Expert Systems With Applications 93 (2018) 395–422

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Classifier ensemble reduction using a modified firefly algorithm: An empirical evaluation

Li Zhang[a,*], Worawut Srisukkham[a], Siew Chin Neoh[b], Chee Peng Lim[c], Diptangshu Pandit[a]

[a] Computational Intelligence Research Group, Department of Computing Science and Digital Technologies, Faculty of Engineering and Environment, University of Northumbria, Newcastle NE1 8ST, UK
[b] Faculty of Engineering, Technology, and Built Environment, UCSI University, Kuala Lumpur 56000, Malaysia
[c] Institute for Intelligent Systems Research and Innovation, Deakin University, Waurn Ponds, VIC 3216, Australia

ARTICLE INFO

Article history:
Received 19 May 2017
Revised 30 September 2017
Accepted 1 October 2017
Available online 3 October 2017

Keywords:
Ensemble reduction
Classification
Firefly algorithm

ABSTRACT

In this research, we propose a variant of the firefly algorithm (FA) for classifier ensemble reduction. It incorporates both accelerated attractiveness and evading strategies to overcome the premature convergence problem of the original FA model. The attractiveness strategy takes not only the neighboring but also global best solutions into account, in order to guide the firefly swarm to reach the optimal regions with fast convergence while the evading action employs both neighboring and global worst solutions to drive the search out of gloomy regions. The proposed algorithm is subsequently used to conduct discriminant base classifier selection for generating optimized ensemble classifiers without compromising classification accuracy. Evaluated with standard, shifted, and composite test functions, as well as the Black-Box Optimization Benchmarking test suite and several high dimensional UCI data sets, the empirical results indicate that, based on statistical tests, the proposed FA model outperforms other state-of-the-art FA variants and classical metaheuristic search methods in solving diverse complex unimodal and multimodal optimization and ensemble reduction problems. Moreover, the resulting ensemble classifiers show superior performance in comparison with those of the original, full-sized ensemble models.

1. Introduction

Ensemble methods have been widely used for improving classification performance. There are many well-known ensemble construction techniques, such as bagging and boosting (Han, Kamber, & Pei, 2011). However, base classifier redundancy is still a challenging research problem, which has received much research attention in the area of computational intelligence. On the other hand, because of the superior search capabilities of evolutionary algorithms, they have been widely used for solving diverse optimization problems. Many state-of-the-art swarm intelligence algorithms are available in the literature for feature optimization and dimensionality reduction, e.g. see Jothi and Inbarani (2016). Motivated by the success of metaheuristic optimization, we aim to employ evolutionary algorithms for discriminant base model selection and to construct optimized classifier ensembles without compromising classification accuracy.

In this research, we propose a modified firefly algorithm (FA) for classifier ensemble reduction. It integrates two new search mechanisms, i.e. an accelerated attractiveness behavior and an evading action, to mitigate premature convergence of the original FA model. The proposed attractiveness operation employs not only the neighboring but also global promising solutions to guide the search process, while the evading action is advised by both local and global worst solutions to lead the search out of gloomy regions. These two new search mechanisms work cooperatively to overcome stagnation, and reach global optimality with fast convergence. The proposed algorithm is useful for identifying discriminant base classifiers to realize ensemble reduction, and obtaining the best trade-off between classification accuracy and ensemble complexity. Evaluated with standard, shifted, and composite test functions, Black-Box Optimization Benchmarking (BBOB) testbeds (Hansen, Auger, Ros, Finck, & Pošík, 2010; Hansen, Finck, Ros, & Auger, 2010; Hansen, Auger, Finck, & Ros, 2012) and high dimensional UCI data sets (Bache & Lichman, 2013), the proposed algorithm outperforms other state-of-the-art FA variants, BBOB optimizers, and several classical search methods, significantly, in solving diverse challenging unimodal and multimodal optimization and

* Corresponding author.
E-mail addresses: li.zhang@northumbria.ac.uk (L. Zhang), srisukkham.worawut@northumbria.ac.uk (W. Srisukkham), chee.lim@deakin.edu.au (C.P. Lim), diptangshu.pandit@northumbria.ac.uk (D. Pandit).

https://doi.org/10.1016/j.eswa.2017.10.001
0957-4174/

ensemble reduction problems. The resulting classifier ensembles also achieve competitive performance in comparison with those of the full-sized, unreduced original ensembles.

The paper is organized as follows. Section 2 discusses diverse ensemble techniques and swarm intelligence-based optimization methods. We introduce the proposed algorithm with the new attractiveness and evading search behaviors in Section 3. Section 4 presents the evaluation of the proposed algorithm and other metaheuristic search methods using standard and complex benchmark functions and high dimensional data sets for classifier ensemble reduction. Section 5 draws the conclusions and suggests a number of areas for further research.

2. Related work

In this section, we introduce state-of-the-art ensemble classification techniques and diverse evolutionary metaheuristic search methods.

2.1. Ensemble classification techniques

There are many state-of-the-art ensemble methodologies proposed in recent years. Galar, Fernandez, Barrenechea, Bustince, and Herrera (2012) conducted a comprehensive review of ensemble techniques for solving the class imbalance problem. Their studies indicated the efficiency of integrating random under-sampling techniques with bagging and boosting ensembles for classification of imbalanced data sets. Their work emphasized the positive synergy between sampling techniques and bagging ensemble learning. An ordering-based ensemble pruning technique was also proposed in their recent work (Galar, Fernandez, Barrenechea, Bustince, & Herrera, 2016) to better tackle imbalanced classification problems. Diao, Chao, Peng, Snooke, and Shen (2014) proposed harmony search based ensemble reduction. They employed bagging and random subspaces for base classifier pool generation. Since bagging uses randomly selected diverse subsets of training instances to build the base models, and the random subspaces method uses randomly selected subsets of attributes for base classifier generation, the base models produced by these two methods showed great diversity. Evaluated with several UCI benchmark data sets, their model achieved impressive accuracy and outperformed ensemble classifiers using the full-sized or any randomly selected base models. Farid et al. (2013) proposed an adaptive ensemble model for data stream classification with concept drifting. They employed three base decision tree classifiers for ensemble construction. The weakest base classifier in the ensemble was updated automatically by a stand-by, newly generated base decision tree learner with the knowledge of the updated class information, in order to represent the most recent concepts in the incoming data streams. Evaluated with UCI benchmark data sets, their model achieved great efficiency for concept-drifting data stream classification. Zhang, Zhang, and Hossain (2015) developed adaptive ensemble classifiers for facial expression recognition and facial action intensity estimation. After extracting the initial dynamic motion-based facial features, the minimal-redundancy-maximal-relevance criterion (mRMR) was used to identify discriminative facial features for subsequent intensity estimation of facial action units. A set of ensemble classifiers was integrated with a distance-based clustering algorithm to identify six basic emotions as well as new unseen novel emotion classes. The complementary neural network was used as the base learner in their work, which possessed the capability of providing uncertainty measure for classification of each instance. The base classifiers and clustering algorithm worked collaboratively to inform the arrival of novel emotion classes. Evaluated with the Bosphorus facial expression database and with online real user testing, it achieved impressive performance for expression recognition and novel class detection.

Sun, Tang, Minku, Wang, and Yao (2016) proposed a model known as Class-Based ensemble for Class Evolution (CBCE) to deal with class evolution and concept drifting in data stream mining. Their proposed method generated a base learner for each class. Similarly, it adjusted to the class evolution by constantly updating the base learners to reflect the latest concepts in the data stream. Moreover, their work addressed a side-effect of class evolution, i.e. the class imbalance problem, by employing an undersampling method for the base learners. Their model showed superiority over other existing methods for class evolution adaptation. Guan, Li, and Roli (2015) proposed an ensemble classification method for covariate-invariant gait recognition. They first employed the random subspaces method to generate the base classifiers. They also employed local enhancing and hybrid decision-level fusion to identify more discriminant features and eliminate inefficient base classifiers, respectively. Importantly, their work claimed that feature subsets randomly generated by the random subspaces method were less discriminant, because its feature selection process was conducted in a random manner without using class label information, therefore causing a low recognition accuracy rate with an expensive computational cost. As such, their work employed two local enhancing supervised learning methods, i.e. two-dimensional linear discriminant analysis (LDA) and IDR/QR, to address the above problems and identify more discriminant features. A majority voting strategy was used to generate the final classification result based on the outputs of the base models. Huang et al. (2016) proposed a deep Convolutional Neural Network (CNN) based ensemble classifier for image retrieval. Both AlexNet and Network in Network (NIN) were deployed to extract the initial image features. Subsequently, weighted average feature vectors based on the outputs of both AlexNet and NIN were generated. Evaluated with CIFAR-10 and CIFAR-100 databases, the proposed aggregate ensemble model outperformed single CNN for image classification and retrieval tasks. Liew, Loo, and Obo (2016) constructed an ensemble reduction method by using the Genetic Algorithm (GA) and a Bayesian Information Criterion fitness evaluation mechanism. Fuzzy extreme learning machines were used as the base learners. Evaluated with the database for Emotion Analysis using Physiological Signals, their model achieved the best trade-off between classification accuracy and ensemble complexity. Sołtys, Jaroszewicz, and Rzepakowski (2015) utilized ensemble models such as bagging and random forests for uplift modeling, whereas Pietruczuk, Rutkowski, Jaworski, and Duda (2017) proposed two theorems for determining the optimal ensemble size for data stream classification based on both classification accuracy and memory requirements.

2.2. FA and modified FA models

Swarm intelligence-based algorithms have been extensively studied for diverse optimization problems. As one of the more recently proposed swarm intelligence-based metaheuristic search methods, FA shows natural search capability of dealing with multimodal optimization problems as compared with other algorithms such as GA and Particle Swarm Optimization (PSO) (Yang, 2009). FA employs the following principles to perform the search process. Firstly, each firefly has a light intensity which denotes the solution quality. Secondly, fireflies with lower light intensities are attracted to neighboring fireflies with higher light intensities regardless of their sex. Finally, the attractiveness decreases as the distance between two fireflies increases (Yang, 2009). A Levy-flight Firefly Algorithm (LFA) has also been proposed by Yang (2010). Instead of using Gaussian distribution, LFA implements Levy flights as random walk to overcome local optima traps.