Student Performance Prediction Method Based on Fuzzy Cognitive Model

Zhenbing Zhao[a], Lei Zhang[b] and Dashuai Hu[c]

School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China

[a]zhaozhenbing@ncepu.edu.cn, [b]ai_zhanglei1212@126.com, [c]1260809084@qq.com

Keywords: Fuzzy cognitive model; Performance prediction; Teaching reform; Implement teaching.

Abstract. With the continuous development of education informationization, the data mining technology has been increasingly interlinked with education teaching; analysis of teaching results based on data mining is playing an increasingly important role in the teaching reform. In order to help students understand their own capabilities and help teachers master students' level, this paper introduced the fuzzy cognitive model, proposed a method for student performance prediction. The method collected students' theoretical examination score and experimental score, combining the cognitive diagnosis model to obtain prediction score both in theoretical examination and experiment. We can find the gap through comparing students' actual score and predicted score, so as to improve the teaching pertinence and quality. The result shows that the method can improve the accuracy of performance prediction and reduce the prediction error effectively, so as to provide strong data supports for the teaching reform.

1. Introduction

ITS(Intelligent Tutoring System) is the teaching system which use machine learning, data mining and artificial intelligence to teach learners knowledge and skills more scientifically and efficiently. It can discover the blind spots in the process of teaching and learning intelligently which teachers and students are hard to find, and develop individualized teaching on the basis of these blind spots. Thereinto, data mining technology has just sprung up in recent years, playing a great role in many fields of national production practice [1]. Searching the hidden useful information from large amounts of data through the algorithm associated with it is the principle of data mining, which also is the appeal of mining valuable rules from education information database, so that we can promote the scientific decision-making in education and teaching system.

Student's score is the final presentation of curriculum evaluation; it can test the effect of teaching and can motivate students to learn better. Effective statistical analysis and prediction on student' score plays a crucial role in many aspects, such as allocate teaching resources reasonably and improve teaching quality extensively. In the process of teaching, each student's mastery degree is different on each problem. However, we hope that teachers can master the different "short board" of each student, make teaching plans more suitable for students, modify the curriculum, and finally, develop the individualized teaching energetically. One of the many applications of data mining technology is student performance prediction; it uses data mining to process student's score and has developed organically in the past few years. Back-propagation (BP) neural network was first adopted by performance prediction model because of its advantages such as self-learning, self-organization and self-adaption. However, since the BP network has inherent limitations and some problems existing in performance prediction, the genetic algorithm is used to optimize the weighting and threshold of BP network [2]. In addition, Xie et al. provided an improved TrAdaboost algorithm to overcome the influence of different data distribution [3].

On the basis of above-mentioned research, this paper proposed a method of student performance prediction based on fuzzy cognitive model, in order to improve the predictive accuracy on especially subjective issues. Our method can find student's "short board" both in objective and subjective issues

quickly and accurately, at which teachers aim to develop individualized teaching, so as to improve teaching efficiency and reduce the learning burden. The experiment proved that this method improved the accuracy of performance prediction successfully, reduced the prediction error and provided strong data support for the teaching reform.

2. Student Performance Prediction Based on Fuzzy Cognitive Model

2.1 Fuzzy Cognitive Model Overview.

For the past few years, the related research and application of ITS has developed like a raging fire, such as the *Cognitively Diagnostic Assessment* proposed by Nichols et al.in 2012 [4], MOOC (*Massive Open Online Courses*) proposed by Anderson et al.in 2014 and so on [5]. A crucial step of these educational researches is the cognitive modelling of students, which aims at discovering the latent characteristics of students and so, to predict their performance on each problem. When what we talk about reflected into the exam, and then became the score students get on each problem. However, problems students faced with may have both objective and subjective types, previous perceptual model (PMF, DINA, IRT, NMF, etc.) performs good on objective problems, but has great limitations in the prediction of subjective issues whose scores are continuous. Therefore, we use FuzzyCDF (*Fuzzy Cognitive Diagnosis Framework*) to fuzzify students' potential features, making the predicted score more close to the students' actual level [6].

In this way, FuzzyCDF is represented as a four-tier hierarchical model: first mining potential properties of each student; then determines students' skill proficiency; next computes students' problem mastery and finally generates students' observable scores on problems by considering slip and guess factors. Throughout this whole process, we propose a Monte Carol Markov chain (MCMC) sampling algorithm to infer the unobservable parameters of FuzzyCDF [7].

2.2 Fuzzifying Skill Proficiency.

In this subsection, we show the first step in cognitive diagnosis: to get the proficiency of students on specific skills (e.g. problem-solving skills like calculation) α .

We decomposed each problem into several skills, which are shown through prior knowledge—Q-matrix proposed by education experts, such as inferential capability, calculation and so on. In the traditional modeling, the student's skill proficiency is assumed to be absolutely mastered (i.e. 1) or nonmastered (i.e. 0) so that this modelling can fit objective problems with absolutely right or wrong responses. However, for a subjective problem which has a partially correct response, the above modelling on skill proficiency cannot fit well. Thus, to address this issue, we adopt fuzzy systems in cognitive model and redefine the skill proficiency of a student by the following assumption: the proficiency of a student on a skill is the grade of membership of the student in a fuzzy set which is related to the skill. The implication of this definition is that the proficiency of a student on a specific skill depends on the student's potential properties, the difficulty and discrimination of the skill. Generally, the major difference between this definition and the traditional models is that we redefine the skill proficiency from binary variables to the continuous ones, so that we can predict more reasonable results according to the student's ability.

2.3 Fuzzifying Problem Mastery.

On the basis of fuzzified skill proficiency α in Section 2.2, we can further fuzzify the problem mastery η of students, that is, the probability of being able to solve the problem. The CDMs model proposed by Pardos et al. in 2008 assumed that problem mastery is a result of interaction of student's proficiency on required skills of this problem [8]. This paper categorized the skill's interaction on problems into conjunctive and compensatory. Conjunctive means that a student must master all the required skills to solve a problem, while compensatory means that he is possible to solve a problem as long as he masters any skill required by the problem. Generally speaking, an objective problem has a unique standard answer and cannot be answered correctly unless the student masters all the required skill. Thus the objective problems skill's interaction is usually assumed to be conjunctive. In contrast, a subjective problem is a free-response one and the student can write not only the final answers but

also the solving process including writing equations, deducing, calculating and so on. That is, the more skills a student masters, the higher score he will get on this problem. Therefore we assume that the skill's interaction on subjective problems is compensatory.

Based on the above discussion, if the skills' interactions between each other on a problem is conjunctive (or compensatory), the mastery of a student on this problem is the grade of membership of this student in the intersection (or union) set of the fuzzy sets.

2.4 Student's Score Prediction.

With the problem mastery defined in Section 2.3, we can now predict students' scores on problems.

Two exceptions, slip s and guess g, which affect the score prediction were first proposed by Baker et al. in 2008 [9]. In an examination, the score of a student on a problem not only depends on the student's problem mastery, but also influenced by many accidental factors. For instance, a student who is unable to solve the problem can get a correct response by guessing an answer; meanwhile, he may get a wrong response as a consequence of carelessness while he could have done it right. Here, we assume that each problem has its own slip and guess factors and consider these two exceptions into our model.

3. Teaching Case Analysis

3.1 Construction of the Dataset.

The course of DSP (Digital Signal Process) was taken for example. We collected the performance information from a session of 90 students in our school, including the theory test scores and the experimental scores, each item is a percentage meter and the specific introduction of data are shown in Table 1. We normalized the data to make the calculation more convenient. The brief summary of our dataset is shown in Table 2.

<div align="center">Table 1 Course Data Introduction</div>

Problem	Full Score	Category
1	18	
2	8	
3	20	
4	7	Theory Test
5	7	
6	10	
7	12	
8	18	
9	40	
10	40	Experiment
11	20	

<div align="center">Table 2 Dataset Summary</div>

Problem	1	2	3	4	5	6	7	8
Detail	Blank-filling	Judgment	SAQs	Concept	DFT	System Function	FIR	IIR

Problem	9	10	11
Detail	Basic Experiment	Basic Experiment	Compositive Experiment

Our dataset is represented by a Score matrix and a given Q-matrix by education experts. Fig. 1 shows the preview of our dataset, where each column for each subfigure stands for a problem and each row above and below represents a student and a skill, respectively.
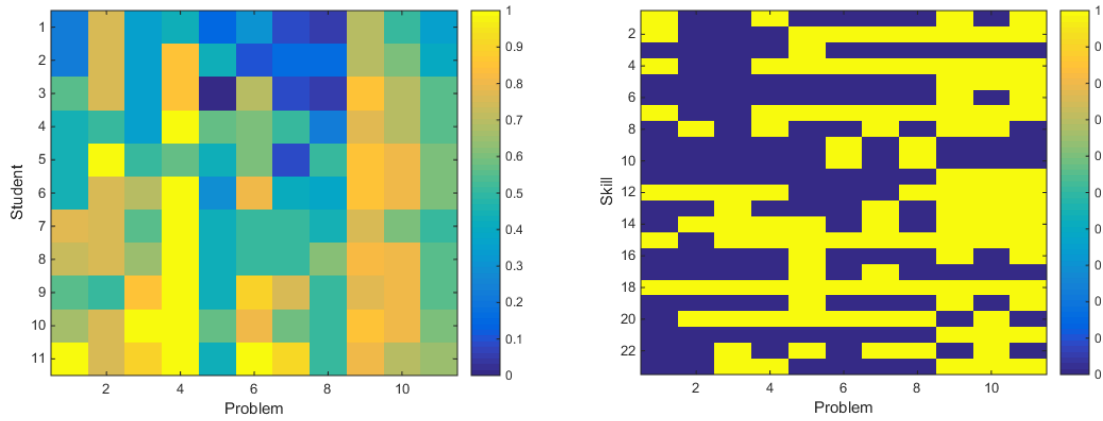
Fig. 1 The preview of the dataset

3.2 Experiments.

The experiments were implemented on the basic of dataset introduced in section 3.1. First of all, in order to get the model parameters, 80% of the data in dataset was used to train FuzzyCDF. And then, we got the predicted scores of 90 students in 11 problems. Here, we calculated the average predicted scores of each 11 problems, that is, the most ideal teaching performance considering the students' potential properties. Finally, we selected a number of students from different scales to find the gap through comparing their actual scores and the average predicted scores, analyzed the teaching effect and put forward the suggestion on teaching reform.

In this experiment, we set the minimum value of iterations of Algorithm to 40 to guarantee the convergence of the Markov chain. Our experiments were implemented by using Matlab 2014b on a Core i7 machine with Ubuntu 14.04 and 4 GB memory.

3.3 Performance Prediction and Analysis.

We selected 11 students in total from different scales to compare actual scores and predicted scores. In order to focus on the students whose score is a bit poor, stu1~stu6 were selected from students whose theory test score is below 60 randomly, while stu7~stu11 belongs to 60~68, 68~76, 76~84, 84~92, 92~100 band, respectively. Fig. 2 (a) shows the contrast of actual score and average predicted score of students in low sublevel, while Fig. 2 (b) describes the students in high sublevel.

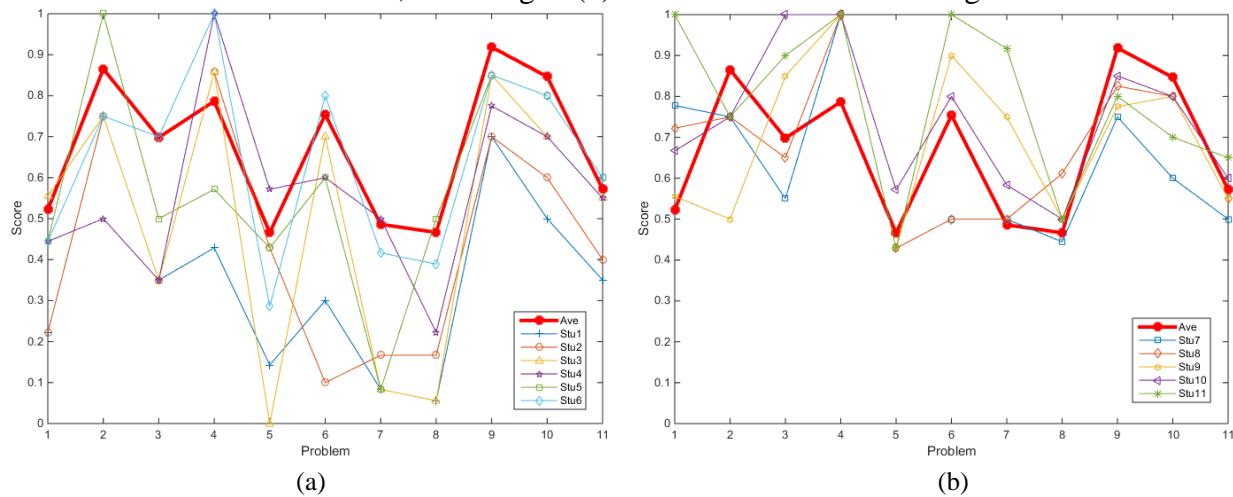(a)   (b)

Fig. 2 Contrast of Actual Score and Average Predicted Score

From this figure, we observed that, actual scores of students in high sublevel are basically flat with the average, some problems' score even higher than it, such as problem4 and 6. However, some actual scores of students selected from low sublevel are far below the predicted average level, such as problem5, 7 and 8. This implicates that there might be shortcomings in the course of teaching, especially on these several problems (problem5, 7, 8). The skills related to these problems can be learned through checking the Q-matrix, so that teachers can modify the curriculum and develop teaching reform to make it more individualized.

On the other hand, it is easy to observe that students' experimental scores (problem9~11) in different levels have no much difference, although their theory test scores (problem1~8) are irregular. This indicates that the experiment scores of students who have poor theoretical scores are still considerable. The reason might be that the setting of experiment is unreasonable, which leads to the low discrimination of students in different levels.

4. Conclusion

In this paper, we introduced a fuzzy cognitive diagnosis framework, FuzzyCDF, which can handle both objective problems and subjective problems simultaneously. We built our own database using the data of DSP course in our school, analyzed the characteristics of each student through mining these data and then, analyzed the teaching efficiency. This method can help us find the gap between student's actual score and predicted score, so as to improve the teaching pertinence and quality.

5. References

[1]. K.Y.Yan, "Data mining application in the evaluation of classroom teaching and research in colleges and universities," Guangzhou: Zhongshan University, 2014.

[2]. Y.Chen, "Research and implementation of result prediction based on genetic neural network," Modern Electronics Technique, vol. 39, 2016, pp. 96-100.

[3]. X.Y.Xie, "Predicting Student Grades Ranking Based on Improved TrAdaboost Algorithm," Computer and modernization, vol. 2, 2016, pp. 122-126.

[4]. P. D. Nichols, S.F. Chipman, "Cognitively diagnostic assessment," London: Routledge, 2012.

[5]. A. Anderson, D. Huttenlocher,et al.. "Engaging with massive online courses," In Proceedings of the 23rd international conference on World Wide Web, 2014, pp. 687-698.

[6]. R. Wu, Q. Liu, Y. Liu, "Cognitive Modelling for Predicting Examinee Performance," Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, July 25-31, 2015, pp. 1017-1024.

[7]. C. Andrieu, "An introduction to MCMC for machine learning," Machine Learing, 2003, vol. 50. pp. 5-43.

[8]. Z.A. Pardos, J. E. Beck, "The composition effect: Conjunctive or compensatory?," International Conference on Educational Data Mining, Montreal, Canada. June 20-21, 2008, pp. 147-156.

[9]. R. Baker, A. T. Corbett, "More accurate student modelling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing," In Intelligent Tutoring Systems, Springer, 2008, pp. 406–415.