

Data mining and analysis based on online teaching platform

Wanjiang Han^{1, a}, YiZhen Wang², Yi Yang²

¹School Of Software Engineering, Beijing University of Posts and Telecommunication, Beijing 100876, China

²International School, Beijing University of Post and Telecommunication Beijing, 100876, China ^aemail: hanwanjiang@bupt.edu.cn

Keywords: SPOC; Online Teaching Platform; Data Analysis; Data Mining

Abstract. With the rapid development of the network and multimedia technology. students no longer simply rely on the teacher's teaching to obtain knowledge. This paper takes Beijing University of Posts and Telecommunications as an example to introduce the online course platform, which provides a wealth of teaching resources, and provides the premise for the flipped classroom teaching mode. At the same time, through the analysis of online platform data, we can mine the relevant information, find the association rules and provide a reference for learning model.

Introduction

The rise of online teaching mode, but also attracted the high tide of the University's reform, flip the classroom gradually accepted by everyone, and gradually be welcomed. Knowledge teaching is through the network in the "class", knowledge internalization is in the "class" after discussion, communication to complete, and in the "after class" students or to complete a higher level of "knowledge exploration". This paper is based on the "software project management" Curriculum of Beijing University of Posts and Telecommunications, which is based on the teaching mode, which is based on the mixed teaching mode, which not only provides the online curriculum resources, but also provides the analysis of the students learning process and results. This is a big data, through large data association analysis, it can provide some conclusions for better guidance and adjustment of online teaching[1][2][3][4].

Online Teaching Platform

The software project management course of Beijing University of Posts and Telecommunications has been carried out through the iClass.bupt.edu.cn platform, and has begun to try out the teaching reform, in the course design, students evaluation and other aspects of the curriculum design, student evaluation and so on. Through continuous summary of experience, we hope to provide a teaching mode can be promoted. This course is about the SPOC model of University of California at Berkeley.

Course Preparation

This course is a compulsory course for undergraduate students majoring in software engineering, and they have mastered the development of language, database and other software development tools, at the same time, students need to comply with the relevant requirements of online learning, complete project practice. During the 16 week of the course, the students follow the teaching plan, combined with their own rhythm, progress and way to study the basic content of the network anytime and anywhere. The classroom time is used more and more to carry out the deep interaction between teachers and students. Each chapter of the online test is a automatic score, automatic scoring function allows students to submit homework, in a higher score, but also to improve their knowledge and skills to master. Each practice group according to their own situation, development and management of the project, after the successful completion of the results can be submitted to online platform.

Platform Resources

This course uses iClass.bupt.edu.cn to improve the online course platform, including curriculum, curriculum, teaching syllabus, curriculum plan, network classroom, online testing, curriculum practice, curriculum practice, exchange and interaction, curriculum case analysis, curriculum resources and so on. See figure 1.

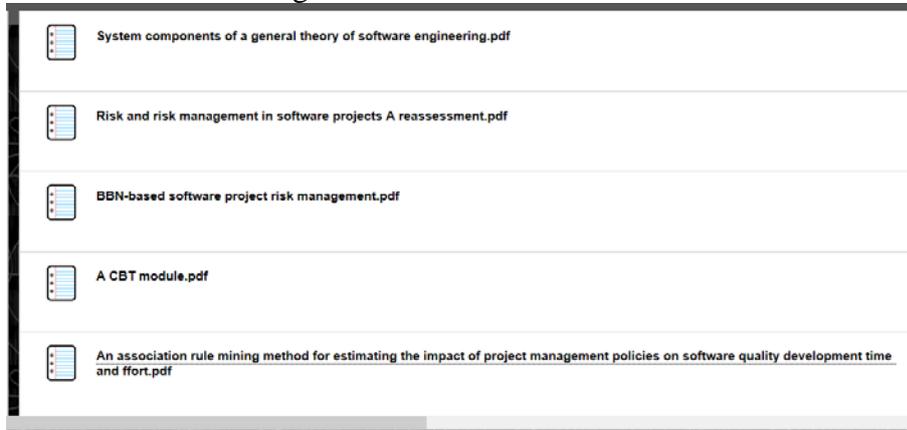


Fig.1. Online course platform

Online Learning

Students learn through the network, the network classroom to provide a simulation environment, including curriculum roadmap and all courses of video, in order to facilitate students to learn, watch video, each chapter of the teaching video in accordance with the knowledge point of the video clips, that is, the basic control of visual time in 5-20 minutes. In the learning process can be through the network platform to question, discuss, express views. At the same time online testing, complete the classroom work.

Data Mining

Data Mining is the Knowledge Discovery phase and it is the process of extracting implicit, useful, previously unknown, non-trivial information from data. The techniques involved in Data Mining are grouped as Classification, Clustering, Association Rules and Sequences. Association Rule Mining uncovers relationships among data in a database. The quality of an Association Rule is measured by using its support and confidence values and several efficient methods are developed.

Association Rule Mining

Association Rule mining was Proposed by Agrawal et al in 1993. It is an important data mining model studied extensively by the database and data mining community. Assume all data are categorical. Its task is to find certain relationships among a set of data (item set) in the database.

An association rule is a pattern that states when X occurs, Y occurs with certain probability.

Association rules mining is interested in finding frequent rules that define relations between unrelated frequent items in databases, and it has two main measurements: support and confidence values. Confidence values are measurements of rule's strength, while support value corresponds to statistical significance.

Frequent items are defined as the itemset that have support value greater than or equal to a minimum threshold support value, and frequent rules as the rules that have confidence value greater than or equal to minimum threshold confidence values. These threshold values are assumed traditionally to be available for mining frequent itemsets. Association rules Mining is all about finding all rules whose support and confidence exceed the threshold, minimum support and minimum confidence values.

Association rules mining proceeds on two main steps. The first step is to find all Frequent

itemsets with adequate supports and the second steps is to generate Association rules by combining these frequent itemsets.[5][6]

Apriori Algorithms

Apriori algorithm has been shown as a classical association rule in mining algorithms that have been cited so far, which is used to find frequent item sets in a database and to generate Association Rules from the frequent item sets

The core issue of the Apriori Algorithm is how to regenerate all the frequent itemsets. It uses an iterative approach layer by layer, applying K-Itemset to search for (K+1)-Itemset. Firstly, the algorithm identifies the collection of Frequent 1-Itemset, denoted by L1. Collection of Frequent 2-Itemset "L2" is computed from L1. We use the same method to acquire L_{k+1} ($K>1$) until it cannot find L_k . Doing this repeatedly, so all the frequency are set out. These rules generated must meet the minimum support and minimum confidence.

Apriori Algorithm adopts recursive method to generate the required frequent sets. Input values are transaction database D and minimum support threshold value min_sup.

Data analysis from the online platform

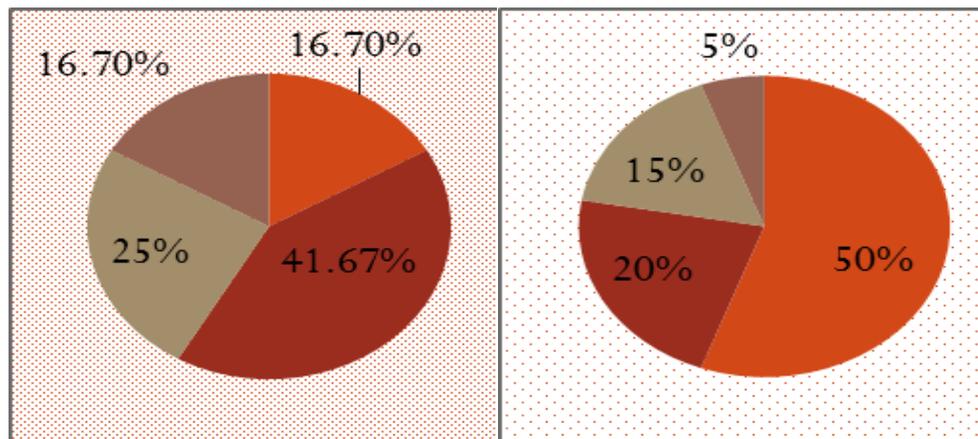
Online platform is not only a learning platform, but also a platform to analyze the learning process. The platform can track and collect a large amount of data.

The Data from the online was taken as a raw data and the preprocessing phase of the KDD process was applied on the data to generate transformed data that could be analyzed to extract knowledge from the data. The process consists of few steps including Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Generation of Patterns, and Knowledge Interpretation.

Learning double-loop is different from the. A practice of double loop learning organization will correct its own mistakes, but it will do many more important things, including the analysis of the causes of their mistakes. Double loop learning needs to analyze the organization itself in the feedback loop in the use of a large number of data. Online education many interesting attempt is about double loop learning, such as mentioned earlier, many neighbors, such as Salman Khan (Salman Khan) of the Khan Academy, in the use of data about how students learn, he not only in the teaching design, but in the design of Khan college itself. The new education of the big data model is a typical double loop learning.

Make full use of big data, the integration of the Internet and education will be truly realized. As the Analects of Confucius said "the labor wants a good job, we must first of its profits", only perfect and mature platform also is soft and hardware system as the basis, Internet Education and the new education model can work well.

We analyzed the learning time of students, compared with learning time between these students whose grade were 90 points or more and 60 points. It is found that the students with high scores are more independent learning time, such as the practice of the project, the poor students' autonomous learning ability, which is basically dependent on the traditional teaching mode. These data indicate that the SPOC teaching mode is more suitable for the students with strong self management. See Figure 2.



(a) Learning time distribution of 90 grades students (b) Learning time distribution of 60 grades students

Fig.2. Learning time comparison

Conclusion

Flipped classroom teaching mode of the Internet + in the education industry, specific, "software project management" course online education provides a wealth of teaching resources, improve the teaching process, flip the classroom to improve the traditional teaching mode, students can also teach, the teacher is more guidance, realize the mixed teaching. Through the data analysis of the online platform, we can dig out the best strategy to provide guidance for students.

References

- [1] Deng Pei, Internet + education to the air outlet in where? China investment advisory network, 2015-10.
- [2] Zhang Qiliang, Wang Aichun. Study on the new type of teaching mode based on "flipped classroom". 2014.04.
- [3]Ma Xiulin, Zhao Guoqing, Wu Tong. University information technology public class flipped classroom teaching empirical research [J]. Distance Education Journal, 2013, (1): 79-85.
- [4]conye chin. Online education "after MOOC era" -- SPOC analysis. Tsinghua University Education Study [J], Vol. 35, No. 1, February 2014.
- [5]Xu Xiaofei. To seize the opportunity to promote the computer and software engineering professional teaching reform of the J. Chinese university teaching, in 2014 first.
- [6]Wan-Jiang Han , Research on the Problem Model of GUI based on Knowledge Discovery in Database,2013