ISSN:0975-9646
G. Anuradha et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3365 - 3370

College InfoBank: A system for efficient visualization and analysis of College Data

G.Anuradha[1] , Saylee Chitre[2], Juilee Chitre[3], Kristel Alphonso[4]

Department of Computer Engineering
St.Francis Institute of Technology,
Mumbai, Maharashtra, India

***Abstract*-A large amount of data gets collected as a result of processes like admissions, lectures and examinations in educational institutions like colleges. This data is used only for operational purposes and its potential for providing vital insights into the trends and prevailing conditions remains largely unutilized. Also, its usage for forecasting future in terms of academics remains largely unexplored. The advantages of mining this crucial information can prove very beneficial for an institution like a college. In this paper, we propose a system for an engineering college, customized for its needs and designed as per its features, which would help the administration gain qualitative knowledge from the quantitative data accumulated. The proposed system has three major modules: student performance analysis, student performance prediction and teaching quality evaluation. The system will have input as the student and teacher database from the college and analyze the students and teachers using the concepts of classification, genetic algorithms based neural networks for mining, optimizing and training respectively.**

***Key words*-Data mining, classification, genetic algorithm, neural networks**

I. INTRODUCTION

In the current education system, the student and teacher/faculty data is used only for record-keeping or at the most it is used for statistics, it is not analyzed in the true sense i.e. its potential to infer the prevailing trends is not utilized.

The student success rate can be used as an indicator of college effectiveness by the accrediting agencies. Student Performance Analysis will help the college know about its performance and help the potential students judge a particular college.

Depending on the Student Performance Analysis, we can predict the performance of an individual or a complete batch of students in the next academic year. The task of prediction can be thought as classifying students into different categories depending on their grades, area of interest, strengths, weaknesses, etc. Upon the identification of students' weakpoints, support mechanisms such as orientations, advising, mentoring, etc. may be employed to boost student persistence and thus increase success percentage.

The student performance depends largely on the way knowledge is imparted to them. Classroom teaching is the key link throughout college teaching, but also is the core of teachers' work. Its teaching quality influences directly the quality of the whole college. Therefore, from the perspective of the educational value, using a scientific method to evaluate the quality of classroom teaching comprehensively, reasonably and effectively; would be a boon to colleges. However, because of their heavy workload, complicated statistics and other reasons, human experts often make a mere formality of the work and thus the teaching quality evaluation is not accurate. Therefore, it is necessary to seek a new scientific assessment method.

II. RELATED WORK

Stamos and Andreas [1] reported that their system was more challenging than the previous systems attempting to predict enrollment rates. The reasons for the same being that less data was used for each student and it was the first time that someone had applied sensitivity analysis on an ANN in order to identify the factors that result into a successful graduation outcome.

Our system will analyze individual student performance based on various parameters and based on the outcome of the analysis student success rate will be predicted. The prediction will be done by applying sensitivity to an ANN and results will be displayed in various text and pictorial representations. Individual analysis can be grouped together and the performance of the current batch can be predicted and compared with the past batches of students.

Juan and Changjun [2] reported that GA-base neural networks for assigning grades to teachers can overcome neural networks one-sidedness to some extent and fuzzy neural networks are not completely black-box operation, experts or teaching managers can adjust fuzzy computing rules according to actual experience, which can solve blindness problem of neural network to some extent.

Our system will evaluate teaching quality using genetic algorithm with a modification of parameters to suit Indian conditions and is tailor-made for Engineering colleges.

III. PROPOSED SYSTEM

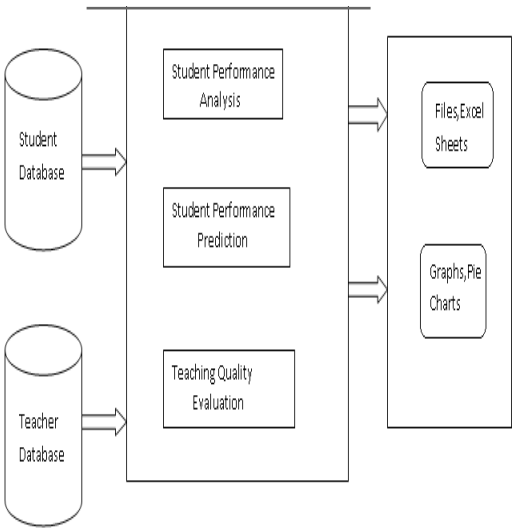


Fig. 1 Modules in the proposed system

The proposed system shown in figure 1 is described as per each module in the following paragraphs:

A. Student Performance Analysis:

Here, we plan to analyse the performance of a batch of students as well as that of an individual student. For a batch of students, the class average will be the studied parameter. The performance of the various batches over the academic years will be displayed in the graphical form. This would be a clear indicator of the competency level of the different classes of students and would make it very evident as to whether the performance is improving , relatively stable or degrading.

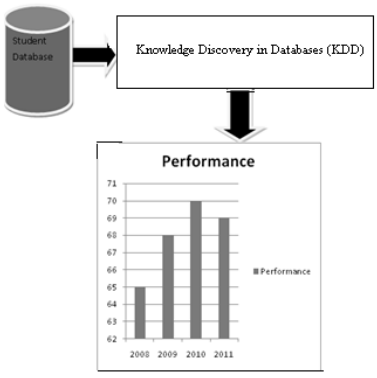


Fig. 2. Student Performance Analysis

For an individual student, the progress report would track his / her academic journey and would be an indicator of his/her intellectual growth in the college.

We will also calculate the Success Rate of the department.

$$\text{Success Rate} = 20 * \frac{\sum_{i=1}^{n} SI(i)}{n} \quad (1)$$

Where

$$SI = \frac{NSC}{NSA} \quad (2)$$

NSC= No. of students who cleared the program in the minimum period of course duration.
NSA= No. of students admitted in the first year of that batch and laterally admitted in 2nd Year.

The Academic performance can be calculated as
Academic Performance = 2 * API (3)
Where API = Academic Performance Index

Average of all the successful students on a 10 point Cumulative Grade Point Average (CGPA) System.

$$= \frac{\sum_{i=1}^{n} CGPA(i)}{n} \quad (4)$$

OR

$$\frac{\frac{\sum_{i=1}^{n} P(i)}{n}}{10} \quad (5)$$

Where "i" is the i^{th} student and "n" is the number of students in the class.

While this forms the basis of performance evaluation, we plan to have sub-modules which will give a very comprehensive and accurate view of the student's profile.The sub-modules are :

1) Student Profile Evaluation:
This will be an inference based module that will conclude details regarding the student's academic profile in the following manner:
(Here, we have assumed the student belongs to the Computer Engineering branch and have assumed the syllabus is as per Mumbai University)

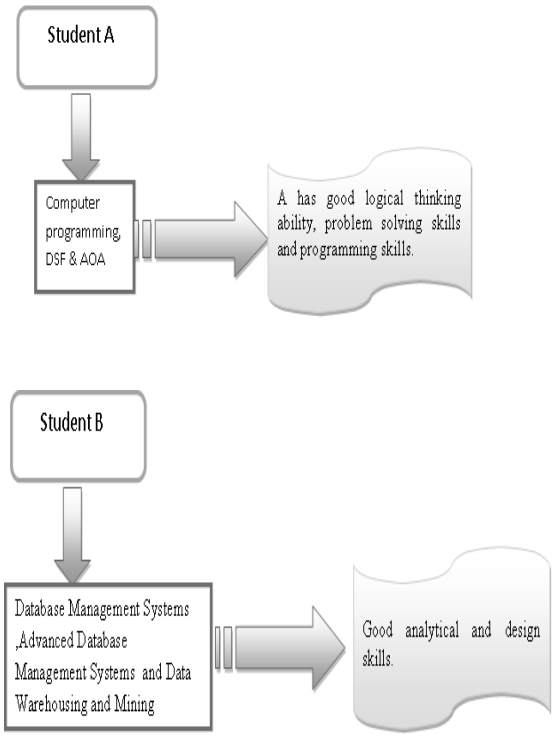


Fig. 3 Student Profile Evaluation

Here, the algorithm we propose to use will be either ID3 or Naïve Bayes Classifier.
This way, the college can provide students feedback regarding their skill-set and help them choose a prospective career specialization field.

2) Weak Area Identification:
This sub-module will enlighten students about their weak areas. Based on the previous semester performance the algorithm predicts those subjects which requires more concentration for the students. As an example, if a student had secured just passing marks (i.e 40) or had a ATKT(Allowed To Keep Term) in a subject like Basic Electrical and Electronics Engineering and/or Electrical Devices and Circuits then he/she should concentrate more on such logical subjects like Analog and Digital Circuits, Digital Signal and Image Processing in further semesters.

B. Student Performance Prediction:
This module uses a novel algorithm Future Grade Algorithm (FGA) proposed by us, based on heuristics to predict the final year marks and also the prospective career field.

1) Final Year Marks Prediction:

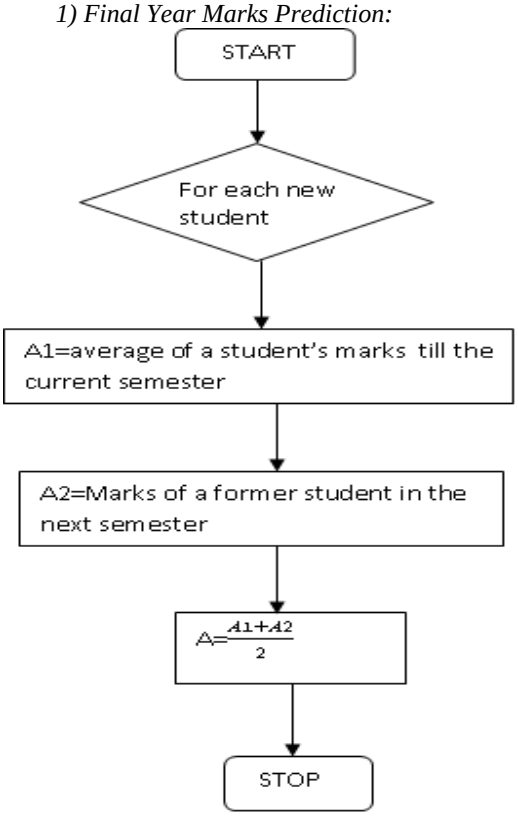


Fig. 4 Flowchart of the proposed algorithm FGA

As per the proposed FGA algorithm in Figure 4, former student will be chosen such that the former students' previous marks closely resemble the current student marks. If there exists many similarity profiles then calculate the average of all of them. Similarity profiles can be with respect to similar marks/ranks in the previous semesters, and/or compatible CET scores.

'A'= the current student's forecasted marks

After prediction results are matched with the actual outcomes. Based on the similarity the tuning factor α which is 0.5 in our case could be modified in the future or suitable weights could be assigned to the factors A1 and A2 respectively. In this way, the working of the FGA is similar to a neural network.

2) Student Prospective Career field Prediction:
Based on the results of the Module A.1 Student Profile Evaluation module as per Figure 3, the proposed system can predict prospective career fields and job positions for the students as per his/her potential. This sub-module will make suggestions of the following kind:

If a student has good visualization skills, good programming skills , then the student could work as a developer in one of the following positions:
Games Developer
Website Developer
Front-end designer etc. and could consider
doing specialization in fields like:
Computer Graphics
Virtual Reality
Human Computer Interaction and so on.

This submodule also can be implemented using Naïve Bayes Algorithm.

3) Assessment of Student Placement and Higher Studies:

In addition to the prediction module, the analysis of the placement scenario can also impart knowledge of prevailing trends to us.

Assessment of the placement scenario for the college can be done in the following manner:

Assessment Points = $\frac{20*(X+1.25*Y)}{N}$ (6)

Where
X = Number of students placed,
Y = Number of students admitted for higher studies with valid qualifying scores/ranks,
N = Total number of students who were admitted in the batch including lateral entry
subject to Max. Assessment Points = 20.

C. Teaching Quality Evaluation:
This module helps in evaluating the teaching quality to stimulate the teaching learning process in the college

1) Gradation of Professor teaching using feedback questionnaire:
Using the GA-based model for teaching quality evaluation proposed by Zhang and Zhu [2] , and with some modification in the parameters to evaluate teachers , we plan to implement this module. These parameters will be calculated depending on the feedback collected from the questionnaires. The questionnaires are subjective varying from university to university.
The comprehensive evaluation indicators for teaching evaluation from the view point of students are as follows:

TABLE I. PARAMETERS FOR EVALUATING A PROFESSOR.

1	SUBJECT KNOWLEDGE
2	UNDERSTANDABILITY
3	PUNCTUALITY
4	APPROACHABILITY
5	TEACHING AIDS USED
6	CONTROL OVER CLASS
7	COMPLETION OF SYLLABUS
8	DOUBT-SOLVING CAPABILITY

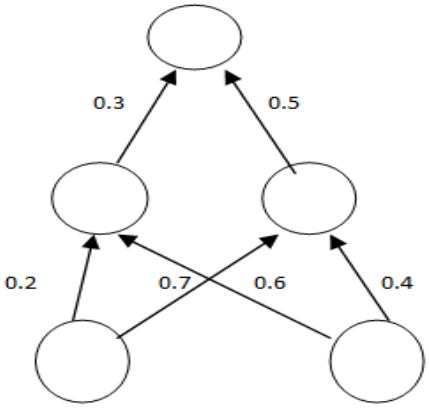
TABLE I. PARAMETERS FOR EVALUATING A PROFESSOR.

Training data set contains records having values of these parameters along with the grade given by the human expert. The neural network will be trained over the training data set using genetic algorithm.

This is a novel technique for training a neural network, the advantages of which have been explored by Zhang and Zhu [2]. As Genetic algorithm is a randomized approach, with an appropriately large number of epochs, the weights can converge to optimal or near optimal values.

However, the operators of Genetic Algorithm have to be adapted to support weight modification in the neural network. This is explained in the subsequent paragraphs.

Consider the following neural network model:



The weights chromosome will be:
(0.2 , 0.6 ,0.7 ,0.4 ,0.3 ,0.5)

Fig. 5 Encoding a network on a Chromosome

The components of the algorithm are as follows:
1) Chromosome Encoding: The weights (and biases) in the neural network are encoded as a list of real numbers (See Fig. 5)
2) *Evaluation Function*: Assign the weights on the chromosome to the links in a network, run the network over the training set of examples, and return the error.
3) *Initialization Procedure*: The weights of the initial members of the population are chosen at random.
4) *Operators:* The following are the genetic operators to be used:
Mutation: For each entry in the chromosome, it will be replaced with a random value with a fixed probablility of 0.1.

(0.4, 0.3 , 0.2 , 0.9 , 0.2, 0.7)
Becomes (0.4, 0.3 , 0.2 , 0.6 , 0.1, 0.7)

Crossover: It will take alternate elements of each of the two chromosomes as follows:

(0.4, 0.3 , 0.2 , 0.9 , 0.2, 0.7)
And (0.8, 0.1 , 0.5 , 0.3 , 0.4, 0.6)
Gives
(0.8, 0.3 , 0.5 , 0.9 , 0.4, 0.7)

In this way, the weight updation of the neural network can be performed. The GA for training the neural network can be applied over several epochs till the error is less than the acceptable error measure.
Then, the GA-based neural network is ready to be applied on the test data. The test data will contain values of the feedback parameters for the professor. Using the neural network, the final grade can be calculated. The transfer function used in the neural network can be:

$$\tan sig(n) = 1/(1+e^{-n}) \quad (7)$$

As the accuracy using this form of neural network is quite high [2], so the judgments of teaching quality evaluation module are predicted to have good compatibility with human-expert rating.

2) Faculty Qualification Assessment:
The level of expertise of the faculty members can be calculated using:

$$\text{Assessment} = 3 * FQI \quad (8)$$

Where FQI = Faculty Qualification Index

$$= \frac{10x+6y+4z}{N} \quad (9)$$

Where
x = No. of Faculty Members with Ph. D in Engineering
y = No. of Faculty Members with M. E / M. Tech
z = No. of Faculty Members with B. E / B. Tech

3) Faculty Retention Ratio:
The faculty retention ratio give the college an idea about the stability and permanency of the faculty in the college.

$$\text{Assessment} = \frac{4*RPI}{N} \quad (10)$$

Where RPI = Retention Point Index
= Points assigned to all Faculty
Where Points assigned to a faculty = 1 point for each year of experience at the Institute but not exceeding 5.

4) Assessment of Faculty Research Publications:
The assessment of the research work of the faculty members will help gain insight into the amount of research work done by the faculty members and to determine its future potential as well.

$$\text{Assessment of FRP} = \frac{4*\sum_{i=1}^{N} RPP}{N} \quad (11)$$

Where
RPP= Research Publication Points scored by the faculty member.
N=Total number of faculty members
A faculty member scores at most 5 Research Publication Points depending upon the quality of the research papers published in the past 3 years.

IV. MERITS

The proposed system has innumerable advantages. The educational administrators and faculty will know the reasons for students not graduating, within the appropriate time period. This will help them in taking corrective actions. The system provides the management with a good teacher evaluation method which eliminated human intervention.
Since GA is used in generating the proper gene pool as an input to neural network only required input parameters are used in training. This increases the network performance.

V. LIMITATIONS

The system is not integrated with any college operation system at this stage and thus, automatic data feeding to the system is yet to be realised. The system is restricted to the Computer Engineering Department at present.

VI. CONCLUSION

Thus, the application of data mining techniques and neural network concepts to college data analysis can provide very keen

insights to administrators, professors and students as well. The results of the system can allow the authorities to focus on areas that need attention and then address the same. The prediction module of the system will help in taking action to prevent any undesirable academic outcome.

VII. FUTURE WORK

The system could be integrated with the college's general transaction processing system to provide for automatic data entry into the system. The proposed system has been designed for Computer Engineering Department. However, using expert knowledge about different subjects covered in various areas of study, it could be extended to other departments as well. The proposed system which is mainly intended to function as per University of Mumbai standards can be extended to any University following any accreditation pattern.

REFERENCES

[1] Stamos.T.Karamouzis, Andreas Vrettos (2009) p 671-675 "Sensitivity Analysis of Neural Network Parameters for Identifying the Factors for College Student Success'. World Congress on Computer Science and Information Engineering.

[2] Juan Zhang, Changjun Zhu (2009) p 33-35"GA-based Neural Network model for Teaching Evaluation". 2nd International Conference on Power Electronics and Intelligent Transportation System.

[3] FanXiu-juan, Liao Qing (2009) p 231-234 "Research and Application of Algorithm for Comprehensive Quality Evaluation of College Students" Second International Conference on Intelligent Computation Technology and Automation

[4] Nor Eleena Yusoff, Daud Mohamad , Zaidah Ibrahim, Syed Ahmad Aljunid (2010)p 1099-1103 "Application of Back Propagation Neural Network in Forecasting University Program"International Conference on Science and Social Research(CSSR) .