorithm selects the similar user set A(u) of the target user u and the similar service set A(s) of the service based on M_u and M_s.

3. The algorithm predicts the QoS value of the target user u for the service i based on A(u), that is, $q_{u,i}(u)$. At the same time, the algorithm predicts the QoS value of the target user u for the web service i based on A(s), that is, $q_{u,i}(s)$. Finally, the algorithm combines $q_{u,i}(u)$ and $q_{u,i}(s)$ to predict the QoS value of the target user for u to the web service i, that is, $q_{u,j}$.

4. Based on the size of Q, the algorithm recommends the Web service with the best QoS value to the target user.

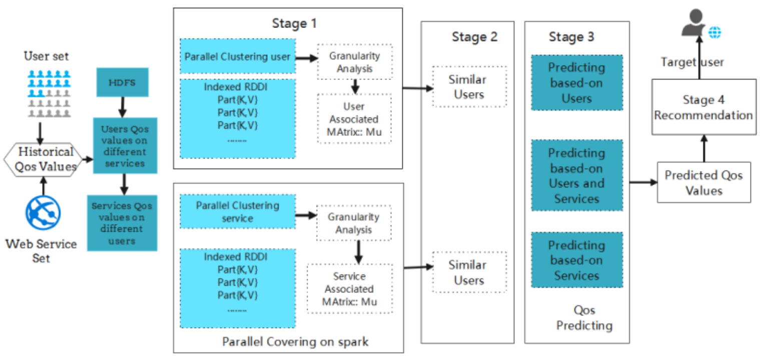

Figure 1 Recommendation framework based on C-K-means algorithm

We first divide the users in U into a data set of user experience skill talents who have had QoS values under each individual Web service. For any web service icS, this method converts all users in U who have QoS values for web service i and their QoS values into IndexedRDD. The algorithm performs parallel overlay clustering on

IndexedRDD, and obtains preliminary clustering results, and then performs granularity analysis on the clustering results. If the cluster granularity is not appropriate after clustering, it will affect the accuracy of the prediction. Therefore, according to the split operation and aggregation operation in the granularity analysis method, the relationship between the clusters is adjusted to obtain the final clustering result close to the optimal value. It groups users with similar QoS values into one category. As shown in Figure 2, there are 4 services in total. In order to describe the clustering process more concisely and clearly, we experiment with the one-dimensional attribute of User Rating attribute. For example, in si, according to the QoS evaluation values of all users (users) on s, the algorithm performs coverage clustering on these users, and groups users with similar QoS evaluation values to s into one category, and obtains {unyu6,us}, { ue} and {u2, u4, u3,us, u}. After the coverage clustering is over, the algorithm combines with the granularity analysis to split the clusters with larger standard deviations and merge the similar clusters, as shown in Figure 3.

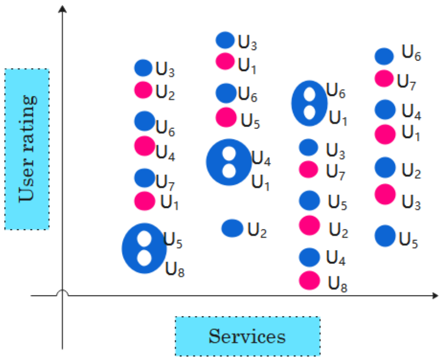

Figure 2 User-based traditional coverage

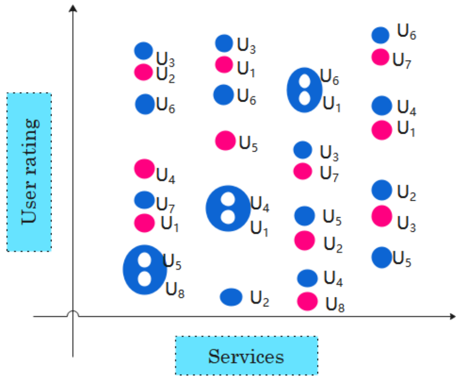

Figure 3 User coverage based on granular analysis

Before clustering, the algorithm first initializes M_u. More, the initial value of all elements in M_u is 0. After that, the algorithm performs a parallel coverage algorithm on the set of users U_s who has called the service s. After the clustering is over, the algorithm performs granularity analysis, and updates the elements in M according to the results of the clustering that has undergone granularity analysis. Each coverage clustering needs to update the elements in the matrix in time. After all coverage clustering ends, the algorithm analyzes the coverage information obtained and generates M_u. Similarly, our method also clusters the QoS values of all the Web services that the user has called for each user. For any user, Web services with similar QoS values are grouped together. The specific process of the service-based parallel overlay clustering algorithm is similar to that of the parallel user-based overlay clustering algorithm, which also obtains MS.

In coverage clustering, u_{ij} represents the number of times that users u_i and u_j are clustered into the same coverage. After obtaining Mu, the algorithm then performs an ascending operation on the elements of each row in Mu, and selects the first k users to form the similar user set $A(u_i)$ of the target user u, and $A(u_i) = \{u_a | u_{ia} \geq u_{ik}, u_{ia} > 0, a > 0, u_a \in U, u_{ik} \in Mu\}$.

We use an example to explain the determination of the similar neighbor set A(u) of the target user. Equation 17 gives a user incidence matrix model. We take user u_9 in matrix M_u as an example for analysis. When selecting similar users of user u_9, we find that the priority relationship of other users relative to u_9 is: $u_8 > u_5 > u_2, u_4, u_6 > u_1, u_3, u_7$. When selecting similar users based on their priorities, there are multiple ways to assign weights to different users. For example, we can assign a set of weights $w_1, w_2, \ldots, w_9$ to users $u_1, u_2, \ldots, u_8$. Among them, $u_8 > u_5 > u_2, w_4, w_6 > w_1, w_3, w_7$, and $\sum_{i=1}^{k} \omega_i = 1$. C-K-means selects users who have been covered in the same coverage more frequently in the incidence matrix M_u as candidates for similar neighbors. When it performs an ascending order operation on the number of times in each row of M_u, it sets Top-k=3, and obtains the similar user set of the target user according to the result of the ascending order. This article takes user u_9 as an example. Similar user set $A(u_9) = \{u_8, u_5, u_2\}$ uses user coverage information to select similar users and obtain effective similar users. This can reduce the impact of untrusted user information on prediction accuracy.

$$M_u = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 3 & 1 & 1 & 2 & 0 & 0 & 1 \\ 0 & 3 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 & 2 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 2 & 2 \\ 1 & 2 & 0 & 0 & 1 & 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & 2 & 0 & 3 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 2 & 0 & 1 & 0 & 3 \\ 0 & 1 & 0 & 1 & 2 & 1 & 0 & 3 & 0 \end{bmatrix} \quad (17)$$

Contrary to finding A(u), the algorithm finding A(s) is to cover and cluster the Web services called by each different user. Therefore, after the coverage clustering is over, the algorithm performs an ascending operation on each row element in the matrix Ms, and selects the first k web services to form the similar service set $A(s_i)$ of the target service si, and $A(s_i) = \{s_b | s_{ib} \geq s_{ik}, s_{ib} > 0, b > 0, s_b \in S, s_{ik} \in Ms\}$.

After determining k_u users similar to the target user u (that is A(u)), the algorithm aggregates the QoS values they perceive on the Web service and predicts the missing QoS values of the target user for the Web service. That is, formula (18) is used as a user-based parallel coverage clustering method to predict the QoS value of user u for Web services s:

$$q_{us}(u) = \frac{\sum_{t=1}^{k_u}\left(r_{a_u(t)} \times n_{a_u(t)}\right)}{N_u} \quad (18)$$

Similar to user-based parallel coverage clustering, after determining k services similar to the Web service (that is A(s)), aggregate the QoS values perceived by the user on the Web service, and predict the target user's missing QoS for the Web service value. And it uses formula (19) as a service-based parallel coverage clustering method to predict the QoS value of user u for web service s:

$$q_{us}(s) = \frac{\sum_{t=1}^{k_u}\left(r_{a_s(t)} \times n_{a_s(t)}\right)}{N_s} \quad (19)$$

Among them, $a_s(t)$ represents the service in A(s), $r_{a_s(t)}$ represents the QoS value of user u for similar service $a_s(t)$, $na_s(t)$ represents the number of times that service s and its similar service $a_s(t)$ are clustered in the same coverage, and N_s represents the total number of times that all services in service s and A(s) are clustered in the same coverage.

In order to accurately predict the QoS value of users for Web services, and comprehensively explore the information of similar users and similar services, we combine the above two methods to predict the QoS of user u for service s. That is, formula (18) and formula (19) are combined to achieve more accurate prediction results through the fusion factor:

$$q_{us} = \lambda q_{us}(u) + (1 - \lambda) q_{us}(s) \quad (20)$$

The parameter λ enables the prediction to be adapted to different environments. Among them, when $\lambda = 1$, the algorithm uses formula (18) to make predictions. When $\lambda = 0$, the algorithm uses formula (19) to make predictions. If the user u has not invoked any service, and the service s has not been invoked by any user, the average value of all service QoS values is used as the final QoS prediction value. We use formula (21) to summarize the above situation:

$$q_{us} = \begin{cases} q_{us} = \lambda q_{us}(u) + (1 - \lambda) q_{us}(s), A(u) \neq \phi and A(s) \neq \phi \\ q_{us}(u), A(u) \neq \phi and A(s) = \phi \\ q_{us}(u), A(u) = \phi and A(s) \neq \phi \\ null, A(u) = \phi and A(s) = \phi \end{cases} \quad (21)$$

The algorithm uses formula (21) to predict the service quality value, and then selects the service with the best QoS value and recommends it to the target user u to complete the entire service recommendation.

4 INTELLIGENT TECHNICAL AND SKILLED TALENTS TRAINING QUALITY ASSESSMENT SYSTEM

The system adopts multi-layer architecture design, which is divided into Web service layer, core service layer, database layer and user interface layer. Among them, the core service layer can be subdivided into business appearance layer, business rules layer, data access layer, data entity layer and system configuration layer (Figure 4).

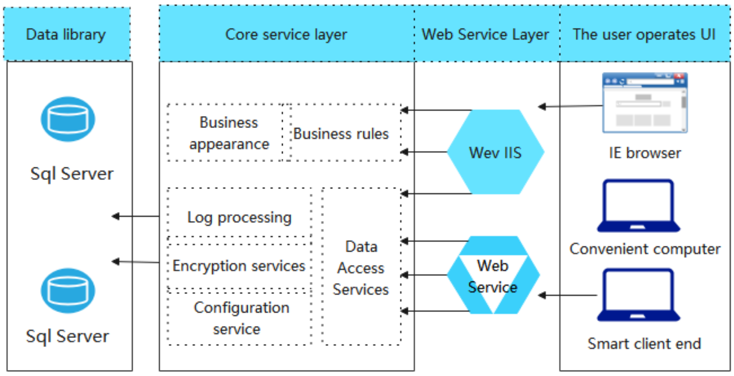

Figure 4 Hierarchical structure

The structure is shown in Figure 5.

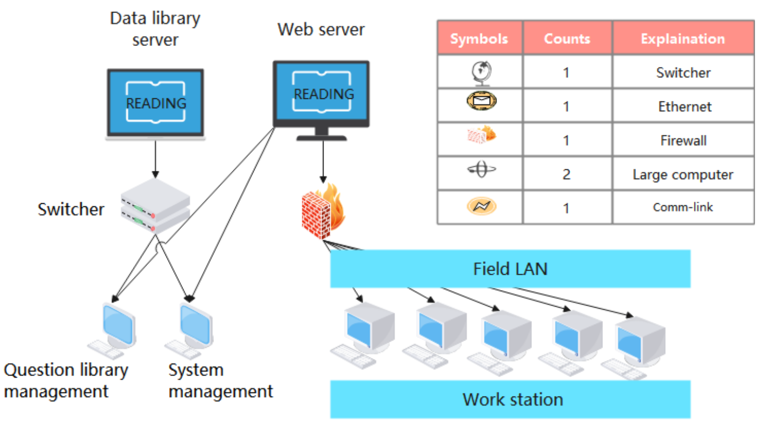

Figure 5 Unified deployment

To build a high-skilled personnel training mechanism, we must first clarify the formation process of high-skilled personnel and the factors that determine the work performance of high-skilled personnel. Based on the actual situation, this paper draws on the theory of competency and the formation mechanism of high-skills to formulate a general idea of the enterprise's high-skilled personnel training mechanism（ as shown in Figure 6）.

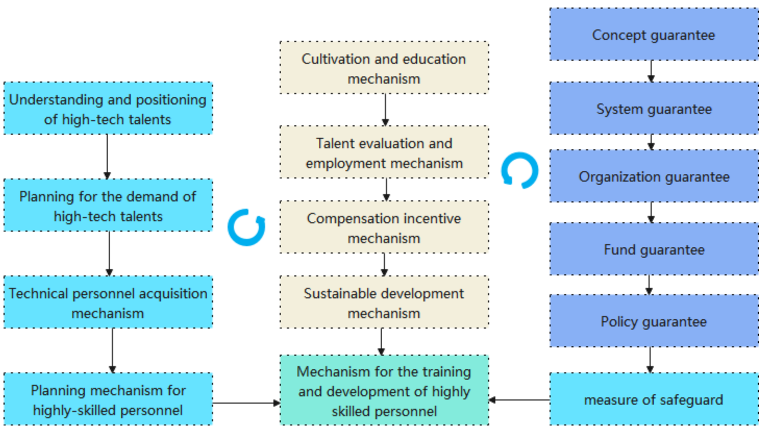

Figure 6 High-skilled personnel training mechanism diagram

This article establishes and completes the high-skilled personnel training and education mechanism, and combines the characteristics of technical and skilled talents to establish a "three-in-one" talent training and education mechanism of "work-study alternation, school-enterprise cooperation, knowledge, skills, and ability quality". The specific content is shown in Figure 7.

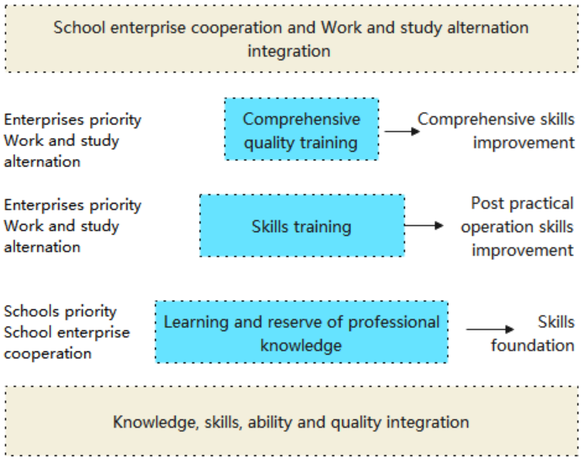

Figure 7 "Work-study alternation, school-enterprise cooperation, knowledge, skills, ability and quality three-in-one" training mechanism diagram

After constructing the above intelligent system, this paper explores the actual effect of the training model of technical and skilled talents based on big data and machine learning. This article combines simulation experiments to evaluate the effects of four aspects: data mining of the training of technical and skilled talents, data recommendation of the training of technical and skilled talents, assessment of the training of technical and skilled talents, and evaluation of the training technology of technical and skilled talents. The results obtained are shown in Table 1 to Table 4.

Table 1 Evaluation of the data mining effect of the technical and skilled talents training model

Number	Data mining	Number	Data mining	Number	Data mining	Number	Data mining
1	96.85	11	96.87	21	96.58	31	94.70
2	92.95	12	95.73	22	87.38	32	87.63
3	93.22	13	90.32	23	87.69	33	92.01
4	89.44	14	94.51	24	90.01	34	93.64
5	86.10	15	88.09	25	93.46	35	86.91
6	86.43	16	96.43	26	87.00	36	86.55
7	93.85	17	90.26	27	92.31	37	90.13
8	87.16	18	93.35	28	86.01	38	86.92
9	90.77	19	96.75	29	86.85	39	91.05
10	96.18	20	93.90	30	96.34	40	93.19

Table 2 Evaluation of the data recommendation effect of the technical and skilled talents training model

Number	Data recommendation	Number	Data recommendation	Number	Data recommendation	Number	Data recommendation
1	88.05	11	83.33	21	84.10	31	85.13
2	92.04	12	85.22	22	91.45	32	83.45
3	89.43	13	80.22	23	83.67	33	79.08
4	82.13	14	85.83	24	88.13	34	91.02
5	80.78	15	83.39	25	88.15	35	88.84
6	79.11	16	86.19	26	92.31	36	92.74
7	87.53	17	88.25	27	91.13	37	82.68
8	81.97	18	87.35	28	92.22	38	82.90
9	81.06	19	84.97	29	84.85	39	86.20
10	88.02	20	81.55	30	80.28	40	85.64

Table 3 Talent assessment and evaluation of technical and skilled talents training model

Number	Performance appraisal	Number	Performance appraisal	Number	Performance appraisal	Number	Performance appraisal
1	87.95	11	77.03	21	77.59	31	80.38
2	75.53	12	82.72	22	87.08	32	70.41
3	70.71	13	75.13	23	69.66	33	85.13
4	81.90	14	87.69	24	80.77	34	87.27
5	82.42	15	69.53	25	88.85	35	84.11
6	72.33	16	69.10	26	88.45	36	78.47
7	83.02	17	70.62	27	80.87	37	76.30
8	79.64	18	76.81	28	88.27	38	79.78
9	80.99	19	69.58	29	85.80	39	88.04
10	80.51	20	77.81	30	78.16	40	75.98

Table 4 Evaluation of the technical evaluation effect of the technical and skilled talents training model

Number	Summary evaluation	Number	Summary evaluation	Number	Summary evaluation	Number	Summary evaluation
1	75.48	11	77.53	21	81.12	31	84.52
2	81.35	12	86.99	22	71.52	32	73.05
3	88.76	13	91.59	23	83.99	33	84.84
4	84.27	14	87.63	24	77.90	34	81.75
5	86.40	15	79.65	25	87.73	35	73.88
6	73.63	16	80.85	26	74.25	36	90.73
7	71.70	17	85.68	27	71.57	37	73.34
8	87.17	18	75.08	28	91.15	38	74.64
9	83.40	19	84.29	29	84.74	39	75.89
10	88.83	20	89.85	30	90.45	40	73.97

From the above research, we can see that the technical and skilled talents training model based on big data and machine learning proposed in this paper basically meets the current society's demand for technical talents.

5 CONCLUSION

Highly skilled personnel are representatives of high-quality laborers. Therefore, to do a good job in the construction of a high-skilled talent team is conducive to creating an atmosphere where "everyone can become a talent" among the workers on the front line of production and service. Moreover, it is conducive to guiding and motivating the majority of workers to become talents, and is conducive to driving the echelon development of skilled workers. It has extremely important practical significance for the realization of the in-depth development of human resources and the improvement of the overall quality of workers. In recent years, the phenomenon of many companies robbing technicians has repeatedly appeared in the newspapers, and there is endless news about high-paying recruitment but not high-skilled talents. These phenomena truly reflect the reality of a serious shortage of highly skilled personnel in our country. Therefore, how to strengthen the training of China's high-skilled talents and transform China's large number of low-quality labors into high-quality skilled talents has become a core issue of China's human resources development. Moreover, it has become a key issue for promoting full employment, promoting enterprise development, and realizing national economic revitalization. This article combines big data technology and machine learning technology to study the training of skilled talents, and proposes a scientific talent training system to provide references for subsequent training of skilled talents.

ACKNOWLEDGE:

1.This article was supported by China Ministry of Education Industry-University Cooperation Collaborative Education Project 2022, Batch II, Business Data Analytics Teacher Training Project No. 220606396233419.

2.This article was also supported by the specific university research of the Ministry of Education, Youth and Sports of the Czech Republic No. SP2023/082.

REFERENCES

[1] Ayentimi, D. T., Burgess, J., & Dayaram, K. (2018). Skilled labour shortage: a qualitative study of Ghana’s training and apprenticeship system. Human Resource Development International, 21(5), 406-424.

[2] Muyia, M. H., Wekullo, C. S., & Nafukho, F. M. (2018). Talent development in emerging economies through learning and development capacity building. Advances in Developing Human Resources, 20(4), 498-516.

[3] Cascio, W. F. (2019). Training trends: Macro, micro, and policy issues. Human Resource Management Review, 29(2), 284-297.

[4] Iscandarov, R. R. (2018). Talent management as a method of development of the human capital of the company. Revista San Gregorio, (25), 107-113.

[5] Timmerman, E. A., Savelsbergh, G. J., & Farrow, D. (2019). Creating appropriate training environments to improve technical, decision-making, and physical skills in field hockey. Research quarterly for exercise and sport, 90(2), 180-189.

[6] Zhu, H. B., Zhang, K., & Ogbodo, U. S. (2017). Review on innovation and entrepreneurship education in Chinese universities during 2010-2015. Eurasia Journal of Mathematics, Science and Technology Education, 13(8), 5939-5948.

[7] Alayoubi, M. M., Al Shobaki, M. J., & Abu-Naser, S. S. (2020). Strategic leadership practices and their relationship to improving the quality of educational service in Palestinian Universities. International Journal of Business Marketing and Management (IJBMM), 5(3), 11-26.

[8] Cheng, X., Su, L., & Zarifis, A. (2019). Designing a talents training model for cross-border e-commerce: a mixed approach of problem-based learning with social media. Electronic Commerce Research, 19(4), 801-822.

[9] Clark, M. E., McEwan, K., & Christie, C. J. (2019). The effectiveness of constraints-led training on skill development in interceptive sports: A systematic review. International Journal of Sports Science & Coaching, 14(2), 229-240.

[10] Rasca, L. (2018). Employee experience-an answer to the deficit of talents, in the fourth industrial revolution. Calitatea, 19(S3), 9-14.

[11] Kolman, N. S., Kramer, T., Elferink-Gemser, M. T., Huijgen, B. C., & Visscher, C. (2019). Technical and tactical skills related to performance levels in tennis: A systematic review. Journal of sports sciences, 37(1), 108-121.

[12] Smirnova, Z. V., Zanfir, L. N., Vaganova, O. I., Bystrova, N. V., Frolova, N. V., & Maseleno, A. (2018). WorldSkills as means of improving quality of pedagogical staff training. International Journal of Engineering and Technology (UAE), 7(4), 4103-4108.

[13] Sarmento, H., Anguera, M. T., Pereira, A., & Araújo, D. (2018). Talent identification and development in male football: a systematic review. Sports medicine, 48(4), 907-931.

[14] Bennett, K. J., Vaeyens, R., & Fransen, J. (2019). Creating a framework for talent identification and development in emerging football nations. Science and Medicine in Football, 3(1), 36-42.

[15] Elferink-Gemser, M. T., & Hettinga, F. J. (2017). Pacing and self-regulation: important skills for talent development in endurance sports. International journal of sports physiology and performance, 12(6), 831-835.

[16] Issurin, V. B. (2017). Evidence-based prerequisites and precursors of athletic talent: a review. Sports Medicine, 47(10), 1993-2010.

[17] Paposa, K. K., & Kumar, Y. M. (2019). Impact of training and development practices on job satisfaction: A study on faculty members of technical education institutes. Management and Labour Studies, 44(3), 248-262.

[18] Ali, N. A. M., & Kasim, N. (2019). Talent management for Shariah auditors: case study evidence from the practitioners. International Journal of Financial Research, 10(3), 252-266.

[19] Rodriguez, J., & Walters, K. (2017). The importance of training and development in employee performance and evaluation. World Wide Journal of Multidisciplinary Research and Development, 3(10), 206-212.

[20] Semenova, L. V., Zaitseva, N. A., Rodinova, N. P., Glagoleva, L. E., Fursov, V. A., Radina, O. I., & Tsutsulyan, S. V. (2018). Improving the training of personnel based on the model of interaction between educational organizations and centers for assessing qualifications. Modern journal of language teaching methods, 8(3), 356-368.

[21] Dawi, A. H., Idris, I., Lambin, R., & Ibrahim, A. B. (2019). Preliminary Study on the Implementation of National Human Development Coaching in Malaysian Skills Diploma. International Journal of Academic Research in Business and Social Sciences, 9(9), 798-810.

[22] Dash, D., Farooq, R., Panda, J. S., & Sandhyavani, K. V. (2019). Internet of Things (IoT): the new paradigm of HRM and skill development in the fourth industrial revolution (industry 4.0). IUP Journal of Information Technology, 15(4), 7-30.

[23] Tri, N. M., Hoang, P. D., & Dung, N. T. (2021). Impact of the industrial revolution 4.0 on higher education in Vietnam: challenges and opportunities. Linguistics and Culture Review, 5(S3), 1-15.

[24] Anufrieva, N. I., & Tsarev, D. V. (2017). Methodological foundations for the artistic performance competence formation of the music students in the educational environment of a university. Contemporary Problems of Social Work, 3(3), 15-23.