



Predicting Student's Performance using Machine Learning

Vrushali A. Sungar
Lecturer
Dr.D.Y.Patil School of
Engineering Lohegaon, Pune

Pooja D. Shinde
Lecturer
Dr.D.Y.Patil School of
Engineering Lohegaon, Pune

Monali V. Rupnar
Lecturer
Dr.D.Y.Patil School of
Engineering Lohegaon, Pune

ABSTRACT

Education plays vital role in a student's life. While choosing any field, number of options available in front of student. Student's marks, aptitude, family background, educational environment are main essential factors while selecting a career path and these factors act as a training set to the learning system for classification. With time educational records are accumulating and increasing rapidly. To handle this data along with new features without forgetting previously learnt knowledge, incremental learning technique is introduced by machine learning. Incremental learning algorithm handles previous knowledge to take future decisions and update the system. Knowledge is represented by combining different classifiers for identification of student's features for his/her career growth. In this paper, ensemble technique is used with incremental algorithm for student's career choice and results over real world data sets are used to validate the effectiveness of this method.

Keywords

Incremental learning, classifiers, machine learning, knowledge

1. INTRODUCTION

Indian society has progressing backbone of education system. The higher quality education is required for expansion and progress of country. Professional education is one of the pillars of higher education [1]. In real world, challenging task is predicting the performance of the students. Data mining, which is the science of digging into databases for information and knowledge retrieval, has recently developed new axes of applications and engendered an emerging discipline, called Educational Data Mining or EDM [2]. This discipline seems to be a lot promising. These Educational Data mining techniques are applied to interpret the information about the students and explore the hidden knowledge from their resumes. Educational data mining can answer number of questions from the patterns obtained from student data such as who are the students at risk? What are the chances of placement of student? Who are the students likely to drop the course? What is the quality of student participation? Which courses the institute should offer to attract more students? Results of educational data mining can be used by different members of education system [3].

Students can use them to identify the activities, resources and learning tasks to improve their skills. The prediction of result or the prediction of category of company

where student can be placed after the completion of course will help to channelize efforts of students for proper progress. It will also help teachers to take corrective measures towards the progress of the student during the course. Also it will help teachers for analysis students strong point and weak points in each and every attributes like self motivated, Managing Relations, Emotional Stability, Value Orientation. For this

purpose, the adopted techniques in data mining are: association rule mining, classification and prediction, and clustering [2].

The goal of this research paper is to foresee the execution performance and placement possibilities of a student by utilizing one of the characterization methods CART, SVM by utilizing already learned data and it will predict the new productive information with the assistance of mapping capacity. Mapping function is heart of proposed architecture. The purpose for to utilize these function is for holding between continuously arrived stream information and previously learn knowledge for future scope to produce new theory like h_1, h_2, \dots, h_n . There are two way to build the incremental module one way is using single instance or it may be multiple instances. Navie Bayes and utgoff's ID5R[8] these two are examples of instance incremental algorithms. These two approaches will used to take further decision by analyzing according to their memory usage. ID5R and Navie Bayes belong to particular boundary. ID5R has unbounded memory whereas Navie Bayes has bounded memory. Any standard machine learning algorithm can be used for multiple instance and single instance. Data mining strategy support for the classification methods among them we had utilized like NB, Kstar, and CART etc. Interesting part in research work is combining classification methods with mapping capacities called ensemble learning. Basically ensemble learning means each classifier will make a different error, and purposefully combining these classifiers errors can reduce the total error and improve accuracy of algorithm [9]. Ensemble learning has many form namely multiple classifiers [10]–[11], dynamic classifier selection [11], classifier fusion [12]–[13], mixture of experts [14], [15], committees of neural networks [16], stacked generalization [17], or composite classifier systems [18]. These all terms are different from each other in way of single instances are executed their hypothesis and multiple instances are executed their hypothesis after combining classifier.

There is pair of two mapping capacities with base classifiers. Utilizing such mix enhances predication execution and capacity of classifier learning for selecting student's features for suitable career choice. Whatever is left of this paper is sorted out as takes after. In Section 2, we give a review of different methodologies recommended for incremental learning algorithm, which were initially proposed for enhancing execution of classifiers. In Section 3, we demonstrate how past information is used for next concealed information these methodologies can be utilized as a part of an incremental learning setting, and present the proposed algorithm. In Section 4, we clarify the benchmark and true databases used to assess the algorithm, alongside results acquired on these databases. We likewise look at the two different dataset for classifier precision and execution on this present reality database. At last, in Section 5, we outline our decisions and point at future examination

points.

1.1 Pair of mapping function with classifiers

In the architecture, as described earlier, the combination of base classifiers and classifiers for the purpose of weight distribution is used. The four pairs of classifiers namely MLP-CART, MLP-SVM, SVM-CART, SVM-SVM are formed for the experiments.

2. BACKGROUND

Supervised technique used to classify the data. Fundamentally, this model consist pair items as input and desired output. Hypothesis is created utilizing predefined classes and then this hypothesis is used to predict the hidden items those are recently presented by framework. At whatever point new occurrences is coming into the framework, framework retrained for this examples, not identify new classes. New classes of instances mean it introduces new features of that instance at each time when new data is introduced. Han and Kamber et al [4] depicts data mining software that permit the users to dissect data from distinctive measurements, order it and outline the connections which are recognized during the mining procedure. Galit et al [5] gave a case study that use students data to examine their learning behavior to predict the outcomes and to inform students at risk before their final exams. Various procedures can be applied to data mining for classification problems, including discriminate analysis, decision trees, artificial neural networks, instance-based classification and so on. These techniques have been applied to an extensive variety of building fields.

At the point when looking at static and dynamic algorithms, it is beneficial to remember the diverse sorts of setting where they may be connected. In a static setting, an algorithm has a fixed collection of examples in hand, and utilizes them to develop a theory, which is utilized from that point for characterization without further alteration. In a dynamic setting, the calculation ceaselessly changes its speculation as it is being utilized; it more than once gets an example, predicts its order, figures out the right characterization, and perhaps redesigns its theory likewise.

The [6] algorithm is a keen ensemble approach for incremental learning enlivened basically by the AdaBoost algorithm makes an ensemble of (weak) classifiers, each prepared on a subset of the accessible examples: in request to center the classifier training on examples conveying novel data, the training examples are drawn from a distribution iteratively upgraded on the premise of the execution of the whole outfit. Besides, as new information arrive, produces extra classifiers, until the group takes in the novel data. This permits to learn new information, even when already concealed classes are presented. Since no classifier is discarded, already gained learning is held. Among the efforts on incremental learning from knowledge discovery and information examination perspectives, various new algorithm and architectures have been created and effectively connected to diverse areas. For instance, an incremental direct discriminant investigation (ILDA) was proposed in to handle the opposite of the inside of class scramble grid issue. In view of ILDA, another algorithm, specifically GSVDILDA, the generalized singular value decomposition LDA, was proposed and effectively connected to the face acknowledgment problem [7]. Some problems tackle by incremental learning for example catastrophic forgetting

The following characterizes the notion of incrementally as it applies to learning algorithms.

1) It ought to have the capacity to take in extra data from new information.

2) It ought not to oblige access to the first information, used to prepare the current classifier.

3) It ought to save beforehand procure learning (that is, it ought not to experience the ill effects of calamitous overlooking).

4) It ought to have the capacity to oblige new classes that may be presented with new information.

A learning algorithm is incremental if, for any given preparing specimen $e_1...e_n$, it delivers a grouping of hypotheses $h_0, h_1...h_n$, such that h_{i+1} depends just on h_i and the present sample e_i . Incremental learning algorithms are likewise called memory less on-line algorithms. For consistency with learning undertakings, the term incremental is favored here.

In machine learning incremental learning are very important approaches to solve real world problems due to following reasons: first reason, before system will fed train samples into use its difficult to collect needful data. When new samples are encouraged, the learning methodology ought to have the capacity of doing a few amendments on the trained framework so that unlearned information encoded in those new examples can be incorporated. Besides, adjusting a trained framework may be less expensive in time cost than building another framework without any preparation, which is helpful particularly progressively applications. The concept of combining classifiers is proposed as a new direction for the improvement of the performance of classifiers.

Hence the research objective of the study is to improve prediction performance and help to student's to improve their academic record. In our report, we are giving the detailed view of the proposed algorithm for student's features classification with the combination of two Mapping function and different base classifiers. System shows that which is best for the student data set. As we all know incremental learning may be a learning mechanism in machine learning, classifier that may incrementally trained the novel knowledge gave by the new information, while not commanding the already procured learning.

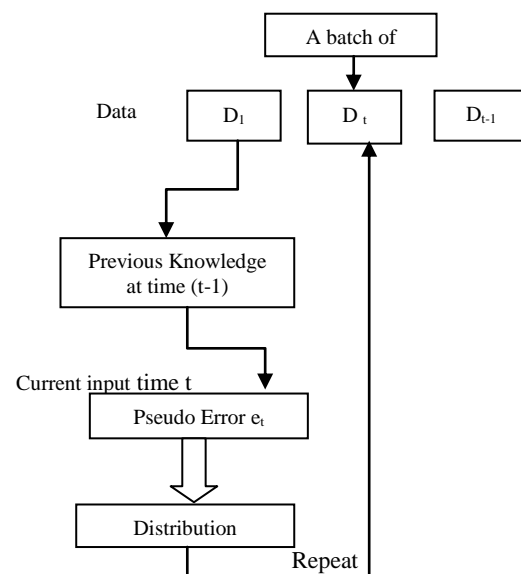


Figure 1: Simulation Flow

3. PROBLEM DESCRIPTION

As an example medical studies, cancer data or financial data analysis applications wherever knowledge area unit perpetually obtained over many years—or any method that generates streaming knowledge area unit samples of real-world applications that may have the benefit of such a classifier. The capability of a classifier to find out underneath these conditions is understood as incremental learning, which suggests capable of learning from continuous information, accumulating expertise over time, and victimization such data to enhance future learning and prediction performance [7].

3.1 Proposed Algorithm

Algorithm 1: Proposed algorithm for students classification

INPUT

Step 1: Initially dataset is divided into equal parts i.e. D_i is dataset which has (X_i, Y_i) features of attributes. Where X_i is training samples and Y_i is corresponding correct classes for $i=1, 2, \dots, K$. T_k is k^{th} randomly selected sample.

Step 2: Call base classifier (SVM or CART) which generate hypothesis h_i . Initially no weight distribution for first data $D_i = 0$. Meanwhile, $D_i = \{\omega_1, \omega_2, \omega_t\}$

Step 3: Calculated Error of h_i

CLASSIFIER OUTPUT WEIGHTS

Step 5: The classifier's weight is calculated based on its accuracy after each classifier is trained.

Step 6: We upgrade the probabilities of each of the training tests showing up in the training set for the next classifier.

Step 7: The main classifier ($t = 1$) is training with equivalent likelihood given to every training sample. After it's training, we process the weight (Beta) for that classifier.

$$B_i = \frac{1}{2} \text{EXP}(1 - \epsilon_i / \epsilon_i)$$

FINAL OUTPUT

Step 8:

$$H(x) = \text{sign}\left(\sum_{i=1}^T \epsilon_i h_i(x)\right)$$

Hence we get final output of classifier we combined weak learner of weighted sum into the previous weight.

Algorithm 1 shows, feature choice method of student's classification. During this algorithm, whole dataset is load for data selection and preprocessing. Input to the current formula is X_i for feature vector of i^{th} sample. Y_i is category label. Hypothesis H_i is generated based on D_i . Incremental learning use previous knowledge to delineate the unseen knowledge for future use. Initially weight is NULL. In second step, call base classifier. Apply either on batch or instances wise stream dataset which can manufacture hypothesis based on D_i . This D_i and D_{i-1} ought to be used for future information to enhance learning capability and prediction performance. Accuracy is calculated when every classifier is trained. Those samples are hard to learn assign more weight to that samples and they are easy to classify.

4. EXPERIMENTAL RESULTS AND ANALYSIS

To predict the student's behavior we need to consider lot of parameters. Personal, social aspects, psychological and other basic environmental variables are important for effective prediction of performance. Most part that included for pertaining

students data based on student's background which plays vital role in classification. Psychometric test was conducted in our institute for 1325 students with 10 features and 7 classes with age group of 16-20. We given some sample questions to students based on this dataset is created. Classes indicate their interest in attribute. Here we consider students dataset for experimental studies. All data is in numeric type.

4.1 Data Preparations

Initially size of the data is 1325 and attributes are 10. Total 7 classes are present in student dataset.

Table 1: Attributes used for each Student

Notion	Attributes	Type
A	Self Awareness	Numeric
B	Empathy	Numeric
C	Self Motivation	Numeric
D	Emotional Stability	Numeric
E	Managing Relations	Numeric
F	Integrity	Numeric
G	Self Development	Numeric
H	Value Orientation	Numeric
I	Commitment	Numeric
J	Altruistic Behavior	Numeric

4.2 Data Selection and Transformation

Table I has 10 attributes. This dataset has total 7 classes respectively high, normal, low, mid, excellent, and ordinary, extraordinary. Name of each attribute and its types has been covered in the table I. Table IV shows that evaluation score sheet of students. Each and every feature has some numeric range. Suppose students range in between 3 and below in particular attributes. Then we can conclude that students may be low in that attribute. Likewise remaining calculations has been done. This framework is divided into two part- CART Classification and SVM classification Batch. In the CART classification Batch parts, 1325 instance divided 10 times using 10 fold classification methods of data mining. Each batch belongs to classes. Batch 131, 263, 395, 526, 658 belongs to class 1, 2, 3 and Batch 790, 921 belongs to class 1, 2, 3, 4, 5. Batch 1053, 1185, and 1316 belongs to class 1, 2, 3, 4, 5, 6, 7. Table II shows that running time in seconds taken by CART and SVM when MLP as mapping function on core processor i5, RAM is 4 GB and 64 bit operating system. In order to authenticate the performance of the proposed framework, two real-world data sets with varied size and number of classes from UCI machine learning repository are used for research purpose. The detailed information of these data sets can be found in Table I. In this experimental study, each dataset randomly selected and divide into 10 chunks with same size. At each T_k , one data is randomly select D_{t-1} for testing purpose and reaming 19 chunk pass to training purpose. For each 20 times dataset is folded for better accuracy. MLP-CART and MLP-SVM act as base classifier. MLP is a mapping function to bind the previously unseen data with newly arrive data for future knowledge. MLP is an non linear function which has 10 hidden layer neurons and one output layer. The number of input neurons is set to be equal to the number of features for each data set. The training epochs of the MLP is set to be 500.

Table II shows the overall running time on student dataset and satellite dataset. Obviously, SVM spends much more time than

CART in learning from streamed data chunks. Table IV shows the buildup time required for CART and SVM on student dataset and satellite dataset respectively. CART take 0.09 ms time to build students dataset and Satellite dataset takes 0.06ms. Whereas SVM model takes 1.79ms on student’s dataset and 226.95ms on satellite dataset.

Table 2: Running Time for Cart and SVM (In Seconds)

Dataset	CART		SVM	
	Training Set	Testing Set	Training Set	Testing Set
Students Dataset	0.17	0.02	1.36	0.03
Satellite Dataset	0.11	0.02	230.83	0.03

Table 3: Time Taken to Build Model

Dataset	CART	SVM
	Time taken to build model	Time taken to build model
Students Dataset	0.09	1.79
Satellite Dataset	0.06	226.95

Table V indicate that predication accuracy of student’s dataset on batch wise with the MLP mapping function and SVM mapping function. Mapping is a heart of algorithm. In MLP mapping function, performance on CART is 93.24 other side SVM is 92.33. In SVM mapping function, instances incorrectly classified more. As compare to SVM mapping function, when we apply MLP as mapping function and CART as base classifier, CART gives better prediction accuracy.

Table 4: Error Calculation of Cart and SVM Classifiers

Parameters	CART	SVM
Correctly Classified	93.2422 %	92.33 %
Incorrectly Classified	6.75 %	7.66 %
Kappa Statistic	0.9197 %	0.0988 %
Mean Absolute Error	0.035 %	0.2057 %
Root Mean Squared Error	0.135 %	0.3036 %
Relative Absolute Error	14.5339 %	85.4877 %
Root Relative Squared Error	38.9218 %	87.529 %

Table 4: Raw Score for Particular Age Group of Student

Factors	A (Self Awareness)	B (Empathy)	C (Self Motivation)	D (Emotional Stability)	E (Managing Relations)	F (Integrity)	G (Self Development)	H (Value Orientation)	I (Commitment)	J (Altruistic Behavior)
High	10 and Above	11 and above	12and above	11 and above	12 and above	7 and above	6 and above	6 and above	6 and above	6and above
Normal	2 to 3	2 to 3	2 and below	2 and below	2 and below	2 to 3	2 to 3	2 to 3	2 to 3	2 to 3
Low	3 and Below	6 and below	8 and below	3 and below	4 and below	3 and below	1 and below	1 and below	1 and below	1 and below
Mid	1 and below	1 and below	1 and below	1 and below	1 and below	1 and below	1 and below	1 and below	1 and below	1 and below
Excellent	4 to 10	7 to 14	9 to 17	4 to 10	5 to 11	4 to 7	2 to 5	2 to 5	2 to 5	2 to 5
Ordinary	3 to 8	5 and above	6 to 9	5 to 7	4 and above	4 to 8	3 to 7	3 to 7	3 to 7	3 to 7
Extraordinary	11and Above	15 and above	18 and above	11 and above	12 and above	8 and above	6 and above	6 and above	6 and above	6and above

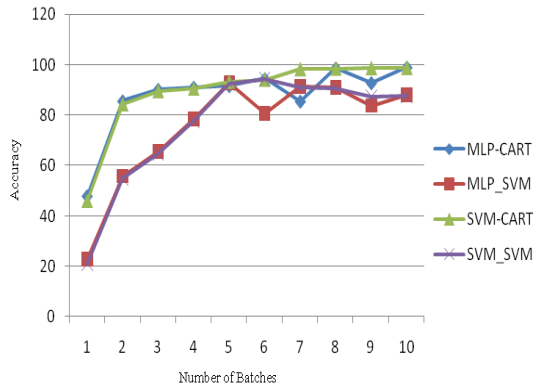


Figure 2: Predication Accuracy on Students dataset using Mapping fuction with CART and SVM as Classifier

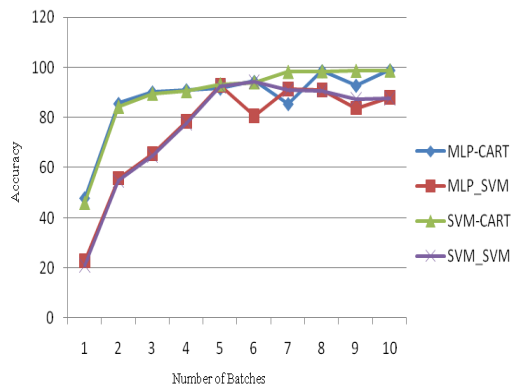


Figure 3: Predication Accuracy on Satellite dataset using Mapping fuction with CART and SVM as Classifier

5. CONCLUSION AND FUTURE SCOPE

To handle large volume of data different data mining technique used to classify trained data. There are different methods or technique used to classify the data like neural network, navies bayes, decision tree, k-stars. time educational records is accumulating and increasing rapidly. To handle this unseen data along with newly introduce feature without forgetting previously learn knowledge, incremental learning technique introduced by machine learning. Our model shows comparative results of MLP and SVM mapping function with SVM and CART as base classifier. MLP mapping function gives good results. Future scope of this system will be to reduce gap between previously unseen data and newly coming data for this purpose mapping function can be used with different dataset and to handle large amount of data.

6. REFERENCES

- [1] Chady El Moucary, "Data Mining for Engineering Schools Predicting Students' Performance and Enrollment in Masters Programs," International Journal of Advanced Computer Science and Applications, Vol. 2, No. 10, 2011.
- [2] R.S.J.D. Baker, "Data Mining for Education," In International Encyclopedia of Education, vol. 7, B. McGaw, P. Peterson, E.Baker (Eds.), 3e, Oxford, UK: Elsevier, pp. 112-118, 2010.
- [3] Vikas Chirumamilla, Bhagya Sruthi T, Sasidhar Velpula, Indira Sunkara, "A Novel approach to predict Student Placement Chance with Decision Tree Induction", Double Blind Peer Reviewed Journal Vol.7, Issue 1, 2014, pp 78-88.

- [4] Surjeet Kumar Yadav , Saurabh Pal ,” Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification”, World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 2, 51-56, 2012
- [5] Galit.et.al, “Examining online learning processes based on log files analysis: a case study”. Research, Reflection and Innovations in Integrating ICT in Education 2007.
- [6] Roozbeh Razavi-Far, Member, IEEE, Piero Baraldi, and Enrico Zio, Senior Member, IEEE, “Dynamic Weighting Ensembles for Incremental Learning and Diagnosing New Concept Class Faults in Nuclear Power Systems”, IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 59, NO. 5, OCTOBER 2012.
- [7] Haibo He, Senior Member, IEEE, Sheng Chen, Student Member, IEEE, Kang Li, Member, IEEE, and Xin Xu, Member, IEEE, ” Incremental Learning from Stream Data”, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 22, NO. 12, DECEMBER 2011 1901.
- [8] Geoffrey Holmes, Richard Kirkby, Bernhard Pfahringer , “A batch –incremental learning for mining data streams”
- [9] Devi Parikh and Robi Polikar, Member, IEEE, “An Ensemble-Based Incremental Learning Approach to Data Fusion”, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 37, NO. 2, APRIL 2007.
- [10] L. Xu, A. Krzyzak, and C. Y. Suen, “Methods of combining multiple classifiers and their applications to handwriting recognition,” *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 3, pp. 418–435, May/June 1992.
- [11] K.Woods, W. P. J. Kegelmeyer, and K. Bowyer, “Combination of multiple classifiers using local accuracy estimates,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 405–410, Apr. 1997.
- [12] L. I. Kuncheva, J. C. Bezdek, and R. P. W. Duin, “Decision templates for multiple classifier fusion: An experimental comparison,” *Pattern Recognit.*, vol. 34, no. 2, pp. 299–314, 2001.
- [13] *Pattern Anal. Mach. Intell.*, “A theoretical study on six classifier fusion strategies,” *IEEE Trans.* vol. 24, no. 2, pp. 281–286, Feb. 2002.
- [14] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, “Adaptive mixtures of local experts,” *Neural Comput.*, vol. 3, no. 1, pp. 79–87, 1991.
- [15] M. J. Jordan and R. A. Jacobs, “Hierarchical mixtures of experts and the EM algorithm,” *Neural Comput.*, vol. 6, no. 2, pp. 181–214, 1994.
- [16] H. Drucker, C. Cortes, L. D. Jackel, Y. LeCun, and V. Vapnik, “Boosting and other ensemble methods,” *Neural Comput.*, vol. 6, no. 6 pp. 1289–1301, 1994.
- [17] D. H. Wolpert, “Stacked generalization,” *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992
- [18] B. V. Dasarathy and B. V. Sheela, “Composite classifier system design: Concepts and methodology,” *Proc. IEEE*, vol. 67, no. 5, pp. 708–713, May 1979.
- [19] V.A.Sungar, “Psychological Impact on Student’s Behavior” , International Journal of Computer Science., ISSN:vol.5, Nov.2017.