Information Sciences 281 (2014) 128–146

Contents lists available at ScienceDirect

Information Sciences

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

Feature subset selection by gravitational search algorithm optimization

XiaoHong Han *, XiaoMing Chang, Long Quan, XiaoYan Xiong, JingXia Li, ZhaoXia Zhang, Yi Liu

Key Laboratory of Advanced Transducers and Intelligent Control Systems, Ministry of Education China, Taiyuan University of Technology, Taiyuan, Shanxi, China

ARTICLE INFO

Article history:
Received 16 March 2013
Received in revised form 9 May 2014
Accepted 19 May 2014
Available online 28 May 2014

Keywords:
Feature subset selection
Classification
Gravitational search algorithm
Learning algorithm
Sequential quadratic programming
Chaotic map

ABSTRACT

A new method for feature subset selection in machine learning, FSS-MGSA (Feature Subset Selection by Modified Gravitational Search Algorithm), is presented. FSS-MGSA is an evolutionary, stochastic search algorithm based on the law of gravity and mass interactions, and it can be executed when domain knowledge is not available. A wrapper approach, over Naive-Bayes, ID3, K-Nearest Neighbor and Support Vector Machine learning algorithms, is used to evaluate the goodness of each visited solution. The key to the success of the MGSA is to utilize the piecewise linear chaotic map for increasing its diversity of species, and to use sequential quadratic programming for accelerating local exploitation. Promising results are achieved in a variety of tasks where domain knowledge is not available. The experimental results show that the proposed method has the ability of selecting the discriminating input features correctly and can achieve high accuracy of classification, which is comparable to or better than well-known similar classifier systems. Furthermore, the MGSA is tested on ten functions provided by CEC 2005 special session and compared with various modified Gravitational Search Algorithm, Particle Swarm Optimization, and Genetic Algorithm. The obtained results confirm the high performance of the MGSA in solving various problems in optimization.

1. Introduction

In the last decade, Feature Subset Selection (FSS) has been applied to the field of classification in the case that the large datasets are often involved. Since redundant or irrelevant features increase the size of search space and make generalization more difficult, the reduction of dimensionality of features has become an imperative task in various areas such as pattern recognition [52,29,66,56,22,25], data mining [12,79,14,23,68], gene selection from microarray data [26,71,2,49,42], text categorization [58,65,36,7,16], multimedia information retrieval [37,45,21,72,84,86], and so on. This dimensionality reduction made by feature subset selection can carry out several advantages for a classification system in a specific task [48]:

– a reduction in the cost of acquisition of the data,
– improvement of the comprehensibility of the final classification model,
– a faster induction of the final classification mode,
– an improvement in classification accuracy.

* Corresponding author. Tel.: +86 134 85325973.
E-mail addresses: hanxiaohong@tyut.edu.cn, jmqchs@sohu.com (X. Han).

http://dx.doi.org/10.1016/j.ins.2014.05.030
0020-0255/

Feature subset selection involves the definition of the most informative and discriminative features in the original data for classification. This can be performed by eliminating the redundant, uninformative, and noisy features. Feature subset selection algorithms can be classified into two categories based on whether or not feature selection is done independently of the learning algorithm used to construct the classifier. If feature selection is done independent of the learning algorithm, the technique is said to follow a filter approach. Otherwise, it is said to follow a wrapper approach. While the filter approach is generally computationally more efficient than the wrapper approach, its major drawback is that an optimal selection of features may not be independent of the inductive and representational biases of the learning algorithm to be used to construct the classifier [11,44]. The wrapper approach, on the other hand involves the computational overhead of evaluating candidate feature subsets by executing a selected learning algorithm on the dataset represented using each feature subset under consideration [3]. Wrapper methods select the subset of features using learning algorithms which can promise better results than filter methods in terms of classification accuracy. [35]. In this paper, we adopt a wrapper method by using an optimization algorithm for feature subset selection.

Feature subset selection can be viewed as a search problem, with each state in the search space specifying a subset of the possible features of the task. Exhaustive evaluation of possible feature subsets is usually unfeasible in practice because of the large amount of computational effort required. Many search techniques have been proposed to solve the FSS problem when there is no knowledge about the nature of the task, carrying out an intelligent search in the space of possible solutions. FSS based on stochastic and evolutionary search algorithms has attracted great attention. Raymer and Punch [64] suggested using Genetic Algorithms (GA) to tackle the problem. Authors in [4,75,78] proposed to use binary Particle Swarm Optimization (PSO) for FSS. Zhang and Sun applied tabu search in this problem [85].

Gravitational Search Algorithm (GSA) is one of the newest evolutionary optimization algorithms inspired by the Newton's Second Law of gravity and motion, which first proposed by Rashedi et al. [62]. The principle of gravitational search algorithm is that each object in the search space attracts every other one with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. Further, Rashedi et al. [63] presented the Binary Gravitational Search Algorithm (BGSA), which is a slightly modified algorithm of the original gravitational search algorithm.

Gravitational Search Algorithm is an effective optimization algorithm compared with other evolutionary algorithms [62]. However, there are two major issues in terms of the search performance of gravitational search algorithm: one is premature convergence happening in original gravitational search algorithm due to rapid reduction of diversity; the other is that gravitational search algorithm converges rapidly in the beginning of the search process while slows down quickly when the global best solution is near the optimum of the local search space. Consequently, more ineffective iterations are done to get a more accurate estimation of the local optima; in addition, it is hard to have a good balance between exploration and exploitation. Faced with these drawbacks, researchers have made much efforts to improve it [6,13,17,24,27,28,30,34,38,40,41,67,69,70,83]. Sarafrazi et al. [67] introduced an operator called "Disruption" originating from astrophysics to improve the exploration and exploitation abilities of the original gravitational search algorithm. Li and Zhou [38] strengthened the algorithm by combining the search strategy of particle swarm optimization and applied it to the solution of optimization problems in parameters identification of hydraulic turbine governing system. Shaw et al. [69] proposed an opposition-based GSA which improves the convergence rate of the gravitational search algorithm by utilizing opposition-based learning for population initialization and also for generation jumping. Yin et al. [83] integrated two strategies, the maximum and minimum of the dth dimension and a new $G(t)$, into the original GSA to add the objects' diversity and make them explore more solution spaces. Chen et al. [6] proposed an improved gravitational search algorithm which accelerates convergence speed through integrating the search strategy of particle swarm optimization and elastic-ball method into the original GSA. Dowlatshahi et al. [17] suggested a discrete gravitational search algorithm by using a Path Relinking strategy instead of the classic way in which the objects of gravitational search algorithm usually move from their current position to the position of other objects. In [70], an opposition-based learning is employed for population initialization and also for generation jumping. In [24] a novel fuzzy gravitational search algorithm was presented for optimal design of multimachine power system stabilizers. Authors in [40] proposed a piecewise function based gravitational search algorithm and applied it in parameter identification of automatic voltage regulator system, in which a piecewise function is designed as the gravitational constant function to replace the traditional exponential equation for the purpose of providing more rational gravitational constant to control the convergence of algorithm. In [34] a chaotic gravitational search algorithm was proposed to improve the forecasting performance. The algorithm employs the chaotic local search in the iteration of the original GSA to search and refine the current best solution. David et al. [13] suggested a new gravitational search algorithm with improved search accuracy which is characterized by the modification of the denominator in the expression of the force acting on an object from the other object. Li et al. [41] proposed a novel gravitational search algorithm to search the better parameters of coarse antecedent membership function around the coarse results, in which chaotic search is embedded in the iteration of original gravitational search algorithm to search and replace the current best solution of gravitational search algorithm.

In this paper, we present a modified gravitational search algorithm by integrating PieceWise Linear chaotic map (PWL) and sequential quadratic programming into original GSA to enhance its performance. First, the piecewise linear chaotic map is combined with gravitational search algorithm to form a PWL_GSA, and then we use sequential quadratic programming to accelerate the local search for the global best found by PWL_GSA to do exploitation. This integration forms a new algorithm, called Modified Gravitational Search Algorithm (MGSA). Since the modified gravitational search algorithm is