Neurocomputing 146 (2014) 113–124

Contents lists available at ScienceDirect

Neurocomputing

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

Multi-objective unsupervised feature selection algorithm utilizing redundancy measure and negative epsilon-dominance for fault diagnosis

Hu Xia, Jian Zhuang*, Dehong Yu

School of Mechanical Engineering, Xi'an Jiaotong University, Xianning West Road, 710049 Xi'an, PR China

ARTICLE INFO

Article history:
Received 30 July 2013
Received in revised form
5 June 2014
Accepted 7 June 2014
Available online 14 July 2014

Keywords:
Multi-objective evolutionary algorithm
Feature selection
Redundancy measure
Negative epsilon-dominance
Fault recognition

ABSTRACT

The multi-objective evolutionary algorithm (MOEA) has shown remarkable capability of selecting feature subset. Most MOEAs use the cardinality of the feature subset as one of its objectives and adopt a strict Pareto dominance relationship to select individuals. However, these techniques limit available solutions and may omit several appropriate but dominated solutions. A multi-objective unsupervised feature selection algorithm (MOUFSA) is proposed to solve these issues. A new objective, which incorporates the correlation coefficient and cardinality of the feature subset, not only evaluates the redundancy of selected features but also provides several objective values for each particular size of feature subset. A relaxed archiving strategy based on negative epsilon-dominance and the box-based method is designed to preserve promising solutions even if they are dominated. Three new mutation operators of different abilities are also presented to enhance the algorithm. Nine UCI datasets and five fault recognition datasets are employed as test objects, and the obtained feature subsets are then used for subsequent classification and clustering. Experimental results show that MOUFSA outperforms several other multi-objective and traditional single-objective methods.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

More and more objects in commercial affairs [1–3], image processing [4–6], and fault diagnosis [7–9] are so complex that many characteristics are required to distinguish one object from the others. Different technologies for signal processing, such as statistical analysis in the time domain, Fourier analysis, and wavelet analysis, are available to accomplish description. Thus, the set of feature candidates that represents a complex object is usually high-dimensional, irrelevant, and redundant. An example of this situation is the fault recognition of a mechanical system. Signals reflecting the running state of machinery can be obtained from many sensors with different types or locations, so each sample in the fault dataset is high-dimensional in the feature space. The complex construction of machinery also makes exact information on the fault source difficult to acquire, so features are always noisy, irrelevant and redundant. If we directly use the original dataset, the amount of calculation will be huge, and the recognition result may be unfavorable.

Feature selection is widely used for processing high-dimensional datasets because it can extract relevant features, reduce the price of collection, and increase effectiveness and efficiency in subsequent classification or clustering. Many researchers have proposed their own feature selection methods, which can be divided into three categories, namely, filter methods [10–12], wrapper methods [13–15] and hybrid methods [16–18]. Filter methods select features based on an evaluation function that involves the properties of the dataset but is independent of any classification or clustering algorithm. Wrapper methods evaluate features on the basis of the performance of the learning algorithm applied. Hybrid methods obtain the optimal feature subset by taking advantage of the other two methods, namely, filter methods for their computational cheapness and wrapper methods for their high accuracy. Regarding the search strategies used for these methods, exhaustive search is inefficient and even unacceptable when the set of feature candidates is large, whereas heuristic search is feasible and preferable. Heuristic search includes greedy search [19,20], genetic algorithm [21,22], and particle swarm optimization [23,24]. These heuristic search algorithms usually evaluate a feature subset based on a single objective, but feature selection is a multiple-objective task, wherein it minimizes the cardinality of the feature subset, selects all relevant features and excludes all redundant features. The weighting method can integrate different objectives into a single function, but the weights among the objectives are difficult to automatically identify. The multi-objective evolutionary algorithm is a more attractive method because it can simultaneously achieve different goals and

* Corresponding author. Tel.: +86 29 82668607 804; fax: +86 29 82664344.
E-mail address: stxiahu@gmail.com (J. Zhuang).

http://dx.doi.org/10.1016/j.neucom.2014.06.075
0925-2312/© 2014 Elsevier B.V. All rights reserved.

does not introduce additional parameters. Several multi-objective feature selection algorithms have been proposed in recent years [25–32]. However, these algorithms have the following disadvantages: (1) most of these algorithms adopt the two-objective model. The first objective is used to evaluate the performance of the trail feature subset, and the second objective often minimizes selected features. The available values of the second objective are decided by the size of feature candidates, so the number of obtained solutions is limited and only one solution corresponds to a particular size of the feature subset in most situations; (2) a strict Pareto dominance relationship is used for the multi-objective feature selection algorithm, indicating that non-dominated solutions are better than dominated ones. However, the objectives we construct seldom completely agree with the real properties of the dataset, so the error may result in the non-dominated solutions being worse than the dominated ones. Even if the non-dominated solutions are better, some dominated solutions may still be useful for decision makers because they have similar performance and their selected features are easy to obtain.

In this study, a multi-objective unsupervised feature selection algorithm (MOUFSA) for the filter method is proposed to solve the above-mentioned issues and enhance performance. We use a two-objective model wherein the first objective is an entropy-based measure [19], and the second objective includes an average correlation coefficient and the cardinality of the feature subset. The second objective reduces the redundancy of the selected features and produces many solutions for each particular size of the feature subset. A new archiving strategy is also proposed to update the evolutionary population, where a negative epsilon-dominance method is used to preserve several promising dominated solutions to the next population and a box-based method is presented to maintain the diversity of the population. Three different mutation strategies are designed to reduce the size of the feature subset, remove redundant features and enhance the diversity of the population, respectively. Experimental results on many UCI datasets and fault recognition datasets on a reciprocating compressor show the superior performance of our proposed method.

The organization of this paper is as follows. Section 2 introduces several related concepts and reviews the genetic algorithm for feature selection. Section 3 describes our proposed multi-objective unsupervised feature selection algorithm. Section 4 tests many UCI and fault recognition datasets, and analyzes the results. Section 5 concludes the paper.

2. Related work

2.1. Multi-objective optimization and epsilon-dominance

Given that a maximization problem can be easily transformed into a minimization problem, we discuss only minimization. Multi-objective optimization minimizes a vector of objective functions $\boldsymbol{F}(\boldsymbol{x})=(F_1(\boldsymbol{x}),\ldots, F_m(\boldsymbol{x}))$ in conflict with each other, where $\boldsymbol{F}\in \boldsymbol{Q}^m \subseteq \boldsymbol{R}^m$, and $\boldsymbol{x}\in \boldsymbol{Q}^n \subseteq \boldsymbol{R}^n$ and $\boldsymbol{x}=(x_1,\ldots, x_n)$ are the vectors of the decision variables. The Pareto dominance relationship is defined as follows: for two points $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$, if $F_i(\boldsymbol{x}^1)\le F_i(\boldsymbol{x}^2)$ for all $i=1,2,\ldots, m$ and $F_i(\boldsymbol{x}^1)<F_i(\boldsymbol{x}^2)$ for at least one objective, $\boldsymbol{x}^1$ is said to dominate $\boldsymbol{x}^2$ ($\boldsymbol{x}^1 \prec \boldsymbol{x}^2$); if both $\boldsymbol{x}^1 \prec \boldsymbol{x}^2$ and $\boldsymbol{x}^2 \prec \boldsymbol{x}^1$ cannot be satisfied, we say $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ are non-dominated in relation to each other. Point $\boldsymbol{x}*$ is a non-dominated solution if no point in the domain of $\boldsymbol{Q}^n$ dominates it. All the non-dominated solutions compose a set of Pareto optimal solutions called Pareto Set (PS), whereas the set of its corresponding images is called Pareto Front (PF).

Epsilon-dominance is a relaxed form of Pareto dominance: for a vector of positive values $\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots, \varepsilon_m)$, point $\boldsymbol{x}^1$ epsilon-dominates a point $\boldsymbol{x}^2$ ($\boldsymbol{x}^1 \prec_{+\varepsilon} \boldsymbol{x}^2$) if $F_i(\boldsymbol{x}^1)-\varepsilon_i \le F_i(\boldsymbol{x}^2)$ for all $i=1,2,\ldots,m$ and $F_i(\boldsymbol{x}^1)$-$\varepsilon_i < F_i(\boldsymbol{x}^2)$ for at least one objective. Correspondingly, if $F_i(\boldsymbol{x}^1)+\varepsilon_i \le F_i(\boldsymbol{x}^2)$ for all $i=1,2,\ldots,m$ and $F_i(\boldsymbol{x}^1)+\varepsilon_i < F_i(\boldsymbol{x}^2)$ for at least one objective, we say $\boldsymbol{x}^1$ negative epsilon-dominates $\boldsymbol{x}^2$ ($\boldsymbol{x}^1 \prec_{-\varepsilon} \boldsymbol{x}^2$).

Laumanns et al. [33] applied epsilon-dominance to the archiving strategy of an MOEA. Diversity was maintained by dividing the objective space into boxes based on the epsilon value, only one solution was retained in each box, and convergence was guaranteed by permitting replacement within a box if the new solution dominates the original one. Deb et al. [34] proposed an epsilon-MOEA that can rapidly converge within a limited computational cost, and diversity would only slightly deteriorate with increasing objectives. However, epsilon-dominance has the following disadvantages: the predefined epsilon value is not competent for different PFs, the extreme points and almost horizontal and vertical parts of PF are easy to lose, and the final obtained solutions are fewer than the theoretical number. Therefore, Hernández-Díaz et al. [35] proposed an adaptive epsilon-dominance method whereby the geometrical shape of the PF curve is approximated and the epsilon value adaptively adjusts according to the geometrical characteristics. Schütze et al. [36] introduced a tight value to measure the distances between a trail solution and archive members to achieve the same goal. Negative epsilon-dominance was first proposed by Schütze et al. [37] to preserve several near-optimal and Pareto-optimal solutions because they can provide many options to the decision maker. This method has been used to solve {0, 1}-knapsack problems [38] and space mission design issues [39]. Table 1 shows the procedure of updating an archive with a trail solution $\boldsymbol{y}$ based on negative epsilon-dominance. The solution $\boldsymbol{y}$ is approved if it is a dominated one but is unfavorable in a range of tolerance decided by the epsilon value, whereas several archive members are removed only if the solution $\boldsymbol{y}$ is appropriate enough. Thus, the negative epsilon-dominance produces a "thick" relaxed PF.

2.2. Review of genetic algorithm for feature selection

The genetic algorithm is effective in global searches and in the solution of NP-hard problems, such as feature selection. The binary-encoded mode is often adopted to represent features, where the *i*th gene corresponds to the *i*th feature, a value of "1" for the gene indicates that the feature is included in the subset, and a value of "0" indicates that the feature is discarded. Various crossover, mutation, and even local search operators are used to produce new individuals. The selection operator is used to construct a population for the next generation and guide the search direction based on the objective values of all individuals. Genetic algorithms for feature selection can be divided into single-objective and multi-objective methods. Single-objective methods have been developed for many years [40–43]. Huang et al. [21] used an advanced mutual information measure as the objective of

Table 1
Archiving strategy based on negative epsilon-dominance.

Input: Archive $\mathbf{A}$, trail solution $\boldsymbol{y}$
if $\exists \boldsymbol{x}\in \mathbf{A}$ such that $\boldsymbol{x} \prec_{-\varepsilon} \boldsymbol{y}$
  $\mathbf{A}' = \mathbf{A}$
Else
  $\mathbf{D}=\{\boldsymbol{x}\in \mathbf{A} \mid \boldsymbol{y} \prec_{-\varepsilon} \boldsymbol{x}\}$
  $\mathbf{A}=\mathbf{A}\cup \boldsymbol{y} \setminus \mathbf{D}$
  $\mathbf{A}' = \mathbf{A}$
end if
Output: $\mathbf{A}'$

Download English Version:

https://daneshyari.com/en/article/406575

Download Persian Version:

https://daneshyari.com/article/406575

Daneshyari.com