Comparative Analysis of Features Extraction Strategies for Classification in Educational Data Mining

P. Kavipriya, K. Karthikeyan

Abstract--- In this era of automation, education has also smartened up itself and is not imperfect to old lecture method. The expected search is on to find out innovative ways to build it more useful and proficient in developing students' performance. Currently, huge of data are gathered in educational databases, other than it remains unutilized. In order to obtain essential benefits from such big data, strong tools are required. Data mining is alarmed with the development of methods and techniques for making use of data analysis and prediction. This is the course of pattern detection as well as extraction where the vast amount of data is concerned. Both the data mining and education industry have emerged some of the reliable systems. In regard to this emerge; we present an approach for handling feature extraction by utilizing data mining algorithms for educational system and systematically evaluate the performance of the algorithms to find best-fit features for the further classification process. Results are discussed for selected papers and a summary of the finding is presented to conclude the paper.

Keywords--- Feature Extraction, Genetic Algorithm, Feature Selection, Classifiers, Support Vector Machine

I. INTRODUCTION

Feature extraction is one of the techniques of data mining field. With the growing competition among educational institutes, it is advantageous to extract some intelligent information from educational data. Educational data mining is a field that exploits statistical, machine-learning, and data-mining algorithms over the different types of educational data. Its majorpurpose is to examine these kinds of data in order to determine educational research problems. The increase in both instrumental educational software and state databases of student's information has created large repositories of data reflecting how students learn. Conversely, the use of the Internet in education has generated aninnovative context recognized as e-learning or else web-based education in thathuge amounts of data about teaching-learning interaction are endlessly generated and universally available [1].

The Data Mining in education converts raw data coming from educational systems into useful information that could potentially have a great impact on educational research and practice and useful to management in a certain decision-making process. The extraction of information is being done from an enormous number of databases. The database is huge in size. It is also growing at an extraordinary rate. The decision needs to be done rapidly and the decisions must be made with maximum knowledge. This leads to the complexity of time and space. A proper feature extraction mechanism is required to properly extract the data [2]. The principal motivation behind dimensionality reduction by feature extraction is as follows,

- Simplification of the data models to make easier interpretation
- Reduction in the training time
- Reducing over fitting

In order to prepare the data for further classification process, we implement several typical methods of feature extraction for educational data. This paper shows the comparison of data feature extraction and classification algorithms and we also compare their characteristics and difference.

II. FEATURE EXTRACTION

The system is utilized to develop the classification performance after the process of feature extraction for the decrease of irrelevant and redundant data which follows the principle of natural selection called Evolutionary Algorithms (EA). The Evolutionary Algorithms are based on the alike philosophy in the biological sciences involves natural selection along with the survival of the fittest [3]. This research paper can be divided into three main areas such as Feature Extraction, Feature Selection and classification methods and compare their experimental results to calculate the better performance.

Ant Colony Optimization Algorithm

Ant Colony Optimization (ACO) is a population-based advance method for solving combinatorial optimization problems which are inspired by the foraging activities of ants also their inherent capability to determine the shortest path from a food source to their nest [4].

Researchers in the ant colony optimization technique has been developing rapidly equally in academia and industry. The ant colony algorithm can be work constantly and adjust to changes in real time. An ACO algorithm is fundamentally a system based on agents that imitate the natural behavior of ants, including mechanisms of cooperation and adaptation. It is a probabilistic technique for solving computational problems to reduce the search path to find the best path through graphs [5]. The activities of ants naturally,

1. The ants are around randomly as well as left a trail of pheromones
2. The ants back to the nest and also left a trail of pheromones after finding a food source. Other ants

Manuscript received February 01, 2019

P. Kavipriya, Asst Professor, Dept of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, Tamil Nadu, India..

Dr.K. Karthikeyan, Head, Department of Computer Science, Government Arts College, Palladam, Tamil Nadu, India.

allow following the trail of pheromones when pheromones were found

3. After the ants back to the nest, the ants come out again to look for food. The pheromones will evaporate in long-term conditions. It is possible the ant lose the trail of pheromones, except it is reinforced by the trail of a large number of ants [7]. The following figure illustrates the process of ACO feature selection,

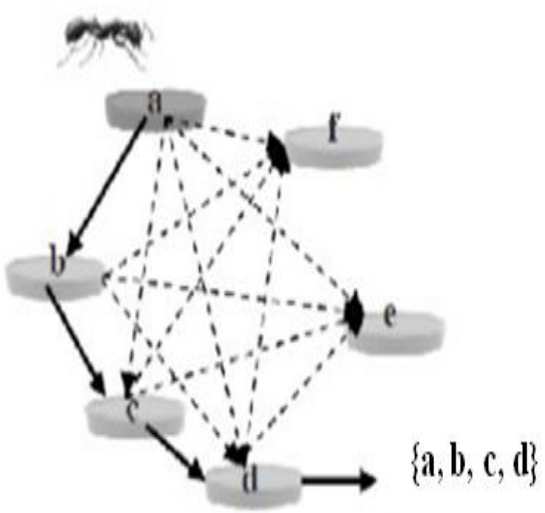


Fig. 2: Graph representation of ACO for Feature selection

The main abilities of the ants are determining the shortest route when they left a trail of pheromones during the journey. Every ant has a maximum value probabilistic to follow the direction of the trail of pheromones. In other situations, if the ants discover difficulty in the shortest path suddenly, the ACO system will respond and find new solutions.

Genetic Algorithm (GA)

The genetic algorithm works based on the principle of natural selection and biological evolution process. It is a random optimization method that changes the solution of problems in data individuals of a gene string structure in genetic space by a certain coding scheme, converts the objective function into fitness value and evaluates the advantages and disadvantages of individuals and as the basis for the genetic operation [8].

The genetic algorithm uses a structured population model in that every genome's reproductive partner is selected from within its local neighborhood.

It aids prolong diversity within the population also encourages local niching, which liable to result in an investigation of several areas of the search space and matches well with the goal of finding numerous patterns in a single run [9]. The Genetic algorithm has the following phases,

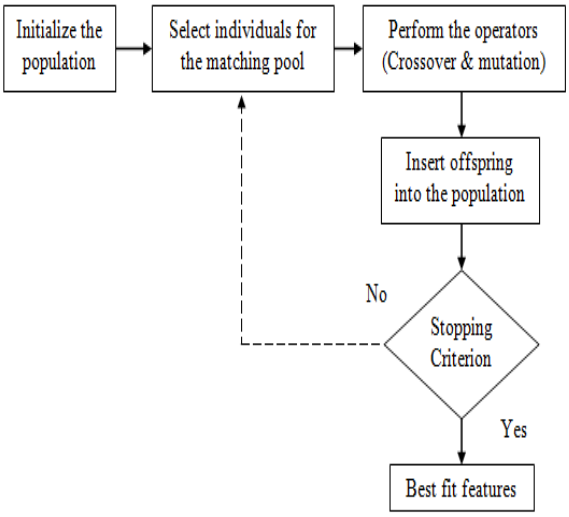


Fig. 2: Phases of GA for Feature selection

III. FEATURE SELECTION

Feature selection can be termed as the process of selecting a particular feature from a huge collection of features. All the features are residing within the data. So we need to select a particular feature from huge set of features that are residing within the dataset Feature selection [10].

The extraction of information can be done easily by feature selection mechanism. For performing a data mining operation we are dealing with redundant feature. These features may be irrelevant. This can lead to the increase in complexity. Hence reducing the number of irrelevant features plays an important role. Feature selection mechanism is done so as to eliminate the redundant features and extract the information. Feature should be selected in such a way that the:

i) The accuracy and performance should not be affected.
ii) Output must be the same.

A better feature selection mechanism helps in facilitating the data visualization, data understanding, reducing the storage requirements and utilization in time and in reducing the dimensionality [10]. Feature selection mechanism deals with four main steps:

- Feature generation
- Feature evaluation
- Stopping criterion
- Validation procedure

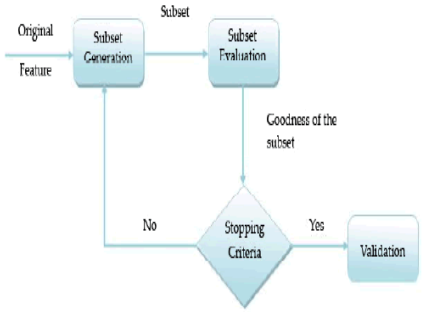


Fig. 3: Feature Selection Process

Feature sub set generation is the process of selection of subset of features from huge collection. Feature evaluation deals with the evaluation of the subset of features in such a way so as to fit the requirements. This evaluation can be done on the basis of dependent and independent measures. Dependence measure evaluates the feature subset by monitoring the performance of algorithm applied on it. Evaluation without the application of any learning algorithm is done in independent measures [12]. The different types of independent measures are:

- **Distance Measures**: Distance measures deals with the measuring of similarity or distance between two points. These points having smaller distance between them have similar features.
- **Information Measures**: Deals with the measurement of information that a selected feature provides. An infrequently occurring event provides more information than frequently occurring event.
- **Dependency Measure**: Dependency measure deals with measuring the dependence between two random variables.
- **Consistency Measure**: The concept of consistency measures was introduced to evaluate the distance of a given feature from consistency state.

Feature selection algorithms can be classified into three categories. They are filters, wrappers and embedded techniques. Filters work by extracting features from the data without any learning mechanism. Wrappers works by application of learning techniques. The embedded techniques deal with a combination of approaches. It combines the feature selection step and classifier construction.

Filters

Filters work without taking classifier into consideration. Filters can be classified into multi-variate and uni-variate. Multi-variate methods are able to find the relationship among features. Uni-variate methods consider each feature separately.

Wrappers

Wrappers considers model hypothesis into account by training and testing in the feature space. Wrappers perform better in selecting features. The dependencies of the features are considered by the wrapper method.

Embedded Methods

Embedded methods integrate the feature selection process into the model training process. They are usually faster than the wrapper methods and able to provide suitable feature subset for the learning algorithm [11].

The followings are the basic filter feature selection algorithms [12],

S.NO	Feature Selection Algorithm
1	χ^2 test
2	Euclidian distance
3	T-test
4	Information gain
5	CFS-correlation based feature selection method
6	MBF- Markov blanket filter
7	FCBF-fast correlation based feature selection

IV. CLASSIFIERS

A variety of data mining algorithms such as Decision Tree Classification, Bayesian Classification, Neural Networks Classification, Clustering, Association Rule Mining, Prediction, Time Series Analysis, Sequential Pattern and Genetic Algorithm and Nearest Neighbor have been used for large educational data sets. A number of the general as well as useful data mining algorithms have been discussed in this section.

Decision tree Classifier

Decision tree in data mining is one of the simplest and easiest methods which are most frequently used by the researchers on their work. The root node of the decision tree is a top node resembles simple question also called as a posture that bears multiple branches called sub nodes with answers for the root node question. In turn each answer related to a set of questions or conditions that help us to predict the data, on which the final decision is made. ID3 and C4.5 are called as induction algorithm of decision tree developed by the researcher called Ross Quinlan. Both algorithm supports greedy method, top-down recursive in divide-and-conquer manner and they does not support backtracking.C4.5 is also known as superset of ID3.The advantages of this technique are, it doesn't require detailed knowledge, it deals with complex data, these are easy to understand, and data Classification becomes simpler, makes learning easier, it produces very accurate end result [13].

Naive Bayes Algorithm

An easy probabilistic classifier which depends on applying Bayes theorem with strong (naive) independence hypothesis is the Naive Bayes classifier. In simple terms, a Naive Bayes classifier adopts a particular feature in the presence (or absence) of a class that is unrelated to any other feature. For example, a vegetable may be considered to be a carrot if it is orange in color, it shaped as in cone, and about 10 to 15 centimeter long in length. If these features match on all other or else match upon the survival of the other features, a Naive Bayes classifier then deems all these properties autonomous part to the probability which this vegetable is a carrot. Based on the exact nature of the probability model, the data set is then trained very effectively by a Naive Bayes classifier in a supervised learning setting. In most of the practical applications, the parameter estimation for Naive Bayes models uses the method of maximum likelihood; in other words, one can also work with the Naïve Bayes model without considering the Bayesian probability or by using any Bayesian methods. In spite of their over-simplified assumptions of a Naive design, the Bayes classifiers will work effectively in solving many complex real-world circumstances.

There are numerous benefits of Naive Bayes classifier for instance; it requires a small amount of training data set to estimate the essential for classification.

Since, the independent variables are assumed only to the variances of the variables of every class that is required to be detected also not the entire covariance matrix, irrelevant features are removed to improves the classification performance, results High Performance and take very less computational time [13].

Neural Networks

A neural network is a pool of neurons like processing units with weighted connection between the units. This invents of numerous elements, called nodes that are connected in between. The connection among two nodes is weighted as well as by the adjustment of this weight; the training of the network is performed. There are lots of benefits of neural networks like adaptive learning ability, self-organization, real time operation and insensitivity to noise. Neural networks are utilized to recognizing patterns or else trends in data as well as well suited for prediction. There are several neural network algorithms such as Back Propagation, NN Supervised Learning, and Radial Base Function (RBF) Network etc [13].

DT-J48

The J48 Decision tree classifier is an implementation of ID3 classifier used to create a decision tree based on the attribute values of the available training data set. So, whenever it encounters a set of items i.e. training set it identifies the attribute that discriminates the various instances most visibly. This feature tells us about the data instances so that we can classify them clearly and the best is said to have the highest information gain. At present, along with the possible values of this feature, if there is some value for which there is no uncertainty, that is, for that the data instances falling within its group have the same value for the target variable, then we terminate that branch and assign to it the target value that we have obtained [13].

Support Vector Machine

A powerful Support Vector Machine (SVM) which was first proposed by Vapnik and it has a great potency of interest in the machine learning research community. Several past studies have reported that the SVM generally has a proficient of delivering the high accuracy in classification when compared to other data classification algorithms. Though, for certain datasets, the achievement of SVM is very subtle in determining the cost parameter and kernel parameters. As in the case of closure, to figure out the most encouraging condition environment the user normally needs to conduct extensive cross validation. Basically this technique is baited to as a model selection. A superior asset of this SVM technique is that, concurrent miniaturize the projected classification error and make best use of the geometric margin, So SVM is also named as paramount Margin Classifiers. It is found on the Structural Risk Minimization (SRM), SVM can be used for both classification and prediction.

There are several advantages of SVM such as it uses maximum marginal hyper plane for classifying linearly separable data, Data can be separated clearly into rations, extends by itself in order to classify the linearly inseparable data. In SVM, given labeled training data (supervised learning), the algorithm outputs an optimal hyper plane which categorizes new examples [14].

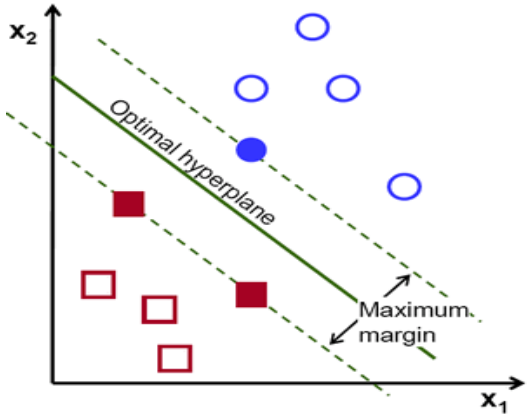


Fig 4: SVM Classifier

The issues that are to be faced in SVM is the hyper plane constructed by SVM is dependent on only a fraction of training samples called Support Vectors (SVs) which recline close to the result border as well as by Removing some training samples that are not applicable to support vectors may possibly to have no effect on the building the appropriate decision function.

V. RESEARCH METHODOLOGY

In the education field, data mining has huge transformational potential and it discovers how people learn, predicting learning and understanding real learning behavior. By obtaining these goals, educational data mining can be used to design better and smarter learning technology and to better inform learners and educators. This paper focused on improving student performance prediction, based on their academic performance characteristics.

The main aim of the research is to evaluate the performance of different feature extraction algorithms on classification algorithms for student educational dataset. To achieve the objective of the research, a student dataset is taken from valid sources, and then different feature extraction and classification algorithms are applied on it. The main steps are described in following section,

Data set Description and Experimental setup

The dataset used in this research is taken from the source https://archive.ics.uci.edu/ml/datasets/Student+Performance, and is comprised of 395 students 33 features. In this paper, the main aim of using the dataset is to identify the best combinations of feature extraction algorithms and classifiers, in order to identify the key performance factors on the academic achievements of students. Waikato Environment for Knowledge Analysis (WEKA) is used as a data mining tool. It has a rich source of Machine learning algorithms [15] [16].

Feature Extraction and Selection Algorithms for Classifiers

In this research work two feature extraction algorithms Ant colony optimization (ACO), genetic algorithm (GA), are evaluated.

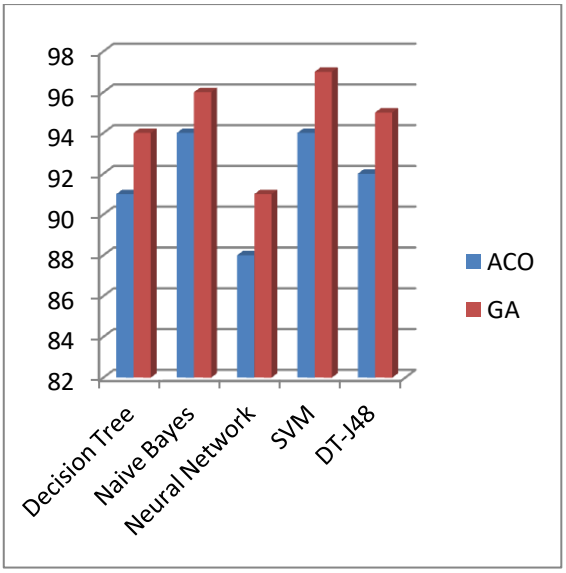
The classification algorithms Decision tree, Naive Bayes Algorithm, Support Vector Machine, DT-J48 are evaluated through the educational data set.

As for classification rule mining, classification algorithm selection is another key issue after feature selection. In addition, the classification problem becomes very hard when the number of possible different combinations of parameters is so high that algorithms based on exhaustive searches of the parameter space rapidly become computationally infeasible. Therefore, we adopted the recently proposed GA and ACO [17]as the classification algorithm since it is feature selection algorithm based on the Darwinian principle of the survival of the fittest and the natural process of evolution through reproduction. Moreover, as a kind of globally optimal algorithm, Feature selection algorithm can apply heuristic approach to find a good-enough solution to the classification problem

VI. RESULTS AND DISCUSSIONS

The investigation in the paper focuses on the performance of the few of feature extraction algorithm along with the classifiers. The results of two feature extraction techniques applied on five classifiers are given in following table.

Feature Selection	Classification Methods (Accuracy Rate in %)				
	Decision Tree	Naive Bayes	Neural Network	SVM	DT-J48
ACO	91%	94%	88%	94%	92%
GA	94%	96%	91%	97%	95%



The above mentioned figure showed some of the actual accuracy rate predicted in previous our research by using some data mining techniques [14]. As we have seen, predicting students' performance earlier is a difficult task because it is a comprehensive problem and because the available data are normally imbalanced. To resolve this problem and to improve the accuracy and quality, the Support Vector Machine algorithm can be used which is showing the greatest accuracy among other techniques.

At present, many researchers are working on the combination of ensemble algorithm to enhance the performance of classification. The main idea of the ensemble algorithm is to combine the results of base algorithm to improve the classification accuracy better than individual models. The generalization ability of an ensemble is usually to improve the classification accuracy than single algorithm. Base classifiers are also called as a weak learner because ensemble algorithm is also applicable to boost the weak learner to give much accurate predictions.

VII. CONCLSION

This paper summarizes various available Feature Selection Methods. There is still lots of work is about to develop to handle the Educational Data Set. All the methods are efficient over the specific domain analysis. Broad classes of Data mining algorithms have been developed and applied successfully to a wide range of real-world domains. In the ensemble learning algorithm, the combination of algorithms plays an important role. The main aim of the ensemble algorithm is to fuse the outputs of several algorithms. A suitable combination algorithm lead to improved accuracy of the classification decision compared to a decision based on any of the individual data sources. In future we will identify the best ensemble method which improves performance of classification on educational data.

REFERENCES

1. Mamta Sharma, MonaliMavani "Accuracy Comparison of Predictive Algorithms of Data Mining: Application in Education Sector" ICAC3 2011, CCIS 125, pp. 189–194, 2011. © Springer-Verlag Berlin Heidelberg 2011.
2. Aparna.U.R, Shaiju Paul "Feature Selection and Extraction in Data mining" 978-1-5090-4556-3, 2016 Online International Conference on Green Engineering and Technologies (IC-GET), IEEE.
3. Sewell, Martin, 2007, "Feature Selection" http://machine-learning.martinsewell. Com/feature-selection/ (viewed January 15, 2012).
4. M. Dorigo and G. Di Caro, "The ant colony optimization meta-heuristic," in New Ideas in Optimization, D. Corne, M. Dorigo, and F. Glover, Eds. New York: McGraw-Hill, 1999, pp. 11–32.
5. Dorigo M., Stützlet T, "Ant Colony Optimization, MIT Press", ISBN 0-262-04219-3, 2004.
6. Ant Colony Optimization (ACO) Submitted by: Subvesh Raichand MTech, 3rd Semester 07204007 Electronics and Communication optimization-23928078.
7. G. Y. Yu, Y. Z. Wang, "Applied Research of improved genetic algorithms," Machinery, vol. 5, 2007, pp. 58-60.
8. M. Pei, E.D. Goodman, and W.F. Punch. Feature extraction using genetic algorithms. Technical report, Michigan State University: GARAGe, June 1997.
9. J. Yang and V. Honoavar. Feature Extraction, Construction and Selection: A data Mining Perspective, chapter 1: Feature Subset Selection Using a Genetic Algorithm, pages 117–136. H. Liu and H. MotodaEds, massachusetts: kluwer academic publishers edition, 1998.
10. Maryam Zaffar, Manzoor Ahmed Hashmani, K.S.SAVITA "Performance Analysis of Feature Selection Algorithm for Educational Data Mining" 2017 IEEE Conference on Big Data and Analytics (ICBDA).
11. K. Patel, J. Vala, and J. Pandya, "Comparison of various classification algorithms on iris datasets using WEKA," Int. J. Adv. Eng. Res. Dev.(IJAERD), vol. 1, 2014.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

12. Aparna.U.R, Shaiju Paul "Feature Selection and Extraction in Data mining" 2016 Online International Conference on Green Engineering and Technologies,978-1-5090-4556-3/16/$31.00, IEEE
13. P.Kavipriya, Dr.K.Karthikeyan "A Comparative Study of Feature Selection Algorithms in Data Mining" International Journal of Advanced Research in Computer and Communication Engineering, ISO 3297:2007, Vol. 6, Issue 11, November 2017.
14. P.Kavipriya, Dr.Karthikeyan." Case Study: On Improving Student Performance Prediction in Education Systems using Enhanced Data Mining Techniques. "- International Journal of Advanced Research in Computer Science and Software Engineering,- Volume 7, Issue 5, May 2017
15. K. Patel, J. Vala, and J. Pandya, "Comparison of various classification algorithms on iris datasets using WEKA," Int. J. Adv. Eng. Res. Dev. (IJAERD), vol. 1, 2014.
16. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," ACM SIGKDD explorations newsletter, vol. 11, pp. 10-18, 2009.
17. P.Kavipriya, Dr.Karthikeyan ,"A Hybridization of Enhanced Ant Colony Optimization (ACO) and Genetic Algorithm (GA) for Feature Extraction in Educational Data Mining," Journal of Advanced Research in Dynamical & Control Systems, Vol. 10, 12-Special Issue, Pages: 1278-1284,2018.