

A Study on the Construction of Data Mining Course-Oriented Data Literacy Training

Xiaohong Shen^{1,2,a,*}

¹School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, Shandong, China

²Shandong Research Center of China-US Digital Media International Cooperation, Jinan, Shandong, China

^asnxhong@sina.com

*corresponding author

ABSTRACT

The data literacy training is becoming important in the big data era and data mining course has been one of the key courses to cultivate data literacy in universities. After analyzing the problems in the actual teaching, the paper discusses the construction of data mining course oriented data literacy training and proposes a new project-driven teaching scheme that includes the case-based classroom teaching and flipped-classroom practice teaching with a multilevel practice system. The new scheme can offer an individualized teaching and promote the learning enthusiasm and initiative of students efficiently.

Keywords: data literacy training, data mining, project-driven, flipped classroom

1. INTRODUCTION

With the rapid development of mobile Internet, cloud computing and Internet of things, more and more data have been generated in various data-centric fields, and people's data literacy is becoming increasingly important. It has reached a consensus that data thinking will ignite the intelligent revolution[1]. Thinking with data can help people learn techniques for turning data into knowledge they can use and has been widely advocated as a scientific literacy and practical methodology. In recent years, the concept and connotation of data literacy education have been deeply studied[2]. Generally, the data literacy education in the big data era can be divided into six aspects: data attitude, data awareness, data knowledge, data skills, data application, data ethics and morality[3].

Data mining aims to extract information and knowledge from a large number of data that people do not know in advance but are potentially useful[4]. It is a step in knowledge discovery in database and big data and includes the process of data collection, data creation, data management, data analysis and value information generation. It is mainly related to computer science, involving multi-disciplinary integration such as statistics, database technology, information retrieval, machine learning and artificial intelligence technologies[5]. Now Data mining course is one of the important courses to cultivate data literacy in universities.

To improve the students' ability of data application and data innovation, the paper has explored the teaching reform on data mining course, especially oriented data literacy training in the big data environment. The rest of this paper is organized as follows. In Section 2, we analyze the problems in the actual teaching process of data mining course. In Section 3, we propose a teaching scheme oriented project-based data literacy training about data mining course and discuss the implementation of our scheme. In Section 4, the conclusion is given.

2. THE PROBLEMS IN THE ACTUAL DATA MINING COURSE

Data mining is an interdisciplinary course and aims to train students to have the ability to find and solve data related problems. It covers a wide range of knowledge and involves a lot of basic theoretical knowledge and complex algorithms, which brings great challenges to teachers. The problems of the actual teaching process in data mining course are analyzed in the following.

(1) The teaching mode is simplistic. Most teaching modes in the data mining course are characterized by the traditional teacher-preaching and student-listening and lack of interaction. The teaching mode is very simplistic and students suffer from the passive learning and cognitive overload. As a result, it is difficult to active students' learning enthusiasm, especially in the face of those obscure, theoretical and complex algorithms.

(2) The practical application of the experiments is poor. The teaching mainly focuses on the theoretical description of mathematical models and algorithms, and neglects the practical application of experiments, which lead to the disconnection between theory and practice. Moreover, those experiments are relatively independent and datasets are small, so it is difficult for students to systematically understand the nature and process of data mining.

The cultivation ability of sustainable development is insufficient. Teachers usually pay more attention to the explanation of basic theoretical knowledge, but rarely emphasize the relationship between knowledge. This teaching approach can not reflect the interdisciplinary characteristics of this course, and does not well meet the requirements of diversified learning and development of students.

3. THE TEACHING SCHEME ORIENTED PROJECT-BASED DATA LITERACY TRAINING

3.1. The Design of Our Scheme

To improve the cultivation ability of data mining course, especially the training of data application and innovation ability of students, we propose a new teaching scheme oriented project-based data literacy training of data mining course. The project-based data literacy training method originates from the outcome-based education (OBE) training method which was proposed by Spady, an American scholar,

in 1981[6]. OBE is famous as the educational concept that learning output drives the whole curriculum activities and the learning of students.

In our scheme, the curriculum contents of data mining are firstly formulated and divided into seven modules according to the real requirement of data literacy training. Then several projects are chosen from the international data mining competitions based on the curriculum contents and the typical teaching cases are produced from these projects to organize the teaching modules used for both classroom teaching and practice teaching. Figure 1 shows the module organization of data mining course. In each module, the relationships of the related knowledge points are analyzed and

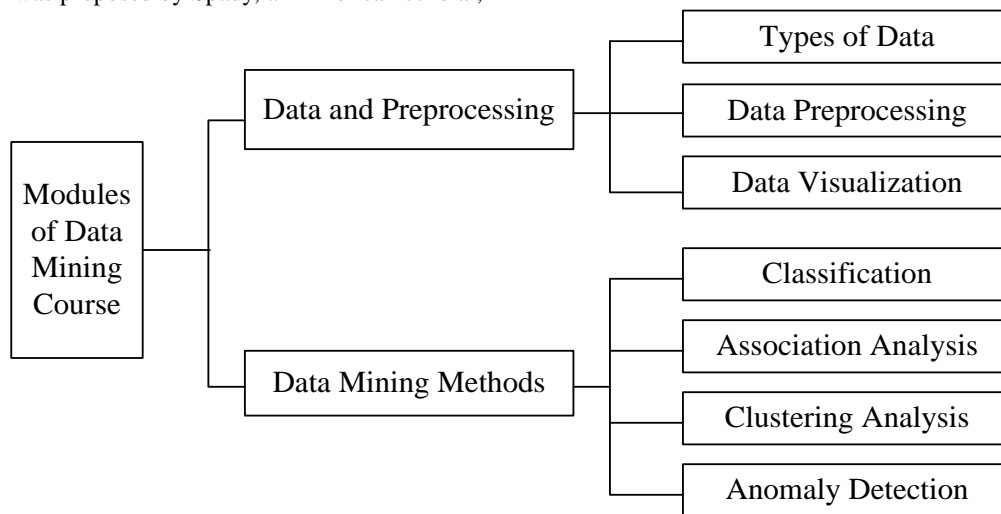


Figure 1 The Module Organization of Data Mining Course

the multilevel child projects are provided to implement personalized teaching. The case-based or project-driven modules can be easily performed as flipped-classroom teaching to promote multilevel practice teaching[7,8].

3.2. The Implementation of Our Scheme

3.2.1. The Case-based Classroom Teaching

The classroom teaching for each module is carried out from the case analysis. The teaching case tells the data problem to

be solved as well as its application background. After that, teacher need to explain the knowledge points of data mining involved in this case, including types of data, data preprocessing, data mining methods and visualization technology. Table 1 gives the cases and related algorithms for each teaching module. Obviously, taking the case as an example can easily help students understand the abstract theoretical knowledge of data mining, which can improve the learning enthusiasm and initiative of students. Finally, teachers expand the relevant knowledge and instruct students to transfer knowledge to solve the new data problems.

Table 1 The Case for Each Teaching Module

Modules	Cases	Related Algorithms/Function	
Basic Topics	Data Preprocessing	Sales Data of Medical Company	Data Cleaning
	Data Visualization	Sales Data of Medical Company	Visualization
	Classification	Spam Emails Classification	C4.5, KNN, Naive Bayesian Classifier
	Prediction	Red Wine Quality Identification	CART
	Association Analysis	Beer and Diaper	Apriori, FP-Growth
	Clustering Analysis	Retail Customer Segmentation	K-Means, BIRCH
	Anomaly Detection	Credit Card Fraud Detection	Logistic Regression
Advanced Topics	Collaborative Filtering	Recommendation System	SVD++
	Text Mining	Character Recognition	CNN+RNN+CTC

3.2.2. The Multilevel Project-driven Individualized Teaching

In the practice teaching part, we present a multilevel project-driven practice system to provide an individualized

teaching for students. The projects involve many data mining areas such as finance, telecommunication, Internet, public service and education. Table 2 describes the projects for multilevel practice teaching.

Table 2 Projects for Multilevel Practice Teaching

Application Area	Project
Finance and Telecommunication Internet	Precise Intelligent Marketing Based on Customer Segmentation
	Stock Trend Prediction Based on Inflection Point Change
	Product Recommendation in E-commerce Website
	Customer Service Perception Evaluation for Information System
Public Service	Credit Rating for Enterprise and Company
	Purchase Behavior Analysis of Cigarette consumers
Education	Online Learning Behavior Analysis
	The Knowledge Map of Education Science Based on Journal Database

3.2.3. The Flipped-classroom Practice Teaching

The flipped-classroom practice teaching works as the student-center education and puts forward based on the cases, which gains an excellent teaching effect to improve the interest and initiative in learning. This part is partitioned into three sections that are design before class, discussion in class and application after class. The learning task list and some learning materials will be provided to students before class to make sure learning objectives and knowledge points. In the class, students are divided into groups to discuss the cases and projects. The common problems are summarized by teacher to offer interactive teaching. After the class, each group need to prepare and make a PowerPoint presentation to share their implementation of the project.

students to easily understand these abstract theoretical knowledge points. Projects are used to carry out the flipped-classroom practice teaching and provide a multilevel practice system. This teaching scheme can provide an individualized teaching for students and improve the enthusiasm and initiative in learning.

4. CONCLUSION

To improve the data literacy training ability, we study the teaching reform of data mining course and propose a project-driven teaching scheme oriented data literacy training. In this scheme, the teaching cases that produced from projects are used in the classroom teaching to help

ACKNOWLEDGMENT

This article was supported by the Teaching Reform Research Project of Shandong University of Finance and Economics of China (Grant No. jy201854).

REFERENCES

[1] Shron M., "Thinking with Data: How to Turn Information into Insights", O'Reilly Media. 2014.
 [2] He H., "Research on the Data Literacy Education in Chinese Academic Library under the Big Data Background", Journal of Modern Information. 2015, 9.
 [3] Hu H., Wu M., "What Information Literacy Education Can Bring to Data Literacy Education", Library Theory and Practice. 2016, 11.
 [4] Tan P. N., Steinbach M., Karpatne A., Kumar V.,

- “Introduction to Data Mining”*, Pearson. 2018.
- [5] Han J.W., Kamber M., Pei J., *“Data Mining: Concepts and Techniques”*, Morgan Kaufmann. 2011.
- [6] Spady W. G., *“Outcome-Based Education: Critical Issues and Answers”*, American Association of School Administrators. 1994.
- [7] Giannakos M., Krogstie J., Chrisochoides N., *“Reviewing the Flipped Classroom Research: Reflections for Computer Science Education”*, Computer Science Education Research Conference ACM. 2014, 11.
- [8] Peterson J. D., *“The Flipped Classroom Improves Student Achievement and Course Satisfaction in a Statistics Course: A Quasi-Experimental Study”*, Teaching of Psychology. 2016, 1.