Swarm and Evolutionary Computation ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at SciVerse ScienceDirect

Swarm and Evolutionary Computation

journal homepage: www.elsevier.com/locate/swevo

Regular Paper

Optimization of stacking ensemble configurations through Artificial Bee Colony algorithm

P. Shunmugapriya [a,*], S. Kanmani [b]

[a] Department of Computer Science and Engineering, Pondicherry Engineering College, Pillaichavady, Puducherry 605014, India
[b] Department of Information Technology, Pondicherry Engineering College, Pillaichavady, Puducherry 605014, India

ARTICLE INFO

Article history:
Received 26 April 2012
Received in revised form
10 March 2013
Accepted 6 April 2013

Keywords:
Classifier
Classifier Ensemble
Stacking
Artificial Bee Colony

ABSTRACT

A Classifier Ensemble combines a finite number of classifiers of same kind or different, trained simultaneously for a common classification task. The Ensemble efficiently improves the generalization ability of the classifier compared to a single classifier. Stacking is one of the most influential ensemble techniques that applies a two level structure of classification namely the base classifiers level and the meta-classifier level. Finding suitable configuration of base level classifiers and the meta-level classifier is always a tedious task and it is domain specific. The Artificial Bee Colony (ABC) Algorithm is a relatively new and popular meta-heuristic search algorithm proved to be successful in solving optimization problems. In this work, we propose the construction of two types of stacking using ABC algorithm: ABC-Stacking1 and ABC-Stacking2. The proposed ABC based stacking is tested using 10 benchmark datasets. The results show that the ABC-Stacking yields promising results and is most useful in selecting the optimal base classifiers configuration and the meta-classifier.

1. Introduction

Classifier Ensemble (CE) has been an interesting topic of research during the past decade. The idea behind Ensemble Learning is to learn a set of classifiers instead of learning a single classifier and then combine the predictions of multiple classifiers [1]. The motivation behind this is to reduce variance and reduce bias so that the results are less dependent on peculiarities of a single training set and the combination of multiple classifiers may learn a more expressive model than a single classifier [2].

Since its proposal, there has been an explosive growth in the CE domain and many ensemble methodologies have been proposed [1]. Stacked Generalization also known as Stacking is one of the most popular ensembles and it employs two levels of construction: base classifier level and meta-classifier level. Stacking is one possible solution where the combiner is trained on unseen data constructed by a cross-validation procedure [3].It is inferred from the literature that, stacking increases the predictive accuracies of the ensemble to a greater extent [1]. On Stacking the ensemble, two important problems arise: selection of appropriate base classifiers which will constitute the ensemble; Selection of suitable meta-classifier to combine the base classifiers [4].

Classifiers give varied performance on the basis of problem domain under consideration. The same classifier may yield good results for one dataset and not for another [5]. A classifier when acting as a meta-classifier may give a varied performance, because it is dealing with meta- instances and not the actual instance. So even after deciding upon the classifiers to be configured into the ensemble, the choice of meta-classifier still stands as a question and again this is a matter of issue [4]. The selection of base classifiers and a suitable meta-classifier to obtain good performance is very difficult. If the option of base-classifiers and meta-classifier to be considered is smaller in number, an exhaustive search could be applied to find the optimal configuration. However exhaustive searches are time consuming and the optimal configurations obtained for one domain may not be optimal for another [6]. An alternative solution could be to use meta-heuristic search which has a smaller search space and a good convergence speed compared to exhaustive search [6].

Nature inspired meta-heuristic search have been widely used in finding optimal solutions. In the recent past, Evolutionary and Swarm Intelligent algorithms like Genetic Algorithm (GA), Ant Colony Optimization (ACO), Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO), Differential Evolution (DE) and Niching Algorithms use meta-heuristics and play a major role in solving combinatorial, constrained and unconstrained problems wherever optimization is required [7–13]. Apart from stand alone, the ensemble form of these algorithms also finds its application in producing optimal solutions [12–16].

This paper proposes the construction of stacking using the Artificial Bee Colony (ABC) algorithm. The ABC algorithm is a swarm intelligent, meta-heuristic search algorithm inspired by the

* Corresponding author. Tel.: +91 9894710267.
E-mail addresses: pshunmugapriya@gmail.com (P. Shunmugapriya), kanmani@pec.edu (S. Kanmani).

2210-6502/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.swevo.2013.04.004

foraging behavior of honey bee swarms. This algorithm was introduced by Karaboga [17] and it serves as a powerful optimization tool. The ABC algorithm has shown a competitive performance (and sometimes superior) compared to GA, PSO and ACO [18–20]. The ABC algorithm is also simple in concept and has only fewer control parameters [21].

In literature, stacking has been constructed using optimization algorithms like GA [22,23] and ACO [6,24]. GA-Stacking and ACO-Stacking have given optimal configuration of base classifiers for stacking with promising results [6,22–24]. However to the best of our knowledge, stacking using ABC algorithm has not been proposed so far and this proposal has shown excellent results. Though stacking configuration has been optimized using GA and ACO, the main idea behind this proposal is: to analyze whether more optimization can be achieved with increased performance of the ensemble stacking; to analyze the performance of ABC whether it shows competitive performance to ACO and provides powerful optimization.

The ABC-Stacking has been proposed and implemented in two levels: ABC-Stacking1 (Base-Level Stacking) and ABC-Stacking2 (Meta-Level Stacking). In base-level stacking, the employed and onlooker bees search for the optimal configuration of base classifiers and the meta-classifier is a fixed learning algorithm. In meta-level stacking, while forming the base classifier configuration, the bees simultaneously keep exploring for the best meta-classifier through the pool of classifiers, each time the stacking of selected classifiers is done.

This paper is organized in 8 sections. Section 2 gives a brief note on CE and the ensemble stacking. The concept of ABC algorithm is explained in Section 3. In Section 4, the previous works in the literature related to Stacking are discussed. Section 5 gives the detailed description of the proposed ABC-Stacking. Experimental environment and the experiments are discussed in Section 6. Computational results of the proposed method in comparison with the existing methods are discussed in Section 7. Section 8 concludes the paper.

2. Classifier ensemble and stacking

CE is a learning paradigm where several classifiers are combined to increase the generalization ability of a single classifier [1]. Fig. 1 shows the block diagram of a CE. In CE, the dataset is fed as input to '*n*' number of classifiers and the classifiers are trained over the entire feature space. The outputs from the classifiers are given to the combination scheme. The ensemble then provides a consensus opinion by combining the individual classifier decisions, such as bagging, boosting or stacking-type algorithms [25]. This combination may be either statistical or algorithmic. One important issue while building CE is, how to select classifiers such that the decision making quality of the ensemble is superior to that of any individual classifier.

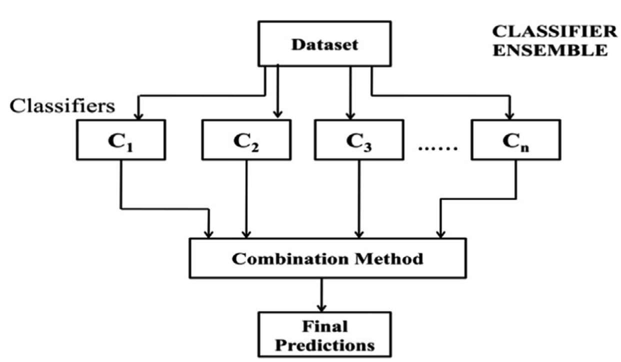

Fig. 1. Classifier Ensemble.

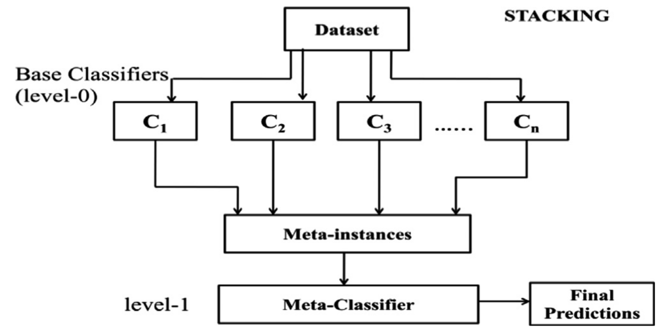

Fig. 2. Stacking ensemble.

Stacking, also known as Stacked Generalization is a well known ensemble approach [3]. It is a method that combines the classifiers (level-0 or base level) by a meta-level (level-1) classifier that predicts the correct class based on the decisions of the base level (level-0) classifiers as shown in Fig. 2. The meta-level classifier is induced on a set of meta-level training data. The meta-level training data are produced by applying a procedure similar to *k*-fold cross-validation [2,26] on the training data. The outputs of the base classifiers for each instance along with the true class of that instance form a meta-instance. A meta-classifier is then trained on the meta-instances. When a new instance appears for classification, the output of all the base learners is first calculated and then propagated to the meta-classifier, which outputs the result.

3. The Artificial Bee Colony algorithm (ABC)

ABC is a swarm intelligent algorithm based on the intelligent foraging behavior of honey bee swarms. This intelligent algorithm was proposed by Karaboga [17] and since its proposal, the algorithm has been employed to solve optimization problems in large number of domains [9,10,27–32]. The foraging behavior model in ABC constitutes food sources and three types of bees: employer bee, onlooker bee and scout. The food sources represent the possible solutions of the optimization problem. The nectar content of the food sources correspond to the fitness of the solutions. The ABC algorithm employs three types of bees in the colony: employed bees, onlooker bees and scout bees. Initially, the food source positions are generated (*N*). The population of the employed bees is equal to the number of food sources. Each employed bee is assigned a food source. Employed bees exploit the food sources and pass the information of nectar amount to the onlooker bees. Both the population of employed and onlooker bees are equal. Based on the nectar information from the employed bees, the onlooker bees exploit the food sources and its neighborhood until the food sources become exhausted. The employed bee of exhausted food sources becomes a scout. Scouts then start searching for new food source positions. The nectar information represents the quality of the solution available from the food source. Increased amount of nectar increases the probability of selection of a particular food source by the onlooker bees [19,20,33].

4. Related works

Stacking is one of the popular ensemble techniques and two important issues arise whenever Stacking has to be implemented [34–40]. They are: the choice of classifiers at the base level (level-0); the choice of a single classifier at the meta-level (level-1). i.e. Whenever Stacking is constructed as a solution for any application

Please cite this article as: P. Shunmugapriya, S. Kanmani, Optimization of stacking ensemble configurations through Artificial Bee Colony algorithm, Swarm and Evolutionary Computation (2013), http://dx.doi.org/10.1016/j.swevo.2013.04.004