

Prediction of College Student Employment Characteristics Based on Deep Manifold Learning

Wenzhi Su*

Big Data Integration and Intelligent Manufacturing, Jiyuan Vocational and Technical College, Jiyuan 459000, China

The employment of college students is of great significance to the development of the national economy. Due to the information gap between enterprises and universities, the use of advanced artificial intelligence technology to analyze and predict student information can provide reliable employment positioning references and effectively alleviate the current tight employment situation for college students. To improve the accuracy of predicting college students' employment trends, a deep learning network based on the geometric structure of data manifolds is proposed for forecasting employment characteristics. First, employment data from a university were collected and normalized, and covariance feature matrices were computed to form a Lie group manifold. Second, by integrating differential geometry theory with deep learning principles, a deep dataset recognition network based on Riemannian manifolds was developed. During model training, feature weight coefficients and learning rates were dynamically adjusted to obtain the optimal network model. Experimental results demonstrate that the deep manifold learning algorithm achieves a 5.52% improvement in prediction accuracy compared to conventional convolutional neural network models, while also exhibiting higher efficiency and robustness. The predictive outcomes of this model can provide valuable insights for university employment departments, improving the foresight, targeting, and precision of employment-related initiatives.

Keywords: college student employment; Lie group manifold; deep manifold learning; employment prediction

1. INTRODUCTION

With the continuous vigorous promotion of higher education, the population of college students in China is becoming increasingly large, which puts increasing pressure on their employment. The current employment situation for college graduates is not overly optimistic. The employment rate of college students, as a key evaluation indicator of a “harmonious society”, is closely related to the stability of social order and whether the economy can achieve sustainable growth. At the same time, the high or low employment rate of college students also intuitively reflects the level of education, quality of education, and social reputation of a university. Therefore, it is of great value to carry out modeling and prediction work on the employment potential of college

students, accurately analyze the employment number of college graduates, and provide practical and reliable reference for universities to carry out employment guidance work [1–3].

In the information society swept by the digital wave, building a scientific and reasonable mathematical model can predict the employment characteristics and patterns of college students in various universities during a specific period of time in the future, providing a valuable decision-making reference for relevant managers. In recent years, numerous scholars at home and abroad have conducted extensive research on prediction problems from different perspectives, using diverse algorithms. Li [4] introduced machine learning technology in his research, constructed a graduate employment prediction model, and proposed an integrated method combining a mutual information and weight-based hybrid feature selection algorithm (HMIGW)

*E-mail: wuxizhiye@163.com

and a XGBoost feature selection algorithm as an effective means of predicting student employment. The study also continuously optimized the algorithm based on feedback from relevant implementation departments. Ma [5] improved the semi-supervised self-training method and applied it to the research on college student employment prediction. Through comprehensive analysis of the prediction results and the actual situation of graduates, corresponding employment guidance strategies were proposed, showing that the model has good application effects. Wang [6] used a decision tree algorithm to construct decision tree prediction models for students' "employment" and "employment area", and applied them to student employment guidance practice. Gao [7] used the natural nearest neighbor classification algorithm to comprehensively predict the employment ability of college students, and constructed a prediction model for college students' salaries based on employment data of graduates from information majors. Wang et al. [8] utilized machine learning algorithms combined with personal work history to achieve personalized job recommendations. Watanabe-Murao et al. [9] conducted an in-depth exploration of career development theory, further confirming its important role in the employment of young students. Xu et al. [10] pointed out that researching and predicting the employment market for graduates has become one of the key components of employment guidance in Western universities. Although there have been many advances in current research on predicting employment for college students, most machine learning methods still consider employment data as a type of data in Euclidean space. In fact, due to the strong intrinsic correlation between the employment attributes of college students, such data often exhibit non-Euclidean geometric characteristics and are more suitable for description using manifold space. Therefore, treating the employment characteristics of college students as manifold structured data, and using neural networks based on manifold structures to learn the intrinsic distribution of employment data, is expected to significantly improve the accuracy and efficiency of employment prediction.

The intelligent prediction method based on machine learning can be found in the latest references [11–12]. The method proposed in this paper first normalizes the employment characteristics of college students and feeds them into a convolutional neural network to obtain shallow feature information. Then, the Euclidean feature information is transformed into a Riemannian manifold data [11–12] structure through covariance pooling. The manifold data is then learned using a manifold structure neural network and, finally, the dataset is identified. The main contributions of this article are as follows: (1) Using manifold structured data to represent the feature information of college student employment data, a complete manifold structured neural network was constructed, which enhanced the model's ability to represent the internal structure of the dataset and effectively improved the accuracy of recognition. (2) The experimental results of college student employment prediction show that the method proposed in this paper has good effectiveness, robustness, and accuracy.

2. PREDICTION METHODS

2.1 Covariance Characteristics and Manifold Space of College Student Employment Data

On the basis of in-depth research on graduates' employment, this study integrated data from multiple systems such as student registration management and employment services, and finally selected a number of 14 key characteristic indicators: number of scholarships, number of internships, comprehensive scores, professional scores, number of clubs involved, professional popularity, number and level of student cadres, political outlook, computer skills, college English level, registered residence type, TOEFL/IELTS/GMAT scores, and gender. The target variables for employment forecasting are divided into five categories: unemployment, state-owned enterprises, foreign-funded enterprises, private enterprises, and continuing education. It was noted that several textual fields in the raw data could not be directly used for neural network training. Furthermore, there are significant differences in the magnitude of each feature value, and directly inputting it into the model may lead to reduced training efficiency or even difficulty in convergence. Therefore, this study normalized the data based on formula (1) and converted it to the [0,1] interval. Text data was encoded using discrete numerical values [0,1], $x = x_1, x_2, \dots, x_{14}^T$ is the final normalized data.

$$x_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

In the formula (1), x_i represents the normalized input value, X_i represents the actual input data value, $X_{\max}$ represents the maximum value of the actual input data, and $X_{\min}$ represents the minimum value of the actual input data.

There are a number of m sample data sets, so the employment vector dataset of college students can be represented as a $m \times 14$ matrix A , where the element a_{ij} represents the j -th feature size of the i -th sample. The covariance feature matrix T is the statistical information of equation (1), that is:

$$T = \frac{1}{m-1} A'^T A' \quad (2)$$

Here, A' is the matrix obtained by subtracting the mean vector from A . T_{ij} is positive, indicating that two features tend to increase or decrease simultaneously. T_{ij} is negative, indicating that when one feature increases, the other feature tends to decrease. If T_{ij} is zero, this indicates that there is no linear relationship between the two features.

Furthermore, the covariance matrix C of the k -th sample data can be constructed as follows:

$$C = \begin{bmatrix} a_{k1} & T_{12} & \cdots & T_{1,14} \\ T_{21} & a_{k2} & \cdots & T_{2,14} \\ \vdots & \vdots & \ddots & \vdots \\ T_{14,1} & T_{14,2} & \cdots & a_{k,14} \end{bmatrix} \quad (3)$$

The covariance matrix C is a symmetric positive definite or semi-positive definite matrix [13] (regularizable $C^+ = C +$

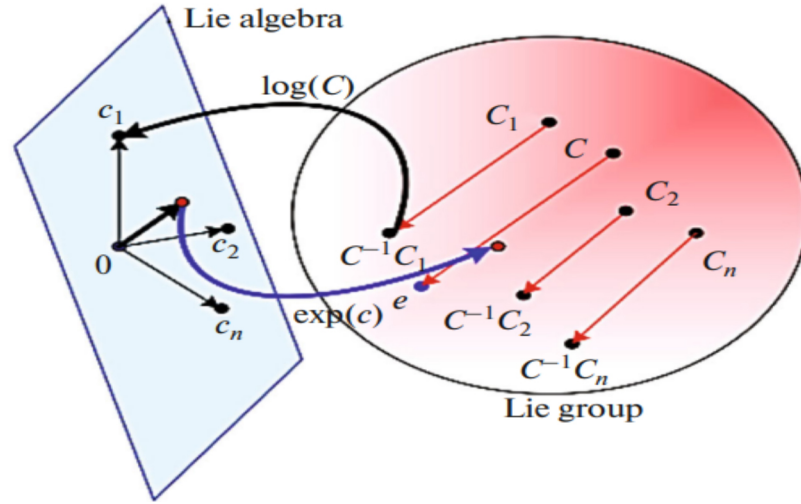


Figure 1 Metric of Lie group manifold space.

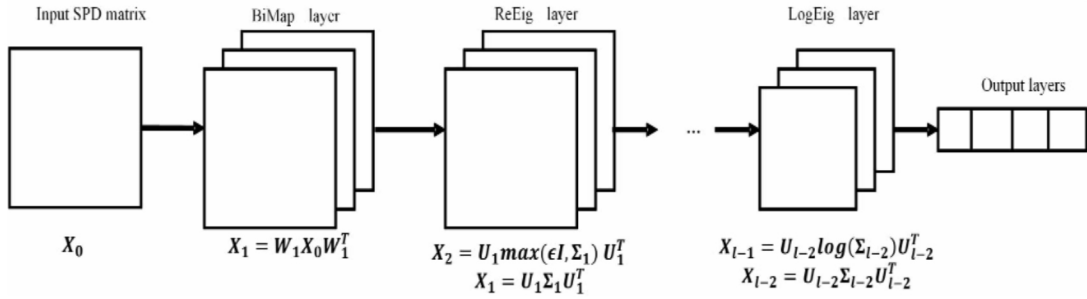


Figure 2 The network structure of SPDNet.

$\lambda \text{trace}(C)E$, E is the identity matrix, λ is the regularization parameter, and trace is the trace of the matrix. The covariance matrix C does not belong to the set of Euclidean space points, but to the Lie group manifold M . Unlike the definition of distance in Euclidean space, distance on a manifold is defined as the length of the shortest curve between two points. As shown in Figure 1, the tangent space of the Lie group manifold is the Lie algebra, which is a Euclidean space. The covariance matrix of the employment characteristics of college students can be expressed as the geodesic distance between any two points on the Lie group manifold M [14]:

$$d(C_i, C_j) = \left\| \log_C(C^{-1}C_i) - \log_C(C^{-1}C_j) \right\|_2, \quad \exp(C) = E, C_i, C_j \in M \quad (4)$$

2.2 Neural Networks on Manifolds

Neural networks based on manifold structures can effectively preserve the geometric characteristics of data and can learn the internal structural features of manifold data while reducing the dimensionality. Figure 2 shows the basic architecture of a deep Riemann neural network (SPDNet)[15–16] based on symmetric positive definite matrices. In this network, the Bilinear Mapping Layer (BiMap) is responsible for converting the input SPD matrix into a low-dimensional SPD

matrix; The eigenvalue correction layer (ReEig) processes the new SPD matrix by means of a nonlinear activation function; the logarithmic layer of eigenvalues (LogEig) is responsible for performing Riemannian geometric operations on the generated SPD matrix. It is worth noting that the BiMap layer and ReEig layer in SPDNet correspond to the convolution operation and nonlinear activation layer in traditional convolutional neural networks respectively, and support multi-layer stacking design. After mapping the manifold data to Euclidean space through the LogEig layer, the network end can access the fully-connected layer to achieve accurate recognition of the target.

The main function of the BiMap layer is to convert the input SPD matrix into a lower dimensional SPD matrix through bilinear mapping B . The conversion process can be represented as:

$$X_k = B^k(X_{k-1}; W_k) = W_k X_{k-1} W_k^T \quad (5)$$

Here, $X_{k-1} \in R^{d_{k-1} \times d_{k-1}}$ is the input SPD matrix of the k -th layer, $W_{k-1} \in R^{d_k \times d_{k-1}}$ is the weight matrix, $X_k \in R^{d_k \times d_k}$ is the output matrix of the k -th layer, $d_k < d_{k-1}$. In order to give the output matrix a manifold structure, it is necessary to ensure that X_k is an effective SPD matrix, therefore, the weight matrix W_{k-1} is a row full rank matrix. The BiMap layer can generate more compact (lower dimensional) and discriminative SPD matrices, similar to the convolutional layers in convolutional neural networks.

Table 1 Normalized data (partial).

Order number	Gender	GPA	Comprehensive score	English level	Computer level	Employment situation
1	1	0.35	0.8	1	0	0
2	1	0.43	0.8	0	1	0
3	1	0.60	0.6	1	1	1
4	0	0.62	0.8	1	1	1
5	0	0.08	0.8	0	1	0
6	1	0.36	0.8	1	1	1
7	1	0.80	0.8	1	1	1
8	1	0.21	0.6	0	0	0
9	1	0.33	0.8	0	1	1
10	1	0.22	0.8	1	0	1
11	0	0.08	0.6	0	0	0
12	0	0.39	0.8	1	1	1

Similar to the ReLU activation function in convolutional neural networks, the ReEig layer introduces a nonlinear transformation mechanism to SPDNet, enhancing the network's representation learning ability. The ReEig layer modifies the SPD matrix through a nonlinear function R , the formula of which is:

$$X_k = R^k(X_{k-1}) = U_{k-1} \max(\varepsilon E, \sum_{k-1}) U_{k-1}^T \quad (6)$$

Here, U_{k-1} and $\sum_{k-1}$ are obtained through eigenvalue decomposition $X_{k-1} = U_{k-1} \sum_{k-1} U_{k-1}^T$, where ε is the rectification threshold, and the max operation obtains the corresponding diagonal matrix by taking the maximum value. When the eigenvalues of the original SPD matrix are all greater than 0, ε is set to a value greater than a number of m minimum eigenvalues and is used to prevent elements in the SPD matrix from approaching non-positive numbers.

To achieve the final predicted response output, the SPD matrix needs to be mapped from manifold space to Euclidean space. The LogEig layer introduces a Lie group structure to the Riemannian manifold of the SPD matrix, enabling the projection into low dimensional Euclidean space and supporting the conventional prediction operations. The calculation process of this layer is defined by the function L as follows:

$$X_k = L^k(X_{k-1}) = \log(X_{k-1}) = U_{k-1} \log(\sum_{k-1}) U_{k-1}^T \quad (7)$$

Here, $\log(\sum_{k-1})$ is a logarithmic diagonal matrix of eigenvalues. The input SPD matrix is located on a non-Euclidean manifold, and the LogEig layer aims to perform Riemann calculations on the SPD matrix to output Euclidean data forms that can be used for Euclidean space calculations.

3. EXPERIMENTAL RESULTS AND ANALYSIS

The data used in this experiment was obtained from the employment management platform of the "Shenyang University of Aeronautics and Astronautics Practical Teaching Management System", and contains relevant information of graduates from three consecutive classes from 2016 to

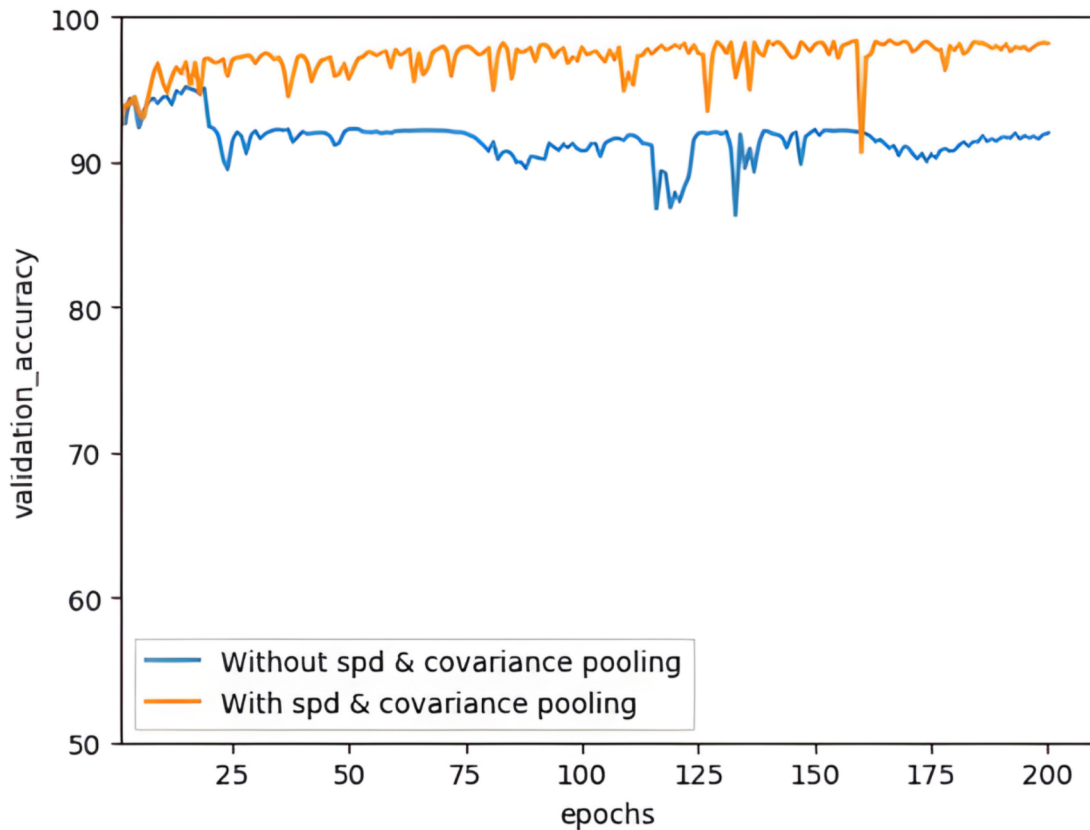
2018. In the data preprocessing stage, we removed or reasonably filled missing values based on their distribution characteristics and attribute meanings; For outliers, data correction or exclusion methods were used for cleaning. After preprocessing, more than 50000 valid samples were retained. According to the experimental requirements, 70% of the total dataset was divided into the training set and the remaining 30% was used as the testing set. Table 1 shows the sample values of the partial dataset after the normalization of each feature.

The deep manifold learning neural network constructed in this study adopts a five-layer architecture, in which the input layer consists of 14 neuron nodes corresponding to the key characteristics of college students' employment obtained by screening: the number of scholarships, the number of internships, comprehensive scores, professional scores, the number of club memberships, professional popularity, the number and level of student cadres, political outlook, computer skills, college English level, registered residence type, TOEFL/IELTS/GMAT scores, and gender. The output layer consists of 5 neuron nodes, representing the five employment states of students: pending employment, state-owned enterprises, foreign-funded enterprises, private enterprises, and continuing education. The model contains three hidden layers as the core structure for feature abstraction and information transmission. In terms of model parameter optimization, this study used ResNet-34 as the backbone structure of a Convolutional Neural Network (CNN) and employed the Stochastic Gradient Descent (SGD) algorithm for parameter updates, with a Momentum of 0.9. The weight decay rate and learning rate gradually decay according to predetermined rules during the training process, which are multiplied by $1 - (\frac{iter}{\max_iter})^{0.9}$ each time, in accordance with the dynamic adjustment strategy. Their initial values are set to 0.0001 and 0.01 respectively. For the experiment, 35000 samples were used as the training set and were randomly trained for 150 rounds. After the model training was completed, the parameter state with the minimum loss function was chosen as the final prediction model.

In target recognition research [17], some commonly used performance evaluation metrics include Detection Rate (DR), Accuracy Rate (AR), and False Alarm Rate (FAR), which are calculated using the following formula:

Table 2 Recognition results of test set.

Number of tests	Detection rate	Accuracy	False alarm rate
15000	98.12%	99.04%	0.36

**Figure 3** Accuracy curves of different methods on the test set.

$$DR = \frac{TP}{P}, AR = \frac{TP + TN}{P + N}, FAR = \frac{NN}{P} \quad (8)$$

Here, TP , TN , P , N are the number of true positive cases, true negative cases, false positive cases, and false negative cases respectively.

The recognition results of the SPDNet test set are shown in Table 2.

Table 2 shows that the algorithm proposed in this paper achieves 98.12%, 99.04%, and 0.36% in detection rate, accuracy, and false alarm rate respectively. This result indicates that, firstly, the SPDNet algorithm exhibits high detection ability and classification accuracy when processing the characteristics of university employment data, indicating its advantage in terms of accuracy in prediction tasks; secondly, the model achieved a relatively low false alarm rate, further demonstrating its strong robustness and stability.

4. RESULTS AND DISCUSSION

To fully validate the advantages of our proposed method, we compared it with the typical convolutional neural network model ResNet-34 [18] that does not include second-order covariance pooling of data. The recognition results are shown in Figure 3.

From Figure 3, it can be seen that the accuracy of the SPDNet method on the test set remains stable at around 98.67%, while the accuracy of the ResNet-34 method remains around 93.15%. The experimental results show that in regard to recognition accuracy, the SPDNet algorithm using the covariance pooling operation significantly outperforms the conventional model that does not include this processing.

The BiMap layer and ReEig layer in the SPDNet structure support stacking usage. To determine the appropriate learning network depth, this paper conducted comparative experiments on SPDNet with different layers. Given the close functional connection between the BiMap layer and the subsequent ReEig layer, it was considered as a complete hierarchical unit in the experiment. The performance comparison under different layer settings is shown in Figure 4.

From Figure 4, it can be seen that the number of model parameters under the three-layer settings is basically the same. Among them, the performance of 2-layer and 3-layer SPDNet is better than that of single-layer structure, but the difference between the two is not significant. In terms of training efficiency, the 2-layer structure takes about 13 minutes per round, while the 3-layer structure takes 17 minutes, significantly increasing the time cost of the latter. In regard to comprehensive performance and efficiency, the 2-layer SPDNet has better balance.

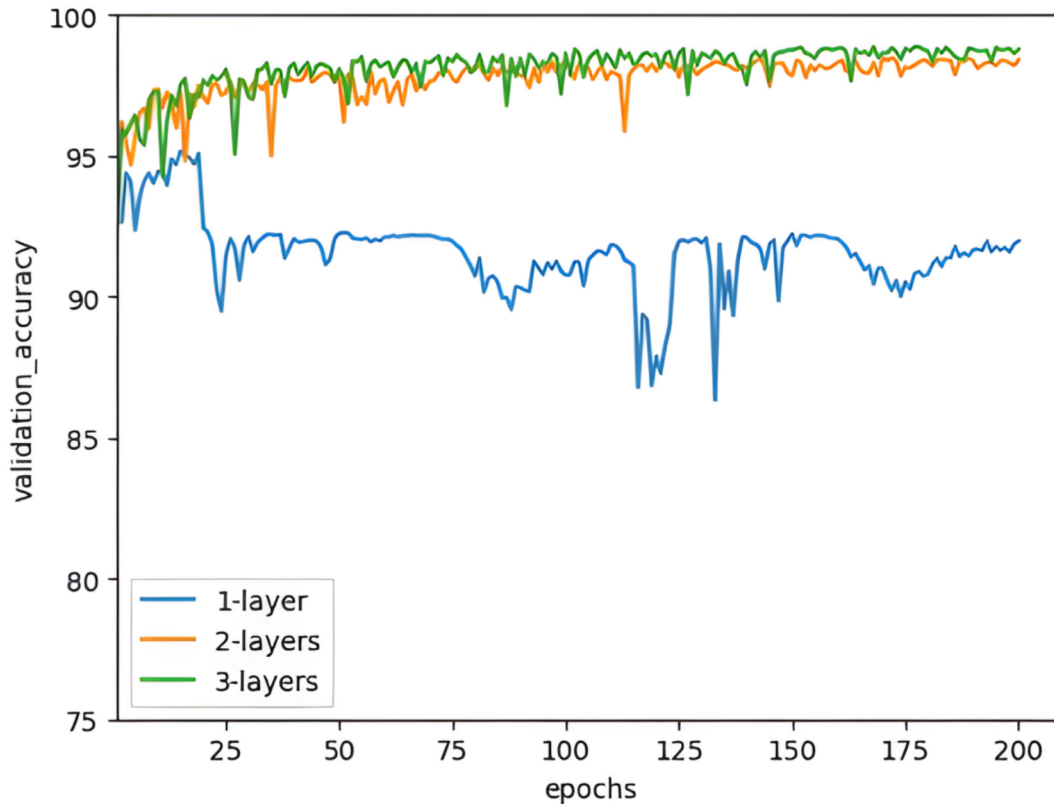


Figure 4 Accuracy curves of SPDNet with different layers.

Based on the analysis results of the employment prediction model, assistance measures such as psychological interventions and skill enhancement can be carried out for students facing employment difficulties to improve their competitiveness in the workplace. Moreover, employment guidance agencies can use professional assessment tools to help students obtain clear and objective self-awareness. In addition, based on the needs of employers and the personal employment intentions of students, universities can organize targeted activities such as employment skills mutual aid groups to effectively improve the quality and satisfaction of students' employment.

5. CONCLUSION

The aim of this study was to identify and classify the characteristics of employment data for college students. Based on the geometric properties of Lie group manifolds possessed by covariance matrices, a concise and efficient deep manifold geometric network model was constructed. This method comprises three processing stages: firstly, the original employment data was normalized and input it into a convolutional neural network to extract shallow features. Subsequently, through the covariance pooling operation, the feature information originally located in Euclidean space was transformed into a symmetric positive definite matrix on the Lie group manifold; Furthermore, utilizing neural networks based on manifold structures to deeply explore the essential features of manifold data, in order to obtain classification

results with stronger discriminative ability. The experiment conducted on the dataset of 50000 graduates from Shenyang University of Aeronautics and Astronautics showed that this method improved the accuracy by 5.52% in target recognition tasks, and demonstrated certain advantages in detection rate and false alarm rate. In addition, this algorithm has a low requirement for neural network layers, which helps to improve computational efficiency and provides feasible technical references for employment guidance in universities. Further research could focus on identifying complex employment data under spatiotemporal distribution, developing employment guidance and management applications, and other directions.

ACKNOWLEDGEMENT

Funding: This work was not supported by any funding project.

REFERENCES

- 1 Na Li. A CiteSpace-Based Analysis of the Evolution of Chinese College Students' Employment Values (1992–2024)[J]. Academic Journal of Humanities & Social Sciences, 2025, 8(4):13–22.
- 2 Jiaqi Xu, Yuan Dai. The Construction of the "College Students Employment Guidance" Curriculum System Under the Rural Revitalization Strategy[J]. Hill Publishing, 2025, 9(2):169–173.
- 3 Peng Zhao. Research on the Impact of Psychological Capital on the Employment Quality of College Students[J]. Journal of Educational Research and Policies, 2025, 7(2):46–48.

- 4 Li Qi. Research and Application of Graduate Employment Prediction Model Based on Machine Learning[D]. Beijing: Chinese Academy of Sciences University, 2019.
- 5 Ma Maoyuan. Research on the Application of Improved Semi Supervised Self training Method in Predicting Employment of College Graduates[D]. Chongqing: Chongqing Normal University, 2019.
- 6 Wang Yaru. Research on the Employment Prediction Model and Application of College Students Based on Decision Tree Algorithm[D]. Wuhan: Huazhong Normal University, 2018.
- 7 Gao Xuan. Application of Natural Neighbor Classification Algorithm in College Student Employment Prediction Model[D]. Chongqing: Chongqing University, 2017.
- 8 Wang J, Zhang Y, Posse C, Bhasin A. Is it time for a career switch? In Proceedings of the 22nd International Conference on World Wide Web[C], 2013, 10(1145):1377–1388.
- 9 Watanabe-Murao A M, Senzaki T A T, Herr E L. Donald Super's Contribution to Career Guidance and Counseling in Japan[J]. International Journal for Educational and Vocational Guidance, 2011, 11(1):1–2.
- 10 H. Xu, Z. Yu, H. Xiong, et al. Learning career mobility and human activity patterns for job change analysis[C]. IEEE International Conference on Data Mining, 2015, 10(1109):1057–1062.
- 11 Dingyan Cai. Consumer Preference Analysis and Marketing Strategy in Digital Economy Based on Deep Learning[J]. International Journal of Engineering Intelligent Systems, 2025, 33(5): 513–524.
- 12 Genchao Wei. Factors Affecting the Quality of Online Open Course Teaching in Universities Based on Big Data Analysis[J]. International Journal of Engineering Intelligent Systems, 2024, 32(6): 659–669.
- [13] Q. Liu, Y. Ge, S. Wang, T. Wang, J Xu. Dynamic manifold-based sample selection in contrastive learning for remote sensing image retrieval[J]. The Visual Computer, 2025, 41(6):4111–4127.
- 14 X. Li, K. Wu, H. Chen, et al. Deep residual PLSR model with manifold optimization and Gaussian filter for enhanced image classification[J]. The Visual Computer, 2025, 41:6087–6102.
- 15 Zhong Zhiyan, Wen Zhiqiang, Ye Degang. Study on Halftone Image Classification Based on the Improved Covariance Matrix[J]. Information Technology and Network Security, 2015, 34(19):54–57(in Chinese).
- 16 Zhang Yun. Mean Shift Based Covariance Matrices Clustering on Differentiable Manifolds[D]. Qinghuangdao: Yanshan University, 2011(in Chinese).
- 17 Huang Z, Van Gool L. A Riemannian network for SPD matrix learning[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2017, 31(1):101–113.
- 18 Chakraborty R, Bouza J, Manton J, et al. Manifoldnet: A deep neural network for manifold-valued data with applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 3:203–214.
- 19 Yuhua Ding, Lisha Hua, Shunlei Li. Research on computer vision enhancement in intelligent robot based on machine learning and deep learning[J]. Neural Computing and Applications, 2021, 1:1–13.
- 20 Huang Z, Gool L V. A Riemannian network for SPD matrix learning[J]. Association for the Advancement of Artificial Intelligence, 2017, 2(4):2036–2042.

