Advances in Social Science, Education and Humanities Research (ASSEHR), volume 183
2018 4th Annual International Conference on Modern Education and Social Science (MESS 2018)

A Study on Academic Early-Warning System Based on Machine Learning

Zi-Jun BAI[1,a,*] ,Gang-Quan CAI[2,b]

[1]School of Opto-electronic and Communication Engineering,

Xiamen University of Technology, Xiamen 361024, China

[2]School of Energy, Xiamen University, Xiamen 361102, China

[a]baizijun@xmut.edu.cn

[b]cgq@xmu.edu.cn

Keywords: academic early-warning system, machine learning, informatization

Abstract. With the expansion of college enrollment, college students' academic problem increasingly prominent. Universities have established effective preventive mechanisms to solve the problems. The most representative measure is academic early-warning system. Take Xiamen University of Technology as an example, this paper tries to put forward the idea of improving the existing academic warning system, and builds up the mathematical model based on machine learning, focusing on the algorithm of the model and validating the model. Finally, discusses the current situation and challenge of colleges and universities. In this paper, the daily teaching management and information system are combined, and the effectiveness and feasibility of computer are used to provide valuable reference for education reform in universities.

Introduction

With the step of enlarging enrollment in colleges and universities, the higher education in our country has changed from elite education to mass education, and the students' academic problems are becoming more and more serious. In order to solve these problems, many colleges and universities have set up the academic early-warning system to remind and help students to complete their studies smoothly [1].

This paper is divided into four parts to expound the academic early-warning system based on machine learning:

First, explaining the academic early-warning system of Xiamen University of Technology for example.

Second, improving the system and providing a data model for machine learning.

Third, discussing the mathematical model of machine learning and implementation.

Four, presenting the problems and challenges which the system will be faced.

Academic Early-Warning System of Xiamen University of Technology

Taking Xiamen University of Technology as an example, the school stipulates that there will be different processing to students according to who's accumulated non-acquired credits from the training programs in the school (hereinafter referred to as "no credits"), as follows:

Table 1 Current academic early-warning system

Indicator	Level of warning
12≤no credits<24	School warning
24≤no credits<32	University warning
32≤no credits<48	Repeat
no credits≥4	Drop out

At the beginning of every semester ,the school office teachers who is responsible for the academic early-warning system will have a manual credit statistic for all student from school to the end of last semester, in order to release warning information.

Here are two problems:

1.It can’t fully reflect the study situation of students in school because the indicators of the academic early-warning system are too single.

2.It is inefficient and error-prone for manual statistics.

It will solve these problems by means of multi-indicator academic early-warning system and machine learning in the following paper.

Improved School Early Warning System

Generally speaking, the exam results could only be one-sided reflection of a certain stage of studying, but not describe a student's learning process timely and comprehensively. The lag result may bring passive early warning and remedial measures, it is also unable to meet the requirements of cultivating high quality talents in colleges and universities [2].

The information mastered by school from student is not only the exam results, but also their family situation, the attendance records, the mental health and so on. It will be able to reflect the student's school situation more comprehensively if using the information and taking all kinds of indicators into the academic early-warning system. For example, there should be a peace of early waring information released if a student has the number of absenteeism exceeds one sixth of a curriculum. In addition, a large proportion of academically poor students are from problem family. Therefore, we should observe and warn those students who are influenced by the family structure, the family culture and the education level of the source land and so on [3].

The academic early-warning system that we want to establish could be able to comprehensively evaluate the students' family situation, daily learning performance, practice performance and various test results. The teachers who in charge of the system can send early warning information to relevant students, parents and leaders in time according to the integration situation, so that the relevant personnel can take timely intervention measures to improve students ' learning status and avoid undesirable consequences. As shown in the following diagram, the improved academic early-warning system will introduce multidimensional indicators, the warning level is divided into focus, junior, intermediate, senior, super.

Table 2 A multi-indicator academic warning system

Level	Indicator	Unit of measurement	Weight	Decisive
focus	The family is located in a remote, relatively backward area of Education	1	1.1	0
	Family structure is special: Parents divorced, single-parent families	1	1.3	0
	Poor family economy: an unfixed source of economy	1	1.3	0
	Partial branch in college entrance examination	1	1.2	0
junior	non-acquired credits: greater than or equals to 12 and less than 24	credits	1.0	1
	the number of absenteeism exceeds one sixth of a curriculum	number of absenteeism	1.3	0
intermediate	non-acquired credits: greater than or equals to 24 and less than 32	credits	1	1
	the number of absenteeism exceeds one fifth of a curriculum	number of absenteeism	1.4	0
senior	non-acquired credits: greater than or equals to 32 and less than 48	credits	1	1
	the number of absenteeism exceeds one fourth of a curriculum	number of absenteeism	1.5	0
super	non-acquired credits: greater than or equals to 48	credits	1	1
	… …	…	…	…

In this table, each indicator has a weight value and a decisive value. The weight value of each indicator is multiplied by the weight to get a score which determines the level of the warning. The decisive value is 0 or 1, it determines the level of the warning if the decisive value is 1, whatever how much the score it is. It enhances the flexibility and rationality of the system by taking into account the quantitative changes and the qualitative change of indicators both [4].

Machine Learning to Establish Early Warning Model

The method of artificial statistics early warning is inefficient and error-prone, and the existing teaching information system can't keep up with the changing demand of the business. The operation problem of the early-warning system will be solved by using machine learning, which could deal with the changing data.

Machine learning (ML) is a multidisciplinary interdisciplinary subject, covering a wide range of disciplines such as probability theory, statistics, approximation theory, convex analysis, algorithmic complexity theory and so on to study how a computer simulates or implements human learning behavior in order to acquire new knowledge or skills and reorganize existing knowledge structures to keep improving their performance. It is the core of artificial intelligence, mainly using induction, synthesis rather than deduction.

Machine learning can transform the disordered data into useful information, and it can get the conclusion what people want through self-study and classification, even excavate the valuable

information that people don't realize, and provide powerful information support for decision. Next we will introduce machine learning and establish the machine learning data model of the academic early-warning system mentioned in section II, and finally verify the correctness of the model.

Machine Learning and Development Environment

In general, the steps to develop machine learning applications are: collecting data, preparing input data, analyzing input data, training algorithms, testing algorithms, using algorithms. The names of each step can directly explain what each step is going to do. Here it only focuses on the combination of machine learning and early-warning systems. The paper uses the Python language to realize the algorithm of machine learning, which is simple and easy to learn, and has a rich third-party library.

Mathematical Model of Academic Early Warning Based On Machine Learning

According to the characteristics of machine learning, we will set up a model of academic early-warning, the following general steps:

(1) Form training samples from the data produced by students in the process of learning and life

(2) Determine the target variable, that is the level of early warning.

(3) Determine the indicator of machine learning combined with the improved academic early-warning system, it is the failure of credit or truancy times for example.

(4) Choose machine learning algorithm, including concrete realization and code compiling.

(5) Train the algorithm of machine learning by samples.

(6) Test the accuracy of the algorithm and improve it.

(7) Use machine learning to identify daily learning life data and get appropriate early warning level.

We need to screen out data from family status, attendance information, disciplinary records, life and so on [5]. As target variable of the system is five warning levels, we could take into account the supervised learning classes when choosing machine learning algorithms, and the K-neighbor algorithm or decision tree can be selected in the supervised learning classes algorithm. Then choosing the more representative indicator according to the early warning system and the actual situation of school. In addition, in order to simplify the model for this paper we will not discuss the performance analysis and evaluation system of the score and the same is the case with other indicator [6]. Finally, we can make a simple training set from the data with the target variables and the indicators (in terms of semester or natural month):

Table 3 Training sample set

	non-acquired credits	number of absenteeism	family situation	disciplinary	cost of life	Harmonious relationship evaluation	categories
1	0	2	2	0	6000	1	1
2	2	15	0	0	6500	1	1
3	13	20	2	1	7300	6	2
4	18	33	2	0	12000	5	2
5	25	52	0	2	18000	3	3
…	…	…	…	…	…	…	…

In the table above, 1 to 5 of the categories represent the level of warning: focus, junior, intermediate, senior, super.

Taking K-nearest neighbor algorithm as an example here, the algorithm is relatively simple, effective, and easy to master. It is characterized by high accuracy, insensitivity to outliers and no data entry assumptions [7]. And then there's the pseudocode for this algorithm.

The following operations are performed in order for each point of the data set of the unknown category attribute:

A. Calculate the distance between the point in the known category data set and the current point.

B. Sort by distance increment order.

C. Select the k points with the minimum distance from the current point.

D. Determine the frequency of the occurrence of the category of the previous k points.

F. Return the highest frequency of the previous K points as the forecast category for the current point.

There are four input parameters in Python code:

def classify0 (inX, dataSet, labels, k)

the inX is the input vector for the classification, the dataSet is the input training sample, the labels is the label vector, K is used to be the number of adjacent neighbors in the K nearest neighbor algorithm. The number of elements in the label vector is the same as the number of rows in the matrix dataset. The Euclidean distance formula is used in the code to calculate the distance between the two vector points.

It leads to the fact that the Euclidean distance between the two vectors is most affected by the cost of living because the cost of living is much larger than other eigenvalues. When dealing with the eigenvalues of these different values, we usually normalize the values, such as processing the range of values from 0 to 1 or -1 to 1. The following formula can convert the eigenvalues of any value range to 0 to 1:

newValue = (oldValue-min) / (max-min)

Finally, we write a small program to test this classifier, in which we randomly enter a student's eigenvalues, such as failing grades or the number of truancy, and then the classifier will tell us which the warning level of the student is:

```
def classifyPersonWarning():
resultList=['focus','primary','intermediate','advanced',’superfine’]
failingCredits = int(raw_input("Failing credits is? "))
absenteeism = int(raw_input("Absenteeism is? "))
familySituation = int(raw_input("Family situation is? "))
violationTimes = int(raw_input("Violation times is? "))
livingExpenses = int(raw_input("Living expense is? "))
relationshipDegree = int(raw_input("Relationship degree is "))
warningDataMat,warningLabels = warning_file2matrix('warningTestSet.txt')
norMat,ranges,minVals = autoNorm(warningDataMat)
inArr = array([failingCredits,absenteeism,familySituation,violationTimes,livingExpenses,relationshipDegree])
classifierResult = classfiy0((inArr-minVals)/ranges,norMat,warningLabels,3)
print "The waring level of this student is :",resultList[classifierResult - 1]
```

In addition, we usually only provide 90% of the existing data as a training sample to train the classifier and use the remaining 10% of data to test the classifier, the 10% of data should be randomly selected. It is estimated that the error rate of the classifier handling the dataset is about 1.3%, which is a fairly good result.

Challenges and Summaries

The improvement of existing academic early-warning system and the establishment of machine learning mathematical model are based on large data. Now colleges and universities invest a lot of resources to build various information systems whose data are not fully opened up, but it can not be able to provide the complete data support for the academic early-warning system mathematical model because of the independent data, which is based on machine learning. That is the challenge we are facing. It is only a stopgap measure that summarizing these dispersed data manually before this problem is resolved. The most fundamental solution is to realize the integration of university

resources so as to provide complete data for the academic early-warning system, tracking the academic situation in time and accurately.

The school still has to invest the corresponding human and financial resources to ensure the success of the academic early-warning system after the establishment of academic early-warning system model based on machine learning. It needs to change the model at any time to ensure that the actual situation, because the school students learning situation is a constantly changing.

Firstly, it requires collaboration between the staff working in the frontline of education and engineers who are familiar with the technology to update the data and models in real time.

Secondly, it will play the real significance of academic early warning if the schools carry out practical and effective helping measures after the publication of the warning results. The aid of academic early warning is not the scope of this paper.

Acknowledgement

This research was financially supported by the special project for the higher educational reform, which is one of the “13th Five-year plan” projects for educational science in Fujian(FBJG20170314).

References

[1] LOU Qi-wei, On Academic Early Warning Mechanism of College Students with Learning Disabilitie, JOURNAL OF SHANXI INSTITUTE OF ECONOMIC MANAGEMENT Vol.24 No.4(2016) 111-114.

[2] ZHU Dong-xin, SHEN Liang-zhong, Research on the Association Rules in College Early Warning, Computer Knowledge and Technology Vol.13 No.23(2017) 196-197.

[3] Lu Peng, Wang Jisheng, Yin Mingjun, Applying of Data Mining Technology in the Academic Early Warning of University, China Educational Technique & Equipment No.30(2009) 120-122.

[4] ZHAO Na, Discussion on the Establishment of College and University Student Academic Five-in-one Early Warning Mechanism under the Credit System, The Guide of Science & Education, No.03(2013) 15-20.

[5] Jin Yifu, Wu Tao, Zhang Zishi, Wang Weidong ,Design and Analysis of Learning Alert System in Big Data Condition, China Educational Technology, No.02(2016) 69-73.

[6] LI Kai-jun, FENG Xiu meng,YAN Shi, WANG Xiao-dong, WANG Jun-lin , Analysis of Test Score and its Guidance on the Educational Management, Medical Recapitulate, No.16(2009) 2552-2554.

[7] Peter Harrington, Machine Learning IN ACTION, Manning Publications Co.(2012).