2012 IEEE fifth International Conference on Advanced Computational Intelligence(ICACI)
October 18-20, 2012 Nanjing, Jiangsu, China

Nearest Neighbor Classification of Pareto Dominance in Multi-objective Optimization

Guanqi Guo, Cheng Yin, Tanshan Yan and Wu Li

***Abstract*—This study investigates the nearest neighbor classification of predicting Pareto dominance relationships in multi-objective optimization. A similarity measurement called the sum of ranked dimensional sequential number is proposed. It transfers the original domains of each decision components into the same integer interval [0:*N*-1], where *N* is the size of sample set. Each decision component of a sample candidate solution is assigned an integer between 0:*N*-1 according to the relative distance from the component to the same dimensional component of a observed candidate solution. The sum of the integers of all decision components of a sample candidate is defined as the similarity measurement. The nearest neighbor classification algorithms using different similarity measurements are tested. The experiments show that the sum of ranked dimensional sequential number is more efficient similarity expression than the Euclidian distance. The nearest neighbor classification uses the proposed similarity is a competent method for predicting Pareto dominance.**

I. Introduction

EVOLUTIONARY algorithms are commonly used to solve multi-objective optimization problems in recent years. A number of well-known multi-objective optimization evolutionary algorithms (MOEAs) such as NSGA-II [1], SPEA-II [2], PAES [3], MOPSO [4] have been proposed. These algorithms determine Pareto dominance relationships among candidate solutions in the population by computing and comparing their objective vectors, recognize all the non-dominated solutions in the population. By iteratively applying the evolutionary operators on the populations, it drives the non-dominated set to converge to the Pareto front asymptotically.

However, for some extremely complicated optimization problems, evaluating an objective vector or a constraint condition is no more than computation of simple numeric function, but a time consuming procedure often taking hours or days [5]. The overwhelming computation cost for evaluating a large number of objective vectors and constraint functions causes the popular MOEAs are almost not competent to large scale complicated multi-objective optimization tasks.

Guo [6] proposed to ease the curse of computation cost by predicting Pareto dominance. Instead of evaluating and comparing objective vectors, the Pareto dominance relationships among candidate solutions are predicted by pattern classification. Based on the assumption that class-conditional probability follows normal distributions, Guo preliminarily implemented a kind of Bayesian classifiers by the criterion of minimizing classification error rate and average risk respectively. The classifier obtained acceptable prediction accuracy on SCH [7] problem. However, in most complex MO problems, the distribution of class-conditional probability is unknown a prior, the assumption of normal distribution causes a number of predicting errors inevitably.

Following the success of binary nearest neighbor classification [8] using similarity measurement based on the weighted sum of binary bit string, this paper proposes the framework of nearest neighbor classification for predicting Pareto dominance and a new method measuring similarity using the sum of ranked dimensional sequential number. The nearest neighbor classification algorithms based on Euclidian distance, the weighted sum of binary bit string and the sum of ranked dimensional sequential number are tested and compared. The rest of the paper is organized as follows. Section II introduces the Pareto dominance concept and describes the framework of nearest neighbor classification of Pareto dominance. Section III reviews binary nearest neighbor classification and proposes the new similarity measurement method. Section IV describes the experiments and Section V concludes this study.

II. Nearest Neighbor Classification of Pareto Dominance

A. Pareto Dominance

A general multi-objective optimization problem can be described as a vector function **F** that maps a space of *n*-dimensional decision vectors to a space of *m*-dimensional objective vectors. Formally:

$$\begin{aligned} &\min \mathbf{y} = \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_m(\mathbf{x})), \\ &s.t.\ \ \mathbf{x} = (x_1, x_2, \cdots, x_n) \in X, \\ &\qquad \mathbf{y} = (y_1, y_2, \cdots, y_m) \in Y. \end{aligned} \tag{1}$$

Where, **x** is called the decision vector, X is the decision space, **y** is the objective vector, and Y is the objective space.

Manuscript received May 9, 2012. This work was supported by the National Natural Science Foundation of China under grant No.60975049 and 30971570, the Natural Science Foundation of Hunan Province under grant key project No.11JJ2037, and the aid program for Science and Technology Innovative Research Team in Higher Educational Instituions of Hunan Province.

G Guo, corresponding author, is with the College of Information and Communication Engineering, Hunan Institute of Science and Technology, Yueyang 414006, Hunan, PRC (phone: 0730-8648851; fax: 0730-8646324; e-mail: gq.guo@163.com)

C Yin is with the College of Information and Communication Engineering, Hunan Institute of Science and Technology, Yueyang 414006, Hunan, PRC (e-mail: yincheng886337@163.com)

T Yan is with the College of Information and Communication Engineering, Hunan Institute of Science and Technology, Yueyang 414006, Hunan, PRC (e-mail: yantaishan163@163.com)

W Li is with the College of Information and Communication Engineering, Hunan Institute of Science and Technology, Yueyang 414006, Hunan, PRC (e-mail: liwu0817@163.com)

978-1-4673-1744-3/12/$31.00 ©2012 IEEE

For arbitrary decision vectors **u, v**$\in X$, iff $\forall i \in \{1,2,...,m\}$: $f_i(\mathbf{u}) \leq f_i(\mathbf{v})$ and $\exists j \in \{1,2,...,m\}$: $f_j(\mathbf{u}) < f_j(\mathbf{v})$, **u** is called to dominate **v,** denoted as $\mathbf{u} \prec \mathbf{v}$; or **v** dominated by **u**, denoted as $\mathbf{u} \prec \mathbf{v}$. Otherwise, **u** and **v** are incomparable, denoted as $\mathbf{u} \sim \mathbf{v}$.

Iff $\neg\exists \mathbf{x}' \in X: \mathbf{x}' \prec \mathbf{x}$, **x** is referred as to a Pareto optimal solution (namely Pareto non-dominated solution or non-inferior solution). The set of all Pareto optimum in the decision space X is referred as to Pareto optimal set, and the set of the corresponding objective vectors is called Pareto front or Pareto optimal surface.

Multi-objective optimization algorithms based on Pareto optimality search the uniformly distributed non-dominated solutions, guide the candidate solution set approximating to the Pareto front, so as to provide the decision-maker with as many optional non-dominated solutions as possible.

B. Nearest Neighbor Classification

The most direct method of determining Pareto dominance relationships is to evaluate and compare the corresponding objective vectors candidate solutions. For saving the computation cost of evaluating a large number of objective vectors, or in the situation lacking the analytic model of objective function, Pareto dominance can be determined by prediction as long as enough samples is available.

For a multi-objective optimization problem, given arbitrary candidate solutions $\mathbf{x}_1$ and $\mathbf{x}_2$ in the decision space, their Pareto dominance relationship is denoted as $\omega(\mathbf{x}_1,\mathbf{x}_2)$. Clearly, $\omega(\mathbf{x}_1,\mathbf{x}_2)$ can be classified into three categories $\prec$, $\succ$ and $\sim$ from the Pareto optimality concept.

Given two candidate solution sample sets S_1 and S_2 respectively with size N, the Pareto dominance relationship $\omega(\mathbf{s}_1,\mathbf{s}_2)$ of the sample $\mathbf{s}_i \in S_1$ and $\mathbf{s}_j \in S_2$ can be determined by evaluating and comparing their objective vectors, where $i, j \in \{1{:}N\}$. By appropriate classifying algorithm, the Pareto dominance of arbitrary observed candidate solutions $\mathbf{x}_1$ and $\mathbf{x}_2$ can be predicted by learning the samples. The framework of nearest neighbor classification algorithm predicting Pareto dominance is described as follows:

Step1. Randomly generating two sample sets S_1 and S_2, each set includes N candidate solutions.

Step2. Determining all $\omega(\mathbf{s}_i,\mathbf{s}_i)$ by evaluating and comparing the objective vector of $\mathbf{s}_i \in S_1$ and $\mathbf{s}_j \in S_2$, where $i, j \in \{1{:}N\}$.

Step2. Find out the most similar sample $\mathbf{s}_1$ to $\mathbf{x}_1$ from S_1;

Step3. Find out the most similar sample $\mathbf{s}_2$ to $\mathbf{x}_2$ from S_2;

Step4. Let $\omega(\mathbf{x}_1,\mathbf{x}_2) = \omega(\mathbf{s}_1,\mathbf{s}_2)$.

The nearest neighbor classification algorithm for predicting Pareto dominance is simple. However, similar to the canonical the nearest neighbor rule in the field of pattern recognition, the classifying accuracy strongly depends on the similarity measurement of the decision vectors, which plays a deterministic part in predicting Pareto dominance.

III. SIMILARITY MEASUREMENT

The Euclidian distance is the simplest and most common method measuring the similarity in the nearest neighbor rule for pattern recognition. But with respect to Pareto dominance prediction, it is not suitable for those decision spaces in which the distance of two decision vectors is dominated by some dimensions. Binary nearest neighbor classification (BNNC) for predicting Pareto dominance encodes each dimension x_i of the n–dimensional decision vector into a string with l binary bits. The value of x_i is calculated by

$$x_i = a_i + \frac{(b_i - a_i) \times \sum_{k=1}^{l} x_{ik} \times 2^{k-1}}{2^l - 1} . \quad (2)$$

Where, $[a_i, b_i]$ is the domain of x_i. The similarity $D(\mathbf{x},\mathbf{y})$ of two decision vectors **x** and **y** is measured using the weighted sum of binary bit string calculated by

$$D(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} \sum_{k=1}^{l} (x_{ik} \oplus y_{ik}) \times 2^{k-1} . \quad (3)$$

Where, $\oplus$ is the logical exclusive or operator.

The analysis and experiments [8] shows that the weighted sum of binary bit string expresses the contribution of each dimension of the decision vector to the similarity measurement more fairly than Euclidian distance. Hence BNNC is able to get higher classification accuracy than the nearest neighbor classification using Euclidian distance (ENNC) in those problems of which the domains of the components of the decision vector differ in orders of magnitude. However, for lacking the advantage of precision of real value representation, BNNC is inferior to ENNC, as the domains of each decision components locate in the roughly identical regions

With respect to multi-objective function optimization problem, for taking advantage of real value representation, we propose a new similarity measurement called the sum of ranked dimensional sequential number. Given a sample set S of size N, the jth component of the ith sample is denoted as s_{ij}. Given a decision vector **x**, first compute $|x_j - s_{ij}|$, where $i{=}1{:}N$ and $j{=}1{:}n$. Then for each dimension, each s_{ij} is assigned a sequential number r_{ij} ranging from 0 to N-1 by ascending order of $|x_j - s_{ij}|$. The similarity $D(\mathbf{x},\mathbf{s}_i)$ of the vector **x** and the sample $\mathbf{s}_i$ is measured using the sum of ranked dimensional sequential number calculated by

$$D(\mathbf{x},\mathbf{s}_i) = \sum_{j=1}^{n} r_{ij} . \quad (4)$$

From the assignment criterion of the dimensional sequential number r_{ij}, it is clearly seen that the most similar sample to **x** is that with minimal $D(\mathbf{x},\mathbf{s}_i)$.

In fact, the characteristic of the weighted sum of binary bit string and the sum of ranked dimensional sequential number is similar. The former transfers the original domains of each dimension into the same intervals $[0{:}2^{l-1}]$, the latter adopts the intervals $[0{:}N{-}1]$. It means that both methods measure the distance of two components on the same dimension by identical coordinate scale, so as to avoid the similarity measurement is dominated by part of dimensions. Certainly, the benefits of using the sum of ranked dimensional sequential number are that it saves the computation cost of encoding and decoding binary string and reserves the precision advantage of real value.

IV. EXPERIMENTS

Five multi-objective optimization problems MOP1~5 are used to test the prediction accuracy of nearest neighbor

classification using different similarity measurement. They are defined as follows:

$$\text{MOP1: } \min\ f_1(x) = x^2,$$
$$f_2(x) = (x-2)^2;$$
$$-10^5 \le x \le 10^5.$$

$$\text{MOP3: } \min\ f_1(x_1,x_2) = (x_1-2)^2/2 + (x_2+1)^2/13 + 3,$$
$$f_2(x_1,x_2) = (x_1+x_2-3)^2/36 + (-x_1+x_2+2)^2/8 - 17,$$
$$f_3(x_1,x_2) = (x_1+2x_2-1)^2/15 + (2x_2-x_1)^2/17 - 13;$$
$$-4 \le x_1, x_2 \le 4.$$

$$\text{MOP4: } \min f_1(x_1,x_2) = x_1,$$
$$f_2(x_1,x_2) = (1+x_2)/x_1;$$
$$x_1 \in [0,0.1], x_2 \in [0,5],[0,50],[0,500].$$

$$\text{MOP5: } \min f_1(x_1,x_2) = x_1,$$
$$f_2(x_1,x_2) = (1+10x_2)[1-(x_1/(1+10x_2))^{\alpha}$$
$$-x_1\sin(2\pi q x_1)/(1+10x_2)],$$
$$q = 4, \alpha = 2,\ x_1 \in [0,1],$$
$$x_2 \in [0,1],[0,10],[0,100],[0,1000].$$

The domains of all the decision components for each of the MOP1~3 have the same order of magnitude. With respect to MOP4 and MOP5, x_1 takes the fixed domain, but x_2 takes various domains which differ in orders of magnitude.

For each of the MOP1~3 and each group domains of x_1 and x_2 for MOP4 and MOP5, two sample sets respectively including 100 candidate solutions are randomly generated. It is used to construct 100×100 Pareto dominance samples determined by evaluating and comparing the corresponding objective vectors. The sample sets are used to classify the Pareto dominance among 40×40 observed solutions randomly initialized. Three kinds of nearest neighbor classification algorithms ENNC, BNNC and RNNC (using similarity measurement by the sum of ranked dimensional sequential number) are respectively tested for 100 trials. Each trial uses newly generated random sample sets, but the fixed 40×40 observed solutions.

For each problem, the average class proportions of the three kinds of Pareto dominance relationships $\prec$, $\succ$ and $\sim$ over 100 trials are listed in Table I. The average classification accuracy rates over 100 trials classifying the 40×40 observed solutions are listed in Table II for MOP1-3 and Table III for MOP4-5.

Though MOP1-3 have different complexity from the view points of the modality of the objective functions, the dimensions of objective vectors and decision vectors, the common characteristic is that the domains of each decision components locate in the same coordinate region. It reflects that each decision components have the roughly identical contribution to the similarity measurement, no matter what kind of distance is used. Hence the three classifying algorithms based on similarity measurement shows the roughly identical prediction accuracy rates as indicated in Table II. The observation that BNNC is slightly inferior to ENNC and RNNC is not hard to be explained, it is caused by the precision loss of encoding and decoding the binary string. The worse prediction performance for MOP2 is caused by the overall complexity originated from the extremely multiple modals of f_2 and the high dimensionality of the decision vector. The additional experiments show that no matter what classification method is selected, the classification accuracy decreases as the modals of objective vector or the dimensions of the decision vector increases. The dimensionality curse of decision vector and the modality curse of objective vector remain to be resolved by further investigation.As indicated in Table III, ENNC, BNNC and RNNC show notable difference in prediction accuracy on MOP4 and MOP5. When the domains of decision components differ in order of magnitude, the prediction accuracy of ENNC is obviously lower than that

TABLE I
AVERAGE CLASS PROPORTION OVER 100 SAMPLE SETS

Problems	$\prec$ (%)	$\succ$ (%)	$\sim$ (%)
MOP1	34.08	33.80	32.12
MOP2	29.75	29.67	40.57
MOP3	26.48	26.31	47.20
MOP4	11.20	11.10	77.70
	14.21	14.17	71.62
	14.10	14.79	71.12
MOP5	25.13	24.85	50.03
	25.15	25.19	49.67
	25.12	24.97	49.91
	24.61	25.69	49.70

TABLE II
AVERAGE ACCURACY RATES OF ENNC, BNNC AND RNNC ON MOP1-3

Problems	Algorithms	Average accuracy rates (%)		
		$\prec$	$\succ$	$\sim$
MOP1	ENNC	98.45	97.57	97.92
	BNNC	97.48	97.61	97.91
	RNNC	98.22	98.84	98.50
MOP2	ENNC	66.51	67.17	58.20
	BNNC	56.35	59.96	52.09
	RNNC	66.27	70.79	57.53
MOP3	ENNC	87.38	84.51	85.44
	BNNC	80.98	82.64	81.21
	RNNC	85.90	84.36	85.97

TABLE III
AVERAGE ACCURACY RATES OF ENNC, BNNC AND RNNC ON MOP4-5

Problems	Domains	Algorithms	Average accuracy rates (%)		
			$\prec$	$\succ$	$\sim$
MOP4	$x_1 \in [0.0.1]$	ENNC	61.30	70.16	90.15
	$x_2 \in [0,5]$	BNNC	71.51	70.84	93.40
		RNNC	76.40	80.66	94.95
	$x_1 \in [0,0.1]$	ENNC	50.63	42.54	78.80
	$x_2 \in [0,50]$	BNNC	70.07	78.76	91.53
		RNNC	85.89	81.83	93.09
	$x_1 \in [0,0.1]$	ENNC	38.41	39.08	73.92
	$x_2 \in [0,500]$	BNNC	80.51	78.26	89.46
		RNNC	83.74	83.62	92.65
MOP5	$x_1 \in [0,1]$	ENNC	88.48	91.40	89.48
	$x_2 \in [0,1]$	BNNC	86.86	83.69	86.07
		RNNC	89.63	89.58	92.07
	$x_1 \in [0,1]$	ENNC	78.74	84.66	83.59
	$x_2 \in [0,10]$	BNNC	87.59	87.87	86.07
		RNNC	89.06	89.90	90.88
	$x_1 \in [0,1]$	ENNC	56.35	58.44	60.51
	$x_2 \in [0,100]$	BNNC	87.53	87.87	86.12
		RNNC	90.13	90.12	90.84
	$x_1 \in [0,1]$	ENNC	50.07	50.26	50.10
	$x_2 \in [0,1000]$	BNNC	83.68	84.30	88.33
		RNNC	90.70	90.59	89.60

of BNNC and RNNC. Under the condition of fixing the domain of x_1, the performance of ENNC degenerates rapidly as the domain of x_2 broadens in orders of magnitude, but BNNC and RNNC exhibit overall consistent or slightly increasing prediction accuracy.

Table I shows the sample set of MOP4 have unbalanced class proportions of Pareto dominance relationships. Table III shows that ENNC is almost unable to get the acceptable accuracy in minor classes such as $\prec$ and $\succ$, especially when the domain of x_2 is greatly broader then x_1. It is clear that BNNC and RNNC are not so sensitive to the distribution of class proportions, they are able to obtain acceptable performance on the minor classes, no matter how to define x_1 and x_2.

V. Conclusions

Multi-objective optimization based on Pareto optimality needs to determine Pareto dominance among the candidate solutions. For extremely complicated optimization problems, population-based evolutionary algorithm may be confronted with the curse of computation cost of evaluating and comparing a large number of objective vectors. This study investigates nearest neighbor classification methods of predicting Pareto dominance relationships. For determining the most similar solution of a candidate, a new kind of similarity measurement called the sum of ranked dimensional sequential number is proposed. The experiments comparing the average classification accuracy of ENNC, BNNC and RNNC show that RNNC and BNNC are more efficient and robust than ENNC, no matter how each decision components of the optimized problem are defined. In addition, RNNC and BNNC are more competent than ENNC to get acceptable prediction accuracy on the problems with unbalance classes of Pareto dominance. RNNC has the precision advantage of real value representation and no need of encoding and decoding decision vector, but for some problems such as combination optimization, BNNC may be more appropriate than RNNC for BNNC uses binary representation of decision vector.

References

[1] K. Deb et al, "A fast and elitist multi-objective genetic algorithm: NSGA-II," *IEEE Trans. Evolutionary Computation*, vol.6, no.2, pp. 182-197, Apr. 2002.

[2] E. Zitzler, M. Laumanns and L.Thiele, "SPEA2: Improving the strength pareto evolutionary algorithm," in *Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems.* K Giannakoglou et al, eds. Berlin: Springer-Verlag, 2002, pp. 95-100.

[3] J. D. Knowles and D. W. Corne, "The Pareto archived evolutionary strategy: A new baseline algorithm for Pareto multi-objective optimization," in *Proc. CEC99*, Vol.1, Piscataway, NJ: IEEE Press, 1999, pp. 98-105.

[4] C. A. C. Coello, G. T. Pulido and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Trans. Evolutionary Computation,* vol.8, no.3, pp. 256-279, Jun. 2004.

[5] P. K. S. Nain and K. Deb, A multi-objective search and optimization procedure with successive approximate models," KanGAL, Indian Institute of Technology, Kanpur, KanGAL Rep.2005002, 2005.

[6] G. Guo et al, "Predicting Pareto dominance in multi-objective optimization using pattern recognition," in *Proc. 2012 Int. Conf. Intelligent System Design and Engineering Application,* Sanya, China, 2012, pp.456-459.

[7] J. D. Schaffer, " Multiple objective optimization with vector evaluated genetic algorithms," in *Proc. Int. Conf. Genetic Algorithms,* New Jersey: Lawrence Erlbaum, 1987, pp.93-100.

[8] G. Guo, C. Yin, T. Yan and W. Li, "Binary Nearest Neighbor Classification of Predicting Pareto Dominance in Multi-objective Optimization," .in *Proc. 2012 Int. Conf. Swarm Intelligence, Part I, LNCS 7331,* Berlin: Springer-Verlag, 2012, pp.537-545.