Student Performance Prediction for Education Loan System

Radhika Kale

Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati

---

Abstract - *In education system evaluation and prediction of student performance is a challenging task. Usually, a degree program consist of term and course architecture where a term is separated in semesters and courses have set of subjects on basis of background of student and interest. Recently, student background and performance are used to provide the financial loan by the banking system for the completion of degree program. Here, banking system service as education loan relies on student academic performance in a term. The student performance prediction model is based on base predictor and ensemble predictor, including data set. A data-driven approach based on latent factor models and probabilistic matrix factorization is used to discover course relevance, which is important for constructing efficient base predictors. This way banking system gets the assurance of successful degree program. Current performance is also important which add the visualization of future student performance. Very few searches have been made for predicting students' performance by completion of degrees, as most searches focus on performance prediction on historical log dataset. Machine learning approaches are used for automatic prediction of students' performance in degree program.*

Key Words*: *Course Relevance, Data driven approach, Latent factor, Prediction of student performance, personalized education, Probabilistic Matrix Factorization

1. INTRODUCTION

Data mining is a major thrust area to predict the future scope in research domain, including education, business, industry, forensic science, health care, cyber security, etc. In education system it is necessary to build a system that can keep track of students' performance which accurately, predicts students' future performance. Existing approach shows difficulties to cover the diversities of student's educational background and it becomes critical for undefined set of course. In case of relevant course student background information is considered which helps to estimate the relevant course in a term. Traditional existing system, Intelligent Tutoring system and Massive open online courses are working on past records. It is difficult to predict the student performance over on-going records so, a prediction model required. Here, it uses machine learning approach for prediction of performance which speeds up the performance and also reduces the computation time. We develop a novel algorithm for making predictions based on students' progressive performance states. This type of model consists of two layers of predictor which is named as base predictor and ensemble predictor for performance evaluation. In the base layer, multiple base predictors make local predictions given the snapshot of the student's current performance state in each academic term. In the ensemble layer, an ensemble predictor issues a prediction of the student's future performance by synthesizing the local predictions results as well as the previous-term ensemble prediction. The cascading of ensemble predictor over academic terms enables the incorporation of students' evolving progress into the prediction while keeping the complexity low. This system also form cluster of courses for finding relevant courses to train the data according to student and the information that a student provides.

2. LITERATURE SURVEY

Rahel Bekele and Wolfgang Menzel proposed the importance of accurate estimation of student's future performance is essential in order to provide the student with adequate assistance in the learning process. To this end, this research aimed at investigating the use of Bayesian networks for predicting performance of a student, based on values of some identified attributes. This present empirical experiment on the prediction of performance with a data set of high school students containing 8 attributes. The paper demonstrates an application of the Bayesian approach in the field of education and shows that the Bayesian network classifier has a potential to be used as a tool for prediction of student performance [2]. C.MARQUEZ-VERA proposed to apply data mining techniques to predict school failure. Several experiments have been carried out in an attempt to improve accuracy in the prediction of final student performance and specifically of which students might fail. The outcomes of each one of these approaches using the 10 classification algorithms and 10-fold cross validation is shown and compared in order to select the best approach to this problem. This research is failed to develop our own classification algorithm using grammar-based genetic programming and cost sensitive classification for comparison versus other classification algorithms [3]. Hao Cen, Kenneth Koedinger, and Brian Junker proposes a semi-automated method for improving a cognitive model called Learning Factors Analysis that combines a statistical model, human expertise and a combinatorial search. This method is used to evaluate an existing cognitive model and to generate and evaluate alternative models. To use the method for datasets from other tutors to discover its potential for model and tutor improvement [4]. Nguyen Thai-Nghe, Tomas Horvath and Lars Schmidt-Thieme proposes to take into account the sequential effect, this work proposes a novel approach which uses tensor factorization for forecasting student performance. With this approach, we can personalize the prediction for each student given the task; thus, it can also be used for recommending the tasks to the students. Experimental results on two large data sets show that incorporating forecasting techniques into the factorization process is a promising approach [5]. Mingyu Feng, Neil Heffernan, Kenneth Koedinger proposed the assessing student math proficiency is to use data that our system collects through its interactions with students to estimate their performance on an end-of-year high stakes state test. This result show that we can do a reliably better job predicting student end-of-year exam scores by leveraging the interaction data, and the model based on only the interaction. Continues assessment systems can perform better as compared to the system proposed in this paper [6]. Man-Ching Yuen, Irwin King, Kwong Sak Leung proposed a Task Recommendation (TaskRec) framework based on a united probabilistic matrix factorization, aiming to recommend tasks to workers in dynamic scenarios. Unlike traditional recommendation systems, workers do not provide their ratings on tasks in crowd sourcing systems, thus they infer user ratings from their interacting behaviors [7].

3. SYSTEM ARCHITECTURE

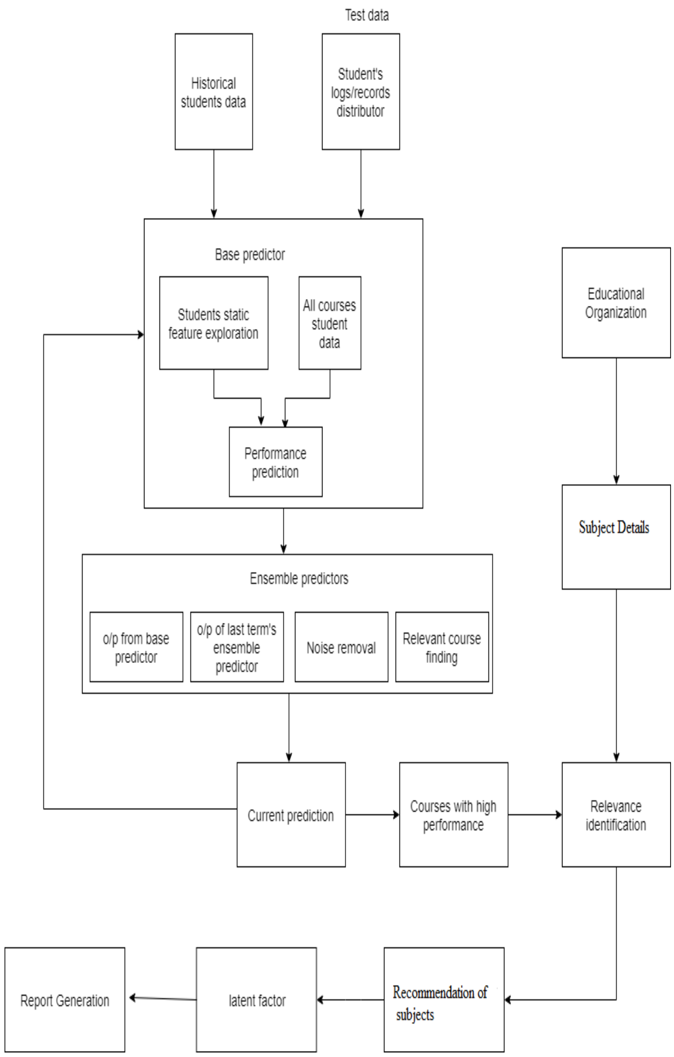


Fig: Block diagram of system

Fig shows the details of bi-layered architecture where system takes student's details at the time of course registration. There is a test data of student historical record i.e. their performance reports and logs. The feature extraction process is done on student's provided data. This data is compared with all students' data from test set of their courses.

In system, student provides the information of his/her educational background. This information along with the reports is used to predict results using two different layers of predictors. Those are as given below:

1. Base predictor layer:

This layer takes the information of students and performs feature extraction. This layer consists of number of predictors which predicts performance individually. For each base predicator denote the prediction result for student given the student's static feature and the current performance state. The base predictors are trained using a dataset consisting of all student data. Learning the base predictor is done offline.

2. Ensemble layer:

This layer does the noise reduction part and finds relevant courses for student's performance prediction. The predicted output is then again passed to the base predictors for adaptation and consideration of student's performance state in future prediction. The ensemble predictor synthesizes the previous ensemble output and output of the base predictors and makes a final prediction.

The ensemble predictor is learned using student data. Learning the ensemble predictors is done online. The output predicted by ensemble layer is further passed to the base predictor layer. This ensures that the current performance state of student is taken into consideration for next prediction.

The naïve bayes classifier is used for classification and prediction purpose. The latent factor model is used for suggest subject which are not covered by previous degree or year.

The probabilistic matrix factorization is used to learn the course correlation which is performed on the student dataset. This model generates a report of performance prediction. This report helps to bank to ensure student's satisfactory and on-time graduation.

4. ALGORITHM

Algorithm 1 Ensemble-based Progressive Prediction (EPP)

Initialization: $L(h^t) = 0, L(f^t) = 0, \forall t.$
for each student i **do**
  Observe backgrounds θ_i, student group g_i
  for term $t = 1$ to T **do** ▷ *Prediction Phase*
    Observe performance state x_i^t
    Extract relevant state $\tilde{x}_i^t$
    Receive prediction $\hat{y}_i^{t-1}$ from f^{t-1}
    Base predictor $h \in \mathcal{H}^t$ predicts $z_{h,i}^t = h^t(\theta_i, \tilde{x}_i^t)$
    Ensemble predictor f^t predicts
      $\hat{y}_i^t = f^t(\hat{y}_i^{t-1}, \{z_{h,i}^t\}_h | v_i^{t-1}, w_i^t)$
  end for
  Observe true label y_i.
  for term $t = 1$ to T **do** ▷ *Update Phase*
    Compute prediction loss $l(\hat{y}_i^t, y_i)$ and $l(z_{i,h^t}^t, y_i)$
    Update $L_i(h^t|g_i) \leftarrow L_{i-1}(h^t|g_i) + l(z_{i,h^t}^t, y_i)$
      $L_i(f^{t-1}|g_i) \leftarrow L_{i-1}(f^{t-1}|g_i) + l(\hat{y}_i^t, y_i)$
    Update weights w_{i+1}^t and v_{i+1}^t according to (4)
  end for
end for

Algorithm 1: [1]

Descriptions of symbol that are used in algorithm are following:

L(h) : Loss of base predictor ,

gi : Group of i student

ht : Base predictor value ,

ɵi : Static function

Zth,i: Prediction value of grade of targeted course,

ft : Ensemble prediction,

wi(ht) : Weight vector of base predictor,

G : SAT scores/Background,

H : Number of base predictor,

Ln(h/g) : Cumulative loss,

l(y' ; y) : Loss function,

Nk(j) : Targeted course of cluster

Algorithm 2: Latent factor

1. Input: Subject List (S), Course Name (c)
2. Output: Latent Factor for Student
3. Latent Factor ← Φ
4. for each Course c ∈ C
5. for each Subject s ∈ c
6. subjects subs ← findRelevantSubjects(s)
7. for each Subject sub ∈ subs
8. if(!(sub ∈ S))
9. Latent Factor ← sub
10. end if
11. end for each
12. end for each
13. end for each
14. return Latent Factor

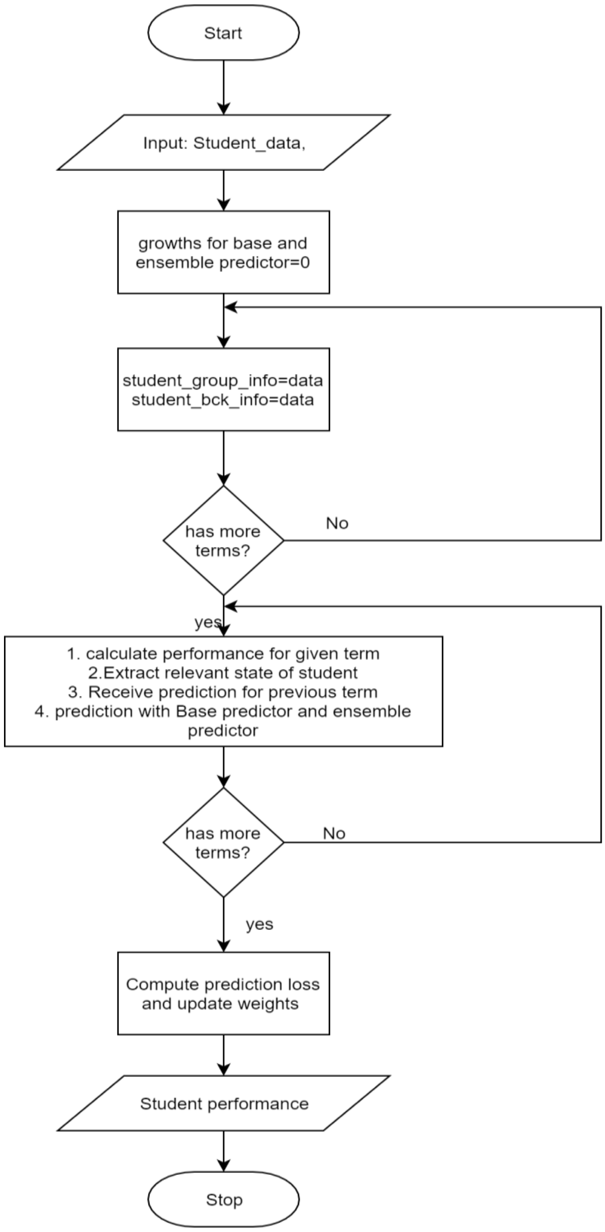
In this system training and testing dataset are used to trained system. The training dataset is of different courses along with their prerequisite courses and subjects. Here in latent factor,find out the relevant subjects for the specific courses and that is the output of the latent factor.

Here in this algorithm, first trained the system using arff, we load the arrff, then we pass the arrff to the classifier to predict the output.

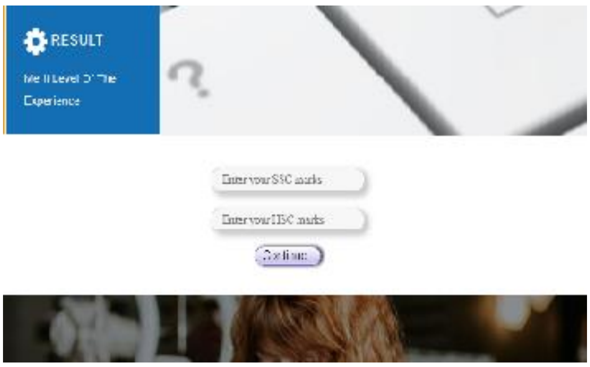
In the above algorithm on line 1 gives subjects and course as a input then we find the relevant subjects relevant to the selected course.

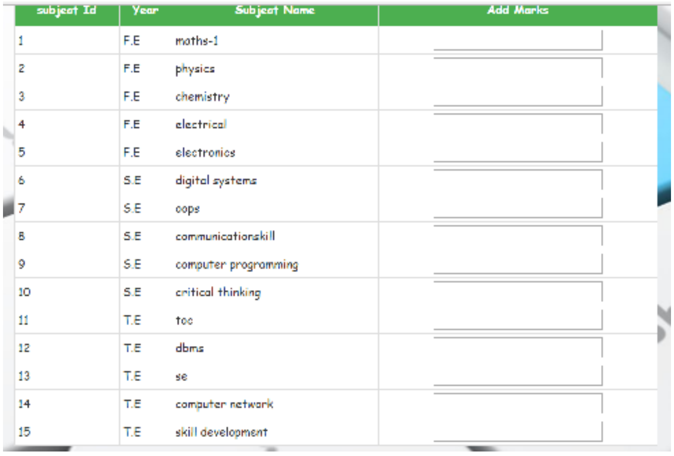
Finally this system is a student performance prediction system on that we can predict student performance for future use.

5. FLOW CHART



6. RESULTS





subject Id	Year	Subject Name	Add Marks
1	F.E	maths-1	
2	F.E	physics	
3	F.E	chemistry	
4	F.E	electrical	
5	F.E	electronics	
6	S.E	digital systems	
7	S.E	oops	
8	S.E	communicationskill	
9	S.E	computer programming	
10	S.E	critical thinking	
11	T.E	toc	
12	T.E	dbms	
13	T.E	se	
14	T.E	computer network	
15	T.E	skill development	

Fig: Prediction result

STU _ID	YEAR	SUBJECT(COURSE)	MARKS	PREDICTION
1		BE(COMP)		
	SSC		86	
	HSC		67	
	F.E	1.Maths-1	75	
	F.E	2.Physics	65	
	F.E	3.Chemistry	48	
	F.E	4.electrical	67	
	F.E	5.electronics	75	
	S.E	1.digital system	64	GOOD
	S.E	2.oops	75	
	S.E	3.communication skill	67	
	S.E	4.computer	65	
	S.E	programming	46	
	T.E	5. critical thinking	67	
	T.E	1.TOC	57	
	T.E	2.DBMS	64	
	T.E	3.SE	76	
	T.E	4.Computer network	67	
		5. Skill development		

7. CONCLUSIONS

A novel method for predicting student's future performance in degree programs given their current and past performance. This performance is also affected on banking services to get loan for student's education. Ensemble-based progressive prediction architecture is developing to incorporate student's evolving performance into the prediction.

REFERENCES

[1] Jie Xu, Member, IEEE, Kyeong Ho Moon and Mihaela vander Schaar "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs" DOI 10.1109/JSTSP.2017.2692560, IEEE Journal.

[2] R. Bekele and W. Menzel, "A Bayesian approach to predict performance of a student (bapps): A case with Ethiopian students", algorithms, vol. 22, no. 23, p. 24, 2005.

[3] C. Marquez-Vera, C. Romero, and S. Ventura, "Predicting school failure using data mining," in Educational Data Mining 2011, 2010.

[4] H. Cen, K. Koedinger, and B. Junker, "Learning factors analysis–a general method for cognitive model evaluation and improvement," in International Conference on Intelligent Tutoring Systems. Springer, 2006, pp. 164–175.

[5] N. Thai-Nghe, T. Horv´ath, and L. Schmidt Thieme, "Factorization models for forecasting student performance," in Educational Data Mining 2011, 2010.

[6] M. Feng, N. Heffernan, and K. Koedinger, "Addressing the assessment challenge with an online system that tutors as it assesses," User Modeling and User-Adapted Interaction, vol. 19, no. 3, pp. 243–266, 2009.

[7] M.-C. Yuen, I. King, and K.-S. Leung,"Task recommendation in crowd sourcing systems," in Proceedings of the First International Workshop on Crowd sourcing and Data Mining. ACM, 2012, pp. 22–26.

[8] C. G. Brinton and M. Chiang, "Mooc performance prediction via click stream data and social learning networks," in 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015, pp. 2299– 2307.

[9] KDD Cup, "Educational data minding challenge," https://pslcdatashop:web:cmu:edu/KDDCup/, 2010.13

[10] C. Marquez-Vera, C. Romero, and S. Ventura, "Predicting school failure using data mining," in Educational Data Mining 2011, 2010.