

Classifying Educational Data Using Support Vector Machines:A Supervised Data Mining Technique

Guleria Pratiyush and Sood Manu

Department of Computer Science, Himachal Pradesh University, Gyan-Path, Summer-Hill, Shimla - 171005, Himachal Pradesh, India; pratiyushguleria@gmail.com, soodm_67@yahoo.com

Abstract

With increase in Educational Institutions there is increase in new trends which results in large data. The data is unstructured which need to be transformed into structured form and to find out meaningful information, effective Mining Tools are required. Educational Data Mining helps in facilitating utilization of resources related to student performance, predicting placement results and finding new educational trends. In this paper placement data of students has been taken and classification approach using SVM is followed on training data for predicting results which not only helps educational institutions to improve student placements from extracted knowledge as well enhances the competitive advantage and decision making by applying data mining techniques. SVM is a supervised learning and effective data mining technique to train the data for pattern recognition and predictions.

Keywords: Classification, Data Mining, Educational Data Mining, Predicting, Support Vector Machines

1. Introduction

Data mining techniques are used to analyze Educational data and useful information is extracted from large amount of unstructured data. With increase in educational institutions, Demand of quality education and placements for the benefit of the students as well as of the institutions has been increased. In¹, author has suggested many classification and clustering Data Mining Techniques which includes K-means, Decision trees, neural networks, SVM, Naive Bayes etc. Using these techniques meaningful and informative knowledge can be discovered which is helpful in delivering quality education.

SVM is one of the DM Technique. It is a supervised learning method and can be used for both regression and classification. According to author in², SVM are one of the most popular classification algorithms for handling high dimensional data. In³, authors have stated that SVM Classifier is known for maximum accuracy and Minimum Root Mean Square error (RMSE).

In this paper, SVM Classifiers are used for the prediction of placement of students as in many cases, students

focus only on their regular curriculum of studies besides on other educational trends which are necessary for overall development of students and their placements. The educational dataset taken for SVM supervised learning technique is shown in Table 1 which consist of some data points that belongs to one of the 2 classes and using SVM, it is to decide that a new data point will occur in which class. Hyperplane will be defined using an equation:

$F(x) = \alpha_0 + \alpha^T x$ where α is known as the weight vector and α_0 is the threshold value. One of the possible representations of the Hyperplane is:

$|\alpha_0 + \alpha^T x| = 1$ where x represents the training set examples which are closer to the hyperplane and they are known as support vectors. In this paper an attempt is done to train and classify the student's dataset and to predict the values for input given.

The paper has three sections. Section 2 consists of data mining of placement data where attributes taken and educational dataset is defined. Section 3 is discussed about Support Vector Machines, support vectors and closest data vectors. Section 4 is a discussion about results

*Author for correspondence

obtained using WEKA classifier and SVM Classifier in MATLAB.

2. Data Mining of Placement Data

2.1 Training Dataset Preparations

We have initially taken a sample size of 200 Graduate Students for classification. Classification⁴ is a DM technique and a supervised learning where training sample set is input to classifier.

2.2 Data Analysis

We have classified the placement results of students taking 6 Attributes i.e. Attendance, GPA, Reasoning, Quantitative, Communication Skills, Technical Skills etc. The prototype of student's placement result is predicted after analyzing performance.

Attributes and Educational Data Set of students is given in Table 1 and Table 2 respectively.

Table 1. Placement cell attributes

Attributes	Description	Coding
ATTD	Attendance	{ Good = "8-10", Average = "6-7.5", Poor = "1-5.5" }
GPA	Performance in semester	
Reasoning	Reasoning Aptitude	
Quantitative	Quantitative Aptitude	
Communication	Communication Skills	
Technical	Technical Skills	{ Yes = "1", No = "0" }
Placement	Placement Performance acts as a "class"	

Table 2. Educational data set

CLASSID	ATTD	GPA	Reasoning	Quantitative	Comm. Skills	Technical Skills	Placement Class
1	10	9.5	9.5	10	10	10	1
2	7	9.5	7.5	9.5	7.5	9.5	1
3	4	9.5	9.5	8.5	7	7.5	0

4	5	9.5	9.5	8.5	7	7.5	0
:	:	:	:	:	:	:	:
200	8	9	8	8.5	7	8.5	1

3. Support Vector Machine

SVM is a discriminant classifier which is defined by a separating hyperplane. SVM constructs a Hyperplane, which is used for classification and regression⁵. It finds nearest data vectors called support vectors (SV), to the decision confinement in the training set and it segregate a given new test vector by using only these nearest data vectors⁶. Steps followed in SVM have been described in Table 3.

Table 3. Steps followed for SVM

Step 1	Let say there are 2 Classes C1 and C2. Now unknown feature vector x either belongs to class C1 or class C2.
Step 2	Define Linear Discriminant Function $g(x) = w^T(x) + b$ where "T" means transpose is weight vector and x is the Input feature vector, "b" is the bias in a 2-dimensional space or vector.
Step 3	In a 2-d space, if a feature vector is a 2-D vector, then this linear equation represents a straight line i.e. $w^T(x) + b = 0$
Step 4	If input feature vector is 3-D feature vector then this linear equation if but equal to zero in 2-D dimension otherwise it becomes plane in 3-D.
Step 5	If the dimensionality of feature vector is more than 3 then it becomes Hyper plane. "w" is the vector perpendicular to the Hyperplane where vector w represents orientation of the Hyperplane in "d" dimensional space.
Step 6	Classification Rules: For every feature vector x , compute linear function. If x lies on positive side of Hyperplane, then $g(x_i) = w^T(x_i) + b$ $w^T(x_i) + b > 0$ -----(1) if this x_i belongs to -ve side of Hyperplane, then $w^T(x_i) + b < 0$ -----(2) if this x_i lies on the Hyperplane, then $w^T(x_i) + b = 0$ -----(3)
Step 7	SVM classifies data by finding the Hyperplane which separates all data points of one class from those of the other class. The Hyperplane which is considered best for an SVM means the one which is having the largest margin between the two classes.

Support vector machine operator consists of kernel types including dot, radial, polynomial, neural, anova etc. Kernel shows the vector similarity in training dataset samples⁷. Functions of kernels are summarized in Table 4. SVM collects of input data and envision, for each given input data, which of the two probable classes embrace the input, making the SVM a non-probabilistic binary linear classifier.

3.1 Multiclass Classification

Multiclass classification makes the inference that each sample is appointed to one label only.

3.2 Multi Label Classification

In this each sample is assigned target labels and several classification tasks will be there.

Table 4. Kernels in SVM

Kernel	Function
dot	The dot kernel shows the inner product of x and y vectors.
radial	In this kernel, gamma value is taken as the performance metric. It is defined by $\exp(-g \ x-y\ ^2)$ where g is the gamma.
polynomial	Polynomial kernel not only considers the input attributes to find the similarity but also their combinations. The polynomial kernel is defined by $k(x, y) = (x \cdot y + 1)^d$ where d is the degree of polynomial and k is the kernel degree.
neural	The neural network consists of input, hidden and output layers. It is defined by a two layered neural net.
anova	The anova kernel is used for analyzing the variations among the different variables and dependencies on their subsets. Here Kernel gamma and kernel degree parameters are adjusted.
multiquadric	The multiquadric kernel is defined as square root of $\ x-y\ ^2 + c^2$. It consists of kernel sigma and kernel sigma shift parameters which are applied for Gaussian or multiquadric combinations.

4. Results and Discussions

After applying the educational data set shown in Table 1 in WEKA Tool and using SVM, Table 4 depicts classifier model with each attribute and its class. It summarizes

the results comprising of each virtue of the dataset and correctly and incorrectly classified instance. The filter applied on the attribute is Nominal to Binary under Supervised Learning Techniques and then we classify the results using 10 fold cross-validations. The classifier model obtained using WEKA tool is shown in Table 4.

Plot Matrix for all the 6 attributes is shown in Figure 1 and Visualization Results between “GPA” and “Quantitative” attribute is shown in Figure 2.

4.1 SVM Classification Using MATLAB

In MATLAB, SVM classify⁸ function consists of arguments i.e. SVMStruct which classifies each row of data in Sample. A SVM classifier structure “SVMStruct” is created using the svmtrain function. Input arguments for SVM classification are shown in Table 5. The svmclassify function derive results from svmtrain so as to classify vectors y according to following equation:

$$= \sum_i \alpha_i K(S_i, y) + b_i$$

where S_i are the support vectors, α_i are the weights, b is the threshold value(bias), and k is a kernel function.

Table 5. Input arguments for SVM classification

SVMStruct	svmtrain function is used to create SVM classifier structure.
Sample	It is a matrix where each row represents an observation and each column corresponds to an attribute. Sample must contain the same number of columns as in the training dataset. Here columns define the dimensionality of the data space.
Showplot	It is to display a plot of the classification and Displays only for 2-D problems. It follows with a Boolean value that true means to display the plot and false to give no display.

Table 6. Classifier model for training set

Schema	Weka.classifiers.functions.support Vector. Reg SMO Improved “weka.classifiers.functions.supportVector. PolyKernel”
Relation	Educational_Data_Set_Sample
Weka.filters.supervised.attribute	Nominal To Binary
Instances	40

Attributes	6{Attendance, GPA, Reasoning, Quantitative, Communication Skills, Technical Skills} Placement acts as a class
Test Mode	10-fold cross-validation
Classifier Model(full training set)	
Number of kernel evaluations	820 (95.594% cached)
Time taken to build model	0.03 seconds
Cross-validation Summary	
Correlation coefficient	0.6131
Mean absolute error	0.2728
RMSE	0.4237
Relative absolute error	53.0773 %
Root relative squared error	82.2502 %

The classification of dataset is done using a trained SVM classifier and the Placement Results obtained based on attributes i.e. “Reasoning” and “Quantitative” attributes are shown in Figure 3 and Figure 4.

In Table 4 classifier model for the training set is shown. It shows the Root Mean Square Value for the Total Instances taken. The Test Mode applied is 10-fold cross validation on the training data to access the predictive performance of SVM Model.

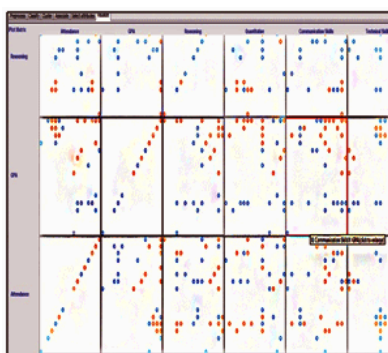


Figure 1. Plot matrix of attributes.

Figure 1 represents the plot matrix in WEKA Tool visualizing all the attributes taken with x and y co-ordinates in 2-D space. Plot matrix visualizes the current dataset in one and two dimensions. The Figure 1 also shows the correlation of any attribute to all other attributes.

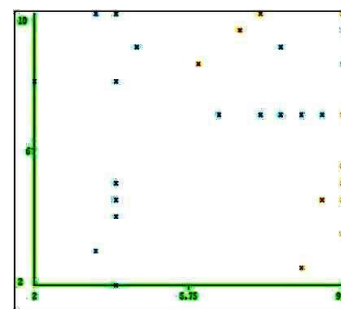


Figure 2. Plot for supervised nominal to binary conversion.

Whereas Figure 2 represents the visualization results of GPA and Quantitative attribute. It visualizes the support vectors for both the classes and categorizes the placement results accordingly.

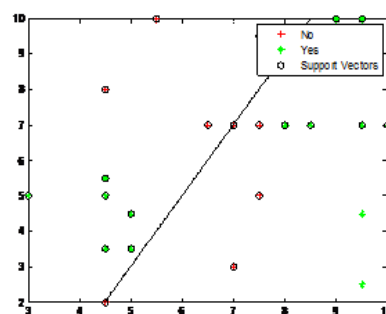


Figure 3. SVM classification for placement data.

Figure 3 for SVM Classification finds a line separating the educational data shown in Table 2 on “Placement” feature for “Yes” and “No” classes, with respect to the values in “Reasoning” and “Quantitative” attributes. Here, svmStructure is obtained using Svmtrain function.

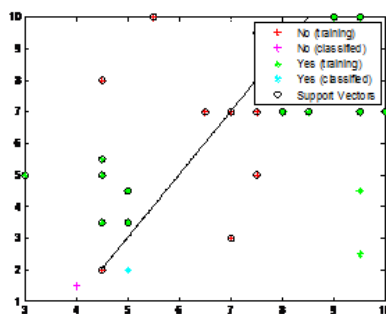


Figure 4. SVM trained classification.

Figure 4 shows the classified and trained values for the placement attribute i.e. “Yes” and “No”. It also shows the support vectors and from it we derive that on the basis

of students marks in “Reasoning” and “Quantitative”, Support vector machine construct a hyperplane showing the closest vectors for “Placement” Attribute.

4.2 Binary Support Vector Machine Classifier

Fitsvm (X, Y) function returns a support vector machine classifier SVMModel, trained by predictors X and class labels Y for one or two-class classification. Figure 5 represents the Trained SVM classifier using the processed data set. The first class (‘No’) is the negative class and the second (‘Yes’) is the positive class.

The support vectors are consideration that occur on or good way off their estimated class boundaries.

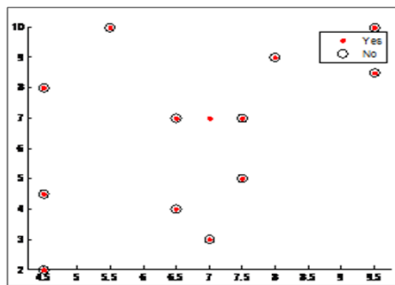


Figure 5. Train binary SVM classifier.

5. Conclusion

Support vector machines are effective in accuracy. In this paper, using SVM data mining technique the dataset is trained for predicting the placement results. Here two classes are defined for placement attribute i.e. “yes” and “no” to analyze student’s data and classified placement results based on certain attributes which helps to improve student placements from extracted knowledge. Using SVM, functional margin is effectuated by Hyperplane which shows the maximum distance to the neighboring training data points of class. In this paper cross-valida-

tion technique is applied to find the best possible values and the dataset is classified into labels. Placement Results classify attributes and gives a better perception of how students should perform and to target on new educational trends apart from their regular studies to get placed. Classification Model in this paper can be further extended for Posterior probabilities of SVM Classifiers, predictions and more accurate results by careful selection of values.

6. References

1. El-Halees Alaa. Mining Students Data to Analyse Learning Behaviour: A Case Study: Available online from: https://uqu.edu.sa/files2/tiny_mce/plugins/filemanager/files/30/papers/f158.pdf.
2. Ertekin S_eyda. Learning in extreme conditions: online and active learning with massive, imbalanced and noisy data; A Dissertation.
3. Agarwal Sonali, Pandey GN, Tiwari MD. Data Mining in Education: Data Classification and Decision Tree Approach. International Journal of e-Education, e-Business, e-Management and e-Learning. 2012 April; 2(2).
4. Mahmood Ali Mohd, Qaseem Mohd. S, Rajamani Lakshmi, Govardhan A. Extracting Useful Rules Through Improved Decision Tree Induction Using Information Entropy. International Journal of Information Sciences and Techniques (IJIST). 2013 January; 3(1).
5. Amarappa S, Sathyanarayana SV. Data classification using Support vector Machine (SVM), a simplified approach. International Journal of Electronics and Computer Science Engineering. ISSN-2277-1956.
6. Tian Xia. Support Vector Machine Based Educational Resources Classification. International Journal of Information and Education Technology. 2016 November; 6(11).
7. The Rapidminer website. Available from: <http://rapidminer.com>.
8. The Mathworks website. Available from: <http://www.mathworks.in>.