

Exploration and Practice of Data Mining Course Teaching Oriented Big Data

Gang Lei^{1, a}, Chen Xiong^{1, b}

¹Software College of Jiangxi Normal University, Nanchang, China

^a497027087@qq.com, ^b502752890@qq.com

Keywords: Data Mining. Course teaching. Big Data. System concept.

Abstract. The course Data Mining, one of the core courses in many majors, has evolved from Data Mining oriented Warehouse, Web Data Mining to the current Big Data Mining. This paper has explored course teaching on discussion in the technology of big data mining and course practice in real data mining. We propose a course teaching plan about Data Mining course oriented big data, analyze features of the course teaching plan with the system concept, and describe the teaching practice to postgraduates majored in Management Science and Engineering. The effect of teaching practice is exciting.

Introduction

The course Data Mining was set up in the 1990s in America. Thanks to Professor Han Jiawei' s book *Data Mining Concept and Techniques*[1,2], the course Data Mining has been offered by most colleges in China and it has been one of the core courses for undergraduates, postgraduates and PhD students from Software Engineering, Computer Science and Technology, and Management Science and Engineering fields, etc.

The domestic course design of Data Mining mostly follows America and the main teaching materials are from Professor Han Jiawei' s book *Data Mining Concept and Techniques* and Pang-Ning Tan' s book *Introduction to Data Mining*[3,4]. For the better course design, domestic experts and teachers have explored and practiced actively, such as the newly literature seminars in Beijing University[5], training in real data mining in Fudan University[5], application data mining to all walks of line in University of Science and Technology Beijing[5] and adjusting the course to students' level in Zhenzhou University[5].

The course Data Mining evolves from Data Mining oriented Warehouse, Web Data Mining to the current Big Data Mining. Professor Philips YU, an Authority Figure in data mining field, Professor at the University of Illinois Chicago distinguished director and director of the Institute of Tsinghua Science Data, stated that with the features of Volume, Velocity, Variety and Veracity, new approaches and technologies in data mining have mushroomed recently[5]. The big data contain considerable and meaningful information, which exerts great influence on the future society. Therefore it is necessary to focus on specific field and then offer its own solutions and measures in data mining.

This paper has explored course teaching on discussion in the technology of big data mining and course practice in real data mining. The main contributions of this paper are:

- (1) We propose a course teaching plan about Data Mining course oriented big data.
- (2) We analyze features of the course teaching plan with the system concept.
- (3) We describe the teaching practice to postgraduates majored in management science and engineering.

The rest of this paper is organized as follows. In Section 2, we propose a course teaching plan. In Section 3, we analyze features of the course teaching plan. In Section 4, we describe the teaching practice. In Section 5, we conclude the paper.

Course Teaching Plan

The course teaching plan mainly consists of teaching plan construction , teaching models and experiments.

Teaching Plan Construction. The teaching plan construction mainly covers teaching objectives, teaching organization and teaching materials.

(1) Teaching objectives. The course Data Mining is classified as an applied curriculum and aims to help students to analyze the data in the specific field after they learned the data mining approaches and tools. Therefore, in the limited course time, students are required to understand and apply classical and frequent-used models and algorithms and be familiar with data mining tools instead of learning all principles, models and algorithms.

(2) Teaching organization. In terms of time, there are 54 lessons totally, 3 lessons per week. In terms of content, there are basic sections, such as Data, Association analysis, Classification, Cluster analysis, Outlier analysis, and thematic sections of big data, such as Data quality, Collaborative recommendation, Crowd sourcing, Privacy protection. The basic sections will be compulsory and the thematic sections will be optional for students.

(3) Teaching materials. The following textbooks will be used as teaching materials: Pang-Ning Tan' *s Introduction to Data Mining*, Han Jiawei' *s Data Mining Concept and Techniques*, Rajaraman' *s Mining of Massive Datasets*[6]. Web courseware and video (video open class), as well as the classic series of foreign literature will be used as supplementary teaching materials.

Teaching models. We apply four models to the course teaching[7].

(1) Model 1. Combining theory with practice. The theory of data mining methods will be elaborated by the lecturers and experiment and literature reading will be in form of seminar after class.

(2) Model 2. Textbook with literature. Classic foreign materials are mainly the curriculum teaching content. Students discuss the relevant data of Surveys & Tutorials as well as the authoritative journals or conferences about data mining

(3) Model 3. Lecture with report. The content of the teaching material is taught by the teacher and the reading report and the experiment report are explained by the students in the class. The teacher's teaching and the student's report account for about 5:1.

(4) Model 4. Knowing with applying. The integration and utilization of mathematics courses and algorithm strengthen students' cognition and understanding of the data mining model. On this basis, considering students' field direction, we train students to apply what they have learned to the real data mining .

Experiment period.

(1) Experiment design. The practical teaching of this course is based on the experimental target, and the training intensity is high. The experimental type includes the verification experiment, comprehensive experiment and the experimental tools includes programming (advanced programming language, R language, Matlab), the use of tools (Weka, SPSS, Sqlserver, Oracle, Excel). The experimental platform consists of general experimental platform, big data platform (Hadoop/MapReduce).

(2) Experimental condition. The experiments consist of verification experiment and comprehensive experiment. For the verification experiment, the data set is an open standard data set, and the data collection of each field is collected by the machine. The experimental source is the realization of the classical data mining algorithm on the data set, or the reproduction of the experimental data in the literature. The comprehensive experiment, according to their respective research directions, can be combined with various types of events held in the same period big data mining related science and technology competition. The experiment was done by students in their laptops in the classroom, the big data environment provided by the college 6 PC machines to form a cluster. Students can make good use of QQ group for discussion.

Features of Teaching Plan

With the system concept, the teaching plan, in the overall framework of professional core curriculum system, makes full use of the existing model statistics and algorithms course, and face big data applications as far as possible. The features of the teaching plan are as follows:

(1) System concept: Data Mining course centered core curriculum system contains Mathematical Modeling, Data Mining, Algorithm Design and Analysis, Pattern Recognition, Big Data Frontier 5 specialized core courses, shown in Fig. 1.

(2) High strength experiment: industry oriented application requirements in teaching practice, with wide knowledge, high strength experiment, experimental standard for high load intensity of training students, give students new learning experience.

(3) The two sides of the classic: teaching practice using classic teaching materials and classic experiments as the main course teaching, the forefront of literature reading and the latest big data platform experiment as an auxiliary.

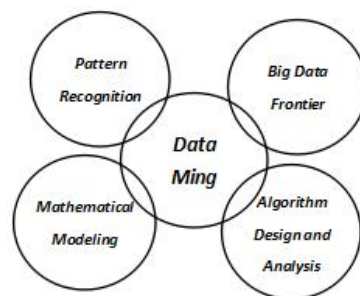


Fig.1 framework of professional core curriculum system

Teaching Practice

With the above teaching plan of the course, the author has implemented teaching practice for 12 postgraduates majored in Management Science and Engineering (3 years) in the class of 2015 in Jiangxi Normal University. The details of the teaching practice are follows:

(1) Course group arrangement: Data Mining course teaching has been arranged in the second semester, when Mathematical Modeling course and Algorithm Design and Analysis course have been also arranged. Pattern Recognition course and Big Data Frontier course have been arranged in the third semester.

(2) Teaching content: Pang-Ning Tan' s *Introduction to Data Mining*, available to beginner, has been chosen as main textbook, whose auxiliary slices have been adopted in the classroom. Basic sections of textbook have taught by teacher, while advanced sections of textbook and thematic sections of big data have been discussed among students and teacher.

(3) Experiment data sets: We have chosen the auxiliary data sets of textbook as ordinary data sets, while big data sets came from three application fields: first was TianChi big data competition[8], second was traveling field, and third was industrial and commercial field.

(4) Student teams: All discussions and reports were carried through as a student team for the unit, and each team was usually made up of two students freely. During discussion and report, teacher was in charge of guiding and commenting.

(5) Exam: The examination paper was told to students at the beginning of semester, that is course paper, in forms of overview, technical report, or innovation thesis.

After one semester of teaching practice, the effect is good, it achieves the expected teaching objectives. Particularly mentioning that several course papers have been published in core journals or international conferences.

Conclusions

This paper has explored course teaching on discussion in the technology of big data mining and course practice in real data mining. We have carefully formulated and effectively carried out a scientific course teaching plan about Data Mining course oriented big data. The future teaching will be pursued.

Acknowledgements

This work was financially supported by the Jiangxi Provincial Research Project of Teaching Reform of Degree and Graduate Education (JXYJG-2015-047): Reform of Data Mining Course Teaching Oriented Big Data, and 2016 Jiangxi Provincial Department of Education Science and Technology Research Projects: Enterprise Entity Resolution in Enterprise Credit Supervision Warning System.

References

- [1] Jiawei Han, Micheline Kamber and Jian Pei: *Data Mining Concept and Techniques*(the third edition) (Morgan Kaufmann Publishers, USA 2011).
- [2] Information on <http://web.engr.illinois.edu/~hanj/>
- [3] Pang-Ning Tan, Michael Steinbach and Vipin Kumar: *Introduction to Data Mining* (Addison-Wesley Publishing Company, USA 2006).
- [4] Information on <http://www-users.cs.umn.edu/~kumar/dmbook/index.php>
- [5] Ming Fan, etc, in: *Technical Report of Data Mining Course Teaching Seminar*, edited by China Machine Press, Beijing, China (2012), in press.
- [6] Anand Rajaraman, Jeffrey David Ullman: *Mining of Massive Datasets* (Cambridge University Press, USA 2011).
- [7] Shengyi Jiang, Xia Li and Qi Zheng: *Computer Education* Vol. 24 (2014), p. 19 (in Chinese).
- [8] Information on <https://tianchi.shuju.aliyun.com>