ISSN 2320-9984
Volume 4 Issue 1 | March 2016 ISSN: 2320-9984 (Online)
International Journal of Modern Engineering & Management Research
Website: www.ijmemr.org

Performance Evaluation of Meta Heuristic Based Feature Selection on Various Learner Algorithm

Neha Asati
Research Scholar
Lakshmi Narain College of Technology
Jabalpur (M.P.), [INDIA]
Email: 31asatineha@gmail.com

Sujeet Kumar Tiwari
Assistant Professor
Department of Computer Science and Engineering
Lakshmi Narain College of Technology
Jabalpur (M.P.), [INDIA]
Email: sujeet.tiwari08@gmail.com

Abstract**—***Feature selection has been a long research area within statistics and pattern detection. It is not surprising that feature selection is as lot of an issue for machine learning as it is for pattern detection, as mutually fields share the common task of categorization. There were lots of issues founded while the research work was going on, which included Memory Management, Compact Data Structure, Multilingual text refining and Domain knowledge integration. The Genetic and PSO Algorithm is used to reduce the features in the selected data. For the Classification process the decision tree classifier has been used as well as Crossover and mutation is applied. The Selected data of PSO which have been taken out is efficient. The selected data is very close to original data. The use of Neural Network has been replaced by the use of Decision tree classifier, which improved the efficiency of the work, along with that PSO is used to apply Crossover. The use of clusters has been avoided to improve the results. The other optimization algorithms can be used to improve the accuracy of the data like learning algorithm. As a Future work, clustering can be done and validity can be measured and tested on the basis of the selected data.*

Keywords:**—***Feature selection, Memory Management, Compact Data Structure, Multilingual text refining and Domain knowledge integration.*

1. INTRODUCTION

Feature selection has been a long research area within statistics and pattern detection. It is not surprising that feature selection is as lot of an issue for machine learning as it is for pattern detection, as mutually fields share the common task of categorization. In pattern detection, feature selection can have an impact on the economics of statistics acquisition and on the accurateness and complication of the classifier. This is also factual of machine learning, which has the additional concern of distilling helpful knowledge from data. Fortunately, feature selection has been exposed to develop the comprehensibility of extracted knowledge.

Unsupervised Methods

The simplest discretization scheme is called equal interval width. This approach divides the variety of observed values for a feature into k equal sized bins, where k is a limitation provided by the user. Dougherty point out that this scheme of discretization is sensitive to outliers that might drastically skew the range. For instance, given the observed feature values

Supervised Methods

Holte presents an easy supervised discretization process that is included in his one-level decision tree algorithm (1R). The

process first sorts the values of a feature, and then attempt to discover interval boundaries such that each interval has a strong majority of one fussy class. The process is constrained to form intervals of some minimal size in order to keep away from having intervals with very few instances. Setiono and Liu present a statistically justified heuristic system for supervised discretization called Chi2. A numeric feature is firstly sorted by placing each observed value into its own interval. The next steps use a chi-square statistic X^2 to establish whether the virtual frequencies of the classes in adjacent intervals are alike enough to justify merging.

PSO Algorithm

A basic variant of the PSO algorithm works by having a population (called a swarm) of candidate solutions (called particles). These particles are moved around in the search-space according to a few simple formulae. The movements of the particles are guided by their own best known position in the search-space as well as the entire swarm's best known position. Formally, let f: $\mathbb{R}n \rightarrow \mathbb{R}$ be the cost function which must be minimized. The function takes a candidate solution as argument in the form of a vector of real numbers and produces a real number as output which indicates the objective function value of the given candidate solution. The gradient of f is not known. The goal is to find a solution a for which $f(a) \leq f(b)$ for all b in the search-space, which would mean a is the global minimum. Numerous variants of even a basic PSO algorithm are possible. For example, there are different ways to initialize the particles and velocities (e.g. start with zero velocities instead), how to dampen the velocity; only update pi and g after the entire swarm has been updated, etc.

Decision tree classifier

The classification technique is a systematic approach to build classification models from an input data set. For example, decision tree classifiers, rule-based classifiers, neural networks, support vector machines, and naive Baye's classifiers are different technique to solve a classification problem. Each technique adopts a learning algorithm to identify a model that best fits the relationship between the attribute set and class label of the input data. Therefore, a key objective of the learning algorithm is to build predictive model that accurately predict the class labels of previously unknown records. Decision Tree Classifier is a simple and widely used classification technique. It applies a straight forward idea to solve the classification problem. Decision Tree Classifier poses a series of carefully crafted questions about the attributes of the test record. Each time it receives an answer, a follow-up question is asked until a conclusion about the class label of the record is reached.

2. RELATED WORK

Waleed H. Abdulla and Nikola Kasabov (2003) have worked on "Reduced feature-set based parallel CHMM speech recognition systems" and designed multi-streams paradigm where they split feature vectors in three independent continuous-density Hidden Markov Model (CHMM) frameworks. They proposed a technique that combines classifiers. Here the three HMM classifiers were applied to speech signals. HMM classifiers had done feature reduction by alleviating the dominance effects of the features and in this way, they reduced the dimensionality of feature vectors.

Alper Unler et al. (2010) have worked on "mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification" and presented a hybrid filter–wrapper feature subset selection algorithm. The filter model was used for feature subset selection and the wrapper model was liable to use the mutual information available from the filter model. They presented a novel feature selection method, which lessen computational cost dramatically. The whole hybrid model performed feature selection and reduction.

Junbo Zhang et al. (2010) have worked on "Composite rough sets for dynamic data mining" and redefined composite rough sets

for feature selection which was a powerful mathematical tool for analyzing various types of data. Zhang et al. proposed an incremental method for dynamic data mining based on neighbourhood rough sets. They accomplished a notable progress after twenty five years of the beginning of the research on feature selection using rough sets. Through rough sets they defined composite information systems that contained attributes of multiple different types, which was liable for feature selection and knowledge discovery.

Harun Uguz (2011) has worked on "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm" and here he did feature selection in two stages and all were filter methods. In the first stage, each term within the document was ranked depending on their importance for classification using the information gain (IG) method. In the next stage, genetic algorithm (GA) and principal component analysis (PCA) feature selection and feature extraction methods were applied separately to the terms which were ranked in decreasing order of importance, and a dimension reduction was carried out. Genetic algorithm was an optimization method mimicking the evolution mechanism of natural selection. GA performs a search in complex and large landscapes and provides near-optimal solutions for optimization problems.

Joaquin Pacheco et al. (2013) have worked on "Bi-objective feature selection for discriminant analysis in two-class classification" and assessed the relevance of formulating the feature selection problem for classification to check and compare the efficacy with the method NSGAFS (non-dominated sorting genetic algorithm). A series Set of all of experiments was run with different databases. There they worked with financial variables and two classes: ‘‘credit-worthy’’ and ‘‘non-creditworthy’’

2. PROBLEM IDENTIFICATION

PCA is partial to re-expressing the data as amalgamation of its basis vectors. A main disadvantage of PCA is that each PC is a linear grouping of all the original variables, therefore leading to a potentially tricky interpretation of the PCs. On the opposing, in a system with a lot of variables PCA may be used to focus the dimension down to a sensible amount of plots, and the major components could be rotated towards a more significant representation. Furthermore, PCA is sensitive with respect to the units of dimension. If the units and the variances of attributes show a discrepancy, then the variables with high variance tend to govern the first few principal components. In this, the data need to be normalized earlier to the PCA transformation.

Memory Management

The next fundamental issue we needed to consider as how to optimize the memory space frenzied when running into the mining algorithm. This includes the way how to make a decision about the information i.e. we must also collect the data from the data streams and how to select a compact in-memory to the data structure that allows to the information to be stored, retrieved and updated efficiently. Fully addressing these challenges for mining algorithm can greatly enhance its performance.

Compact Data Structure

An efficient and compact the data structure are needed to be store, update and retrieving to the collected information. This is due to the bounded memory size and huge amounts of the data streams are coming continuously. Failure in to the developing such as the data structure will largely decrease the efficiency of the mining algorithm because, even if we store the information in to the disks, the additional I/O operations will increase in to the processing time. The data structure needs to be incrementally maintained since it is not possible to rescan entire whole input due to the huge amount of the data and requirement of the rapid online querying speed. A small and

compact of the data structure which is efficient in inserting, retrieving and updating information is most favourable when developing an algorithm to mine association rules for the stream data.

Multilingual text refining

Whereas data mining is principally language self-regulating, text mining involves a important language component. It is indispensable to develop text humanizing algorithms that procedure multilingual text documents and create language-independent in -between forms. While for the most part text mining tools focus on dispensation English documents, mining from documents in former languages allows access to beforehand untapped information and offers a novel host of opportunities.

3. METHODOLOGY

The methodology which has been proposed for the solution of the problems identified in the project is as shown in the Figure 1.

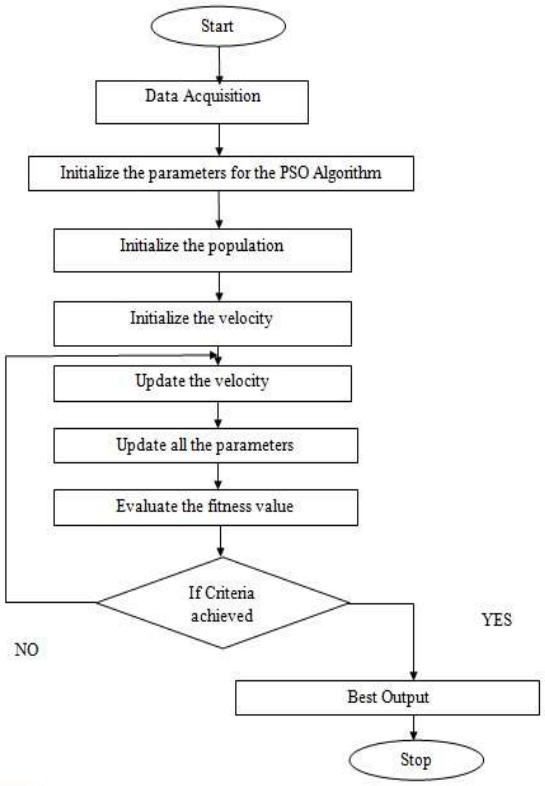

Figure 1 – Flowchart of the methodology by PSO

Step 1:- Data Acquisition

Choose the data from any document and acquire it in a meaningful manner.

Extraction of the data from the various datasets is very significant and it is connected to extraction of error free data.

Step 2:- Initialize the parameters for the PSO Algorithm

Set all the parameters and initialize it.

The parameters have to be used in the PSO Algorithm for further processing.

Step 3:- Initialize the population

Initialize the size of the population.

It includes the declaration of the capacity of the data or information.

Step 4:- Initialize the velocity

Now, initialize the velocity of the datasets.

Step 5: Update the velocity

Update the velocity of the dataset by using the formulae –

$V_{id} = w * v_{id} + c_1r_1 (P_{best} - x_{id}) + c_2r_2 (c_{1best} - x_{id})$

Where V_{id} = calculated velocity

c_1 = weighing coefficient

c_2 = weighing coefficient

Step 6:- Update all the parameters

Update each and every parameter, i.e. X_{id}

$X_{id} = Xid + V_{id}$

Step 7:- Evaluate the fitness value

Evaluate the fitness value through decision tree classifier.

Decision Tree Classifier is a easy and extensively used classification method. It applies a straight forward thought to solve the classification problem.

Decision Tree Classifier has a series of carefully created enquiries about the attributes of the test record.

Each time it obtains an answer, a follow-up question is asked in anticipation of a conclusion about the class label of the documentation is reached

PSO ALGORITHM

initialize the swarm;

while (mission on)

get sensor readings;

if (no object is found) then

move in spiral;

else

evaluate the target position;

report the target position to the host;

set pso on;

while (pso on)

get sensor readings;

update local best;

update global best;

report to host the current position and

get global best;

move to the next best position;

avoid obstacle if present;

if (target reached) set pso off;

end (while pso off)

end if

if (mission completed) set mission off;

end

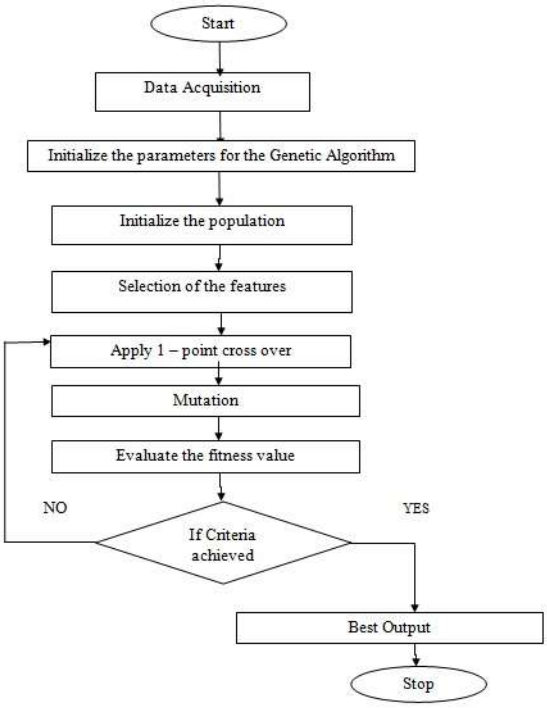

Figure 3 – Flowchart of the methodology by GA

Step 1:- Data Acquisition

Choose the data from any document and acquire it in a meaningful manner.

Extraction of the data from the various datasets is very significant and it is connected to extraction of error free data.

Step 2:- Initialize the parameters for the Genetic Algorithm

Set all the parameters and initialise it.

The parameters have to be used in the Genetic Algorithm for further processing.

Initialize the parameters of genetic algorithm

Parameter Scaling includes the process of defining the characteristics i.e. how many chromosomes should be taken into consideration, how many times it should be illustrated,

and the number of genes which are to be taken.

Step 3:- Initialize the population

Initialize the size of the population.

It includes the declaration of the capacity of the data or information.

Step 4:- Selection of the features in the Genetic Algorithm

A genetic algorithm (GA) is a method for solving both constrained and unconstrained optimization problems based on a natural selection process.

The algorithm repeatedly modifies a population of individual solutions to get best result.

It generates solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection and crossover. The features from the data have to be selected and processed with the help of Genetic Algorithm in order to select the desired meaningful words.

Step 5:Have Cross over with the help of 1 – ***point*** crossover.

In genetic algorithms, crossover is a genetic operator used to vary the programming of a chromosome

It Produces Chromosomes from one generation to the next.

1 Point crossover and 2 point crossover are used in this process, but here we are using 1 – point crossover.

Step 6:- Mutation

Apply mutation on the chromosomes and match the criteria, if yes terminate the loop else continue the same procedure.

The most common form of mutation is to take a bit from a chromosome and alter (i.e., flip) it with some predetermined probability. As mutation rates are very small in natural evolution, the probability with which the mutation operator is applied is set to a very low value and is generally experimented with before this value is fixed.

Step 7:- Evaluate the fitness value

Evaluate the fitness value through decision tree classifier.

Decision Tree Classifier is a easy and extensively used classification method. It applies a straight forward thought to solve the classification problem.

Decision Tree Classifier has a series of carefully created enquiries about the attributes of the test record.

Each time it obtains an answer, a follow-up question is asked in anticipation of a conclusion about the class label of the documentation is reached.

GENETIC ALGORITHM

begin

initialize population with random candidate solution;

evaluate each candidate;

repeat until (termination condition) is satisfied do

select parents;

recombine pair of parents;

mutate the resultant offspring;

select individuals or the next generation;

end;

4. RESULTS

Based on the methodology discussed, we did got some extreme good and improved results while comparing with the previous works, which have been shown with the help of table, screenshot and Graph.

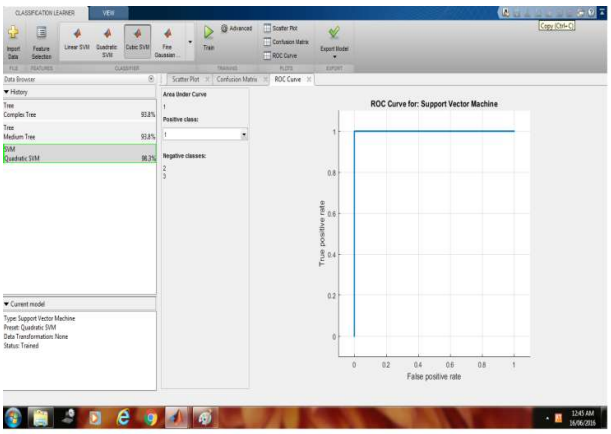

Figure 4 :- Use of ROC Curve for Support vector machine to show the result

Table 1:- Feature Selection

Class	Population Size	Dimension	Feature Selection	Classifier			
			Iteration	Ga		Pso	
				Mean	Std	Mean	Std
Dermi	12	35	3	0..97	0.13	0.97	0
Wine	10	14	3	0.71	0.47	0.94	0
Glass	8	10	3	0.71	0.47	0.94	0

5. CONCLUSION AND FUTURE SCOPE

From the Research work, we came into a conclusion that the Higher data accuracy has been achieved and it is successful to reduce the data so that the storage decreases. The Selected data of PSO which have been taken out is efficient. The selected data is very close to original data. The use of Neural Network has been replaced by the use of Decision tree classifier, which improved the efficiency of the work, along with that PSO is used to apply Crossover. The use of clusters has been avoided to improve the results.

The other optimization algorithms can be used to improve the accuracy of the data like learning algorithm. As a Future work, clustering can be done and validity can be measured and tested on the basis of the selected data.

REFERENCES:

[1] Waleed H. Abdulla, Nikola Kasabov, "Reduced feature-set based parallel CHMM speech recognition systems", Information Sciences, Vol. 156, 2003, pp. 21–38.

[2] Alper Unler, Alper Murat, Ratna Babu Chinnamb, "mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification", Information Sciences, Vol. 181, 2011, pp. 4625–4641.

[3] Junbo Zhang, Tianrui Lia, Hongmei Chen, "Composite rough sets for dynamic data mining," Information Sciences", Vol. 4, 2013, pp. 129-135. http://dx.doi.org/10.1016/j.ins.2013.08.01

[4] Harun Uguz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm" Knowledge-Based Systems, Vol. 24, 2011, pp. 1024–1032

[5] Joaquin Pacheco, Silvia Casado, Francisco Angel-Bello, Ada Álvarez, "Bi-objective feature selection for discriminant analysis in two-class classification", Knowledge-Based Systems, Vol. 44, 2013, pp. 57–64

[6] Ran Li, Jianjiang Lu, Yafei Zhang, Tianzhong Zhao, "Dynamic Adaboost learning with feature selection based on parallel genetic algorithm for image annotation", Knowledge-Based Systems, Vol. 23, 2010, pp. 195–201.

[7] Zhiming Zhang, "On interval type-2 rough fuzzy sets", Knowledge-Based

Systems, Vol. 35, 2012, pp. 1–13

[8] Dewan Md. Farid, and Chowdhury Mofizur Rahman, "Mining Complex Data Streams: Discretization, Attribute Selection and classification", Journal of Advances in Information Technology, Vol. 4, No. 3, August 2013, pp. 129-135.

[9] José M. Carmona-Cejudo, Gladys Castillo, Manuel Baena-García, Rafael Morales-Bueno, "A comparative study on feature selection and adaptive strategies for email foldering using the ABC-DynF framework", Knowledge-Based Systems, Vol. 46, 2013, pp. 81–94

[10] You-Shyang Chen, "Classifying credit ratings for Asian banks using integrating feature selection and the CPDA-based rough sets approach", Knowledge-Based Systems, Vol. 26, 2012, pp. 259–270

[11] Yitian Xu, Laisheng Wang, Ruiyan Zhang, "A Dynamic Attribute Reduction Algorithm Based on 0-1 Integer Programming", Knowledge-Based Systems, Vol. 24, 2011, pp. 1341–1347

[12] Hailiang Chen, Hongyan Liu, Jiawei Han, Xiaoxin Yin, Jun He, "Exploring Ptimization of Semantic Relationship Graph for Multi-relational Bayesian Classification", Decision Support Systems, Vol. 48, 2009, pp. 112–121.