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

Construction of Graduate Student Quality Policy System Based on NMIFS Algorithm

Lu Wang
School of Education, Jiangsu University of Technology, Changzhou,Jiangsu,210012,China
Email: wangwangL1345@163.com

Abstract:

In order to improve the graduate matriculate quality in the Internet era, this paper analyzes the policy system for ensuring the graduate matriculate quality in the Internet era, and improves the policy system for postgraduate student quality by combining intelligent data processing methods. This paper proposes a strategy based on NMIFS dimensionality reduction and threshold optimization. Moreover, this paper analyzes the impact of large-scale cubes on the construction of isolated trees, and proposes a dimensionality reduction strategy based on the NMIFS algorithm. In addition, this paper conducts a dimensionality reduction strategy comparison experiment on the four algorithms in projection transformation and feature selection, which verifies the effectiveness of the NMIFS feature selection dimensionality reduction strategy for the construction of isolated forests. The experimental study verifies that the policy system of graduate matriculate quality proposed in this paper can effectively improve the graduate matriculate quality.

Key words: Internet; security; graduate students; matriculate quality; policy system; NMIFS Algorithm

1 INTRODUCTION

Due to the rapid popularization of higher education, the number of students in colleges and universities has grown rapidly. However, the educational conditions, teaching staff and level, and management mechanism of higher education have not developed synchronously. Therefore, after the "enrollment expansion", there have been problems such as insufficient educational resources, relative decline in matriculate quality, and severe employment situation. These problems make people worry about the quality of higher education.

The training model is completely targeted. "Each model emphasizes certain variables, while ignoring and ignoring other variables. As a result, each model can only partially explain the reality of the organization." How to build a training model that is fully applicable to full-time professional degree and master's graduate education is the theme of this study. The postgraduate training process is a complex structural system formed by the interaction of different training subjects and training elements in different training links. The essence of the postgraduate training mode is a running representation of this complex structural system. The postgraduate training mode involves the proposition of "what kind of people to cultivate" and "how to cultivate" [1]. Based on the basic orientation of professional degrees to "cultivate high-level professionals in specific occupations", the training objectives, training methods, curriculum system, process supervision and management, and quality evaluation are regarded as the basic components of the training mode. The basic training subjects include training units, governments, and industries, and social evaluation institutions are added [2] .The full-time professional degree postgraduate training mode refers to the training objectives, training methods, curriculum system, process supervision and management, and quality evaluation as the basic elements under the guidance of certain educational ideas and theories, and under the interaction of training units, government, industry, social evaluation institutions and other training subjects, The training activities and operation mode formed to achieve the training goal of full-time professional master's degree [3]. The full-time professional degree postgraduate training system refers to an organic training system that, in a certain education system, forms five subsystems, including training target system, training structure system, training control system, training management system, and evaluation system, based on a clear training goal, and reflects the training category of educational activities [4].

"The relevant industry departments involved in professional degrees should gradually consider professional degrees as one of the conditions for the qualification of corresponding professional posts (positions)". There is no doubt that professional degrees have professional attributes. Professional degree postgraduate education is to achieve "specialty" based on "occupation". The professional attribute of full-time professional degree postgraduate education is to carry out educational activities according to the needs of specific occupations, and cultivate compound and application-oriented high-level talents with higher theoretical literacy and strong practical work ability required by specific occupational fields or related occupational fields [5].

From the background, full-time professional degree education is the product of the development of typical economic and social needs. "Production is a process in which all social organizations will transform their input into value added and output." From this definition, education, especially higher education, is a production activity in which colleges and universities add value to their students through training activities and export talents to the society [6].The training objective of cultivating the compound and application-oriented high-level talents required by specific occupational fields has clarified this production attribute. The occupational attribute is intrinsically related to the productive attribute. The acquisition of occupational attribute strengthens its productive attribute, which is also a manifestation of occupational attribute. "Production is not carried out in a closed way, and is always connected with the outside world", so the training process of full-time professional master's degree graduate education with production attributes is not closed, but is closely related to the external environment, especially the industry field or economic organization closely related to the occupation [7]. For the full-time professional degree postgraduate education with particularity, it is necessary to explore the training mode on the basis of considering the occupational attribute and production attribute. The particularity of full-time professional degree postgraduate education is inconsistent with the educational reality, which leads to the contradiction of marginal cultivation in the early stage of education, that is, in the existing academic degree postgraduate cultivation system, professional degree postgraduate cultivation is mainly carried out for fresh graduates [8]. The relatively mature training mode applicable to academic master's degree students and the relatively mature training mode applicable to part-time professional master's degree students are not completely suitable for the training needs of full-time professional master's degree students. In view of the inconsistency between the training needs and the training reality, a new training mode suitable for full-time professional master's degree students is constructed to form a gear coupling effect and become a research source [9]

In the construction of full-time professional degree postgraduate training system, literature [10] believes that it is necessary to build a full-time professional degree postgraduate curriculum system characterized by practice; Set up a full-time professional degree graduate teaching team focusing on ability; Establish a full-time professional degree thesis system with quality as the core; Cultivate the humanistic spirit of full-time professional degree graduates with the focus on management. According to the research on the construction of the training system for a specific specialty, literature [11] believes that the construction of the full-time veterinary master's degree training system must break through the training mode and evaluation mechanism for full-time academic master's students, reform classroom teaching and practical teaching, highlight the cultivation of practical and innovative abilities, and establish a three-level management system of tutors, schools and colleges. Literature [12] believes that in the process of establishing the full-time sports master degree postgraduate training system, it is necessary to clarify the talent training objectives, optimize the curriculum structure, and standardize the teaching content; Broaden the cultivation way and reform the practical teaching method; Strengthen the construction of professional degree graduate tutors. The cultivation of practical innovation ability is the most prominent in the cultivation of professional degree postgraduates. Relevant scholars put forward their own views on how to cultivate the construction system of professional degree postgraduates' practical innovation ability. Literature [13] believes that a series of discussions and innovations have been carried out from the curriculum system, teaching content and methods, teacher team construction, laboratory and practice base construction, research topics and professional skills training, assessment and evaluation standards and methods. The literature [14] believes that from the aspects of practical course teaching, professional practice base construction, double tutor mode, professional practice process management mechanism and assessment index system, degree thesis management, etc., we should explore the full-time professional degree postgraduate training mode and build a practical ability training system. The curriculum system is the basis for training professional degree graduates. In the construction of the curriculum system, scholars generally focus on the research of the teaching system, especially the research of the practical teaching system. For example, Wang Shuqi believes that the curriculum teaching system for full-time professional postgraduate training should start from adapting to the needs of the future career, put forward the curriculum positioning concept of the organic combination of targeted goals and adaptive goals, and put forward the idea of full-time professional postgraduate application-oriented curriculum setting from three aspects: building a large application curriculum system module, increasing the flexibility of postgraduate curriculum setting, and opening cross school courses, So as to adapt to the constant changes in the training process of professional master's students [15]. To build a full-time professional degree graduate practical teaching system, Literature [16]believes that it is necessary to update the practical teaching concept, plan the practical teaching work as a whole, and establish teaching and learning, theory and practice, in class and extracurricular The practical teaching operation mode integrating internal and external teaching. Literature [17] pointed out the problems in the institutional environment of professional degree talents training: the laws and regulations are not perfect, and the legitimacy needs to be strengthened; National policies are treated differently and social recognition is low; The specialty setting lacks planning and the training mode is similar; The evaluation system is not clear and the market integration is low. There are some problems in the professional degree enrollment examination system: the autonomy of university enrollment is relatively small; The contradiction between the difference of enrollment conditions and the same training objectives; The examination methods are inconsistent, and the proportion of the first and second examinations is unscientific.

This paper analyzes the policy system for guaranteeing graduate matriculate quality in the Internet era, and combines intelligent data processing methods to improve the policy system for graduate matriculate quality.

2 DATA PROCESSING FOR MATRICULATE QUALITY EVALUATION

2.1 Data dimensionality reduction

This paper proposes the FSIF-HDLOF method, which uses the optimized iForest pruning method to prune away the high-density spatial data points of the original data set to the greatest extent, and outputs a small outlier candidate set for optimizing the accurate detection of LOF, thereby improving the detection accuracy, efficiency and generalization ability. This chapter is the pruning stage of the FSIF-HDLOF method, and proposes an isolated forest pruning method (FSIF) based on data dimensionality reduction and threshold optimization. The implementation of the pruning method is shown in Figure 1.

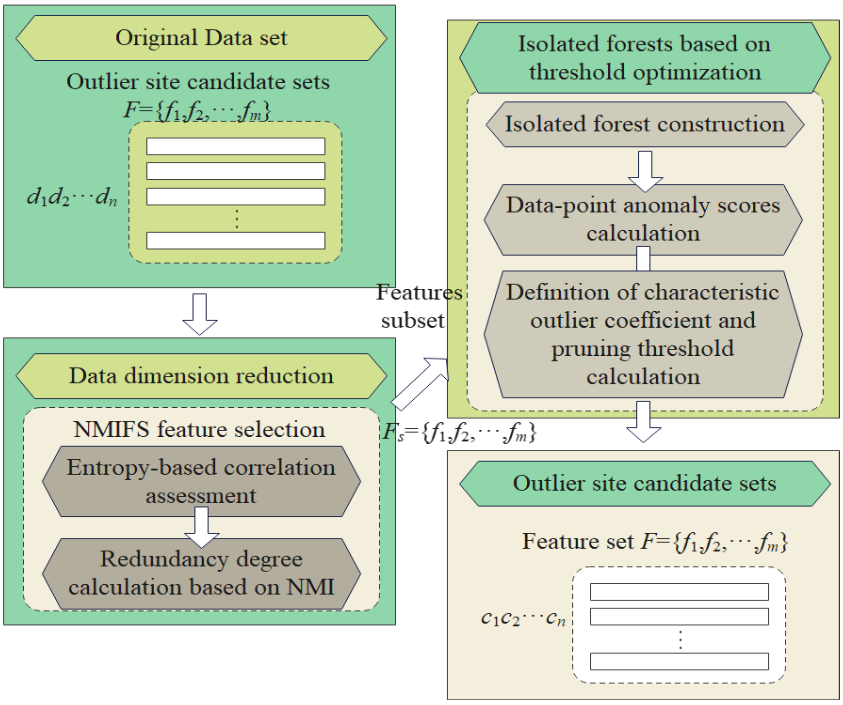


Figure 1 Schematic diagram of pruning method based on data dimensionality reduction and feature outlier coefficient

Here, the symbols above and below are defined:

1) The input original data set is $D: D = \{d_1, d_2, \cdots, d_n\}$, n is the total number of data points of D, OutliersN is the total number of outliers of D, d_i is the i-th data point of D, $i \in (1, n), d_i = \{x_{i1}, x_{i2}, \cdots, x_{im}\}, x_{ij}$is the jth

eigenvalue of data point d_i. The feature set of D is $F: F = \{f_1, f_2, \cdots, f_m\}, m$ is the total number of features of data set D, f_j is the j-th feature of data set D, $j \in (1, m), f_j = \{y_1, y_2, \cdots, y_n\}, y_i$ is the value of the j-th feature in the i-th data point, which is equivalent to x_{ij}. The feature subset F_s of D represents some features selected from F, which have the lowest redundancy and can effectively identify outliers in D. $F_s = \{f_1', f_2', \cdots, f_s'\}, s \in (1, m), f'_j = \{y_1', y_2', \cdots, y_n'\}, y_i'$is the value of the j-th feature in the i-th data point, which is equivalent to x_{ij}.

2) The outlier candidate set OC represents the set of remaining data points after pruning some obvious normal data points in D through the pruning stage. $OC = \{c_1, c_2, \ldots, c_l\}, l \in (1, n), l$ is the total number of data points for OC.

3) The outlier set O represents the set of outliers detected by the FSIF-HDLOF method. $O = \{p_1, p_2, \ldots, p_h\}, h \in (1, l), h$ is the total number of data points for O.

2.2 Optimization strategy for data dimensionality reduction

Mass in a massive cube usually has two meanings. One is the data scale, that is, the training data is sufficient to support the construction of complex algorithm models. The second is the data quality, that is, the data set full of noise or redundant information greatly affects the construction of the model. There are two problems with multidimensionality. 1) The features of different dimensions are related to each other. The concept of similarity measurement such as Euclidean distance no longer exists, and the data set characteristics are different under the views of different dimensions. As shown in Figure 2, the distribution of the same dataset under different dimensional views is different, and outliers A and B can only be separated from other data objects in a specific view on the left or right. Therefore, when projecting multidimensional data into a low-dimensional space, the most critical is how to select a specific subset of features to distinguish outliers from other data objects. 2) Due to the sparseness of data on multi-dimensional space and the presence of noise and irrelevant features, the distance between data objects cannot be measured in a meaningful way, making the outlier detection task more challenging. Therefore, the outlier detection technology needs to be improved according to the actual situation such as the dimension and scale of the dataset.

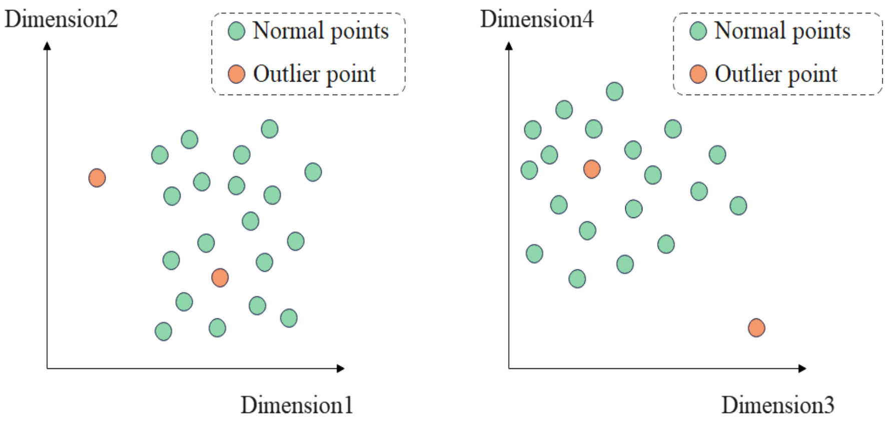


Figure 2 Schematic diagram of the distribution of data points in different dimension views

Information entropy is an evaluation index of feature correlation. The correlation Rel(f) of the feature f and its entropy value H(f) has the following relationship:

$$Rel(f) \propto \frac{1}{H(f)} \# \quad (1)$$

According to the knowledge of information theory, the information entropy measures the relative uncertainty of random variables, and the entropy H(f) of the discrete random feature f is:

$$H(f) = -\sum_j p(y_i) \log\left(p(y_j)\right) \# \quad (2)$$

It can be seen from the above that the most relevant feature for outlier detection is the feature with the smallest entropy value.

Computational Mutual Information (MI) based on NMI is a measure of the interdependence between two random variables, defined as:

$$I(x,y) = H(x) + H(y) - H(x,y)\# \quad (3)$$

For a given discrete random variable x and y, MI can be calculated by the following formula. Among them, p(x,y) is the joint probability distribution, and p(x), p(y) are the marginal probability distributions:

$$I(x,y) = \sum_i \sum_j p\left(x_i, x_j\right) log \frac{p(x_i,x_j)}{p(x_i)p(x_j)}\# \quad (4)$$

On the basis of MI, the redundancy of individual features and feature subsets is calculated by the following formula.

1) The normalized mutual information (NMI) between feature f_i and feature f_j, that is, the redundancy is:

$$R(f_i, f_j) = NMI(f_i, f_j) = \frac{I(f_i, f_j)}{min\{H(f_i), H(f_j)\}}\# \quad (5)$$

2) The average redundancy of feature f_i relative to feature subset feature F_s, that is, the average redundancy of feature f_i relative to each feature $f_j \in F_S$ is:

$$AR(f_i, F_s) = \frac{1}{|F_S|} \sum_{\forall f_j \in F_S} R\left(f_i, f_j\right)\# \quad (6)$$

3) The redundancy of feature subset F_S is the average of the average redundancy of each feature $f_j \in F_s$:

$$AAR(F_s) = \frac{1}{|F_S|} \sum_{\forall f_j \in F_S} A\,R(f_i, f_j)\# \quad (7)$$

2.3 Pruning method of isolated forest based on threshold optimization

The isolation forest algorithm uses subsampling and limiting tree height to achieve low linear time complexity of the algorithm, which can process large-scale data sets and effectively detect global outliers. Local outliers are easily masked by normal data clusters of similar density, and their detection is highly dependent on the neighborhood density. Therefore, this paper first starts from the global perspective, uses the isolated forest to prune data points in high-density space, and leaves a candidate set of outliers with a relatively marginal distribution, and the local outlier factors are further detected from the local density of the data(figure 3).

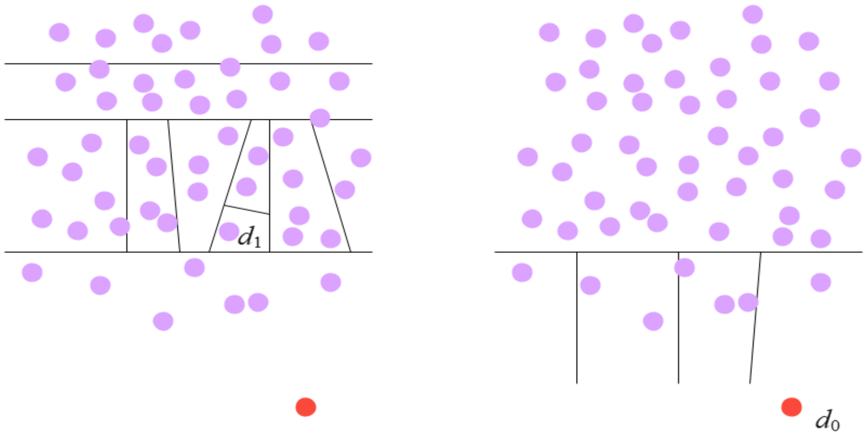


Figure 3 Normal points and outliers in two-dimensional data space

The feature outlier coefficient $Disp\_coe(f_i)$ of feature f_j is:

$$Dis\,p_{coe(f_j)} = \sqrt{\frac{\sum_{j=1}^{n}(y_j - \bar{y})^2}{n\bar{y}^2}}\# \quad (8)$$

The feature outlier coefficients of the m features of the data set are calculated in turn, and the data set feature outlier coefficient vector D_f is obtained, denoted as:

$$D_f = \{Disp\,p\_\,coe(f_1)\,, Disp\,p\_\,coe(f_2)\,, \cdots, Dispcoe(f_m)\}\# \quad (9)$$

According to formula (10), D_f is normalized, and the normalized data set feature outlier coefficient vector $D_{Norm,}$ is obtained, as shown in formula (11):

$$NDisp\_\,coe(f_i) = \frac{Disp_{C}oe(f_i)}{\sum_{i=1}^{m} \text{Disp}\_\,coe(f_i)}\# \quad (10)$$

$$D_{Norm} = \{N\,Dis\,p\_\,coe(f_1)\,, N\,Dis\,p\_\,coe(f_2)\,, \cdots, N\,Disp\,p\_\,coe(f_m)\}\# \quad (11)$$

The pruning threshold θ_D can be obtained by calculating the feature outlier coefficient vector. The anomaly score of each data point is calculated by the isolation forest algorithm and sorted in descending order, and the first $n * \theta_D$ data points with larger anomaly scores are classified into the outlier candidate set.

$$\theta_D = \frac{\alpha \sum D_{Norm} - Top(s)}{s}\# \quad (12)$$

We assume that outliers are located in low spatial density and are very rare. When constructing an isolated tree, outliers are quickly divided into leaf nodes, so they are closer to the root node and have smaller depth values. The anomaly score for a data point d on an isolated tree is defined as:

$$ScoreAtt(\,d) = 2^{-\frac{h(d)}{t.size}}\# \quad (13)$$

Traversing the isolated tree iTree after construction, we can know the number of layers of the tree where the data point d is located, that is, the depth h(d) of each data point:

$$h(d) = e + C(t.size)\# \quad (14)$$

C(t.size) is the correction value, which represents the average path depth of the binary tree constructed by t.size data points. The calculation formula is as follows, where ξ is the Euler constant:

$$C(t.size) = 2\,ln(\,t.size - 1) + \xi - \frac{2(t.size-1)}{t.size} \quad (15)$$

Traversing an isolated tree can calculate the anomaly score for data point d. Considering the sampling subspace and the random selection of features, the reliability of ScoreAtt(d) is very low. Therefore, the construction of the isolated forest can calculate the average path depth of each data point in multiple trees, and then obtain the average anomaly score Score(d) of the data point d. Among them, E(h(d)) represents the mean value of the data point d in the path depth of multiple iTrees, and C(w) is calculated as shown in formula (15), which represents the average path depth of the binary tree constructed by w data points.

$$Score(\,d) = 2^{-\frac{E(h(d))}{C(w)}}\# \quad (16)$$

2.4 Experimental results and analysis

To comprehensively evaluate the algorithm performance, all experiments in this paper are conducted on the five synthetic datasets and five real datasets introduced below.

1) Synthetic datasets: This paper selects five real classification datasets from the University of California Irvine repository (UCI) machine learning repository. First, this paper removes the category labels from the original data set, and secondly treats all data points as normal data points, and then performs random shift preprocessing. This paper uses the following standard contamination procedure to generate outliers: randomly select a subset of data points and shift all their eigenvalues by v standard deviations. The selected processed data points are marked as abnormal, and the remaining data records are marked as normal.

Yeast dataset: It is used to predict the localization site of proteins, including 8 attributes such as the signal sequence identification method of mcg and gvh.

EMG Physical Action (EMGPA) dataset: It contains electrode information generated by 10 normal and 10 aggressive actions of human activity.

Magic Gamma Telescope (MGT) data set: It records the registration of high-energy gamma particles simulated by imaging technology in a gamma telescope, including 10 attributes such as ellipse major and minor axes and gamma signals.

Avila dataset: It is extracted from 800 Latin Bible (AvilaBible) images and contains 10 attributes such as column spacing, top and bottom margins, and line spacing ratio.

KEGG Metabolic Relation Network (KEGG) dataset: It contains 23 network attributes such as edge integers, network diameters, and shortest paths.

In order to evaluate the performance of the pruning method based on data dimensionality reduction and threshold optimization (FSIF) proposed in this chapter, this paper conducts two parts of experiments on synthetic datasets and real datasets: performance evaluation of dimensionality reduction strategy and performance evaluation of pruning method. This section introduces their evaluation indicators and parameter settings.

This paper conducts an experimental comparison of dimensionality reduction methods from the following two aspects: NMIFS and Variance Threshold (VT) in feature selection. Projection transformation uses principal component analysis (PCA) and factor analysis (FA) algorithms. The above four dimensionality reduction methods are combined with the use of isolated forests, which are abbreviated as PCA-IF, FA-IF, VT-IF, and NMIFS-IF respectively. They are compared with the outlier detection effect obtained by the original isolated forest algorithm (in the experimental stage of this paper, IF is used to simplify the description of iForest) on the original data set. Different feature dimensions of the dataset will affect the effect of the isolation forest algorithm to detect outliers. Therefore, for different datasets, this paper selects five representative values after dimension reduction, including the original dimension (except the Smtp dataset), to find the best feature subset, so that the detection effect of the isolated forest is the best.

This section evaluates the performance of PCA-IF, FA-IF, VT-IF, and NMIFS-IF algorithms in terms of F-Measure metrics and dimensionality reduction experiment time cost.

The F-Measures of the above four algorithms on synthetic datasets and real datasets are shown in Figures 4 and 5.

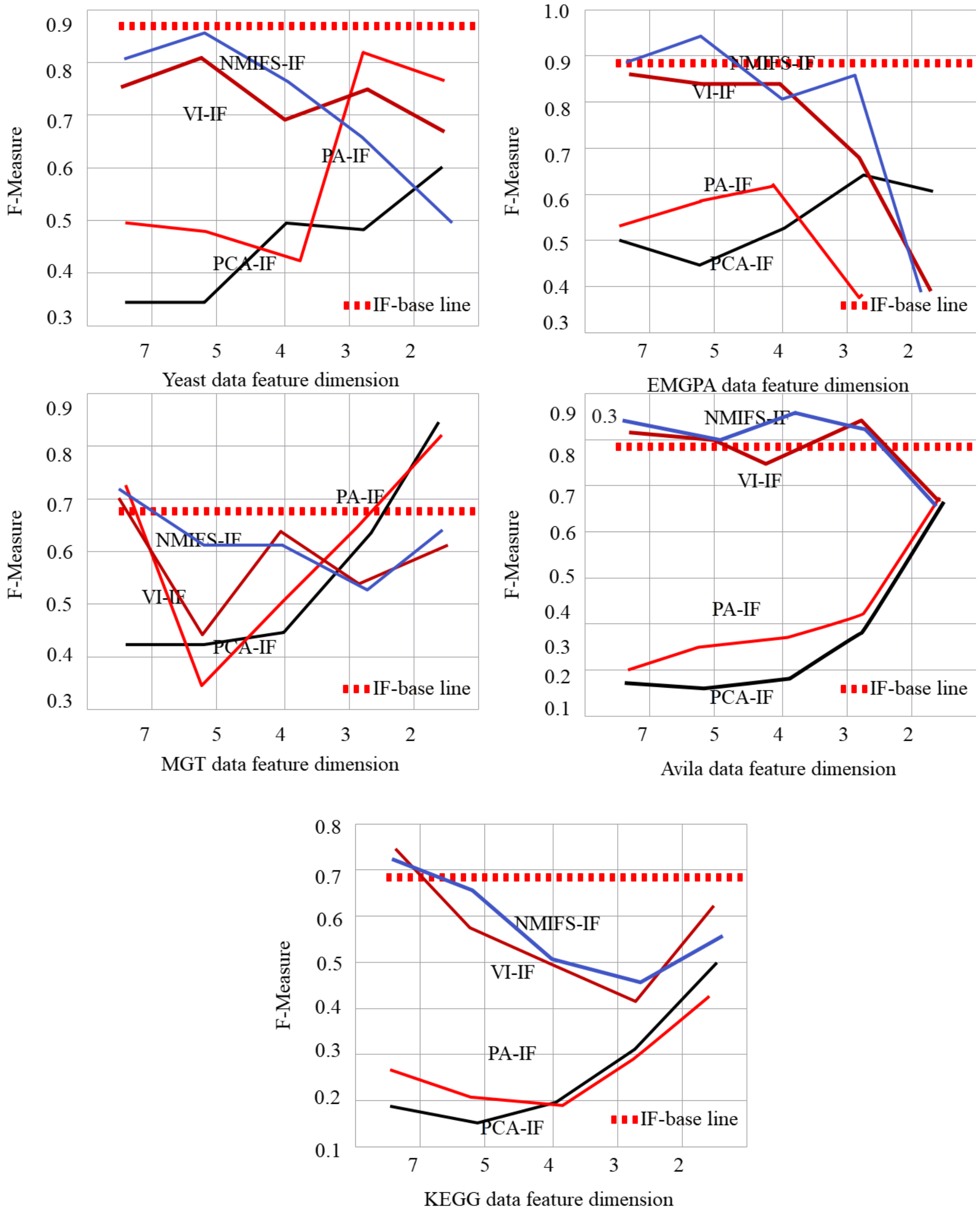


Figure 4 Comparison of F-Measure results of dimensionality reduction strategies on synthetic datasets

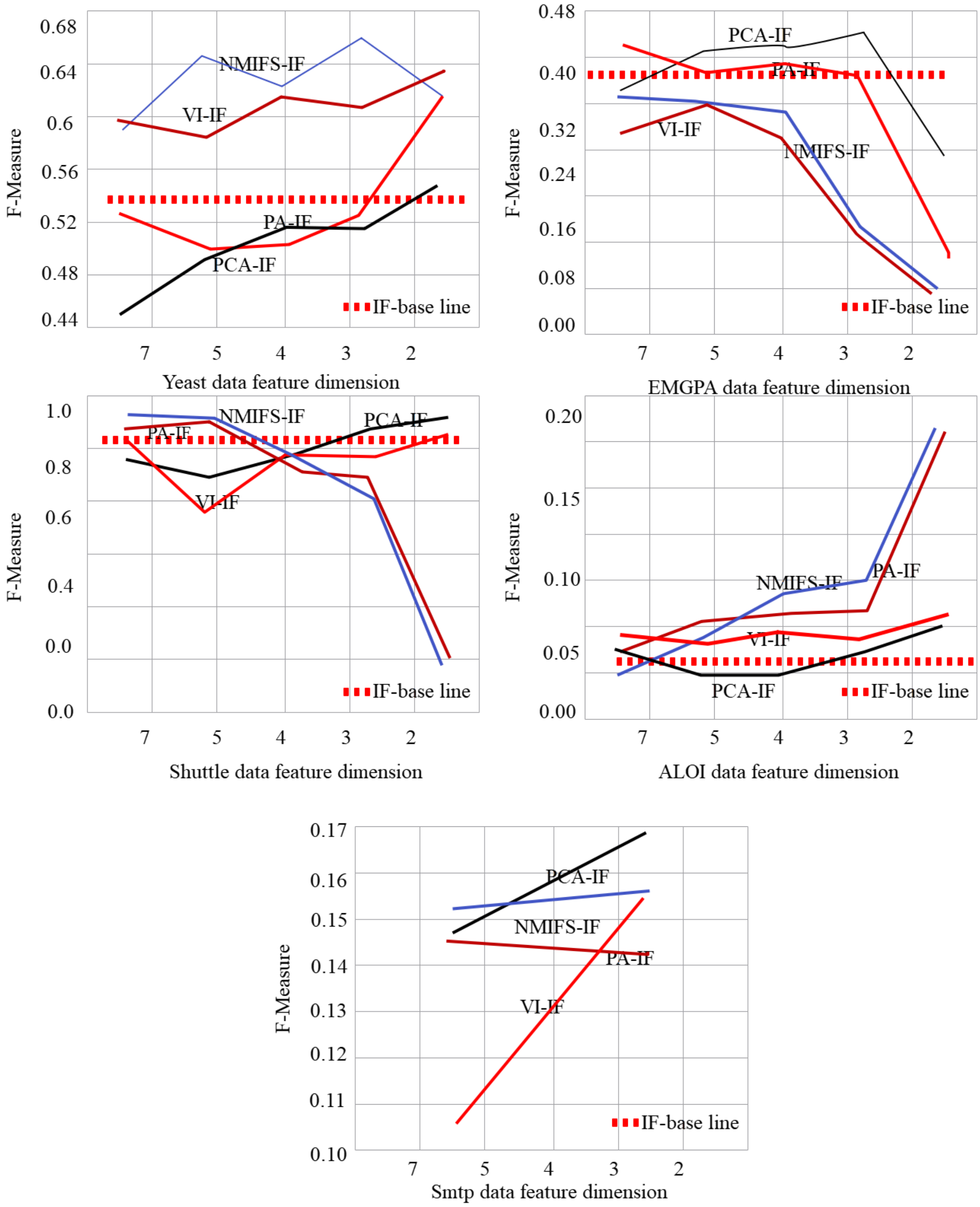
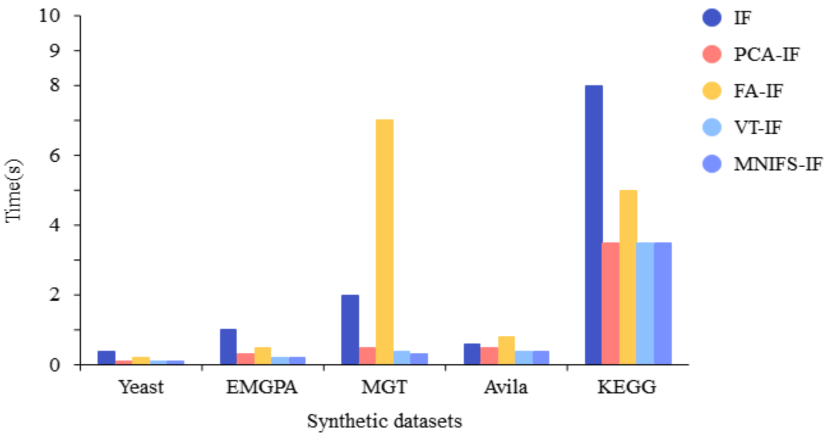


Figure 5 Comparison of F-Measure results of dimensionality reduction strategy on real data sets

In Figure 4, except for the Yeast and KEGG synthetic datasets, the F-Measure of the algorithm using the dimensionality reduction strategy is higher than the IF algorithm based on the original dataset in some dimensions. On the whole, the advantage of dimensionality reduction strategy is not obvious.

Figure 5 shows the F-Measures of PCA-IF, FA-IF, VT-IF, and NMIFS-IF in different dimensions on real datasets. On the remaining 4 datasets except the Shuttle dataset, the dimensionality reduction strategy produces better results for outlier detection in the isolation forest. Two algorithms of feature selection class, VT-IF and NMIFS-IF, outperform PCA-IF and FA-IF of projective transformation class on Satellite, Shuttle and ALOI.

Taking the dimension corresponding to the best detection effect of NMIFS-IF on each data set as an example, this paper calculates the time-consuming of the whole process experiment of IF, PCA-IF, FA-IF, VT-IF, and NMIFS-IF. It includes two parts: dimensionality reduction and detection, and the results are shown in Figure 6.



(a)

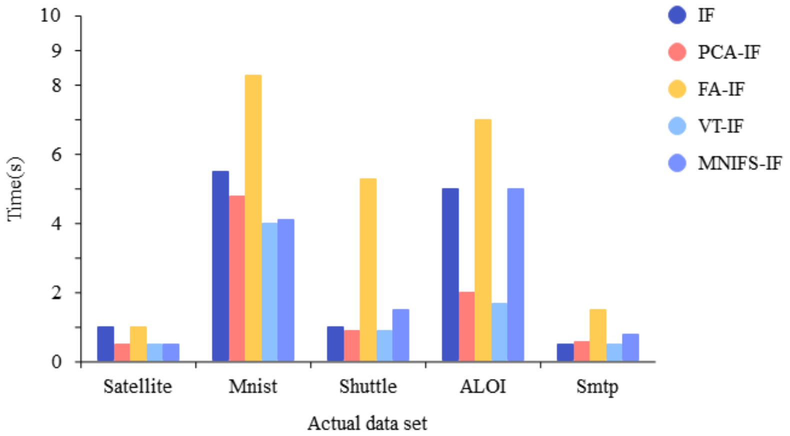


Figure 6 Time-consuming comparison chart of the whole process experiment of dimensionality reduction strategy on synthetic and real data sets

3 THE POLICY SYSTEM FOR GUARANTEEING GRADUATE MATRICULATE QUALITY IN THE INTERNET ERA IS PERFECT

The above methods are applied to the analysis of the policy system improvement of graduate matriculate quality. The construction of a full-time master of engineering training quality evaluation system can be started from three aspects. On the macro level, it is closely related to China current economic development, social needs, and related educational system and policy construction. At the mesoscopic level, discipline construction and school development are the internal driving forces for the construction of the evaluation system, while market orientation and competitive pressure are the external driving forces. The contents of enrollment selection, teaching resources, tutor team, engineering practice, training and awarding, and graduate tracking constitute the specific indicators of the evaluation system from the micro level (as shown in Figure 7).

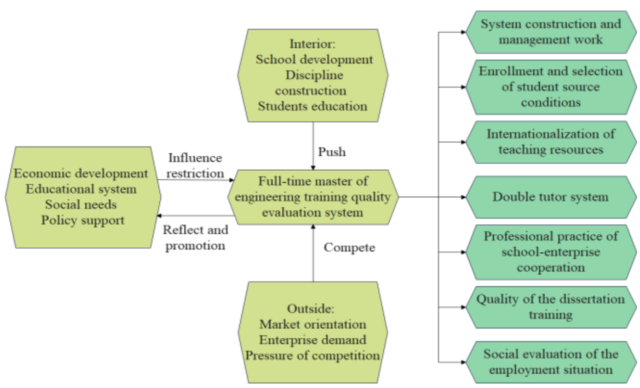


Figure 7 Evaluation system for the quality of master's training

Combined with some experience in the development of foreign professional degrees, focusing on the evaluation of the professional quality and professional quality of full-time professional degree graduate students can meet the needs of the country, society, and individuals for talent training. The specific evaluation indicators are shown in Figure 8.

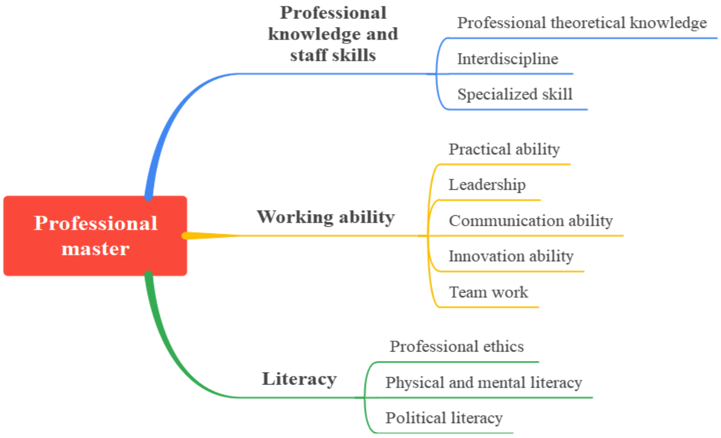


Figure 8 Comprehensive map of master training

The proposed adaptation strategy for the construction of the full-time professional degree postgraduate training system is as follows. One is to improve the training objectives based on diversified needs. The second is to introduce a market-oriented investment mechanism to increase the enthusiasm of training subjects to participate. The third is to reform the training mode and build an active and effective training quality assurance system. The fourth is to strengthen the construction of teaching staff and improve the personnel appraisal and promotion system. The policy system of graduate matriculate quality proposed in this paper is simulated and tested, and the test results shown in Table 1 are obtained.

Table 1 The detection effect of the policy system of graduate matriculate quality

Number	System evaluation	Number	System evaluation
1	83.927	16	81.464
2	86.894	17	79.405
3	86.673	18	85.526
4	82.983	19	81.910
5	88.479	20	81.630
6	80.017	21	81.188
7	86.116	22	80.778
8	86.122	23	85.829
9	84.904	24	87.326
10	87.656	25	84.206
11	85.243	26	81.800
12	84.907	27	87.026
13	84.435	28	83.434
14	88.550	29	84.664

15	79.893	30	84.189

The above research verifies that the policy system for graduate matriculate quality proposed in this paper can effectively improve graduate matriculate quality.

4 CONCLUSION

How to ensure the quality of higher education teaching has received extensive attention from all walks of life. To ensure the quality of higher education, the quality of majors is a key factor to be considered. The reason is that majors, as the basis for running schools and basic units of education and teaching, are the educational and teaching entities for cultivating specialized talents. The quality of the major reflects the education quality, management level and even the running level of the whole school from one aspect. Therefore, how to ensure high-quality professional education has also attracted the attention of educational administrative departments and colleges and universities, and professional quality assurance has also become a research hotspot. This paper analyzes the policy system for guaranteeing graduate matriculate quality in the Internet era. The experimental study verifies that the policy system of graduate matriculate quality proposed in this paper can effectively improve the graduate matriculate quality.

REFERENCES

[1] Liang, D., Liu, Q., Zhou, K., Jia, W., Xie, G., & Chen, T. (2020). IP4M: an integrated platform for mass spectrometry-based metabolomics data mining. BMC bioinformatics, 21(1), 1-16.

[2] Gupta, S., & Sabitha, A. S. (2019). Deciphering the attributes of student retention in massive open online courses using data mining techniques. Education and Information Technologies, 24(3), 1973-1994.

[3] Hu, H., Zhang, G., Gao, W., & Wang, M. (2020). Big data analytics for MOOC video watching behavior based on Spark. Neural Computing and Applications, 32(11), 6481-6489.

[4] Zhang, H., Huang, T., Liu, S., Yin, H., Li, J., Yang, H., & Xia, Y. (2020). A learning style classification approach based on deep belief network for large-scale online education. Journal of cloud computing, 9(1), 1-17.

[5] Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: A systematic evidence map. International journal of educational technology in higher education, 17(1), 1-30.

[6] Jensen, P. (2018). Semantically enhanced 3D: a web-based platform for spatial integration of excavation documentation at Alken Enge, Denmark. Journal of Field Archaeology, 43(sup1), S31-S44.

[7] Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., & Göker, M. (2022). TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic acids research, 50(D1), D801-D807.

[8] Sallow, A. B., Sadeeq, M., Zebari, R. R., Abdulrazzaq, M. B., Mahmood, M. R., Shukur, H. M., & Haji, L. M. (2020). An investigation for mobile malware behavioral and detection techniques based on android platform. IOSR Journal of Computer Engineering (IOSR-JCE), 22(4), 14-20.

[9] Korot, E., Guan, Z., Ferraz, D., Wagner, S. K., Zhang, G., Liu, X., ... & Keane, P. A. (2021). Code-free deep learning for multi-modality medical image classification. Nature Machine Intelligence, 3(4), 288-298.

[10] Qiu, L., Liu, Y., Hu, Q., & Liu, Y. (2019). Student dropout prediction in massive open online courses by convolutional neural networks. Soft Computing, 23(20), 10287-10301.

[11] Edastama, P., Dudhat, A., & Maulani, G. (2021). Use of Data Warehouse and Data Mining for Academic Data: A Case Study at a National University. International Journal of Cyber and IT Service Management, 1(2), 206-215.

[12] Nachman, B. R., & Brown, K. R. (2020). Omission and othering: Constructing autism on community college websites. Community College Journal of Research and Practice, 44(3), 211-223.

[13] Mandal, D., Kumar, V., Bhattacharya, A., Rao, Y. S., Siqueira, P., & Bera, S. (2018). Sen4Rice: A processing chain for differentiating early and late transplanted rice using time-series Sentinel-1 SAR data with Google Earth engine. IEEE Geoscience and Remote Sensing Letters, 15(12), 1947-1951.

[14] Karkar, A. J., Fatlawi, H. K., & Al-Jobouri, A. A. (2020). Highlighting e learning adoption challenges using data analysis techniques: University of Kufa as a case study. Electronic Journal of e-Learning, 18(2), pp136-149.

[15] Van Horne, S., Curran, M., Smith, A., VanBuren, J., Zahrieh, D., Larsen, R., & Miller, R. (2018). Facilitating student success in introductory chemistry with feedback in an online platform. Technology, Knowledge and Learning, 23(1), 21-40.

[16] Raj, S., & Ray, K. C. (2018). A personalized point-of-care platform for real-time ECG monitoring. IEEE Transactions on Consumer Electronics, 64(4), 452-460.

[17] Jones, K. M. (2019). Learning analytics and higher education: a proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. International Journal of Educational Technology in Higher Education, 16(1), 1-22.