PASS: An Expert System with Certainty Factors for Predicting Student Success

Ioannis Hatzilygeroudis[1,2], Anthi Karatrantou[1], and C. Pierrakeas[3]

[1]Department of Computer Engineering & Informatics, University of Patras
GR-26500 Patras, Greece
[2]R. A. Computer Technology Institute, P.O. Box 1122, GR-26110 Patras, Greece
[3]Hellenic Open University, 23 Saxtouri Str., GR-26221, Patras, Greece
{ihatz@cti.gr, ihatz@ceid.upatras.gr}

Abstract. In this paper, we present an expert system, called PASS (Predicting Ability of Students to Succeed), which is used to predict how certain is that a student of a specific type of high school in Greece will pass the national exams for entering a higher education institute. Prediction is made at two points. An initial prediction is made after the second year of studies and the final after the end of the first semester of the third (last) year of studies. Predictions are based on various types of student's data. The aim is to use the predictions to provide suitable support to the students during their studies towards the national exams. PASS is a rule-based system that uses a type of certainty factors. We introduce a generalized parametric formula for combining the certainty factors of two rules with the same conclusion. The values of the parameters (weights) are determined via training, before the system is used. Experimental results show that PASS is comparable to Logistic Regression, a well-known statistical method.

1 Introduction

In the last decades, there has been extensive use of computer-based methods in education, either for administrative or for pedagogical purposes. Those methods can be distinguished in traditional and artificial intelligence (AI) methods. Various forms of regression analysis are representatives of the traditional methods, whereas the expert systems approach is a common representative of the AI methods. Both have been used in various applications in the education domain, e.g. admission decisions [1, 2], academic advising [3], academic performance prediction [1]. In this paper, we use them in a somehow different application: prediction of a student success in the national exams for admission in a higher education institute.

It is obvious that the ability to predict a student's success in the entry examinations to the higher education could be useful in a number of ways. It is important for the teachers as well as the directors of a secondary education school to be able to recognize and locate students with high probability of poor performance (students at risk) in order to provide extra help to them during their studies. So, it is useful to have a tool to assist them in this direction. This is the objective of the work, which is presented here.

M.Gh. Negoita et al. (Eds.): KES 2004, LNAI 3213, pp. 292–298, 2004.
© Springer-Verlag Berlin Heidelberg 2004

We use two methods, an expert system approach and a well-known statistical method, namely logistic regression, to achieve our objective. Logistic regression is used for comparison reasons. In the expert system, we introduce and use a modified version of the MYCIN's certainty factors [4]. We call the expert system PASS (**Pre**dicting **A**bility of **S**tudents to **S**ucceed). Our aim is to use PASS as an education supporting tool, mainly addressed to high school teachers for the above mentioned purpose. The design of PASS is based on an analysis of demographic, educational and performance data of students from an available database.

2 Modeling Prediction Knowledge

2.1 The Problem

Our work concerns students of technical and vocational secondary education in Greece. In evening schools of that level, students attend a three years program (grades A, B and C) and choose one of the *specializations* offered, such as electrolgy, mechanics, nursing etc. Students attend a variety of courses: general education courses, specialization courses etc. Each year has two semesters. At the end of each semester, students are given marks representing their performance. To enter a Technological Educational Institute (TEI), which is a higher-level institute, the students should pass corresponding national exams. The exams include tests in three courses.

It is important to notice that the number of students who succeed is very low, which means that the students of this type of high schools need some help. Thus, it is very important for a teacher to be able to recognize, as early as possible, the students who (a) have a high possibility to succeed, in order to help and encourage them during their studies, (b) have a low possibility to succeed, in order to treat them properly during their studies

So, apart from teacher's personal opinion, a tool that could make predictions about the possibility that a student has to pass the national exams would be of great assistance. It would be also useful for school directors and curriculum designers, because it can offer them useful information about how to organize the school program.

2.2 Specifying the Parameters

Knowledge acquisition in such problems mainly consists in specifying the parameters (input variables) that play some role in predicting the success of a student. To this end, we interviewed some teachers with long experience. We also analyzed data from a student database, which contained 201 records of students, who took the entry examinations during the last three years.

We finally resulted in the following parameters, as being important for predicting the students success: sex, age, specialization, grade A (average mark of all first year courses), grade B (average mark of all second year courses), grade SC (average mark of the three courses to be examined in the national exams at the end of the first semester of the third year). Marks in courses follow the 1–20 scale.

Another crucial point was to determine the values of the parameters/variables, like age, grade A, grade B and grade SC. The variables and their decided values are as follows: specialization: electrology, mechanics, electronics, office clerks, nursing, sex: male, female, age: normal (<20 years), overyeared (≥20 years), grade A/B/SC:

moderate (≥10 and <12.5), good (≥12.5 and <15.5), very good (≥15.5 and <18.5), excellent (≥18.5 and ≤20).

2.3 Modeling the Process

Two other questions that had to be answered were, when (i.e. at which stage of studies) and how (i.e. which parameters should be taken into account) prediction should be made. After some discussions with two experts (teachers), we decided that there should be two predictions made at two different points of a student's studies cycle. A first (initial) prediction is made at the end of the second year. A second (intermediate) prediction is made just after the end of the first semester of the third year. Then, a final prediction is made by combining the two predictions. Thus, the student and the teacher have a semester's time (till exams time) to make improvements to the student's performance.

The *initial prediction* is based on the specialization, sex, age, grade A and grade B of a student (initial input variables). This prediction gives a first indication of the possibility of a student to pass the exams. So, the effort the student must make during the next (final) year of studies can be specified.

The *intermediate prediction* is based on the specialization, age and grade SC of the student (intermediate input variables). Then, the *final prediction* is produced by combining the initial and intermediate ones.

Given that predicting the success of a student is not a clear-cut decision, but includes a large degree of uncertainty, we decided to represent that uncertainty. So, initial and intermediate predictions have some degree of uncertainty and these uncertainties are combined to produce final prediction. This is reflected on the expert system's inference flow (see Fig. 1).

3 The Expert System

3.1 PASS Architecture

PASS has been implemented in CLIPS [10]. It consists of a knowledge base, which includes a rule base and a fact base, and the CLIPS inference engine. Facts in the fact base are organized in CLIPS templates. Each template stores information about problem variables. There are three templates: *student template* (which stores information about student data), *result1 template* (which stores information about the initial prediction) and *result2 template* (which stores information about the intermediate prediction).

Rules in the rule-base are organized in three main groups: *initial prediction rules*, *intermediate prediction rules* and certainty *factors combining rules*. Apart from them, however, there are two other groups: *initialization rules* and results *printing rules*. Each group of rules deal with what its name indicates. In the Appendix, there are examples of each template and rule group in CLIPS.

Inference flow in PASS reflects the prediction process model and is illustrated in Fig.1. Inference flow via rule groups is implemented by giving different priorities to the rules of different groups.

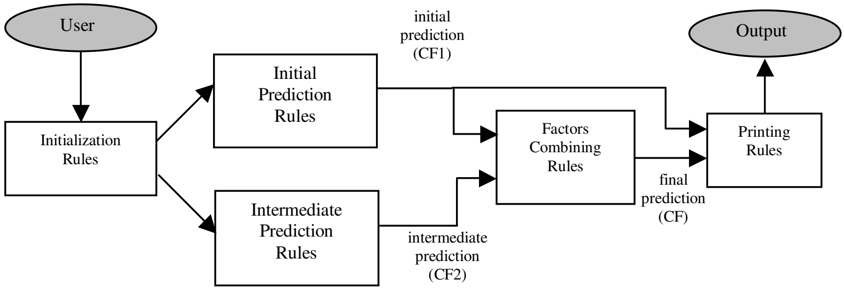

Fig. 1. Inference flow in PASS

3.2 Handling Uncertainty

We use some kind of certainty factors (CFs) in PASS. CFs have been quite popular among expert system developers since their appearance, because they are associated with a simple computational model that permits to estimate the confidence in conclusions being drawn. CFs were first introduced in MYCIN, a medical diagnosis expert system [4]. A CF is a number between –1 (definitely false) and +1 (definitely true), which measures the expert's belief (positive number) or disbelief (negative number) to a conclusion.

In PASS, we use CFs that are always positive numbers between 0 (definitely false) and 1 (definitely true), which seems more natural. Any CF less than 0.5 indicates disbelief, whereas any CF equal to or greater than 0.5 indicates belief in a conclusion. Given that CFs are always positive, when we have the following rules

R1	R2
if e1	if e2
then h (CF1)	then h (CF2)

and they are fired, the combined certainty CF of h, according to MYCIN theory, is given by the formula

$$CF = CF1 + CF2\ (1\text{-}CF1) = CF1 + CF2 - CF1 * CF2 \quad (1)$$

However, the above formula didn't give satisfactory results in our case. Therefore, we use a generalized version of (1):

$$CF = w1 * CF1 + w2 * CF2 + w * CF1\ CF2 \quad (2)$$

where w1, w2 and w are numeric *weights* that should satisfy the following equation:

$$w1 + w2 + w = 1 \quad (3)$$

to assure that $0 \leq CF \leq 1$.

To use formula (2), however, CF1, CF2 and w1, w2, w should be first determined. CF1 and CF2 are calculated as the conditional probabilities p(h/e1) and p(h/e2) respectively, using statistical data from the student database. We use the same statistical data as a *training set* to determine w1, w2 and w. This is done according to the following *weights determination process*:

1. Give initial values to w1, w2, w.
2. Apply the expert system to the training data set.
3. If the results are unsatisfactory, change (some of) the values of w1, w2, w and go to step 2.
4. If the results are satisfactory, stop.

This is an ad hoc process, at the moment. Actually, using a trial-and-error method, we try to optimize the results of the system. Optimization may mean, for example, to have balanced results as far as correct and incorrect predictions are concerned. Changes to the values are guided by the effort to eliminate undesirable results.

4 Logistic Regression Analysis

Logistic Regression (LR) is a statistical method suitable for making predictions. It can model a dependent variable y as a function of a number of independent variables x_1, x_2 … x_k. LR is typically used when the independent variables include both numerical and nominal measures and the outcome (dependent) variable is binary (or dichotomous), having only two values (0-1 or Success-Failure), as in our case. It gives the probability that the outcome, i.e. a student's potential exams result in our case, is successful. The mathematical model for that probability is:

$$E(y) = \frac{1}{1 + \exp\left[-\left(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k\right)\right]} \quad (4)$$

where $E(y)$ represents the probability that y = success (where y represents the exams result), x_1, x_2 … x_k are the independent variables (parameters) that y depends on (specified in Section 2.2), b_0 is the intercept and b_1, b_2, …, b_k are the *regression coefficients*. To calculate $E(y)$, the regression coefficients should be first determined. This is made by using student data from the student database. The well-known software package SPSS is used to make all the above calculations.

A major advantage of using LR, with regards to other statistical methods, is that it requires no assumptions about the distribution of the independent variables. Another advantage is that the regression coefficients can be interpreted in terms of relative risks or odds ratios.

5 Experimental Results and Evaluation

To evaluate PASS, we made the following experiment. From the 201 records of the student database (mentioned in Section 2.2), we used 153 of them (randomly selected) as a *training set* to specify the weights of formula (2), according to the process outlined in Section 3.2. We also used the same set to determine the regression coefficients (via SPSS). Afterwards, we applied PASS to the remaining 48 student records, considering them as the *testing set*. We also applied LR to the same set.

To evaluate PASS and compare it with LR, we used three metrics, commonly used for evaluation of predictions: *accuracy*, *sensitivity* and *specificity* (abbreviated as *Acc*, *Sen* and *Spec* respectively), defined by the following formulas:

$$Acc = (a + d)/(a + b + c + d),\ Sen = a/(a + b),\ Spec = d/(c + d) \quad (5)$$

where, *a* is the number of positive (success) cases correctly classified, *b* is the number of positive cases misclassified as negative (fail), *d* is the number of negative cases correctly classified and *c* is the number of negative cases misclassified as positive. Accuracy shows how good the system is in predicting correctly a student's performance (success or failure). Sensitivity shows how good the system is in predicting a success. Specificity shows how good the system is in predicting a failure. A result in PASS is considered as 'success' when CF ≥ 0.5 ($E(y) \geq 0.5$ respectively in LR).

Table 1. Initial prediction evaluation results

	LR (initial)	PASS (initial)	LR (final)	PASS (final)
Accuracy	0.71	0.75	0.79	0.75
Sensitivity	0.84	0.94	0.97	0.88
Specificity	0.44	0.38	0.44	0.50

The evaluation results are presented in Table 1. Looking at them, we can see that LR and PASS have about the same accuracy. However, LR is more balanced in the initial prediction, whereas PASS is more balanced in the final prediction. LR works more in favor of success in the final prediction than PASS does. This means that misclassifies more failures than successes, which is not desirable. We prefer to have misclassifications of successes than of failures (it's pedagogically more useful). Also, there are some interesting details not illustrated in the tables. For example, an interesting point is how many predictions are marginal (i.e. $0.45 < \text{CF} < 0.55$, $0.45 < E(y) < 0.55$ respectively). There were three such cases in PASS, whereas five in LR.

Another interesting point is that determination of weights in PASS is a quite flexible process. We can define different sets of weights by giving priority to a different metric. For example, the weights resulted after the training phase and used for the above results were w1 = 0.2, w2 = 0.2 and w = 0.6. These values were due to the fact that in the weights determination process we considered as optimized output the one offering a better balance between Acc, Sen and Spec. We also determined the weights based on an optimization policy that gave priority to sensitivity. The weight values were now w1 = 0.2, w2 = 0.4 and w = 0.4. Using these values in PASS, the results of its application to the testing set were the same as those of LR (as far as the values of the three metrics are concerned).

6 Conclusions

In this paper, we present an expert system, called PASS (Predicting Ability of Students to Succeed), which is used to predict how certain is that a student of a specific type of high school in Greece will pass the national exams for entering a higher education institution. Prediction is made at two points. An initial prediction is made after the second year of studies and the final after the end of the first semester of the third year of studies. Predictions are based on various types of student's data. The aim is to use the predictions to help students during their studies towards the national exams.

PASS is a rule-based system that uses a type of certainty factors. We have introduced a generalized parametric formula for combining the certainty factors of two rules with the same conclusion. It is a weighted combination of the two certainty factors and their product. The weights are determined in a training period, prior to the system's use, by using a training set of student data. Experimental results show that PASS is comparable to Logistic Regression, a well-known statistical method. However, PASS is more flexible and gives more balanced results than LR.

Although experimental results show that PASS is of acceptable performance, we feel that it cannot be further improved, even if we increase the training set size. This is mainly due to the fact that the student cases often create a strong non-linear set, which cannot be handled by the type of classification offered by PASS. Therefore, some more advanced classification methods are needed, such as fuzzy rules [5], neural networks [6, 7, 2] or hybrid representations, such as neurules [8] or neuro-fuzzy ones [9]. Use of such methods is a direction of our further work.

References

1. Moore J. S.: An Expert System Approach to Graduate School Admission Decisions and Academic Performance Prediction. Omega International Journal of Management Science 26(5) (1998) 659-670.
2. Walczak S., Sincich T.: A Comparative analysis of regression and neural networks for university admissions. Information Sciences 119 (1999) 1-20.
3. Murray W. S., Le Blanc L. A.: A Decision Support System for Academic Advising. Proceedings of the 1995 ACM Symposium on Applied Computing (SAC'95) (1995) 22-26.
4. Shortliffe E. H., Buchanan B. G.: A model of inexact reasoning in medicine. Mathematical Biosciences (1975) 23(3/4) 351-379.
5. Yager R. R., Zadeh L. A. (Eds): An Introduction to Fuzzy Logic Applications in Intelligent Systems. Kluwer Academic Publishers (1992).
6. Medsker L. R., Liebowitz J.: Design and Development of Expert Systems and Neural Computing. Macmillan College Publishing Company (1994).
7. Cripps A.: Using Artificial Neural Nets to Predict Academic Performance. Proceedings of the 1996 ACM Symposium on Applied Computing (SAC'96) (1996) 33-37.
8. Hatzilygeroudis I., Prentzas J.: Using a Hybrid Rule-Based Approach in Developing an Intelligent Tutoring System with Knowledge Acquisition and Update Capabilities. Journal of Expert Systems with Applications 26(4) (2004) 477-492.
9. Nauck D., Klawonn F., Kruse R.: Foundations of Neuro-Fuzzy Systems. John Wiley & Sons (1997).
10. CLIPS: http://www.ghg.net/clips/CLIPS.html.