Higher Education and Practice Vol. 1 No. 2, 2024

Research on Curriculum Design of Data Mining in Different Majors

Yong Luo[1], Yingyang Chen[2], Junwei Gan[1*]
[1]*School of Economics and Management, Sichuan Tourism University, Chengdu, Sichuan, China*
[2]*Innovation and Entrepreneurship Institute, Sichuan Tourism University, Chengdu, Sichuan, China*
**Corresponding Author.*

Abstract: With the increasing attention paid to the application value of big data, data mining has become an important major course not only in big data, computer and information technology, but also in most majors in colleges and universities. This course involves the explanation of classic algorithms of data mining as well as the application of classical algorithms of data mining, so it usually has both theoretical and practical hours in the composition of teaching form. Because of the differences in the knowledge structure of students in different majors, there should be differences in the theoretical teaching and practical teaching design of data mining courses among different majors. The teaching content should be selected according to the characteristics and needs of students in different majors, and should also consider the development and trends of big data technology. In addition, in order to improve the teaching quality and effectiveness of data mining courses, it is also necessary to use modern teaching methods and technologies to support teaching, such as online teaching resources, flipped classroom teaching, etc. In short, it is necessary to carry out targeted teaching design for data mining courses in different majors to match the basic situation of students in different majors, ensure better teaching effect and play a better role in different fields of specialization.

Keywords: Data Mining; Theory Teaching; Practical Teaching

1. Introduction

Data mining technology, in the era of big data, involves analyzing data to unearth its internal patterns and using these patterns to serve various fields in society [1].The characteristics and value of big data have made big data mining increasingly important [2].Data mining is an interdisciplinary field that involves machine learning, pattern recognition, inductive reasoning, statistics, databases, high-performance computing, and other areas. Each of these fields can be further studied. For applied undergraduate students, simplifying complex reasoning while highlighting applications is particularly important. Students from different majors have significant differences in their prior coursework, understanding of knowledge points, and ability to operate software. Li Ting and Zhang Jizhou [3] also suggest that the focus should be different for students from different majors. If the students' understanding of reasoning and software operation is not considered, teaching according to a uniform standard will inevitably lead to students' inadequate understanding and thinking the course is difficult with little effect. Data mining courses mainly involve theoretical and practical parts. The theoretical part mainly involves the principle of algorithms, while the practical part focuses on algorithm implementation.

In universities, data mining courses can be offered in multiple majors to cultivate students' knowledge application and practical skills in their major areas of study. The ability of data analysis and mining has become an essential basic skill for applied talents. Therefore, it is very necessary to explore how to develop data mining courses that are more suitable for students with different majors' backgrounds and better utilize their values in different professional fields to cultivate innovative applied talents with strong practical abilities and data science thinking [4]. There are many achievements in the teaching and exploration of data mining, such as Hailin Li [5], which discusses how to design data mining courses in universities from the perspectives of the

influence of big data, the specificity of big data, and the strong interdisciplinary nature of data mining, and explores how to cultivate data awareness, strengthen theoretical systems, innovate teaching strategies, and deeply study scientific research. It provides curriculum design methods to solve the problems brought by the abstraction of data mining courses in the big data environment and improve teaching quality. Ren He et al. [6] designed a "Health Data Mining" course for medical schools based on the innovation dimension, with a focus on three aspects: innovative talent dimension design, innovative talent dimension-based course design methods, and scenario-based course knowledge module design. They provided a detailed account of the course design ideas. Qin Huani [7] explored the multi-dimensional teaching reform of data analysis and mining courses. Wu Jinmeng et al.[8] discussed how to better integrate data mining theory and practice in the actual teaching process. Chen Yanan[9] discusses how to use this teaching tool well and build a high-quality secondary vocational mathematics classroom based on the background of "big data". This article proposes some issues that should be noticed when facing multiple majors in data mining courses for compound and innovative applied talents training.

2. A Difference of Data Mining Theory Course in Different Major Course Designs

Understanding the algorithm principles of Data Mining courses requires students to have strong mathematical skills and analytical abilities. A large number of abstract definitions and algorithms can easily intimidate students [10]. To minimize the difficulty of students' learning and enhance their interest in learning, different data mining algorithms should be selectively explained based on students' majors and the total course hours. In addition, examples should be introduced as much as possible during the explanation, and the examples should focus on students' majors, using familiar professional backgrounds to alleviate the fear of abstract theories. Taking the logistics engineering major and trade economy major, which have opened Data Mining courses in my college, as an example, specifically:

The talent training plan for the 2019 logistics engineering undergraduate major includes 64 hours of data mining courses, divided into 32 hours of big data and data mining theory and 32 hours of big data and data mining experiments. Among them, 32 hours of big data and data mining theory are pure theoretical teaching, while 32 hours of big data and data mining experiments are pure experimental teaching. The total hours for the trade economy major are 32, including 16 theory hours and 16 practice hours. Therefore, in the theoretical teaching process, the logistics engineering major teaches more algorithms than the trade economy major. From the industry perspective, both majors require a certain level of data prediction ability, so both majors teach the algorithm principles of nearest neighbor analysis and decision trees. The logistics engineering major involves location selection issues, so it involves clustering algorithms, while the trade economy major involves trade data analysis, resulting in a higher demand for association rules. In addition, considering the class hours and students' basic conditions, students in the logistics engineering major also learned about neural networks and some combined algorithms such as random forests. In terms of selecting case data, more logistics-related data is chosen for students in the logistics major, such as order delivery-related data. In the trade economy major, more shopping-related data is chosen as an example to bring it closer to students' professional fields.

3. A Difference Discussion of Data Mining Practice Course in Different Major Course Designs

3.1 Classification of Data Mining practical Platforms Based on Operation Modes

In various data mining tasks and considering diverse user requirements, data mining platforms can be classified into distinct categories, each with unique features and suitable scenarios.

Graphical User Interface (GUI) Platforms provide an intuitive graphical interface, allowing users to perform data mining tasks through drag-and-drop operations. Such platforms are particularly suitable for non-technical individuals or beginners who can explore data and construct models without writing code. For instance, RapidMiner is a powerful GUI platform supporting the entire

process from data cleansing to model evaluation, enabling users to quickly get started without coding.

Programming Interface Platforms are designed for users with programming experience, offering more flexibility for customization and control. By using programming languages like Python or R, users can delve into data exploration and conduct advanced analyses. Jupyter Notebooks represent a typical programming interface platform, supporting multiple programming languages and providing an interactive computing environment for data scientists and developers.

Automated Data Mining Platforms aim to simplify the data mining process by automatically selecting algorithms and adjusting model parameters, reducing the technical threshold for users. DataRobot is a representative automated platform capable of automatically executing feature engineering, model selection, and tuning. It is suitable for rapid modeling scenarios, especially for non-professionals such as business analysts.

Big Data Data Mining Platforms focus on handling large-scale datasets and distributed computing environments. Apache Spark MLlib is a robust big data platform that leverages the Spark parallel computing framework, supporting distributed machine learning and data mining tasks suitable for high-performance enterprise applications.

Cloud-Based Data Mining Platforms are built on cloud computing architecture, providing users with flexible computing and storage resources. Google Cloud AI Platform is a typical cloud-based platform supporting end-to-end machine learning and data mining in the cloud, suitable for teams requiring elastic computing and collaboration.

Open Source Data Mining Platforms have open-source code, allowing users to modify and customize as needed. scikit-learn is a widely used open-source platform providing simple and efficient tools for data mining and machine learning within the Python ecosystem, suitable for community participation and contribution.

Integrated Data Mining Platforms offer users comprehensive data mining tools and functionalities, covering everything from data preprocessing to model construction. KNIME is a powerful integrated platform supporting the entire data mining process through a graphical user interface and numerous plugins, suitable for users needing to accomplish multiple tasks on a single platform.

3.2 The Advantages and Disadvantages of Different Practical Platforms.

This section compares seven types of data mining practical platforms from Section 3.1 across four dimensions: advantages, disadvantages, programming skills required, and whether they are open source. The comparison results are shown in Table 1.

3.3 Selection of Different Professional Practice Platforms

In the process of completing the data mining practice course, students of different majors should choose a programmable practice platform or an interface operation practice platform. Combining the advantages and disadvantages of various practice platforms in Table 1 with the characteristics of students' knowledge foundation, we can better select suitable practice platforms to improve students' learning effectiveness and practical abilities. The selection should be based on the talent training plan of different majors and whether the students have programming foundation. Taking the 2019-level logistics engineering and trade economics in my college as an example. Students in the 2019-level logistics engineering major have a programming foundation and learned C programming language before. The practice course has 32 hours of practice, which is more than that of the 2019-level trade economics major, which has only 16 hours of practice. Therefore, students in the 2019-level logistics engineering major can choose a more challenging practice platform that requires programming, while students in the 2019-level trade economics major can choose a visual practice platform to avoid wasting time on programming and debugging code. Finally, students in the 2019-level logistics engineering major have no time to implement data mining algorithms.

4. Conclusion

The ability of data analysis and data mining has become an essential professional skill for many majors in the current era. How to enable students from different majors to effectively apply this skill is a critical consideration in the talent cultivation process.

Table 1 Comparison Results of Seven Practical Platforms

Platform Type	Advantages	Disadvantages	Programming Skills Needed	Open Source
GUI Platforms	- User-friendly with a visual interface - Quick learning curve for non-technical users - Suitable for basic data exploration and modeling	- Limited flexibility in complex tasks - May lack advanced customization options	Low	Depends
Programming Interface Platforms	- High flexibility for customization - Ideal for advanced analyses and coding enthusiasts - Access to a wide range of libraries and frameworks	- Steeper learning curve for non-programmers - May be challenging for those without coding skills	High	Depends
Automated Data Mining Platforms	- Streamlines the data mining process - Reduces the need for extensive domain knowledge - Suitable for rapid prototyping and model iteration	- Limited transparency in model selection and tuning - May not handle highly specialized tasks well	Low	Depends
Big Data Data Mining Platforms	- Handles large-scale datasets and distributed computing - Scalable for enterprise-level applications - Supports parallel processing for faster computations	- Complex setup and configuration for some platforms - Resource-intensive, may require significant computing	Moderate to High	Depends
Cloud-Based Data Mining Platforms	- Flexible computing and storage resources - Enables collaborative work in a cloud environment - Easy access to cloud-specific tools and services	- Dependency on internet connectivity and cloud service - Costs associated with cloud usage	Low	Depends
Open Source Data Mining Platforms	- Customizable and modifiable source code - Active community support and continuous development - Cost-effective and often free to use	- May have a steeper learning curve for beginners - Some platforms may lack comprehensive documentation	Moderate to High	Yes
Integrated Data Mining Platforms	- Comprehensive suite of tools for end-to-end workflow - Streamlines the entire data mining process	- May be overwhelming for simple or specific tasks - May have a steeper learning curve for beginners	Low to Moderate	Depends

Note: The "Open Source" column indicates whether the platform's source code is open and accessible for users to view, modify, and distribute. The "Depends" entry indicates that openness can vary among platforms within the same category.

For applied undergraduate students, it is particularly important to simplify complex reasoning while emphasizing practical applications. Students from different majors may have significant differences in their prior coursework, understanding of knowledge points, and ability to operate software. Therefore, it is essential to customize the teaching approach and content to meet the needs and capabilities of each individual student.

To effectively teach data mining, it is recommended to use a combination of classroom lectures, hands-on exercises, and case studies. Classroom lectures should provide an overview of the field and introduce fundamental concepts and algorithms. Hands-on exercises allow students to apply their knowledge in a practical setting, while case studies demonstrate the real-world applications of data mining techniques.

Moreover, it is important to incorporate

modern teaching methods and technologies into the data mining curriculum. Online resources, flipped classroom teaching, and collaborative learning environments can enhance student engagement and facilitate effective learning. By providing opportunities for students to actively participate in class discussions, collaborate on projects, and interact with peers, the teaching experience becomes more engaging and effective.

In conclusion, teaching data mining effectively requires a tailored approach that considers the differences in students' backgrounds and capabilities. Emphasizing practical applications while simplifying complex reasoning is essential for undergraduate students. By incorporating modern teaching methods and technologies, engaging students in active learning, and providing opportunities for collaboration and real-world applications, data mining courses can be made more accessible and effective for students from various backgrounds.

Acknowledgments

This work was supported by grants from the Sichuan Tourism University's School-level Teaching Reform Project (No. JG2021026, No. JG2021005, Sichuan Tourism University [2021] No. 215), Logistics Engineering construction project for Sichuan Province's higher education institutions' provincial curriculum ideological and political demonstration courses (Sichuan Education Department [2022] No. 199), Cold Chain Logistics construction project for Sichuan Province's online and offline blended first-class courses (Sichuan Education Department [2021] No. 493).

References

[1] Ding Yitao. Exploration of Data Mining Course Reform in the Era of Big Data. Technology Wind, 2021(27): 27-29.

[2]LAI Jian-hui, WANG Yang, LUO Tian-tian, LIU Di, CHEN Yan-yan.Discussion on the Training System of Big Data Mining and Application Ability for Traffic Engineering Major. Education and Teaching Forum, 2021(28): 169-172.

[3]LI Ting, ZHANG Ji-zhou. Data Mining Course Construction Research for Undergraduate in the Big Data Environment. Educational Modernization, 2017(40): 230-232.

[4]Yang Xiuzhang, Wu Shuai, Xia Huan, Zhou Jisong, Yu Xiaomin, Fan Yufeng.Research on teaching reform for the data mining and analysis course in big data era.Computer Era, 2021(9):107-111.

[5] Hailin Li .Teaching exploration of data mining in big data environment. Computer Era, 2014(2):54-55.

[6] REN He, XIAOZhengguang, SHAO Zeguo, LI Ping, LING Chen. Exploration of curriculum design of“HealthDataMining”of Big Data Specialty in medical colleges and universities. CHINA EDICALEDUCATION TECHNOLOGY, 2023, 37(6): 700-704+714.

[7] Qin Huani. Exploring Multidimensional Teaching Reform of Data Analysis and Mining Courses. Journal of Higher Education, 2023(32): 111-115.

[8] Wu Jinmeng, Ma Lei, Zhang Yaozong.Reform and Optimization of the Teaching Mode for the Course of Data Mining and Big Data Analysis[J]. The Guide of Science & Education,2023(14):95-97.

[9] Chen Yanan.Research on Improving the Quality of Secondary Vocational Mathematics Classroom Teaching in the Context of Big Data.Mathematics Learning and Research, 2021(35): 98-100.

[10] CAO Yuan-nan, XU Wei-xia. Discussion on the Teaching of Data Mining. Education and Teaching Forum, 2022(4):172-175.