International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

# A Hybrid Approach for Integrating Genetic Algorithms with SVM for Classification and Modelling Higher Education Data

**Kamiya Malviya, Prof. Anurag Jain**

Bhopal, Madhya Pradesh, India

**Abstract:** ***Higher education institutions are hub of research and future development acting in a competitive scenario, with the basic goal to generate, gather and share knowledge. In this research work my objective is to explore data mining technique like classification, clustering on higher education data, my objective is to integrate genetic algorithm (GA) andsupport vector machines algorithm(SMO),for integration of two classifiers we use ensemble stacking, which is a fusion of classifiers. We present a generalize and powerful hybrid methodology of spectral clustering which originally operates on SMO and genetic algorithm classifiers, and further develop algorithms for classification on the basis of minimum attribute selecting and normalization of dataset. It could be concluded that the proposed GA-SMO classifier approach improves the classification accuracy and gives the better results, than other methods.***

**Keywords:** SMO, GA, Classification, Ensemble, Feature Extraction

## 1. Introduction

Educational Data mining (EDM) is an recent developing area, having growing method for studying the special types of data that come from educational system and using those methods to better understand students performance [1]. The classification of higher education data has become an increasingly challenging problem; many institutions do not have sufficient information to give guidance to students, therefore they are not able to give suitable advice to the students.We also observe that there is no perfect grouping of courses to recognize which type of course is most suitable to be offered to which type of student and Classification of this largest amount of data is time consuming and take excessive computational effort, which may not be for predicting the academics performance of students. For this, we develop an approach to pre-processing reducing the size of the training dataset, by removing noise points, outliers and insignificant points, which are not important for classification, Then we classifying the data by Sequential Minimal Optimization (SMO) algorithm is applied on the reduced dataset for optimize the support vector machine parameters, and optimize the results by genetic algorithm (GA). After we compare our work with the traditional SMO technique to show its improvement in terms of classification efficiency and other measure.

## 2. Methodology

### 2.1 Data Pre-Processing

Real data is often incomplete, inconsistent, and lacking in certain behaviours and is sometimes contain many errors. Data pre-processing prepares raw data for further processing [2]. Transformation of data includes dimensional reduction techniques like feature selection and feature extraction. Feature Selection is the method of finding the "best" subset of features from the initial 'N' features in the datasets; this reduces the dimensionality of feature sets, removes redundant, irrelevant data. It brings a speeding up a data mining algorithm, improving the data quality [3]. Feature Extraction defines a transformation from pattern space to feature space such that the new feature set gives both better separation of pattern classes and reduces dimensionality of datasets. Thus feature extraction is a kind of feature selection, it is a superset of feature selection; feature selection is a special case of feature extraction [5].

### 2.2 SMO (sequential minimal optimization)

A support vector machine (SVM), first introduced by Vapnik in 1995. SVMs are a set of supervised learning methods used for classification, regression and outliers detection in both linear and nonlinear data [10].For training SVMs we have three basic algorithms: Chunking, Osuna's algorithm, and SMO [7].

J. C. Platt proposes an algorithm for training support vector machines: Sequential Minimal Optimization (SMO). Training a SVM requires the solution of a very large quadratic programming (QP) optimization problem, SMO breaks large QP problem into a series of smallest possible QP problems, these small QP problems are solved systematically, that avoids using a time-consuming numerical QP optimization as an inner loop. The memory requirement for SMO is linear in the training setsize [7].This is simple, easy to implement, faster, and has better rising properties for difficult SVM problems than the standard SVM training algorithm.

Optimization problems are rapidly solved using SMO algorithm. Consider a binary classification problem with a dataset $(x_1, y_1)$... $(x_n, y_n)$, where $x_i$is an input vector and $y_i$, belongs to {-1, +1}, is itscorresponding binary label. Support vector machine helps solve the binary form of quadratic programming problem as follows:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j K(x_i, x_j) \alpha_i \alpha_j,$$

Subject to'