

Rule Minimization in Predicting the Preterm Birth Classification using Competitive Co Evolution

Jyothi Thomas^{1*} and G. Kulanthaivel²

¹Department of Computer Science and Engineering, Sathyabama University, Chennai - 600119, Tamil Nadu, India; phd.jthomas@gmail.com

²National Institute of Technical Teachers Training and Research, Chennai - 600113, Tamil Nadu, India; gkveldr@gmail.com

Abstract

Objective: Accurate prediction of pre term birth probability in the deliveries of babies with an effective classifier tool is a big challenging task. This paper talks about a novel competitive co evolution rule prediction classifier for extracting minimum number of rules for identifying the preterm birth. **Methods/Analysis:** Competitive Co evolution algorithm is applied to the preterm classifier for deriving the pattern governing the classified dataset. In this approach we have used two classes' namely normal birth dataset and preterm birth dataset as two individuals competing with each other. Fitness of each individual is calculated based on the relative fitness of the other population. **Findings:** The dataset consist of 1052 records of preterm birth and normal dataset of 1314 each of five attributes. The experimental result shows the total no of rules needed for training and testing is drastically reduced compared to the total rules. The accuracy is also improved to a greater extent by applying the proposed algorithm. **Applications/Improvement:** The proposed algorithm minimizes the number of training rules to 16 and testing rules to 11 out of total 28 possible rules in the rule set. The accuracy of 0.962938881664499 of correct classification of preterm birth dataset is obtained through method.

Keywords: Co Evolution, Preterm Birth, Reproduction, Rules Extraction

1. Introduction

Pre term birth is the delivery of baby at least three weeks before the delivery date i.e. usually before 37 weeks of pregnancy¹. The pre term birth child has the risk of hearing problems, eyes problem and delay in the development of various organs. There are many causes for preterm birth mainly multiple pregnancies, diabetics, blood pressure, mental stress and smoking tobacco. Assisted reproductive technology for achieving pregnancy is also one of the factors for pre term birth. The probability of early prediction² of pre term is very difficult and no exact test to predict since the causes of pre term is unpredictable. Every year millions of preterm birth occurs all over the world causing a serious concern in health related issues to these new born babies.

Evolution⁸ is the process of evolving things better and better over a period of generations resulting in better future traits. Darwin's theory says "survival of

the fittest" which is the back bone of the evolutionary algorithms. Evolutionary algorithms are applied in optimization³ problems to derive the best possible solution closest to the reality. Competitive co-evolution⁴ is the concept of evolving to traits⁵ better and better over the competition exists between these two traits. The two different populations of species evolves over generations in competing against each other which makes them to evolve better and better as the generations increases. Genetic Algorithm is applied on each population for evolving its traits over the generations. Genetic algorithm contains five phases namely initial population of individuals, selection of parents, Reproduction, mutation and selection of individuals for next generation. Initial stage is the representation of individuals in the population in the form of chromosomes. There are two ways of representing chromosomes namely binary chromosomes and real valued chromosomes. In binary chromosomes

* Author for correspondence

every gene is represented either using '0' or '1' which are the values for the attribute. In real valued chromosomes the gene value is represented as floating value.

The next phase is the selection of parents for reproductions where the individuals are selected from the set of population based on fitness function. Every individual in the population is evaluated using the fitness function. Fitness function⁶ is the mathematical formula mapping of individual with respect to the solution of the problem. There are many selection⁷ strategies for selecting parents for reproduction. Random selection, roulette wheel selection, tournament selection⁸, Boltzmann selection are few selection techniques used for selecting individuals for reproduction. Reproduction⁹ is the process of exchanging the genetic material between the parents to evolve new offspring's. Mutation¹⁰ is applied once in n generations to enhance the speed of convergence of reaching the optimized results at a faster rate. Mutation is

the sudden biological change in the chromosomes genetic value by slightly triggering unknown changes which modifies the genetic composition of the chromosome.

2. Methodology

Competitive Coevolution¹¹⁻¹³ algorithm is applied to the preterm classifier for deriving the pattern governing the classified dataset. The competitive co evolution algorithm improves the quality of both the population of two groups which compete each other. In this approach we have used two classes' namely normal birth dataset and preterm birth dataset as two individuals competing with each other. Fitness of each individual is based on the relative fitness of the other population. Figure 1 contains the proposed competitive co-evolution preterm classifier algorithm which generated the pattern for classification.

```

Competitive co evolution preterm classifier
{
  Initialize a set of n individuals representing preterm birth datasets as population C1 ;
  Initialize a set of m individuals representing normal birth dataset as population C2;
  while(!stopping condition())
  {
    for each individual xi =1,2,3.....ni ? C1
    Evaluate the relative fitness RF(xi) with respect to the individuals of population C2;
    Select Parents from C1 based on the relative fitness using selection strategy;
    Perform reproduction by crossing over of parents to produce offspring's;
    Perform mutation on the offspring's;
    Evolve next generation of individuals for the population C1;
    For each individual xi=1,2,3.....mi ? C2
    Evaluate the relative fitness RF(xi) with respect to the individuals of population C1;
    Select Parents from C2 based on the relative fitness using selection strategy;
    Perform reproduction by crossing over of parents to produce offspring's;
    Perform mutation on the offspring's;
    Evolve next generation of individuals for the population C2;
  }
  Select the top 10% of individuals in the population C1 as the solution.
  Extract Rules from the individuals in the selected population;
}

```

Figure 1. Pseudo code for the competitive co evolution preterm classifier algorithm.

3. Implementation

The prediction of identifying the class of preterm birth dataset is based on extracting the rules. The competitive co evolution technique is applied for the classification of pre term birth datasets. The proposed Competitive co evolution preterm classifier algorithm extracts rules for the prediction of preterm birth class. The various steps implemented in the algorithm as shown in Figure 1 are discussed below.

Steps:

1. Initialization: In this phase two populations C_1 and C_2 representing the preterm birth and normal class respectively are randomly generated with n and m individuals. Each individual represents a real valued chromosome of five attributes randomly generated using uniform distribution.
2. Measure of Relative Fitness: Every individual in each population C_1 and C_2 , the relative fitness is measured with respect the opponent individuals in the other population. The relative fitness of each individual in C_1 is measured using the opponents of individuals in C_2 . For each individual x_i belongs to C_1 , 50 opponents are randomly selected and tournament is conducted with each selected individual opponents in C_2 . The result of the tournament is calculated in terms of score which is 1 if x_i wins (i.e. x_i fitness is greater than the opponent) and 0 if x_i losses (i.e. x_i fitness is less than the opponent). The Equation 1 represents the measure of relative fitness of every individual in the population.

$$\text{Relative fitness } R F(x_i) = \sum_{j=1}^n (S_{ij}) \tag{1}$$

The relative fitness is the summation of the outcome of all the matches conducted in the tournament by summing the results S_{ij} . The result is either win or loss represented by 1 or 0 respectively based on the fitness of the opponent as per the Equation 2.

$$\text{Where } S_{ij} = \begin{cases} 1 & \text{if } f(x_i) > f(y_j) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Where y_j is the individual from the opponent population

3. Reproduction: Reproduction is the process of generating new offspring's by exchanging genetic composition with the parent chromosomes. Parents are selected for reproduction based on the relative fitness measured in the above step. Crossing over of genetic materials is simulated as per the given Equation 3.

$$x_j(t) = (1-\gamma) x_{1j}(t) + \gamma x_{2j}(t) \tag{3}$$

Where $x_{1j}(t)$ and $x_{2j}(t)$ are the j th component of the two parents respectively and γ is a constant with a value of 0.5.

4. Mutation: Mutation is performed by altering or modifying the offspring once in five generations to speed up the convergence of solution. Next generation of individuals are evolved for the population C_1 .
5. The steps 2 to 4 are performed on the individuals of the other population C_2 with C_1 as the opponents until the stopping condition is satisfied.
6. Extraction of Rules: Repeat step 2 to 5 until the stopping condition is satisfied. Select the top 10% of individuals from the population C_1 and rules are extracted in the form of IF---THEN statements.

Rule 1: IF (range₁ >= a₁ && a₁ <= range₂) then class=pre-term

Rule 2: IF ((range₁ >= a₁ && a₁ <= range₂) AND (range₁ >= a₂ && a₂ <= range₂)) then class=pre-term.

Rule n: IF ((range₁ >= a₁ && a₁ <= range₂) AND (range₁ >= a₂ && a₂ <= range₂) AND.....((range₁ >= a_k && a_k <= range₂)) then class=pre-term

Where a₁, a₂, a_n are attributes of the preterm birth dataset. Totally 28 rules are generated by combining the attributes.

4. Experiments and Results

Experiments are conducted using pre term birth datasets. The dataset consist of 1052 records of preterm birth and normal dataset of 1314 each of five attributes. Out of 1052 records 80% of records selected in random is used for training and the remaining 20% of datasets for testing. The preterm birth datasets represent the population C_1 and the normal dataset represents the population C_2 as per the algorithm. At the end of the algorithm top 10% of the best individuals representing the population C_1 is taken as the solution. Rules are extracted from the solution for

predicting the preterm classifier in the form of IF THEN Statement using the attributes in the condition. The general formats of the rules are

IF ($\min_1 < a_1 < \max_1$) AND ($\min_2 < a_2 < \max_2$) AND ($\min_n < a_n < \max_n$) THEN class = Preterm.

Where $a_1 = a_1, a_2, \dots, a_n$ number of attributes.

Totally 28 rules are generated from the solution set in the form of IF...THEN statement as defined above.

The experimental result shows the total no of rules needed for training and testing is drastically reduced compared to the total rules. The accuracy is also improved to a greater extent by applying the proposed algorithm.



Figure 2. Optimisation of rules interms of training and testing based on number of iterations.

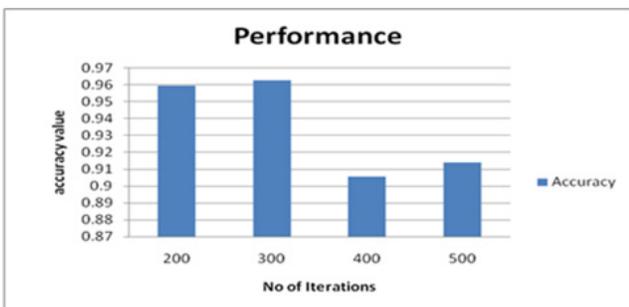


Figure 3. Performance interms of accuracy.

In Figure 2 the no of rules required for training and testing decreases as the iteration increases. It is found that a minimum of 16 rules for training and 11 rules for testing is required for the classification of preterm birth datasets out of totally 28 existing rules. The rules are pruned and the minimization of rules is obtained when the iteration is 400.

In Figure 3 the performance of the algorithm is

measured in terms of accuracy of prediction. Accuracy of prediction is given by the formula

$$\text{Accuracy} = \frac{(\text{True Positive} + \text{True Negative})}{(\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative})} \quad (4)$$

Maximum accuracy of 0.9629 is obtained in the proposed algorithm which is very effective in terms of classification of preterm birth datasets.

5. Conclusion

The proposed competitive co evolution preterm classifier extracts the prediction rules for classifying the pre term datasets. The accuracy of 0.9629 which is obtained at iteration=200 is the best performance of this algorithm. The proposed algorithm also tries to minimize the no of prediction rules needed for classifying under training and testing data. The classification of preterm birth can also be implemented by using the splitting of attributes into two groups as opponents and the attributes value can be competitively evolved over the generations using genetic algorithm.

6. References

1. Sonek JD, Iams JD, Blumenfeld M, Johnson F, Landon M, Gabbe S. Measurement of cervical length in pregnancy: comparison between vaginal ultrasonography and digital examination. *Obstetrics Gynecology*. 1990; 76(2): 172-75.
2. Back T. Selective Pressure in Evolutionary Algorithms: A Characterization of Selection Mechanisms. *Proceedings of the First IEEE World Congress on Computational Intelligence; Conference on Evolutionary Computation*; Orlando: FL, 1994; 1: 57-62.
3. Angeline PJ, Pollack JB. Competitive Environments Evolve Better Solutions for Complex Tasks. *Proceedings of the Fifth International Conference on Genetic Algorithms*; 1993. p. 1-8.
4. Nguyen S, Zhang M, Johnston M, Tan KC, Automatic Design of Scheduling Policies for Dynamic Multi-objective Job Shop Scheduling via Cooperative Coevolution Genetic Programming. *IEEE Transactions on Evolutionary Computation*. 2014; 18 (2): 193-208.
5. Holland JH. ECHO: Explorations of Evolution in a Miniature World. Ed: J.D. Farmer and J. Doyne. *Proceedings of the Second Conference on Artificial Life*; 1990.
6. Le Ly D, Lipson H. Optimal Experiment Design for Coevolutionary Active Learning. *IEEE Transactions on Evolutionary Computation*. 2014; 18 (3):394-404.
7. Catley C, Frize M, Walker CR, Petriu DC. Predicting High-

- Risk Preterm Birth Using Artificial Neural Networks. *IEEE Transactions on Information Technology in Biomedicine*. 2006; 10(3): 540–49.
8. Michalewicz Z. *Genetic Algorithms + Data Structures = Evolutionary Programs*. Springer: Berlin. 1996.
 9. Cohen WW. Fast Effective Rule Induction. *Proceedings of twelfth conference (ML 95)*; 1995. p. 115–23.
 10. Wei F, Wang Y, Zong T. A Novel Cooperative Coevolution for Large Scale Global Optimization. *2014 IEEE International Conference on Systems, Man, and Cybernetics*; San Diego, CA: USA. 2014. p. 738–41.
 11. Chandra R, Bali K. Competitive two-island cooperative coevolution for real parameter global optimization. *IEEE Congress on Evolutionary computing (CEC)*; Sendai. May 2015. p. 93–100.
 12. Thomas J, Kulanthaivel G. Classification of Preterm Birth with Optimized Rules by Mutating the Cognitive Knowledge of PSO. *Indian Journal of Science and Technology*. 2015 July; 8(15):1–5.
 13. Nguyen S, Zhang M, Johnston M, Tan KC, Automatic Design of Scheduling Policies for Dynamic Multi-objective Job Shop Scheduling via Cooperative Coevolution Genetic Programming. *IEEE Transactions on Evolutionary Computation*. 2014; 18 (2): 193–208.