

A Dynamic Early Warning Method of Student Study Failure Risk based on Fuzzy Synthetic Evaluation

Chunqiao Mi^{a,b,*}, Qingyou Deng^c, Jing Lin^{a,b}, and Xiaowu Deng^{a,b}

^a*School of Computer Science and Engineering, Huaihua University, Huaihua, 418000, China*

^b*Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, Huaihua, 418000, China*

^c*Human Resource Department, Huaihua University, Huaihua, 418000, China*

Abstract

As more and more students fail in course studies, higher education is now facing challenges regarding increasingly lower course completion rates as well as overall graduation rates. However, failures in course studies is a comprehensive result of various factors, which is characterized by uncertainty. To deal with this issue, fuzzy sets theory and fuzzy logic are advantageous compared with traditional methods. In this study, based on dynamic student study process data, a fuzzy synthetic evaluation method for dynamic early warning student study failure risk is provided. For each student, three specific early warning factors: 1) student course participation, 2) assignment earned points, and 3) student attendance record, are selected as risk indicators, and the overall risk level is determined by a fuzzy synthetic evaluation approach, which can dynamically give the situation of risk as the evaluation time point changes. Finally, our obtained results show that the employed method is good for identifying at-risk students and exploring the risk reasons by showing the degrees of each early warning factors to the overall risk level. It is of significance for educators to timely apply corresponding strategic pedagogical interventions to help at-risk students avoid academic failure.

Keywords: dynamic early warning; study failure risk; fuzzy synthetic evaluation

(Submitted on February 1, 2018; Revised on February 26, 2018; Accepted on March 28, 2018)

© 2018 Totem Publisher, Inc. All rights reserved.

1. Introduction

Currently, along with the large-scale enrollment expansion of higher education, the quality of student study has begun to drop greatly. Early warning of student study failure has become a basic need of college and university education, but quantifying failure risk is a very difficult issue. Fortunately, with the development and application of modern information technology in education, varieties of educational data about student study process have been gradually collected in colleges and universities, so further study on early warning methods based on data science and machine learning technologies [1] is becoming feasible.

There have been some related researches in recent years. Chen et al. [8] and Ma et al. [16] used decision trees and association rules in identifying students who are potentially weak in learning. Minaei et al. [20] and Morris et al. [22] used discriminant analysis for predicting online course completion success and student final outcome grades. Campbell [6], Bravo et al. [5] and Leah et al. [17] detected underperforming students using a method combining factor analysis and logistic regression. Barber and Sharkey [4] and Sandeep et al [25] reported predictive models to identify academically at-risk students. Baker et al. [3] and Jay et al. [2] studied early indicators of student success and failure. Douglas et al. [10] and Geraldine et al. [12] predicted student failing risk based on interaction data. Kris et al. [7] used the naive bayes algorithm to predict academic performance based on intelligence, motivation and study habits factors. Madhavi et al. [26], Mollica et al. [21], and Khalaf et al. [13] used the bayes classification to determine student study performance. Martin et al. [18] and Kevin et al. [14] used a classification system to identify poor performers during current course studying. Yang et al. [29] and Mi et al. [19] applied an artificial neural network approach to student grade prediction and learning failure risk early

* Corresponding author.

E-mail address: michunqiao@163.com

warning. Yang et al. [30] used principal component analysis and multiple linear regression for predicting student academic performance. All these researches covered many machine learning methods, but there is little report about fuzzy set method application in education. These attempts provide valuable inspirations for this study. Most of the current researches mainly focus on assessment or prediction of static risk result rather than dynamic risk analysis during the student studying process, and the risk reason constitution analysis is also insufficient. Since the warning information that can be collected and used in the very early stage of student study is very little and uncertain, it is difficult to use conventional methods to do early warning in this study. The fuzzy comprehensive method is good at dealing with such problems, and it is feasible for early assessment of student study failure risk. We can also easily explore the risk reasons with fuzzy synthetic evaluation based on dynamic student study process data.

Therefore, in order to better deal with the problems existing in current early warning management in college and university education, which include: 1) very late warning times, usually long after the final examinations, 2) one single evaluation factor (often only focus on evaluation of the final grades), and 3) lack of dynamic analysis and risk reason analysis (only doing static descriptive analysis but without dynamic analysis and going back to the cause of the problem). The objective of this study is to establish a dynamic method for student study failure risk analysis and early warning based on fuzzy synthetic evaluation, which can timely identify the at-risk students in different course progresses before the final examination and easily explore the risk reasons contributed by different factors. It is very valuable for improving the efficiency and effectiveness of early warning education management in current colleges and universities.

2. Materials and Methods

2.1. Data Description

In this study, a total of 87 students during the course of “Introduction to Computer Science” with a 17-week time span in the Fall semester of 2016-2017 were chosen as our study sample. They were from three majors including, computer science and technology, software engineering, and network engineering. For each student, daily study process data during the course were collected from an online study system named BDedu developed by ourselves. These data were used to calculate the value of evaluation factors, which affect study outcome performance. Besides, in order to validate our obtained calculation results, the course final outcome grade point data was also collected from the educational administration management system, which uses the hundred percentage point system with 60 points as the passing grade.

The accuracy of study failure risk assessment depends on the significance of the chosen factors, with respect to their effects on course final outcome grade. There are many factors related to student study outcome performance in a course, of which student participation is the most important one that reflects the students' attention to the course. In addition, the assignment points can reflect student mastery of course knowledge points, and the attendance record is related to the time that student spent in the course. Therefore, in this study, we chose the following three specific factors to assess student study failure risk: 1) student course participation (denoted as SCP); 2) assignment earned points (denoted as AEP); and 3) student attendance record (denoted as SAR).

2.2. Data Calculation

In order to make dynamic risk evaluation, a time point variable t_i was set up, where i is the course week number that is used to control the dynamic assessment progress. Then for each student, in order to calculate the value of factor $SCP(t_i)$, we collected his or her visiting logs from the online study system, and counted the times that he or she accessed the specific course resources before time point t_i , which was long before the final examination. It could be a proxy to reveal the student's effort, and so, it can be used to quantify factor SCP. The calculation was shown in Equation (1), where B is the begin date of the course, E_{t_i} is an end time point of our calculation before the final examination, and T_d is the total visiting times that the student accessed the specific course at date d .

$$SCP(t_i) = \frac{\sum_{d=B}^{E_{t_i}} T_d}{\{SCP(t_i)\}_{max}} \quad (1)$$

In order to quantify $AEP(t_i)$ and $SAR(t_i)$ of each student, the percentage points earned scheme was used. $AEP(t_i)$ was determined by Equation (2) and (3), where P_i is the points earned in the i th assignment, K is the full points of the i th assignment, and N_{t_i} is the total number of assignments during the specific course time period before time point t_i . $SAR(t_i)$ was similarly calculated according to Equation (4) and (5), where R_j is the j th attendance record, and M_{t_i} is the full attendance number during the specific course time period before time point t_i . educators can dynamically determine this time point variable t_i , the it, and then the dynamic early warning can be realized.

$$AEP(t_i) = \frac{\sum_{i=1}^{Nt_i} P'_i}{N} \quad (2)$$

$$P'_i = \begin{cases} \frac{P_i}{K} * 100\%, & \text{submitted} \\ 0, & \text{not submitted} \end{cases} \quad (3)$$

$$SAR(t_i) = \frac{\sum_{j=1}^{Mt_i} R_j}{M} * 100\% \quad (4)$$

$$R_j = \begin{cases} 1, & \text{attended} \\ 0, & \text{not attended} \end{cases} \quad (5)$$

Based on these equations, the values of factors $SCP(t_i)$, $AEP(t_i)$ and $SAR(t_i)$ were calculated using a program developed by Python 3.6. In order to illustrate the modelling process, in the following sections, we take $t_i=17$, which is the last week of the course as an example, and the descriptive statistics of the obtained calculated results are shown in Table 1.

Table 1. The descriptive statistics of the calculated values of factors $SCP(t_i)$, $AEP(t_i)$ and $SAR(t_i)$ when $t_i=17$

Factor	Min	Avg	Max	Std	Sample count
SCP	3%	41%	100%	20%	87
AEP	16%	65%	73%	15%	87
SAR	29%	93%	100%	11%	87

2.3. Risk Determination based on Fuzzy Synthetic Evaluation

The fuzzy set theory was first introduced by Zadeh in 1965 [32]. Because it is good at dealing with uncertainty and imprecision decision boundaries, nowadays fuzzy synthetic evaluation models have been used in many areas [9,15]. In this study, for each student, study failure risk assessment using fuzzy synthetic evaluation includes the following steps. In each step, the calculations are illustrated with $t_i=17$ as an example.

2.3.1. Determination of Evaluation Factor and Criteria

The evaluation factor set $U = \{u_i\}$, $i = 1, 2, \dots, n$, where n is factors number. In our study, as described in above section, we specifically chose $n = 3$ and $U = \{SCP, AEP, SAR\}$. For each evaluation factor, the evaluation criteria set $V = \{v_j\}$, $j = 1, 2, \dots, m$, where m is criteria categories number, v_j is the j th criteria threshold value. In our study, a three-level risk classification scheme of red (R), yellow (Y), and green (G) was developed. So, the outputs of study failure risk assessment were classified into three levels (categories): level 1 (R, serious risk), level 2 (Y, moderate risk), and level 3 (G, no risk). So, in this study $V = \{v_1, v_2, v_3\}$, and $m = 3$ were chosen. However, there is no uniform method to determine the value of v_i , and it should refer to the context of specific issue. In the education domain, the number 60 (points) is usually used as the passing grade. For factor AEP, we chose 60% as the no risk threshold, and made $v_3=60\%$, $v_2=40\%$, and $v_1=20\%$ respectively. In the university where the first author works, there is a regulation that says if a student's total absence percent during a course is higher than 30% (namely the percentage points earned is lower than 70%), then his or her final examination qualification of the course will be cancelled. So, for factor SAR, we chose 70% as the serious risk threshold, and made $v_1=70\%$, $v_2=80\%$, and $v_3=90\%$ respectively. At last, for factor SCP, we chose a statistic method, namely the quantile method, to determine v_j for factor SCP, and made $v_1=1$ st quartile (25% in this study), $v_2=$ median (50%), and $v_3=3$ rd quartile (75%) respectively, according to the theory of statistics. At last, all determined criteria values are shown in Table 2.

Table 2. The Criteria threshold values of each evaluation factor

Factor	Criteria threshold		
	v_1	v_2	v_3
SCP	25%	50%	75%
AEP	20%	40%	60%
SAR	70%	80%	90%

2.3.2. Determination of Membership Functions and Fuzzy Relationship Matrix

In the fuzzy set theory, for each criteria of each factor, set $A = \{f_A(x), x \in X, f_A(x) \in [0, 1]\}$, where X is a domain, x is a concrete value of X , f is a membership function mapping X into $[0,1]$. $f_A(x)=1$ means x is a full member in set A , $f_A(x)=0$ means x is not a member in set A , and $f_A(x) \in (0, 1)$ means x is a partial member in set A . There are many shapes for

membership function f_A , such as S curve, Gauss curve, and half trapezoid curve. In our study, according to the domain expert's experience, the half trapezoid curve and triangle curve were chosen. So, on the basis of determined evaluation criteria thresholds shown in Table 2, the membership functions of each factor at each criteria level can be determined using the following Equations (6), (7), and (8), where $f_{ij}(x_i)$ is the calculated membership degree value of evaluation factor i at criteria level j (in this study $i = 1, 2, 3; j = 1, 2, 3$). Then, the fuzzy relationship matrix D can be determined as Equation (9) by substituting the calculated value of each evaluation factor and the determined criteria threshold values into the above determined membership functions, where $d_{ij}=f_{ij}(x_i)$.

$$f_{ij}(x_i) = \begin{cases} 0, & x_i \geq v_{i(j+1)} \\ \frac{v_{i(j+1)} - x_i}{v_{i(j+1)} - v_{ij}}, & v_{ij} < x_i < v_{i(j+1)} \\ 1, & x_i \leq v_{ij} \end{cases} \quad \text{when } j = 1 \quad (6)$$

$$f_{ij}(x_i) = \begin{cases} 0, & x_i \geq v_{i(j+1)} \text{ or } x_i \leq v_{i(j-1)} \\ \frac{x_i - v_{i(j-1)}}{v_{ij} - v_{i(j-1)}}, & v_{i(j-1)} < x_i < v_{ij} \\ \frac{v_{i(j+1)} - x_i}{v_{i(j+1)} - v_{ij}}, & v_{ij} \leq x_i < v_{i(j+1)} \end{cases} \quad \text{when } j = 2 \quad (7)$$

$$f_{ij}(x_i) = \begin{cases} 0, & x_i \leq v_{i(j-1)} \\ \frac{x_i - v_{i(j-1)}}{v_{ij} - v_{i(j-1)}}, & v_{i(j-1)} < x_i < v_{ij} \\ 1, & x_i \geq v_{ij} \end{cases} \quad \text{when } j = 3 \quad (8)$$

$$D = \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix} \quad (9)$$

2.3.3. Determination of Evaluation Factor Weights and the Relative Importance Set

The evaluation factor weights set $W = \{w_i\}$, $i = 1, 2, \dots, n$, where n is factors number (in this study $n=3$), and w_i is the relative importance of factor i . In multi-factor analysis, the determination of factor weights is a very important issue. There are some frequently used subjective methods, like analytic hierarchy process, expert experience method, and so on. There are also some frequently used objective methods, such as coefficient of variation, factor analysis, and information entropy method. In student study failure risk construction, different factors have different contributions. In order to make it as objective and precise as possible, the weight set here was determined by a coefficient of variation (CV for short) method combined with an expert experience method. At first, the CVs of factor SCP, AEP, and SAR were calculated, and the values of 0.50, 0.23, and 0.12 were obtained respectively. Then, combining them with education domain expert's experience, we made some fine tuning and chose $W = \{w_{SCP}, w_{AEP}, w_{SAR}\} = \{0.5, 0.3, 0.2\}$ for this study. Then, the relative importance set $B = W * D = (b_1, b_2, \dots, b_j, \dots, b_m)$, where $*$ is an operator of fuzzy composite, and b_j is determined by Equation (10). Therefore, in this study the relative importance set B can be concretely determined by Equation (11).

$$b_j = \sum_{i=1}^n w_i \cdot d_{ij} \quad (10)$$

$$B = W * D = [w_{SCP}, w_{AEP}, w_{SAR}] * \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix} = [b_1, b_2, b_3] \quad (11)$$

2.3.4. Obtainment of Final Evaluation Risk Result

For each student, the final result of study failure risk evaluation is determined by Equation (12), where the maximum value of the three relative importance values (b_1, b_2, b_3) determines the final risk level of the student. If there are two or three equal maximum values, the first one is chosen.

$$\text{Final-Risk-Level} = L_j \text{ subject to } b_j = \max \{b_1, b_2, b_3\} \quad (12)$$

3. Results and Discussion

When all data were calculated and ready, by the procedure described above, the result of student study failure risk evaluation could be obtained for each student. Randomly taking one student for example, when time point $t_i=17$, his factor data value {SCP, AEP, SAR} = {23%, 24%, 86%}, and his actual final outcome grade was 33 points (hundred percentage point system). Our fuzzy evaluation model included 3 factors {SCP, AEP, SAR} and 3 criteria levels $\{v_1, v_2, v_3\}$ for each factor. Based on Equations (6) – (8), membership functions of factor SCP, AEP, and SAR at 3 criteria levels were established, with factor SCP at level 1 as an example shown in Equation (13). Then, by substituting the student's factor data into the membership functions of each factor at each level, the fuzzy matrix D was obtained for this student, which is shown in Equation (14).

$$f_{SCP1}(x_{SCP}) = \begin{cases} 0, & x_{SCP} \geq 50\% \\ \frac{50\% - x_{SCP}}{25\%}, & 25\% < x_{SCP} < 50\% \\ 1, & x_{SCP} \leq 25\% \end{cases} \quad (13)$$

$$D = \begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.78 & 0.22 & 0.00 \\ 0.00 & 0.43 & 0.57 \end{bmatrix} \quad (14)$$

With $W = [0.5, 0.3, 0.2]$ determined above, the relative importance set $B = W \cdot D = [0.73, 0.16, 0.11]$ was obtained. Then, based on Equation (12), the final risk level of this student belongs to Level 1 (red, serious risk) because the membership degree to Level 1 ($b_1 = 0.73$) is the max. The obtained result is consistent with his actual final grade of 33 points. By the same procedure, we can get the final evaluated risk levels for other students, and the obtained classification results (predicted) of all students and their actual final grade categories (observed) are shown in Figure 1 and Figure 2. In the observed categories, R (red, serious risk) means actual final grade < 50 points, Y (yellow, moderate risk) means $50 \leq$ actual final grade < 60 points, and G (green, no risk) means actual final grade ≥ 60 points.

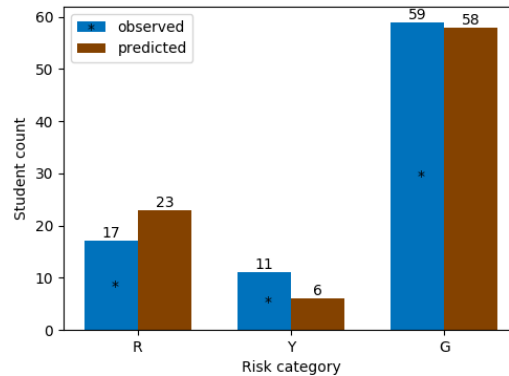


Figure 1. The observed and obtained predicted risk results of all students

		Predicted		
		R	Y	G
Observed	R	14	2	1
	Y	6	3	2
	G	3	1	55

Figure 2. The risk classification details of all students

From Figure 1, we can see that in the observed, there were 17 R students, 11 Y students, and 59 G students, while the model resulted in 23 R students, 6 Y students, and 58 G students. Generally speaking, the model used in our study accurately placed students in the 'R', 'Y' or 'G' category 82.8% ($= (14+3+55)/87$) of the time, which is shown in the light blue background in Figure 2. Besides, the model resulted in 'Type II' errors (classifying an R student as Y or G, or classifying a Y student as G) at a rate of only 5.7%: 5 students out of 87 were predicted to be performing well, while their final course grade placed them in the R or Y category, which is shown in the red diagonal background in Figure 2. However, of these 5 students, only 3 actually failed the course (achieving a final grade < 60 points), indicating a 'predictive failure' rate of only 3.4 % (three students out of 87). The model used in our study also resulted in 11.5% 'Type I' errors by placing 10 students in the R or Y category even though these students actually passed the course (achieving grade > 60 points), which is shown in the yellow dotted background in Figure 2. However, considering the importance of identifying students at risk of failure early in their course study process, the Type I error is of relatively less concern. In other words, it is relatively better to

mistakenly identify a student as at risk of failure than to neglect a student who is in need of additional help and learning support. So, in general our model is promising in evaluating student study failure risk.

During the process of fuzzy synthetic evaluation, we cannot only get the final risk level of each student, but can also reveal some other important concerned information, like the contribution of each evaluation factor to the final risk result. For the same student described above, we can not only see that his synthetic evaluated risk level is level 1, but can also see what the contributions of the three factors are $\{w_{SCP} \cdot d_{11}, w_{AEP} \cdot d_{21}, w_{SAR} \cdot d_{31}\} = \{0.5 \cdot 1.00, 0.3 \cdot 0.78, 0.2 \cdot 0.00\} = \{0.50, 0.23, 0.00\}$. From here, we can see that the major risk factor for this student is fewer course participation since the contribution of SCP is the largest one (0.50). Besides, from the relative importance set $B = W \cdot D = \{b_1, b_2, b_3\} = \{0.73, 0.16, 0.11\}$, we can see the membership degree values to other levels. This information is very useful in understanding the construction of student failure risk, and it is also very valuable for educators to apply corresponding teaching interventions in advance.

In order to explore the dynamic effectiveness of our model used in this study, we further examined the Type I errors and Type II errors as time point t_i changed. Since at the very beginning of the course there are very few processes and behaviour data about student study, we begin at time point $t_i = 5$. The obtained results are shown in Figure 3, from which we can see that at the very beginning, the total accuracy is very low (just nearly 40%) and the Type I error and Type II error are both very high (about nearly 30%). This is due to fewer information data at the very beginning. As time goes on, there are more and more procedural data about student study behaviour, so the total accuracy becomes higher and higher, and the Type I error and Type II error become lower and lower. But, 13 weeks later, namely, when time point t_i is bigger than 13, the growth trend of the overall accuracy has become relatively gentle, and the drop in the Type I error and Type II error also shows the same trend. So, we can see that the 13th week is a time critical point, after which the early warning model has a high credibility. This also tells us that the more data about student study behaviour we can collect, the more accurate the result is, and that at least 13 weeks of procedural data is needed to make a relatively reliable risk assessment result.

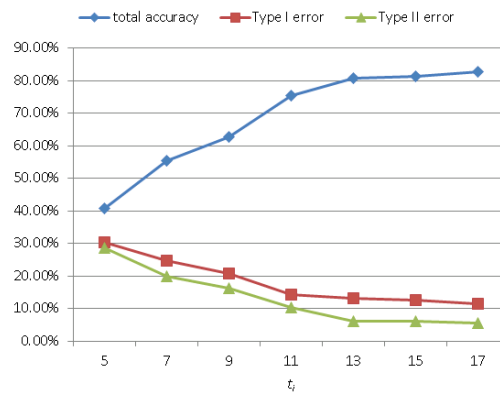


Figure 3. The total accuracy, Type I error and Type II error in different time point t_i

4. Conclusions

The obtained results showed that this dynamic early warning method based on fuzzy synthetic evaluation is promising when the time point is later than 13 weeks in identifying the majority of the at-risk students who failed or almost failed the course. Fuzzy synthetic evaluation of the overall study failure risk of a student gives us a clear detailed picture of risk situation by showing the membership degree values of each factor and which is the major risk factor. With such important information, educators can give a timelier corresponding strategic teaching intervention to help students avoid academic failure, which is very significant in early warning education practice. The model is also dynamically adaptive as time point variable changes. For early warning practice, the earlier the time is, the greater the warning meaning is. But, for the early warning model, the later the time is, the higher the warning accuracy is. So, it is very important to find a balanced time point that is not late for early warning practice, but also has relatively high accuracy. Through empirical research in this study, we find that the 13th week is an important point in assessing student failure risk in course study, which has important academic significance in the dynamic early warning field.

However, to improve the model and get better risk assessment results, further study is still needed [23]. For example, student background data and some more implicit indicator factors [24] in the very early stage of student course study could be considered based on more experiments. Secondly, in addition to external behavioural data about student studying process, some internal psychological data about student studying emotion such as eye movement data [31,33,34] could be integrated into the model. At last, the combination of different data mining algorithms [11,27,28] can be considered in order to enhance the accuracy. In sum, the obtained results of this study show that the dynamic early warning method based on fuzzy

synthetic evaluation is not only a promising approach for assessing student study failure risk, but is also of practice significance for improving early warning education management and early warning research.

Acknowledgements

This study is supported by the Hunan Provincial Educational Science 13th Five-Year Planning Program (No.XJK016QXX003), the Hunan Provincial Philosophy and Social Sciences Foundation (No.17YBQ087), the Program of Hunan Provincial Social Science Achievements Evaluation Committee (No.XSP18YBC182), the Hunan Provincial Natural Science Foundation (No.2017JJ3252), and the teaching reform project "Research on the individualized teaching reform of software engineering major under the background of new engineering". The authors of this study are very thankful to all of them. We are also very grateful to the reviewers and editors who give constructive comments and helpful suggestions for this work.

References

1. H. Almarabeh, "Analysis of Students' Performance by Using Different Data Mining Classifiers," *International Journal of Modern Education & Computer Science*, Vol. 9, No. 8, pp. 9-15, 2017.
2. J. Bainbridge, J. Melitski, A. Zahradnik, E. Lauría, S. Jayaprakash, and J. Baron, "Using Learning Analytics to Predict At-Risk Students in Online Graduate Public Affairs and Administration Education," *Journal of Public Affairs Education*, Vol. 21, No. 2, pp. 247-262, 2015.
3. R. S. Baker, D. Lindrum, M. J. Lindrum, and D. Perkowski, "Analyzing Early At-Risk Factors in Higher Education E-Learning Courses," in *Proceedings of the 8th International Conference on Educational Data Mining*, pp. 150-155, 2015.
4. R. Barber, and M. Sharkey, "Course Correction: using Analytics to Predict Course Success," in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, Vancouver, Canada, New York, NY: Association of Computer Machinery, pp. 259-262, 2012.
5. J. Bravo, S. Sosnovsky and A. Ortigosa, "Detecting Symptoms of Low Performance using Prediction Rules," in *Proceedings of the 2nd Educational Data Mining Conference*, Universidad de Cordoba, Cordoba, Spain, pp. 31-40, 2009.
6. J. P. Campbell, "Utilizing Student Data within the Course Management System to Determine Undergraduate Student Academic Success: an Exploratory Study," *Doctoral dissertation*, Purdue University, pp. 31-61, 2007.
7. K. A. Capao, A. D. Cantara, A. M. Ceniza, P. M. J. Eduardo, S. B. Polinar, and J. M. Tero, "Predicting Academic Performance with Intelligence, Study Habits and Motivation Factors using Naive Bayes Algorithm," *International Journal of Engineering Research & Technology*, Vol. 5, No. 3, pp. 182-185, 2016.
8. G. D. Chen, C. C. Liu, K. L. Ou, and B. J. Liu., "Discovering Decision Knowledge from Web Log Portfolio for Managing Classroom Processes by Applying Decision Tree and Data Cube Technology," *Journal of Educational Computing Research*, Vol. 23, No. 3, pp. 305-332, 2000.
9. M. J. Chen, Q. Wang, Y.P. Zhang, and S. G. Duan, "An Exploration of Quality Evaluation System for University Classroom Teaching based on Fuzzy Mathematics," *Journal of the Hebei Academy of Sciences*, No. 1, pp. 1-4, 2015.
10. D. Detoni, C. Cechinel, R. A. Matsumura, and D. F. Brauner, "Learning to Identify At-Risk Students in Distance Education Using Interaction Counts," *Revista de Informática Teórica e Aplicada*, Vol. 23, No. 2, pp. 124-140, 2016.
11. A. Elbadrawy, A. Polyzou, Z. Ren, S. Mackenzie, K. George, and R. Huzefa, "Predicting Student Performance Using Personalized Analytics," *Computer*, Vol. 49, No. 4, pp. 61-69, 2016.
12. G. Geraldine, M. Colm, O. Philip, and H. Markus, "Learning Factor Models of Students at Risk of Failing in the Early Stage of Tertiary Education," *Journal of Learning Analytics*, Vol. 3, No. 2, pp. 330-372, 2016.
13. A. K. Hamoud, A. M. Humadi, W. A. Awadh, and A. S. Hashim, "Students' Success Prediction based on Bayes Algorithms," *International Journal of Computer Applications*, Vol. 178, No. 7, pp. 6-12, 2017.
14. C. Kevin, and A. David, "Utilizing Student Activity Patterns to Predict Performance," *International Journal of Educational Technology in Higher Education*, Vol. 14, No. 1, pp. 1-15, 2017.
15. H. Lin, "Based on Fuzzy Mathematics in College Teaching Quality Evaluation System," *Journal of Hunan Institute of Science & Technology*, No. 4, pp. 25-27, 2012.
16. Y. Ma, B. Liu, C. K. Wong, P. S. Yu, and S. M. Lee, "Targeting the Right Students using Data Mining," in *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining*, Boston, Massachusetts, USA, pp. 457-464, 2000.
17. L. P. Macfadyen, and S. Dawson, "Mining LMS Data to Develop an Early Warning System for Educators: a Proof of Concept," *Computers & Education*, Vol. 54, No. 2, pp. 588-599, 2010.
18. H. Martin, Z. Zdenek and Z. Jaroslav, "Ouroboros: Early Identification of At-Risk Students without Models based on Legacy Data," in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, Vancouver, Canada, pp. 6-15, 2017.
19. C. Mi, X. Peng, and Q. Deng, "An Artificial Neural Network Approach to Student Study Failure Risk Early Warning Prediction Based on TensorFlow," *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, Vol. 219, pp. 326-333, 2018.
20. B. B. Minaei, and W. Punch, "Using Genetic Algorithms for Data Mining Optimization in an Educational Web-Based System," in *Proceedings of Genetic and Evolutionary Computational Conference*, Chicago, Illinois, USA, pp. 2252-2263, 2003.
21. C. Mollica, and L. Petrella, "Bayesian Binary Quantile Regression for the Analysis of Bachelor-to-Master Transition," *Journal of Applied Statistics*, Vol. 44, No. 15, pp. 2791-2812, 2017.

22. L. V. Morris, S. Wu and C. Finnegan, "Predicting Retention in Online General Education Courses," *The American Journal of Distance Education*, Vol. 19, No. 1, pp. 23-36, 2005.
23. S. M. Muthukrishnan, M. K. Govindasamy, and M. N. Mustapha, "Systematic mapping review on student's performance analysis using big data predictive model," *Journal of Fundamental and Applied Sciences*, Vol. 9, No.4S, pp. 730-758, 2017.
24. F. Razaque, N. Soomro, S. A. Shaikh, S. Soomro, J. A. Samo, N. Kumar, and H. Dharejo, "Using Naive Bayes Algorithm to Students' bachelor Academic Performances Analysis," in *Proceedings of 4th IEEE International Conference on Engineering Technologies and Applied Sciences*, Afyonkarahisar, Turkey, pp. 1-5, 2017.
25. M. J. Sandeep, W. M. Erik, J. M. L. Eitel, R. R. James, and D. B. Joshua, "Early Alert of Academically At-Risk Students: an Open Source Analytics Initiative," *Journal of Learning Analytics*, Vol. 1, No. 1, pp. 6-47, 2014.
26. M. Tripathi, and A. K. Agarwal, "Probabilistic Determination of Student Performance using Naive Bayes Classification Algorithm," *International Journal of Engineering Science and Computing*, Vol. 7, No. 8, pp. 14749-14752, 2017
27. M. Xu, Y. Liang, and W. Wu, "Predicting Honors Student Performance Using RBFNN and PCA Method," *Lecture Notes in Computer Science*, Vol. 10179, pp. 364-375, 2017.
28. J. Xu, K. H. Moon, and M. V. D. Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 11, No. 5, pp. 742-753, 2017.
29. T. Y. Yang, C. G. Brinton, W. C. Joe, and M. Chiang, "Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 11, No. 5, pp. 716-728, 2017.
30. S. J. H. Yang, O. H. T. Lu, A. Y. Q. Huang, J. C..H. Huang, H. Ogata, and A. J.Q. Lin, "Predicting Students' Academic Performance Using Multiple Linear Regression and Principal Component Analysis," *Journal of Information Processing*, Vol. 26, pp. 170-176, 2018.
31. J. Yi, "Research and Application of Learning Status Evaluation based on Eye Movements," *Master dissertation*, Shanghai Jiao Tong University, pp. 6-44.
32. L. A. Zadeh, "Fuzzy Sets," *Information Control*, No. 8, pp. 338-353, 1965.
33. Q. Zhang, and L. Yang, "Learning Measurement Progress and Trends in e-Learning—based on Eye Movement Application Perspective," *Distance Education and Online Learning*, Vol.11, pp. 68-73, 2016.
34. Q. Zhang, F. Wu, and S. Lai, "Information Architecture of Learning Analytics Dashboards via Meta-analysis of the Eye Movement Data," *Open Education Research*, Vol. 23, No. 6, pp. 94-103, 2017.

Chunqiao Mi received his Ph.D. degree in the College of Information and Electrical Engineering, China Agricultural University, China in 2012. Now, he is an Associate Professor at the School of Computer Science and Engineering, Huaihua University, China. He mainly teaches data related science and technology and software engineering. He has hosted many research projects, such as projects from the National Natural Science Foundation, the Hunan Provincial Natural Science Foundation, the Hunan Provincial Philosophy and Social Sciences Foundation, the Hunan Provincial Social Science Achievements Evaluation Committee, the Hunan Provincial Educational Science 13th Five-Year Planning Program, the Education Department of Hunan Province, and so on. He has also published more than 30 academic papers, obtained 5 items of software copyright, and 4 items of big data analyst relevant professional skill certificates in related areas. His current research interests include data science and educational informatization, geographic information science and agricultural disaster risk management, and big data technology in education and agriculture.

Qingyou Deng received her B.A. degree in the School of Public Administration, Hunan Normal University, China in 2007. Now, she is an Assistant Researcher of the Human Resource Department at Huaihua University, China. She mainly engages in researching data related technology application in education management and teacher training. She has participated in many research programs and projects, such as projects from the Hunan Provincial Philosophy and Social Sciences Foundation, the Hunan Provincial Social Science Achievements Evaluation Committee, and so on. She has also carried out some practical projects in related areas. Her current research interests include data analysis technology, educational information technology, teacher training technology, education management process optimization and risk management, and financial management of higher education.

Jing Lin received his M.A. degree in the Faculty of Information Engineering and Automation at the Kunming University of Science and Technology, China in 2009. During 1995-1998, he worked as a software engineer with the Shenzhen Jinbada Communication Company, where he was in charge of developing a detecting system for telephone device fault. During 1999-2005 he was a software manager in the Shenzhen Ncking Technology INC., LTD, analyzing and designing hospital information systems. Since 2009, he has been an Associate Professor at the School of Computer Science and Engineering, Huaihua University, China. He has hosted many projects, such as research projects from Huaihua University, the Education Department of Hunan Province, the Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, and so on. He has also published more than 10 academic papers and obtained 2 items of software copyright in related areas. His current research interests include intelligent information system, information safety, and data mining.

Xiaowu Deng received his M.S. degree in the College of Information and Electrical Engineering, Shandong University of Science and Technology. Now, he is a lecturer at the School of Computer Science and Engineering, Huaihua University, China. He mainly engages in the teaching of embedded system technology. He has hosted many projects, such as research projects from Huaihua University, the Education Department of Hunan Province, the Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, the Internet of Things Association of Hunan Province, and so on. He has also published more than 10 academic papers and obtained more than 10 items of software copyright and patent right in related areas. His current research interests include embedded system, membrane computing, and machine learning.