MULTICLASS CLASSIFIER DESIGNED USING STEPWISE CROSSOVER AND SPECIAL MUTATION TECHNIQUE

Pankaj Patidar[1] Bhupendra Verma[2] Arpit Bhardwaj[3]

[1]Department of Computer Science Engineering, TIT, Bhopal
`pankaj_pp20@rediffmail.com`
[2]Department of Computer Science Engineering, TIT, Bhopal
`bk_verma3@rediffmail.com`
[3]Department of Computer Science Engineering, SDITS, Khandwa
`arpitbhardwaj17@gmail.com`

ABSTRACT

A Multiclass classifier is an approach for designing classifiers for a m-class (m>=2) problem using genetic programming (GP). In this paper we proposed three methods named Triple Tournament Method, Special Mutation Method and a Step Wise Crossover method. In Special Mutation technique we are generating the two child from single parent and selecting the one child on the basis of fitness and also applying the elitism on the child so that the mutation operation does not reduce the fitness of the individual and in Stepwise Crossover we select the two child for the next generation on the basis of size, depth and fitness along with elitism on each step from the six child which is generated during crossover.
To demonstrate our approach we have designed a Multiclass Classifier using GP by taking few benchmark datasets. The results obtained show that by applying Stepwise crossover together with Special Mutation improves the performance of the classifier. In Triple Tournament Method, we select the two individual for the crossover operation on the basis of size, depth and fitness

KEYWORD

Elitism,Triple Tournament, Special Mutation, Stepwise Crossover.

1. INTRODUCTION

Genetic programming (GP) is an automated method for creating a working computer program from a high-level problem statement of a problem. GP has already spawned numerous interesting applications such as [1]–[2]. GP has been used by many authors [3] to design classifiers or to generate rules for two class problems. Genetic programming has seldom been used for Multi-class classification purposes. Previously, it was achieved by separating the output manually for different classes or expanding an n-class problem to n two-class problems. Multiclass or multi-label classification is the special case within statistical classification of assigning one of several class labels to an input object. Genetic programming is a collection of methods for the automatic generation of computer programs that solve carefully specified problems, but highly abstracted principles of natural selection. It is the compounded breeding of (initially random) computer programs, where only the relatively more successful individuals pass on genetic material (programs and program fragments) to the next generation. Genetic

programming starts from a high-level statement of the requirements of a problem and attempts to produce a computer program that solves the problem. The most commonly employed method for selecting individuals in GP is tournament selection. An approach takes an integrated view of all classes when the GP evolves. A Multitree representation of chromosomes is used. GP begins by generating some initial population. But in our paper we have proposed a special type of tournament known as Triple tournament. In Triple tournament first we randomly selects 9 individual from the population and from those selected individual we select 5 on the basis of size and from those 5 individual we select three on the basis if depth and from those 3 we select the 2 individual on the basis of fitness.

In this we also propose a stepwise crossover technique and a special mutation operation that reduces the destructive nature of conventional genetic operations. Mutation affects an individual in the population. It can replace a whole node in the selected individual, or it can replace just the node's information. Mutation is an important part of the genetic search as help to prevent the population from stagnating at any local optima. Mutation operations often end up decreasing the fitness of an individual because the new material is untested. They tend to be used less than crossover; however, they are still important to a successful GP run. In special Mutation technique we generate the two individual from the selected parent and from the two generated individual we reject one individual on the basis of fitness and the remaining individual is now compared with the parent if the fitness of the parent is better than the selected individual than with a probability of 0.5 we transfer the parent to the next generation. Stepwise crossover enables the algorithm to extract the best genes from different individuals and recombine them into potentially superior children. An individual which has high fitness rate and low size and depth must be inherited.

2. PREVIOUS WORK

Chaudhari,N.S., Purohit, A., Tiwari, A.[1] This paper presents an approach for designing classifiers for a multiclass problem using Genetic Programming (GP). The proposed approach takes an integrated view of all classes when GP evolves. An individual of the population will be represented using multiple trees. The GP is trained with a set of N training samples in steps. A concept of unfitness of a tree is used in order to improve genetic evolution. Weak trees having poor performance are given more chance to participate in the genetic operations, and thus improve themselves. In this context, a new mutation operation called non destructive directed point mutation is used, which reduces the destructive nature of mutation operation. The approach is being demonstrated by experimenting on some datasets.

D J Nagendra Kumar, Suresh Chandra Satapathy, J V R Murthy [2] In this paper we propose a simple scalable genetic programming multi-class ensemble classifier of higher accuracy. A formula is derived to obtain the maximum number of nodes permitted in a GP classifier. A wrapper approach for feature selection mechanism based on GP classifier is adopted in our work.

Mengjie Zhang and Will Smart [3] describes two different methods, an online gradient descent scheme and an offline gradient descent scheme, are developed and compared with the basic GP method on three image data sets with object classification problems of increasing difficulty. The results show that both the online and the offline gradient descent GP methods outperform the basic GP method in both classification accuracy and training time and that the online scheme achieved better performance than the offline scheme.

3. PROPOSED WORK

In this paper we have designed a Multiclass Classifier using Step Wise Crossover and Special Mutation technique.

3.1. INITIALIZATION

In genetic programming first we have to generate initial population for performing operations in it. The terminal and function sets are important components for generating initial population.
Each of the trees for each individual is initialized randomly using the function set F which consists of arithmetic functions and the terminal set T containing feature variables and constants. The function set F and terminal set T used here are as follows:
F = {+, -, *, %} and
T = {feature variables, R}
Where R contains randomly generated constants in [0.0 to 10.0]. We have initialized trees using the ramped half-and-half method

3.2. FITNESS MEASURE

The fitness function determines how well a program is able to solve the problem. GP is guided by the fitness function to search for the most efficient computer program to solve a given problem.

Fitness = n/N

n= Samples classified correctly
N= Number of samples used for training during evolution

3.3. SELECTION

We select the parent for crossover on the basis of triple tournament selection method. For mutation the parent is selected randomly and for reproduction we use roulette wheel selection method. In special mutation we select one best child on the basis of fitness and reject another one. In stepwise crossover we select best child and remove remaining child step by step.

3.4. ALGORITHM FOR TRIPLE TOURNAMENT

1. First we randomly select 9 individual from the population.
2. Now on the basis of depth we select 5 out of 9 individual the individual with smaller depth are selected.
3. Now on the basis of size we eliminate 2 individual from 5 and the individual with smaller size are selected.
4. Now finally from these 3 individual we select two individual with higher fitness.

3.5. SPECIAL MUTATION

Special Mutation is a special technique in which we randomly selected an individual from the population and from this randomly selected individual we generate the two children. From this two generated children one is rejected on the basis of fitness the children with the lower fitness is rejected now we compare the selected individual with the parent and compare its fitness with the parent if the fitness of parent is better than the child than with a probability of 0.5 parent is transferred to the next generation otherwise children is transferred to the next generation. By

applying this special mutation technique we are sure that the generated individual does not reduce the fitness and also provide the diversity among the individuals.

3.6. ALGORITHM FOR SPECIAL MUTATION METHOD

1. First we generate the two children from the randomly generated individual.
2. Reject the one individual on the basis of fitness the individual with greater fitness are select.
3. Now we apply the elitism we compare the selected individual with parent and if the fitness of parent is better than the child than with a probability of 0.5 we transfer the parent to the next generation otherwise we transfer the child to the next generation.

3.7. STEPWISE CROSSOVER

The two individual which are selected from the triple tournament performs the Stepwise Crossover. The two individual will generate the six children. From these six generated children two are rejected on the basis of depth and size, children with the smaller depth and size are selected. Now from these 4 selected children eliminate two on the basis of fitness, children with the greater fitness are chosen. Now apply the elitism on the generated children and compare the fitness of the children with the parent if the fitness of the parent is greater than the children than with the probability of 0.5 parent is transferred to the next generation otherwise children is transferred to the next generation.

3.8. ALGORITHM FOR STEPWISE CROSSOVER

1. Randomly select individual from the population for triple tournament selection.
2. Select the best two individuals of the triple tournament for the crossover operation.
3. Randomly select the subtree or a node from a parent and place it at three different positions in another parent and generate the three children and the same process is repeated with another parent also, so the total number of generated children are six.
4. Now select the best 4 children from these 6 child on the basis of depth and size the children with the smaller depth and size are chosen.
5. Finally the two children are chosen on the basis of fitness the individual with the higher fitness are selected.
6. Now we apply the elitism on the selected individual if the fitness of the parent is better than the child than we transfer the parent with a probability of 0.5 to the next generation otherwise we transfer the child to the next generation.

3.9. GP Algorithm with Step Wise Crossover and Special Mutation

1. GP begins with a randomly generated population of solutions of size N.
2. A fitness value is assigned to solution.
3. A genetic operator is selected probabilistically.
4. Reproduction replicates the principle of natural selection and survival of the fittest.
5. If it is the crossover operator, then we apply the Step Wise Crossover.
6. If the selected operator is mutation, we apply a special mutation.
7. Continue 3. Until the new population gets   solutions.
8. This completes one generation.
9. Step 3 to 7are repeated till a desired solution is achieved. Otherwise, terminate the GP operation after a predefined number of generations.

4.EXPERIMENTAL RESULTS

We have designed a MultiClass Classifier to demonstrate our results. We have used Java 6.0 as a front end tool and Oracle 10g as a back end tool to develop our project. We have used 4 real data sets for training and validating our methodology. These are IRIS, WBC, BUPA, WINE, Table I gives a brief description about all the data sets used.

Data Sets

1) IRIS Data: This is the well-known Anderson's Iris dataset. It contains a set of 150 measurements in four dimensions taken on Iris flowers of three different species or classes.The classes are Iris Setosa, Iris Versicolour and Iris Virginica. The four features are sepal length, sepal width, petal length, and petal width. The data set contains 50 instances of each of the three classes.
2) Wisconsin Breast Cancer (WBC): This data set has 683 data points distributed in four classes. Each data point is represented by ten attributes. They describe characteristics of the cell nuclei present in the image.
3) BUPA Liver Disorders (BUPA): It consists of 345 data points in six dimensions distributed into two classes on liver disorders. The first 5 variables are all blood tests which are thought to be sensitive to liver disorders that might arise from excessive alcohol consumption.
4) WINE: This dataset contains 178 instances and 13 attributes. These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

Table 1. Datasets

Name of Data Set	No of classes	No of Features
IRIS	3	4
WBC	2	9
BUPA	2	6
WINE	3	13

PARAMETERS
Table II describes the common parameters used for all the data sets.

Table 2. Common Parameters for all datasets

Parameters	Values
Probability of Crossover Operation	75-80%
Probability of Reproduction Operation	10%
Probability of Mutation Operation	10-15%

Population Size	100-400
Minimum Tree Depth	2
Maximum Tree Depth	6
Number of Generations	20-60

RESULTS

The training and testing of the classifier generated is done using real data sets. To demonstrate our approach we have designed a MultiClass Classifier using genetic programming. We have compared the outcome of our results with the conventional crossover and mutation method shown in Table III and found that our method outperforms the conventional crossover and mutation method and improves the accuracy of the classifier with a fair amount.

Table 3. Comparison of Conventional Crossover and Mutation Method with
Stepwise Crossover and Special Mutation Method

NAME OF DATASETS	Conventional Crossover and Mutation Method		Stepwise crossover and Special Mutation Method	
	Training Accuracy	Generalization Accuracy	Training Accuracy	Generalization Accuracy
IRIS	85.89%	81.56%	92%	90%
WBC	80%	78%	88%	85.23%
BUPA	81%	79.58%	86.4%	84.11%
WINE	86%	84.45%	91.78%	90%

CONCLUSIONS

In this paper, we have proposed a Triple Tournament technique to select the parent for the crossover operation which helps in improving the overall fitness and applying the limit of size and depth on the individual which are selected for crossover operation. To demonstrate our approach we have designed a MultiClass Classifier and presented the results on different datasets. To describe the usefulness of our approach, we have compared our method with the conventional crossover and mutation method and obtained satisfactory results in terms of accuracies.We also proposed a Stepwise crossover technique in which we generate the 6 child from two parent and step wise eliminate the 4 child on the basis of size, depth and fitness and apply the elitism on the two children and compare with parent. We are also using a unique special Mutation technique in which we are generating two children and applying the fitness limit on the two generated children and also applying the elitism on the selected individual due to which the mutation operation also can't reduce the fitness of the classifier and it also provides the diversity among the individuals.

REFERENCES

[1] Chaudhari, N.S., Purohit, A., Tiwari, A, "A multiclass classifier using Genetic Programming" Control, Automation, Robotics and Vision, ICARCV 2008.10th International Conference.

[2] Mengjie Zhang and Will Smart "Genetic Programming with Gradient Descent Search for Multiclass Object Classification".

[3] W. Banzhaf, P. Nordin, R. E.Keller, and F. D. Francone, Genetic Programming: An Introduction. San Mateo, CA: Morgan Kaufmann, 1998.

[4] R. Poli, "Genetic Programming for image analysis," in Proc. 1st Int. Conf. Genetic Programming, Stanford, CA, July 1996,pp. 363– 368.

[5] D.Agnelli, A. Bollini, and L. Lombardi, "Image classification: an evolutionary approach," Pattern Recognit. Lett., vol.23, pp. 303–309,2002.