Adult and Higher Education (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/aduhe.2023.050103
ISSN 2523-5826 Vol. 5 Num. 1

Exploring the Cultivation of Innovative Talents in the Era of Big Data and Cloud Computing

Yajie Li*, He Li

School of Science, Beijing University of Posts and Telecommunications, Beijing, 100876, China
**Corresponding author*

Keywords: Computer technology, artificial intelligence, business analysis, data analysis, learning design, innovation ability

Abstract: Higher education shoulders the important mission of high-level talent training and innovation and creation, and is the cornerstone of national development and social progress. The rapid development of computer technology in the future requires a large number of composite talents in statistics, data analysis and computer application. With the goal of academic success and the spirit of scientists, we have designed a training program for data visualization and artificial intelligence applications, designed a training program for the integration of business analysis and computing technology, and conducted research on the training mode of cross integration of science and engineering. Through the curriculum practice of multivariate statistical analysis and data analysis, and through conscious and organized learning design, diversified learning activities are organized. From students' homework and classroom performance, it is fed back that students can comprehensively use R, SPSS and other computer software to conduct data mining, and students can obtain information from multiple perspectives from real problems, and think at a higher level, so that students can exercise their innovation ability, it has enhanced the interdisciplinary ability.

1. Introduction

With the development of cloud computing, big data, the Internet of Things, and artificial intelligence, the national government is in the process of digital transformation. In 2015, the State Council issued the Action Plan for Promoting the Development of Big Data. In the future, there is an urgent need for a large number of compound talents who understand data mining technology, models, algorithms, economy, products, and commerce. It is imperative to cultivate, research, and practice data analysis and innovation talents.

It is a multi-disciplinary coordinated development to solve complex problems from the perspective of different disciplines and learn from other disciplines to serve the research and teaching of this discipline. Lehmann discussed the integration of knowledge across conceptual boundaries and its role in learning [1], Janssen et al. discussed the plan for teachers to integrate information technology [2], Cruz Ram ír ez, S.R. et al. discussed the use of NAO robots as a teaching method [3]. Moradi, M. and others discussed the influence of serious games on learning

3D computer graphics [4]. Brown, R.B. explored the interdisciplinary model of environmental protection and disease prevention [5]. Azer, S.A. and others conducted group interaction in the learning tutorial based on questions [6]. Xie, T. and others explored the structural relationship between computer self-efficacy, perceived immersion and intention to use virtual reality training systems [7]. Awofala AO et al. discussed that attitudes towards computers, computer anxiety and gender are determinants of computer self-efficacy of pre service science, technology and mathematics teachers [8]. Bogusevschi D et al. used 3D virtual learning environment for physics teaching [9]. Boden et al. believed that interdisciplinary education faced many obstacles, both structural and multidisciplinary knowledge challenges and cognitive challenges, namely, the challenge of integrating and even creating new knowledge from traditional knowledge systems [10]; Some scholars seek corresponding knowledge nodes from the knowledge map, answer the questions raised by students, and effectively improve the quality of talent training [11]; The topic filtering algorithm is used to filter irrelevant problems and improve efficiency [12,13]; Establish quality monitoring feedback module [14]: combine automatic question answering with manual processing [15]. In order to continue to deepen the comprehensive reform of higher education, focus on promoting the connotative development of education, and cultivate new people of the times to take the responsibility of national rejuvenation, this paper mainly studies the innovative talent training model based on data analysis of cross integration of science and engineering.

2. Explore Innovative Talent Training Mode

We adhere to the principle of educating people, centering on the all-round development of morality, intelligence, physique, beauty and labor, and taking the effectiveness of moral cultivation as the fundamental criterion for testing education; Adhere to demand orientation, take root in China, and comprehensively improve the ability of education to serve national and regional development; Adhere to the guidance of innovation, enhance the sense of mission and responsibility of students, and comprehensively improve the ability of knowledge innovation and practical innovation; Adhere to the reform drive and accelerate the construction of a high-quality, efficient and open education system.

2.1. Operational Cross Fusion Culture Mode

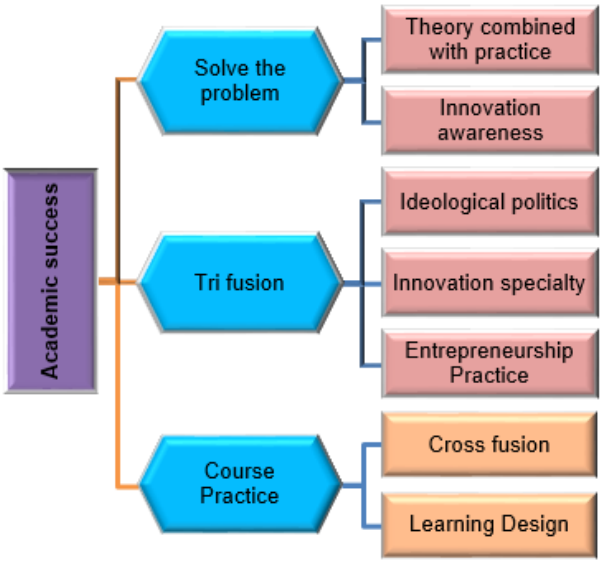

Figure 1: Data analysis and innovative talent training mode

We explore the training mode of data analysis and innovation talents in the cross integration of science and engineering, take academic success as the goal, take training students how to solve problems optimally as the path, take "ideological and political education + innovation professional education + entrepreneurship practice education" as the research method, take the data analysis course as the practice platform, and take students as the center to test the effect, and conduct research on the training mode of data analysis and innovation talents in the cross integration of science and engineering, to discuss the specific scheme of operable cross fusion culture, see Figure 1.

2.2. The Combination of Theory and Practice is an Effective Way

What is academic success? What are the key factors for academic success? From the practical experience of postgraduate education in various colleges and universities, students' cognitive ability, non cognitive ability and individual characteristics play an important role in their academic success, that is, academic success depends on a variety of factors, including cognitive level, as well as emotional intelligence, self-efficacy, responsibility, communication ability, career planning ability and other non cognitive abilities.

One of the feasible ways of innovation and entrepreneurship in colleges and universities is academic entrepreneurship, which is an important way for researchers to transform their scientific and technological achievements into real productive forces. The more academic scientists with each entrepreneurial orientation, the greater their scientific identity center. Academic entrepreneurship is conducive to scientific self-development. In the future, compound talents are needed, integrating different information and laws from different disciplines and applying more innovative research methods and tools has become a necessary condition for success. Today's social and economic development level not only needs the promotion of knowledge or technology in a certain discipline, but also depends on the collaborative drive of interdisciplinary and interdisciplinary knowledge or technology. The traditional talent training model of a single discipline can no longer meet the development needs of the knowledge economy era. Cross disciplines are based on different majors and can be divided into borrowing, synthesis, interdisciplinary and conceptual cross disciplines according to the degree of integration. For example, the cultivation of interdisciplinary doctoral students in the United States has provided an IDE (Interdisciplinary Doctoral Education) practice model to break down the barriers between traditional colleges and disciplines. By strengthening the flexibility of the training mechanism, a series of reforms such as interdisciplinary research projects have been carried out, so as to cultivate students with more innovative spirit and broad academic vision.

The combination of theory and practice is an effective way to solve problems optimally. For example, Tycho, a Danish astronomer, spent 30 years observing the position of the planets precisely, but he was not good at theoretical analysis. In 1600, he invited Kepler, a German, as an assistant. Kepler, on the contrary to Tycho, was not interested in observation and was very talented in theoretical research. By analyzing Tycho's data, he first assumed that the sun revolved around the earth, with a large error, which was inconsistent with the observation, it is also assumed that Mars moves in a circle around the sun, and the calculation results are not ideal. It is better to innovate boldly and put forward the hypothesis that "Mars' motion orbit is elliptical, and the sun is at a focus of the ellipse". The results are consistent with the observation data. In this way, Tycho's accurate observation and Kepler's profound research led to the discovery of the three laws of planetary motion, which is an example of the combination of theory and practice, it is an important way for excellent researchers to succeed. We have designed the following two schemes.

3. Innovative Talent Training Program

3.1. Training Program for Data Visualization and Artificial Intelligence Application

In recent years, from traditional applications such as face recognition, object recognition, beauty software, to automatic driving, medical image recognition, and the application of image data has shown a wider and wider application prospect in practice. In the academic report on the official account "Bear Club", Wang Hansheng, a teacher from Guanghua School of Management, Peking University, explained the analysis and application of image data, and put forward reasonable suggestions on the future development trend of image data in artificial intelligence.

This scheme is applicable to the target and requirements of graduate students who are interested in "talking with data", "statistical visualization" and "image processing". Requirements: master the basic knowledge of data science. Have studied basic courses in statistics, big data, machine learning, etc. in at least one related direction. Be able to skillfully use at least one of R, Python, SAS, and SPSS, and be able to read and write data, clean up, process, draw common graphics, and build common models.

The learning objectives of this program are: 1) master the basic methods of data visualization. 2) Master the writing specification of data analysis report and the visual presentation method of report writing. 3) Understand the basic characteristics and common processing methods of image data. 4) Master the basic methods of image data transformation and mining; Have the ability to apply the analysis results of image data to solve practical problems. 5) Master the basic application of data visualization and image data processing in the field of artificial intelligence.

The training content of this program consists of three contents: data visualization, image datalization, and their application in the field of artificial intelligence. Among them, data visualization includes 1) visualization of statistical principles, 2) learning of common statistical indicators, 3) visualization of qualitative data, 4) visualization of quantitative data, 5) visualization of time series data: Internet of Vehicles data, 6) visualization of proportional data, 7) visualization of network structure data, and 8) forming a data visualization report. Among them, image data processing includes 1) basic introduction and reading of image data, 2) basic processing of image data: basic operation, 3) basic processing of image data: basis and realization of mosaic, 4) basic processing of image data: image transformation and magick package, 5) analysis of image data: gray difference and definition analysis of image, 6) analysis of image data: information entropy and saturation analysis of image, 7) Case: Image based PM2.5 data analysis and prediction. Among them, applications in the field of artificial intelligence include: 1) deep learning is nonlinear regression analysis, 2) images are matrices (or vectors, or Tensors), 3) color pictures can be changed into sketches, 4) AI can score facial features, 5) AI can recognize handwritten digits (MNIST dataset), 6) AI can automatically recognize gender, 7) DL core technology - convolution+pooling, 8) NLP example: machine poetry.

Digital image data can be represented by matrix, so matrix theory and matrix algorithm can be used to analyze and process digital images. The pixel data of the image is a matrix $= (x_{ij})$, and the element x of the matri x_{ij} corresponds to the pixel of the grid image, and the value is the pixel value.

The pixel value of the image is in the range of 0-1 for double (floating point number) and 0-255 for uint8. The smaller the x_{ij} value of the matrix element, the darker the color, and vice versa.

The change of image is to change X into Y by transforming $Y=f(X)$. If floating point number is used, the pixel data matrix of the image is $= (x_{ij})$, $0 \le x_{ij} \le 1$.

For example, let $Y_1 = X/2$, the image becomes dark.

For example, let $Y_2 = (X+1)/2$, the image becomes lighter and has a "beauty effect".

For example, let $Y_3 = 1 - X$, the dark color of the image becomes light, and the light color becomes dark.

For example, for the identification problem, let $Y_4 = \begin{cases} 1, & \text{if the same person} \\ 0, & \text{not the same person} \end{cases}$, then the image classification problem according to $Y_4$is an identity recognition problem.

For example, for license plate recognition, let $Y_5 = \begin{cases} \text{province} \\ \text{number} \\ \vdots \end{cases}$, then the image classification problem according to Y_5 is the license plate recognition problem.

For example, for handwritten digit recognition, let $Y_6 = \begin{cases} 0 \\ \vdots \\ 9 \end{cases}$, then the image is classified according to the 10 classification problems of Y_6, which is a handwritten digit recognition problem.

There are many image data sets, among which the MNIST data set is the classic one. The purpose of this data set is to realize the recognition of handwritten digits through algorithms. It consists of 28 pixels × these pictures are composed of 28 pixel handwritten numerals, and only contain gray value information. Usually, these handwritten digital images are classified into 0~9, a total of 10 categories. In 1998, Yan LeCun et al. published the paper "Gradient Based Learning Applied to Document Recognition", first proposed LeNet-5 network, and realized handwritten font recognition using the above data set.

3.2. Training Program for Business Analytics Data Innovation Professionals

Business Analytics is a rapidly rising branch in the field of data science practice in recent years. It focuses on commercial business and needs a good technical foundation. BA is a complex interdisciplinary subject, which requires three dimensions of skills: data science theory, programming, and business practice. The first two are the basic knowledge necessary for data science practice. The third dimension focuses on solving practical problems with professional knowledge, which requires the ability to connect professional knowledge of various disciplines to solve practical problems; it is the most critical finishing touch in the ability composition of business analysts. Statistics and computers are particularly important foundations, especially statistical knowledge.

This program is applicable to students who are interested in business analysis and are looking for or engaged in business analysis internship. Requirements: master the basic knowledge of data science. Have studied basic courses in statistics, big data, machine learning, etc. in at least one related direction. Be able to skillfully use at least one of R, Python, SAS, and SPSS, and be able to read and write data, clean up, process, draw common graphics, and build common models.

The learning objectives of this program include: 1) becoming bilingual in business and data, identifying data problems from business, and finding appropriate application scenarios for data results. 2) Build a two-way transformation idea of "data business" through the basic theory of statistics, and establish data thinking and data sensitivity. 3) Learn to draw and speak from pictures, and have the ability to spread their conclusions and achievements, establish and expand their influence. 4) Through practical cases, we can go through the process of coming from the business and returning to the business completely, and master the general idea of using data to solve practical problems.

The training content of this program consists of three parts: data thinking, visualization and report writing, and business analysis practice. Among them, data thinking refers to the two-way transformation thinking between knowing and establishing "knowledge business". It involves 6

aspects. 1) Uncertainty and mathematical expression, 2) Law of large numbers and central limit theorem, 3) Statistical inference (parameter estimation and hypothesis testing), 4) Principle of regression analysis, 5) Linear regression analysis, 6) Logical regression analysis. Among them, visualization and report writing are to let data speak and expand the influence of data work results. Consider 8 aspects: 1) standard report writing, 2) standard statistical chart and common statistical indicators, 3) description of qualitative data, 4) description of quantitative data, 5) staged exercise, 6) topic selection and background introduction, 7) refined description text, 8) complete data analysis report. Among them, business analysis practice refers to the application of the basic methodology of data to solve business problems to solve a real problem

4. Course Practice

The implementation of the design scheme requires a practice platform. We take courses such as "multivariate statistical analysis" and "data analysis" as the practice platform, decompose the scheme for the implementation of the scheme and the weekly teaching plan, study and design important knowledge points of statistics and data analysis, and integrate value shaping, knowledge teaching and ability training. The design idea is the integration of reality, data and research, Practice feedback improvement in teaching.

Learning design is a process and method to create a high-quality learning environment and experience for students. Through conscious and organized contact with teaching materials, learning activities and interaction, students can obtain information, acquire skills and practice higher level thinking. The focus of curriculum design is to provide students with the best learning experience in an environment that supports and appreciates learning and intellectual development. Effective design should enable learning to have a positive impact and produce appropriate expected results. Effective design includes the following key elements: 1) what do you want students to learn? And how to measure the content of learning, 2) select a series of activities, assignments and materials to help students learn.

Optimize the course content, pay attention to the leading edge and method teaching, strengthen the postgraduate's understanding of the innovation process, consolidate the basic theory, strengthen the core of the specialty, enrich the cutting-edge knowledge, highlight the cultivation of innovation ability and humanistic quality, such as the learning design of the "principal component analysis" dimension reduction method in multivariate statistical analysis, which is closely related to machine learning, artificial intelligence and other hot fields such as face recognition. Learning design starts with a classic economic case, introduces statistical principles, and then introduces the application of face recognition. In the latest serial "Hunting Crime Manual", the simulation portraitist found 11 face templates (the first principal component) under the face changing videos through the muscle movement rules of the face changing videos in the new fraud case of AI face changing.

Through the smart classroom, we will build an online and offline learning platform to deepen the reform of teaching methods. For example, we will put course materials in the online platform "Beijing University of Posts and Telecommunications Teaching Cloud Platform", which includes courseware, data, extracurricular reading and other materials, such as video resource "Alberto Cairo Visualization", pay attention to the frontier and internationalization, and cultivate the knowledge acquisition ability, academic identification ability Independent research ability and ability to solve practical problems; Extracurricular reading, such as F J. Anscombe's 1973 paper "Graphs in Statistical Analysis", studying this paper can realize the role of data visualization [16].

Build an activity system with rich contents and various forms to expand the academic vision of graduate students; Stimulate graduate students' enthusiasm for scientific research practice and academic innovation; Create a free, equal, open and active academic atmosphere. For example, we

organize a topic discussion. The topic is to think about problems for a long time after class, and the answers to the questions are not unique, which can trigger discussion and stimulate students' vitality. Students can give answers after thinking. Based on the weekly teaching objectives and the latest progress of the discipline, we will conduct auditions and lectures at the same time.

From students' homework and classroom performance, it is fed back that students can obtain information from multiple perspectives, think at a higher level, and exercise their innovation ability. With the rapid development of computer network technology, epidemic situation and other factors, online education has become an important means of teaching. For example, a student uses principal component analysis to give a "quality evaluation of online teaching", and looks for multiple indicators from the online course website, such as "number of participants", "number of courses", "number of teachers", "number of comments", "score", "average length of each teaching video" "The number of exercises in each section", "the number of students' questions", "the number of students' discussions", "the average completion rate of each course", etc. After discussion, it is believed that the "number of participants" can correctly reflect the quality of the curriculum, so it is selected as the basic indicator, and other indicators are tested for Kendall correlation with the basic indicators, excluding indicators with weak correlation. A comprehensive evaluation of the curriculum quality of the school is given by the MOOC of the University of China in the course Signal and System, as shown in Table 1.

Table 1: Comprehensive evaluation of online teaching of a course

Signal and System		
School	Principal Component Score	Comprehensive Evaluation
Beijing University of Posts and Telecommunications	1.749	1
Hunan University	0.932	2
Xi'an University of Electronic Science and Technology	0.911	3
Beijing Jiaotong University	0.874	4
Southeast University	0.8	5
…	…	…

A student discussed the living quality of a certain area from 2010 to 2014 (indicated as 1-5 respectively). The survey indicators include air quality, urban population, economic level, number of college students, and average life span. He drew a Facebook with computer R software. According to Figure 2, this city had the highest living quality in 2014.

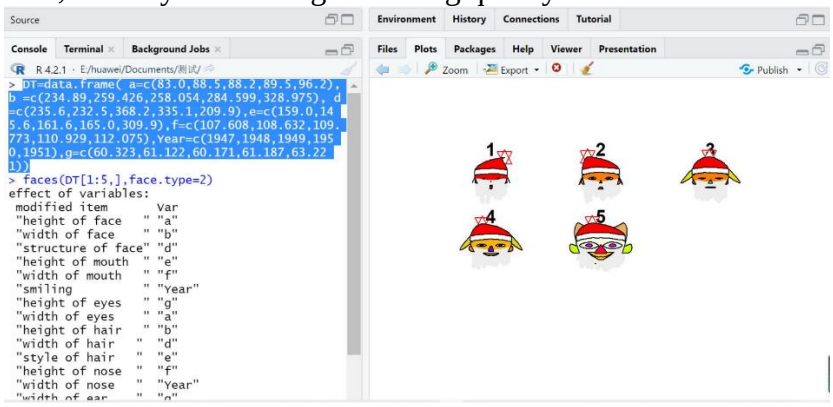

Figure 2: Student's Data Visualization Works

5. Conclusion

In 2021, our school put forward the "Seven Action Plans" for postgraduate students. Centering on the core point of comprehensively improving talent training ability, we will launch a number of ideological and political demonstration courses, famous teachers and teams of ideological and political teaching courses. We will set up projects to build cutting-edge courses, basic courses and professional courses for many years in a row, carry out construction in a solid and orderly manner, and effectively improve the course construction and teaching quality. Our school has successively held artificial intelligence forums, such as the third issue of "Human Machine and Things Fusion Group Intelligent Computing", and invited Professor Guo Bin from the School of Computer Science of Northwestern Polytechnical University to give a report to the students of the school. Human Machine and Things Fusion Group Intelligent Computing involves the Internet of Things, artificial intelligence, biology, community ecology and other disciplines. The report discusses the basic theory, characteristics, scientific challenges, and research progress of CrowdHMT. Through the action plan guide and focus on cutting-edge research, we will continue to deeply explore the data analysis and innovative talent training mode of science and engineering cross integration, which will broaden our research ideas and practice path, and we will continue to work hard.

Acknowledgment

Thank you for the 2022 Education and Teaching Reform Project of Beijing University of Posts and Telecommunications (500522511), 2021 Education and Teaching Reform Project of Beijing University of Posts and Telecommunications (2021JXYJ13), 2021 Postgraduate Education and Teaching Reform and Research Project of Beijing University of Posts and Telecommunications (2021Y029).

References

[1] Lehmann, T. Student Teachers' Knowledge Integration across Conceptual Borders: The Role of Study Approaches, Learning Strategies, Beliefs, and Motivation. Eur J Psychol Educ, 37, 1189–1216 (2022).
[2] Janssen, N., & Lazonder, A. W. Supporting Pre-Service Teachers In Designing Technology-Infused Lesson Plans. Journal of Computer Assisted Learning, 32(5), 456–467(2016).
[3] Cruz-Ram ́ırez, S.R., Garc ́ıa-Mart ́ınez, M. & Olais-Govea, J.M. NAO Robots as Context to Teach Numerical Methods. Int J Interact Des Manuf, 16, 1337–1356 (2022).
[4] Moradi, M., Noor, N.F.B.M. & Abdullah, R.B.H. The Effects of Problem-Based Serious Games on Learning 3D Computer Graphics. Iran J Sci Technol Trans Electr Eng, 46, 989–1004 (2022).
[5] Brown, R.B. Transdisciplinary Model for Environmental Protection and Primordial Prevention of Disease. J Environ Stud Sci, 12, 898–904 (2022).
[6] Azer, S. A., & Azer, D. Group Interaction in Problem-Based Learning Tutorials: A Systematic Review. European Journal of Dental Education, 19(4), 194–208(2015).
[7] Xie, T., Zheng, L., Liu, G. et al. Exploring Structural Relations Among Computer Self-Efficacy, Perceived Immersion, And Intention To Use Virtual Reality Training Systems. Virtual Reality, 26, 1725–1744 (2022).
[8] Awofala AO, Olabiyi OS, Awofala AA et al. Attitudes toward Computer, Computer Anxiety and Gender as Determinants of Pre-Service Science, Technology, and Mathematics Teachers' Computer Self-Efficacy. Digital Educ Rev, 36, 51–67(2019).
[9] Bogusevschi D, Muntean C, Muntean GM. Teaching and Learning Physics Using 3d Virtual Learning Environment: A Case Study of Combined Virtual Reality and Virtual Laboratory in Secondary School. J Comput Math Sci Teach, 39(1), 5–18(2020).
[10] Boden D, Borrego M, Newswander L K.Student Socialization in Interdisciplinary Doctoral Education. Higher Education, 6,741-755(2011).
[11] Chen X, Jia S, Xiang Y.A review:Knowledge Reasoning over Knowledge Graph. Expert Systems with Application, 141(5), 112948(2020).
[12] Lin J, Zhao Y, Huang W, et al. Domain Knowledge Graphbased Research Progress Of Knowledge Representation.

Neural Computing and Applications, 33(2), 681-690(2021).
[13] Qiu X, Sun T, Xu Y, et al. Pre-Trained Models for Natural Language Processing: A Survey. Science China Technological Sciences, 63(10), 1872-1897(2020).
[14] Shao B, Li X, Bian G.A Survey of Research Hotspots and Frontier Trends of Recommendation Systems from the Perspective of Knowledge Graph. Expert Systems with Applications, 165, 113764(2021).
[15] Liu W, Zhou P, Zhao Z, et al. K-bert: Enabling Language Representation with Knowledge Graph. Proceedings of the AAAI Conference on Artificial Intelligence, 34(3), 2901-2908(2020).
[16] Anscombe, F.J. Graphs in Statistical Analysis. American Statistician, 27, 17-21(1973).