Various Collaborator Selection Pressures for Cooperative Coevolution for Classification

Catalin STOEAN

Faculty of Mathematics and Computer Science,
Department of Computer Science,
University of Craiova, Romania
catalin.stoean@inf.ucv.ro

Abstract. In recent work, the state-of-the-art cooperative coevolutionary paradigm has been tailored to handle classification tasks. Problems that are solved by means of a cooperative coevolutionary technique have to be decomposed into several components. Cooperative coevolution implies the use of several populations, each population having the aim of finding the solution for a component of the considered problem. Populations evolve separately and they interact only when individuals are evaluated. Interactions are made with the aim of obtaining complete solutions by collecting individuals from each of the populations. In this respect, there are several ways of selecting the individuals from each population. In present paper, a classification problem is thus considered and the main task is to determine the optimum choice for the collaborator selection pressure parameter, i.e. the way the individuals are selected from each population in order to form complete solutions.
Keywords: cooperative coevolution for classification, collaborator selection pressure, proportional selection, Iris data set.
Math. Subjects Classification 2000: 68T05, 68T20, 92D10.

1 INTRODUCTION

Since its recent history, cooperative coevolution has proven to be a very powerful means of solving optimization problems. The solution of the considered problem is decomposed into several components and each of these components is treated by an evolutionary algorithm (EA). The EAs evolve separately and interactions between populations exist only in the moment when fitness is computed; when an individual is evaluated, collaborators from all the other populations are selected in order to form a complete solution that can be measured.

The first class of problems that was considered for solving by means of a cooperative coevolutionary algorithm represented the optimization of several difficult multi-modal functions [1], [2]. Obtained results indicated that a cooperative coevolutionary algorithm outperformed a genetic algorithm.

More recently, another successful application of cooperative coevolutionary algorithms was obtained through the development of a rule-based control system for agents; two species were considered, each consisting of a population of rule sets for a class of behaviours [3], [4].

There are also personal attempts to solving a classification problem by means of cooperative coevolution [5], [6]. The work here goes further in improving the classification tool based on cooperative coevolution as a much deeper analysis of the coevolutionary parameters is achieved especially through the consideration of a proportional selection scheme instead of a random selection for the collaborator selection pressure parameter. Comparison to previously obtained results indicates the fact that, through the use of this selection mechanism, the novel cooperative coevolution for classification gains indeed a lot in improvement.

The paper is organized as follows: next section presents the basic ideas behind the cooperative coevolutionary model, while in section 3 the approach for classification is described. Section 4 presents obtained experimental results and the paper ends with a section of conclusions and ideas for future work.

2 COOPERATIVE COEVOLUTION. BASICS

The first step that has to be done when a problem is intended to be solved by cooperative coevolution is to find a proper decomposition of the solutions to the problem into components. Then, for each component a population (or species) is considered; each population evolves independently, except for the moment when the evaluation process takes place. As each individual in a population represents a component of the solution to the problem, collaborators have to be selected from all populations in order to assemble a solution that may be evaluated. Therefore, at each generation, when an individual c from a population is evaluated, collaborators from all the complementary populations are selected and a complete solution is formed; the evaluation of the complete solution is assigned to c.

Algorithm 1 outlines the steps that are followed by a cooperative coevolutionary algorithm. It starts with the initialization of each population. In order to measure the fitness of a certain individual for the first evaluation, a random selection of individuals (collaborators) from each of the other populations is performed and obtained solutions are evaluated. After this starting phase, each population is evolved using a canonical EA.

The way of choosing collaborators represents the main issue in this process. Consequently, there are three attributes regarding selection that have to be decided when a cooperative coevolutionary algorithm is constructed [2]:

1. **Collaborator selection pressure** refers to the way individuals are chosen from each population in order to form complete solutions to the problem, i.e. pick the best individual according to its previous fitness score, pick a random individual or use classic selection schemes in order to select individuals from each of the other populations.
2. **Collaboration pool size** represents the number of collaborators that are selected from each population.

Algorithm 1 Cooperative coevolutionary algorithm

$t = 0$;
for each species s **do**
  randomly initialize population $\text{Pop}_s(t)$;
end for
for each species s **do**
  evaluate $\text{Pop}_s(t)$;
end for
while termination condition = false **do**
  $t = t + 1$;
  for each species s **do**
    select population $\text{Pop}_s(t)$ from $\text{Pop}_s(t - 1)$;
    apply genetic operators to $\text{Pop}_s(t)$;
    evaluate $\text{Pop}_s(t)$;
  end for
end while

3. **Collaboration credit assignment** decides the way of computing the fitness of the current individual. This attribute appears in case the *collaboration pool size* is higher than one. There are three methods for computing this assignment:
   (a) *Optimistic* - the fitness of the current individual is the value of its best collaboration.
   (b) *Hedge* - the average value of its collaborations is returned as fitness.
   (c) *Pessimistic* - the value of its worst collaboration is assigned to the current individual.

In performed experiments, for the first time in proposed approach to classification, proportional selection was employed for the collaborator selection pressure parameter and, for reason of comparison, results of random selection of individuals [5], [6] were also mentioned.

3 COOPERATIVE COEVOLUTION APPROACH FOR CLASSIFICATION

The solution of a classification problem is regarded as a set of if-then conjunctive rules in first order logic. The final rule set is imagined as to have one rule for each category of the classification problem. A natural decomposition of the problem solution is to assign rules of a certain kind to a population; thus, the number of species equals the number of classes.

The data set is split into a set used for training the cooperative technique and a set for the testing step. In the process of evolving the rules, only information regarding data from the training set is available. After the evolution process is complete, rules are tested against the test set.

When the quality of an individual (rule) from a certain population is measured, one rule from each of the other populations is considered in order to form a complete set of rules. The rule set is applied to the training data and obtained accuracy is assigned as the fitness evaluation of the initial individual.

An individual c has the exactly same encoding as a sample from the data set to be classified, i.e. it contains the corresponding attributes, $c = (c_1, c_2, ..., c_m)$, where m is the number of indicators of samples in the data set. As stated before, individuals represent simple if-then rules having the condition part in the attributes space and the conclusion in the classes space. Each population evolves individuals (rules) with the same class.

In the classification process, the similarity between individuals and samples from the data set has to be computed. In this respect, distance between an individual and a sample from the data set has to be defined. In the experiments, normalized Manhattan was chosen as the distance measure (1). The distance does not depend on the class of the individual/sample.

$$d(c, x_i) = \sum_{j=1}^{m} \frac{\mid c_j - x_{ij} \mid}{b_j - a_j} \qquad (1)$$

where a_j and b_j represent the lower and upper bounds of the j-th attribute. As usually the values of attributes belong to different intervals, the distance measure has to refer their bounds.

3.1 EVALUATION OF INDIVIDUALS

Different values for the collaboration pool size parameter (denoted by n), were chosen. Therefore, in order to evaluate an individual from a certain population – that is a rule of a certain outcome – a collaborator from each of the other populations is selected n times. Every time, the set of rules is applied to the entire training collection. Obtained accuracy represents the fitness of the current individual.

The way the fitness of an individual is computed is described in Algorithm 2 – the percent of correctly classified samples from the training set (variable *correct* in the algorithm specifies the number of samples that were correctly classified).

For each training sample s multiple sets of rules are applied in order to predict its class. Naturally, all rules within a set have different classes. Scores for sample s for each of the possible outcomes are computed in the following manner: when a set of rules is considered, a certain outcome is established for s. The score of that outcome is increased by unity. Each of the n sets of rules are applied to s. Finally, the class of s is concluded to be the class that has the highest score.

There may appear situations when, for a certain sample, the same maximum score is obtained for more classes. In this case, one class has to be decided and it was considered to choose the first one in the order of outcomes.

Algorithm 2 Fitness evaluation of an individual c

for $i = 1$ to n **do**
  select a collaborator from each population different from that of c (the way of selecting depends on the collaborator selection pressure parameter);
  for each sample s in the training set **do**
    find the rule r from the set of all collaborators that is closest to s; increase the score of the r's class for s with one unity
  end for
end for
$correct = 0$;
for each sample s in the training set **do**
  if the real class of s equals the class that had the higher score for s **then**
    s is correctly classified;
    $correct = correct + 1$;
  end if
end for
$accuracy = 100 * correct$ / number of training samples;

As herein all combinations of rules count in the determination of accuracies, this new choice of assignment [6] is closer to the classical hedge type.

3.2 SELECTION AND VARIATION OPERATORS

For the EA selection (that takes place within each population), a proportional selection was used in the experiments but any other selection scheme may be successfully employed.

As stated before, for the collaborator selection pressure parameter, random selection of individuals was considered on the one hand (as in [5], [6]) and, on the other hand, a proportional selection scheme was for the first time considered for this task. This proportional selection is based on the scores individuals obtained the last time they were evaluated; in the first generation, collaborators are randomly selected. As envisaged in the experimental results section, the use of a selection scheme instead of random selection, as collaborator selection pressure, indeed improves the final accuracy of classification.

3.3 APPLICATION OF OBTAINED RULES TO THE TEST SET

The coevolutionary algorithm provides in the end the rules that are to be applied to the test set: there are as many populations of individuals (rules) as many classes the classification problem encodes. n times (we recall that n represents the collaboration pool size), one individual (rule) is selected from each population - the selection type is the same that is used for the collaborator selection pressure; each time, obtained rule set is applied to the test set and accuracy is computed in a similar manner to Algorithm 2 with the only difference that the training set is replaced by the test set.

4 EXPERIMENTAL RESULTS

Experiments are conducted on Fisher's Iris data set from the UCI repository. The data set contains 150 samples (iris flowers), each with 4 attributes (length and width of petals and sepals), all numerical, and 3 classes (three types of Iris). The collaboration pool size was varied from 1 up to 7 collaborators and, as collaborator selection pressure was concerned, tests were performed for random and for proportional selection.

In all conducted experiments, in each run of the algorithm, data set is split into 100 randomly chosen samples as training set and the rest of 50 samples as test set.

EA parameters were manually tuned and they are presented in Table 1. The population size parameter in the first column refers to one population only (all populations have the same cardinal). As there are three populations that coevolve, the overall number of individuals that appear in the algorithm is 450. Next two parameters denote the crossover and mutation probabilities. Mutation strength is denoted in the table by ms: in fact, the mutation perturbation value is obtained for each gene by dividing the size of the attribute the gene represents to the ms value in the table. Last parameter represents the predefined number of generations.

Table 1. Parameter values of involved EAs in cooperative coevolution for classification

Population size	p_c	p_m	ms	No. of generations
150	0.4	0.6	150	300

For each choice of considered cooperative coevolutionary parameters - collaborator selection pressure and collaboration pool size - 30 runs of the algorithm were performed and the average accuracy, the standard deviation, the average number of fitness evaluation calls were computed, as well as the number of times mutation and crossover took place. However, obtained results for random selection as collaborator selection pressure are the ones from [5], where the values of the EA parameters are not identical to the ones presented in Table 1: there, population size was 100, crossover and mutation probabilities were 0.6 and 0.5 respectively, mutation strength was set to 100 and the total number of generations was 200.

Results from Table 2 are the average obtained after 30 runs when one collaborator was chosen and proportional selection was employed. Only very minor differences were noticed when more collaborators were considered.

The average results obtained after 30 runs in each of the seven configurations of parameters (only collaboration pool size is varied), when just proportional selection is used for the choice of collaborators, are illustrated in Table 3.

Table 2. Descriptors independent of the cooperative coevolutionary parameters choices

Fitness evaluations	Times mutation	Times crossover
188 956	131 537	26 978

Table 3. Average results obtained after 30 runs for different values of collaboration pool size using proportional selection as collaborator selection pressure

Collaboration pool size	Time (seconds)	Test Accuracy			Standard deviation
		Avgerage	Minimum	Maximum	
1	65	93.73	84	100	3.85
2	133	93.0	82	100	4.69
3	198	92.8	84	98	4.05
4	306	93.53	86	100	3.85
5	331	95.06	86	100	3.05
6	397	93.86	82	100	4.23
7	452	95.4	88	100	3.28

Obviously, runtime increases proportionally to the value of the collaboration pool size as there are several computations to be performed when more collaborators are considered.

Results obtained using random selection are presented in detail in [5]. However, they are directly confronted with the ones from Table 3 in Figure 1. It can be easily noticed that the new results are significantly better in almost all cases. Moreover, in case random selection is used, the standard deviations are slightly higher than the ones obtained in case proportional selection is employed for the collaborator selection pressure: this indicates the fact that proposed approach is more stable when an EA selection scheme is used.

At the same time, obtained results can be directly compared to those obtained by other techniques in the literature. In [7], a similar way of selecting the training and test sets was used. The difference to present approach is that 80% of the samples from the Iris data were used for training and the rest for testing and that average accuracies are obtained after 500 runs. The worst accuracy (93.47%) resulted from the application of nearest-neighbor on the random recursive partitioning dissimilarity matrix and the best (96.31%) followed the employment of linear regression.

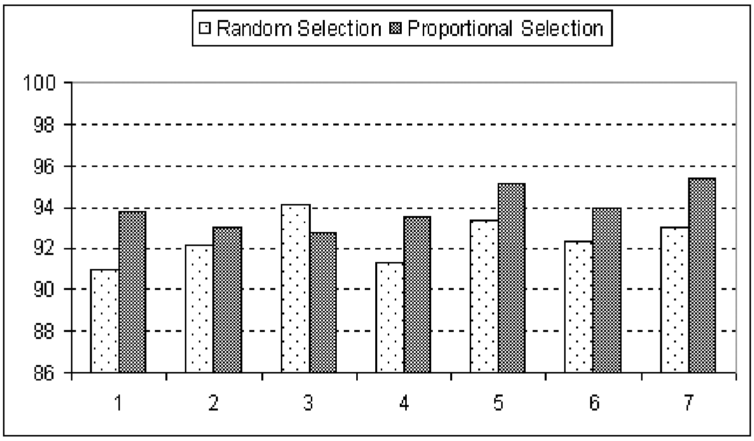

Fig. 1. Average accuracies obtained in 30 runs using the random and proportional selection as collaborator selection pressure. The x axis denotes the collaboration pool size parameter while the y axis denotes obtained average accuracy.

Table 4. Results obtained for Fisher's Iris data set by other algorithms

Technique	Accuracy	Std. dev.
kNN	95.63	3.3
RRP	93.47	4.2
Classification Trees	94.96	4.1
Linear Regression	96.31	3.4

5 CONCLUSIONS AND IDEAS FOR FUTURE WORK

A classification technique based on cooperative coevolution is presented in current paper and several cooperative coevolutionary parameter settings are considered. For the first time an EA selection scheme is used instead of a random choice of individuals for the collaborator selection pressure. The method is tested against a widely known benchmark problem and obtained results, which outperform the ones previously obtained by the cooperative coevolutionary approach for classification, are outlined.

In case more collaborators (than 7) are considered, there are higher chances that obtained results may be better but, at the same time, the runtime is noticeably increased; the task is thus to find a proper balance between accuracy and runtime when the solving tool for a problem is the cooperative coevolutionary approach for classification.

However, improvements still remain to be done as there may be several rules for one category, not only one, as is the case considered in present approach. In that situation, the replacement of the canonical EA with a multimodal one may transform the approach into an even more competitive one. Plus, more fine tuning for the EA parameters or automatic tuning may drive search to some significantly improved results.

References

[1] **M. A. Potter, K. A. De Jong**, A Cooperative Coevolutionary Approach to Function Optimization, *Proceedings of the Third Conference on Parallel Problem Solving from Nature*, Springer, pp. 249-257, 1994.

[2] **R. P. Wiegand, W. C. Liles, K. A. De Jong**, An Empirical Analysis of Collaboration Methods in Cooperative Coevolutionary Algorithms, *Proceedings of GECCO 2001*, pp. 1235-1245, 2001.

[3] **L. Panait, S. Luke**, Collaborative multi-agent learning: A survey, Technical Report GMU-CS-TR-2003-01, Department of Computer Science, George Mason University, 2003.

[4] **M. A. Potter, L. A. Meeden and A. C. Schultz**, Heterogeneity in the Coevolved Behaviors of Mobile Robots: The Emergence of Specialists, *Proceedings of The Seventeenth International Conference on Artificial Intelligence*, Morgan Kaufman, pp. 1337-1343, 2001.

[5] **Catalin Stoean, Mike Preuss, D. Dumitrescu, Ruxandra Stoean**, A Cooperative Coevolutionary Algorithm for Multi-class Classification, *Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2006*, submitted, 2006.

[6] **Catalin Stoean, D. Dumitrescu, Mike Preuss, Ruxandra Stoean**, Cooperative Coevolution for Classification, *Bio-Inspired Computing: Theory and Applications, BIC-TA 2006*, submitted, 2006.

[7] **S. Iacus, G. Porro**, Missing Data Imputation, Classification, Prediction and Average Treatment Effect Estimation via Random Recursive Partitioning, UNIMI - Research Papers in Economics, Business, and Statistics. Statistics and Mathematics, http://services.bepress.com/unimi/statistics/art7, 2006.