Teaching Reform Research on Big Data Analysis and Application with Python Language

FENG LI, LINGLING WANG*
School of Management Science and Engineering
Anhui University of Finance and Economics
Bengbu 233000, CHINA

Abstract: - In order to promote the improvement of teaching quality, this paper focuses on the analysis of big data analysis and its application with Python Language. Firstly, expounds the overview of the Python course. Then, introduce different strategies by improving the teaching reform model through the gradual optimization of teaching resources, case-based teaching, strengthening students' practical ability, and diversified assessment. Adopt reasonable reform measures can enhance the ability of teacher-student interaction and achieve the goal of improving the quality of teaching.

1 Introduction

Big data technology is mainly to analyze the big data area, so as to obtain more intelligent, deep and valuable information [1]. Big data analysis ie inseparable from data quality and data management, high-quality data an effective data management, no matter in academic research or business applications. However, in many application sides, the original data is often incomplete and inconsistent, which seriously affects the efficiency and effect of data analysis [2]. Therefore, the application of big data technology is becoming more and more extensive, gradually infiltrating into all walks of life, which puts forward new requirements for the talent training mode of colleges and universities [3].

Big data analysis and application courses should be based on big data programming language, with big data computing framework as the core, big data algorithm as the sublimation and big data application as the foothold [4]. Meanwhile, big data should combine with the development trend of The Times, and integrate, accept and tolerate professional knowledge in other fields. Big data programming language is a basic tool for big data practitioners to learn and work. Only with the foundation of big data programming language can we have a deeper understanding of the core ideas of big data computing framework and understand the design concept of big data architecture from the source level. Only by understanding the design concept of the big data framework, can the big data algorithm based on MapReduce be reasonably and efficiently deployed on the big data computing platform. After fully understanding and mastering big data programming language, framework design and big data algorithm, we can flexibly use various big data skills and methods to solve practical problems in production and scientific research and create value for the society.

At present, the courses related to big data and analysis are still in the early stage of construction, teaching cases are not rich enough, and the construction of curriculum system still needs to be improved, but relevant curriculum construction is being actively explored. Big data courses mainly cover big data storage, big data algorithms and big data platform construction, involving the use of Linux operating system, programming languages such as Python, R and Java, and computing frameworks such as Hadoop, Spark, Storm, HBase and Hive, data mining, artificial intelligence, machine learning and other big data algorithms as well as big data visualization and many other technologies [5]. In the actual application process of big data, a variety of technologies are often needed, which greatly increases the learning difficulty of students.

2 Overview of Python Language

Python language is a free, open source, cross-platform advanced dynamic programming language [6]. It supports various programming methods and has a large number of powerful built-in objects,

standard libraries and extension libraries. Powerful programming functions can be realized by directly calling the built-in functions or standard libraries. By its nature, Python is both an "object-oriented" language and an "interpreted" language [7]. Python is relatively easy to get started. Its syntax is similar to That of English, and programs can be executed directly through the interpreter, but it consumes a lot of hardware resources.

In the application sides, the Python Language is particularly well suited for data analysis and processing. Matplotlib is a 2D drawing tool that is often used to chart data with a few simple lines of code [8]. Pandas is an open source tool for manipulating complex two-dimensional and three-dimensional arrays and for manipulating data in relational databases [9]. Python's powerful and rich libraries and data analysis capabilities make it well suited for the field of artificial intelligence. In neural networks, deep learning, Python can find mature packages to call. And Python is an object-oriented, dynamic language for scientific computing, which makes Python a favorite for artificial intelligence. The ease of call and the power of scientific computation remain Python's strongest competitiveness in the field of AI.

Python aims to train students to make use of Python language in the application of their major, and plays an important role and position in the curriculum system and major construction of machine learning, pattern recognition, computer vison and so on. The language is an interpretive, object-oriented computer programming language for data statistics, analysis, visualization and other tasks, as well as machine learning, artificial intelligence and other fields.

Additionally, it can meet almost all the functional requirements of data processing, statistical model and graph drawing under data mining. A large number of third-party modules support content ranging from statistical computing to machine learning, from financial analysis to biological information, from social network analysis to natural language processing, from various databases, various language interfaces to high-performance computing models.

To sum up, the introduction of the pre-class targeted preview method in the course of big data analysis and application effectively promotes the development of classroom teaching; In the course of teaching system theory, small cases of Python are integrated to deepen students' understanding and mastery of relevant theoretical knowledge. At the end of the course, using the project case development method, students are required to

conduct data mining and analysis for applications in aviation, e-commerce, public service, power and other industries.

In the process of project development, students learned to use octopus and other tools for data collection; Completed data cleaning, attribute reduction, data transformation and other data processing work; Programming to realize the process of data visualization; Completed the model building and analysis of group tasks, discussed common algorithms such as k-means clustering optimal K value selection scheme, and conducted comparative experimental analysis;

For the optimization and application expansion of the model, the author also puts forward his own views and makes a preliminary attempt. Compared with the traditional single theoretical teaching mode, the mixed teaching mode gives full play to the students' main role and awareness of classroom participation, forms a good classroom interaction, greatly stimulates students' interest in learning, improves their hands-on ability, enhances the teaching effect, and achieves the expected teaching objectives.

3 Exploration of teaching reform mode

The online and offline mixed teaching mode is also called the mixed teaching mode or the mixed learning mode. The mixed teaching mode is not a brand new teaching method, but a kind of online and offline mixed teaching mode combining online teaching and traditional teaching.

With the development of Internet technology, the integration of curriculum teaching and information technology is deepening day by day, attracting considerable attention. The online and offline hybrid teaching mode fully exploits the advantages of traditional classroom teaching and digital classroom teaching, which is the organic combination and complementary advantages of the two, so as to merge into an online and offline teaching mode. By introducing this teaching mode into class, students can master the knowledge from simple to profound. Learning through online interactive learning, this platform has a large number of books, journals and important literature materials at home and abroad, and support the construction and learning of online courses [10-12].

3.1 Optimize network teaching resources

With the popularization of the Network and the rapid development of big data technology, the resources in the network increase exponentially and

have become an indispensable part of our work and study [13].

At the same time, excellent online teaching resources are beneficial tools to improve the efficiency of online teaching mode. Teachers should make full use of the Internet and establish excellent network resource platforms and online live broadcasting, such as the integration of Chinese college students MOOC [14], SPOC [15], Superstar learning and Tencent live and other tools, so that students can fully carry out independent learning.

3.2 Case teaching mode

Learn the basic knowledge of Python, such as building the Python development environment, basic Python syntax, program flow control, common data types, and custom functions. By learning the basic principles of programming, students can not only master theoretical knowledge, but also grasp the hot issues of The Times, such as the current COVID-19 epidemic, master python's approach to big data processing, and develop the basic skills of big data analysis.

Combining theory with practice, through a large number of examples, such as Python based data visualization analysis under the global COVID-19 situation, analysis and prediction of stock fluctuations, loss and profit analysis and prediction of listed companies, real estate price fluctuations and other practical problems [16]. Reasonable design of teaching programs. Student is the main learning, the teacher is more of a need to guide and help students, through the reasonable design of teaching plan, such as the introduction of the design and implement the student to the theory of knowledge idea, again carries on the reasonable guidance, let the students interest in knowledge, for mobilizing the enthusiasm of students, our students' autonomous learning initiative.

3.3 Strengthen students' practical ability

The ultimate goal of education is to cultivate the ability of application and practice in learning [17]. On the one hand, students study in university for a deeper grasp of professional theoretical knowledge, on the other hand, to cultivate their practical ability, so as to improve their employability.

In order to improve the students' practical ability, strengthen the practice ability, online, offline mixed teaching mode in the process of implementation, the corresponding training and competitions organized school activities, learning in practice, practice in learning, can help students flexible use of learned knowledge, fully mobilize students' learning enthusiasm, stimulate students' active learning initiative, so as to improve students' practical ability.

3.4 Diversified assessment methods

Relying on online resources, the combination of offline classroom education, using a variety of platforms and technology, improve the inspection way, strengthen the procedural appraisal of students' learning, cultivating students use python solution actual problem ability, realize the whole process of students learning situation feedback, avoid back problem of test. Taking Anhui University of Finance and Economics as an example, its assessment focuses on the process, disperses the overall assessment results to the whole process of the whole semester, including stage test results, final exam results and regular results.

The stage test was conducted twice, and the in-class test was conducted uniformly by using the general examination client. The final exam is arranged by the teachers themselves and can be in the form of large assignments, group defense and other forms. The performance of the final exam is mainly composed of homework, check-in and video completion. Therefore, the total score = final score (40%) + average score of stage test (40%) + video study (10%) + attendance homework (10%).

4 Conclusion

In this paper, Python is used in teaching big data analysis and application courses by virtue of its powerful functions, quick learning and easy to get started, so that students can deepen their understanding of relevant theoretical knowledge in practice, stimulate their strong interest in learning, improve their hands-on ability and improve the teaching effect.

Centering on the exploration of Python course reform, this paper aims to enhance the ability of teacher-student interaction by gradually optimizing online education resources, strengthening students' practical ability, case teaching and diversified assessment, so as to improve teaching quality and teaching effect.

References:

[1] Singh, Nitin. "Big data technology: developments in current research and emerging landscape." *Enterprise Information Systems* 13.6 (2019): 801-831.

[2] Katal, Avita, Mohammad Wazid, and Rayan H. Goudar. "Big data: issues, challenges, tools and good practices." 2013 *Sixth international conference on contemporary computing (IC3).* IEEE, 2013.

[3] Peng, Kanghua, and Jincheng Shi. "Design and Application of Digital Resource Sharing Talent Training Program System for Big Data Technology." *2021 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS).* IEEE, 2021.

[4] Elia, Gianluca, et al. "Assessing learners' satisfaction in collaborative online courses through a big data approach." *Computers in Human Behavior 92* (2019): 589-599.

[5] Irizarry, Rafael A. Introduction to data science: Data analysis and prediction algorithms with R. CRC Press, 2019.

[6] Sanner, Michel F. "Python: a programming language for software integration and development." J Mol Graph Model 17.1 (1999): 57-61.

[7] Srinath, K. R. "Python–the fastest growing programming language." International Research Journal of Engineering and Technology (IRJET) 4.12 (2017): 354-357.

[8] Ari, Niyazi, and Makhamadsulton Ustazhanov. "Matplotlib in python." 2014 11th International Conference on Electronics, Computer and Computation (ICECCO). IEEE, 2014.

[9] McKinney, Wes. "pandas: a foundational Python library for data analysis and statistics." Python for high performance and scientific computing 14.9 (2011): 1-9.

[10] Huang, Yutian, et al. "The Reform and Exploration of Intelligent PYTHON Language Teaching." 2021 16th International Conference on Computer Science & Education (ICCSE). IEEE, 2021.

[11] Li, Min, et al. "Reform of the Multi-platform Blended Teaching Model of Python Programming Based on BOPPPS." 2021 16th International Conference on Computer Science & Education (ICCSE). IEEE, 2021.

[12] Fang, Xiang. "The Teaching Innovation and Research of Python Programming in Financial and Economic Universities." 2020 15th International Conference on Computer Science & Education (ICCSE). IEEE, 2020.

[13] Hashem, Ibrahim Abaker Targio, et al. "The rise of "big data" on cloud computing: Review and open research issues." Information systems 47 (2015): 98-115.

[14] Hu, Hui, et al. "Big data analytics for MOOC video watching behavior based on Spark." Neural Computing and Applications 32.11 (2020): 6481-6489.

[15] Zhou, Shan, et al. "Research and Application of Mixed Teaching Method of Python Programming Based on SPOC." 2019 2nd International Conference on Information Systems and Computer Aided Education (ICISCAE). IEEE, 2019.

[16] Khanam, Fahima, Itisha Nowrin, and M. Rubaiyat Hossain Mondal. "Data visualization and analyzation of COVID-19." Journal of Scientific Research and Reports 26.3 (2020): 42-52.

[17] Eken, Süleyman. "An exploratory teaching program in big data analysis for undergraduate students." Journal of Ambient Intelligence and Humanized Computing 11.10 (2020): 4285-4304.

Sources of funding for research presented in a scientific article or scientific article itself

Report potential sources of funding if there is any