TOWARDS WEAKLY PARETO OPTIMAL: AN IMPROVED MULTI-OBJECTIVE BASED BAND SELECTION METHOD FOR HYPERSPECTRAL IMAGERY

Bin Pan[1], Liming Wang[2], Xia Xu[1] and Zhenwei Shi[1]

[1]Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China
[2]State Key Laboratory of Information Security, Institute of Information Engineering Chinese Academy of Sciences, Beijing 100093, China
{panbin,xuxia,shizhenwei}@buaa.edu.cn

ABSTRACT

Band selection refers to finding the most representative channels from hyperspectral images. Usually, certain objective functions are designed and combined via regularization terms. Owing to the parameters independence and the optimal solutions, multi-objective based methods have presented promising performance. However, the characteristics of the hyperspectral band selection problem make its range to be discrete. In this case, recently proposed weighted Tchebycheff based multi-objective band selection methods could only reach the weakly Pareto optimal, which would result in non-unique solutions. In this paper, we improve the decomposition process of the multi-objective based band selection method via a boundary intersection approach. Compared with weighted Tchebycheff decomposition, the proposed method is able to change the shape of the contour lines between Pareto Front and the ideal point, and this approach is particularly suitable for discrete-range problems. The effectiveness of our improvement is demonstrated by comparison experiments.

Index Terms— Multi-objective optimization, band selection, hyperspectral imagery

1. INTRODUCTION

Hyperspectral imagery (HSI) includes abundant spectral information which contributes to many remote sensing applications. However, in some certain circumstances only part of the information is required [1], *i.e.*, bands are redundancy. Thus band selection methods are developed. There are two manners for HSI band selection: supervised and unsupervised. The former usually targets at some particular applications such as classification, while the latter utilizes the hyperspectral data characteristics which have attracted broader research. In this paper we focus on unsupervised band selection.

Thanks to the National Key R&D Program of China under the Grant 2017YFC1405600 and National Natural Science Foundation of China under the Grants 61671037. *Corresponding author: Liming Wang (wangliming@iie.ac.cn) and Zhenwei Shi.*

Researchers usually defined certain objective functions to constrain the bands correlation [1–3]. However, it is difficult to determine which objective is the most effective. Although objectives can be combined via regularization terms, the regularization coefficients usually require manual setting. To overcome this problem, multi-objective (MO) [4] based methods were used for HSI band selection. MO-based methods could optimize several objectives simultaneously without the setting for regularization coefficients. Moreover, MO-based methods can directly generate a series of subsets corresponding to different numbers of bands in a single run, so as to avoid the setting of band numbers. Therefore MO-based methods are promising for HSI band selection.

Recent works about MO-based hyperspectral band selection are not many, where these proposed by Gong *et al.* [5] and Xu *et al.* [6] are two typical ones. These two methods were both developed under the framework of multi-objective evolutionary algorithm based on decomposition (MOEA/D) [7]. Literature [5] (MOBS for short) introduced the MO based method to hyperspectral band selection for the first time, which was an exploratory work. Method of [6] (RMOBS for short) mainly tried to solve the decision making problem in band selection, where its objectives were not completely equal to MOBS.

However, because the range of band selection problem is discrete, MOEA/D is hard to reach Pareto optimal. In other words, in this case several solutions will reach optimal at the same time. This situation is called weakly Pareto optimal [4]. For some real applications such as band selection, we hope that the solution must be unique. Thus weak Pareto solutions should be avoided.

Targeting at the weakly Pareto optimal problem, in this paper a new multi-objective based band selection method is proposed. Mathematically, the weak Pareto solutions of MOEA/D in discrete range problems are generated in the weighted Tchebycheff decomposition process. This decomposition process will lead to rectangular contour lines between Pareto front and the ideal point. Inspired by this observation, we propose a boundary intersection based multi-objective band selection method (BiMOBS). The key idea of BiMOBS is using a boundary intersection approach to improve the decomposition process in [5] and [6], where the shape of contour lines around ideal point is transferred to be elliptical. By this means the weakly Pareto optimal problem is solved in principle.

The major contribution of this paper is that a new boundary intersection decomposition based MO method is developed, which is able to overcome the problem that weighted Tchebycheff decomposition usually only leads to weakly Pareto optimal in HSI band selection.

2. PROPOSED METHOD

2.1. Objectives

One of the advantages of MO-based band selection methods is that they provide a feasible manner to combine several preferred objectives. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_L]^T \in \mathbb{R}^{L \times M}$ denote the HSI data with L bands and M pixels. Define $\mathbf{b} = [b_1, b_2, \cdots, b_L]$ as a binary vector representation for $\mathbf{X}$, where $b_i = 1$ if the i-th band is selected and $b_i = 0$ otherwise. Since the novelty of BiMOBS exists in the process of optimization, for a fair comparison, here we use the same objectives as [5]:

$$\min_{\mathbf{b}} \mathcal{F}(\mathbf{b}) = [f_1(\mathbf{b}), f_2(\mathbf{b})]$$
$$\begin{cases} f_1(\mathbf{b}) = \|\mathbf{b}\|_0 \\ f_2(\mathbf{b}) = \dfrac{1}{|\mathbf{b}|} \displaystyle\sum_{i=1}^{|\mathbf{b}|} \dfrac{1}{H(\mathbf{x}_{\mathbf{b}}^i)} \end{cases} \tag{1}$$

where $H(\mathbf{x}_{\mathbf{b}}^i)$ is the information entropy of the i-th band. BiMOBS tries to optimize the two objectives $f_1(\cdot)$ and $f_2(\cdot)$ simultaneously. It is worth noting that $f_1(\cdot)$ is non-convex and NP-hard. Usually, convex relaxation or greedy algorithms are used for L0 problem. However, these methods cannot guarantee that the obtained solution is optimal. By comparison, BiMOBS can directly handle L0 problem without any relaxation. This is also another advantage of MO-based methods [5, 6, 8].

2.2. Optimization Process

The optimization process of BiMOBS can be described by 3 steps: (1) Initialize a solution set; (2) Calculate an ideal point according to current solutions; (3) Update the solutions based on the boundary intersection decomposition approach and random flipping strategy. During each iteration the solution set is updated. The iterations stop when the solutions keep stable. The final result is called Pareto (optimal) front which is a solution set. As a conference paper, some backgrounds for MO-based methods are not shown. We recommend the readers find more detailed description about MO from [4].

2.2.1. Initialization

BiMOBS begins with an initial solution set $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_p\}$ called *population*, each of the solution in $\mathbf{B}$ is called *individual*, and p is the population size. For the i-th individual a uniformly distributed direction vector $\boldsymbol{\lambda}_i = [\lambda_1, \lambda_2]$ corresponding to the two objectives with $(\lambda_1 + \lambda_2) = 1$ is assigned. Because the objectives in BiMOBS is sparse, in order to get to the real solution as close as possible, we set all the individuals as zero vectors. After several times update, individuals in $\mathbf{B}$ will be diverse.

2.2.2. The Ideal Point

MO-based methods try to make the population become "better" during the iteration process, thus there should be an ideal point to guide the movement of the population. In BiMOBS, an ideal point $\mathbf{z}^*$ is defined by

$$\mathbf{z}^* = (z_1^*, z_2^*)$$
$$\begin{cases} z_1^* = \min f_1(\mathbf{B}) \\ z_2^* = \min f_2(\mathbf{B}) \end{cases} \tag{2}$$

where z_1^* is the minimum $f_1(\cdot)$ value of all the individuals, and so as z_2^*. Obviously, $\mathbf{z}^*$ is usually a virtual point. The target of BiMOBS is forcing all the individuals to get closer to the ideal point. It is worth noting that because the population keep updating during iterations, the ideal point changes in each iteration.

2.2.3. Update

Based on the current ideal point and population, there must be a criterion which could guide the individuals approach the ideal point. In [5] and [6], weighted Tchebycheff decomposition method was used to evaluate the superiority of current individuals. However, in sparse band selection problem, the range is discrete. In this case the weighted Tchebycheff decomposition may suffer from weakly Pareto optimal (shown in Fig. 1 and discussed in section 2.3).

In this paper, we use boundary intersection approach [7] to improve the weakly Pareto optimal problem. The distance between $\mathbf{b}_i$ and $\mathbf{z}^*$ is defined by

$$g(\mathbf{b}_i | \boldsymbol{\lambda}_i, \mathbf{z}^*) = d_1 + \theta d_2, \quad s.t. \quad \mathbf{b}_i \in \boldsymbol{\Omega} \tag{3}$$

where

$$d_1 = \frac{\|(\mathcal{F}(\mathbf{b}_i) - \mathbf{z}^*)\boldsymbol{\lambda}_i^T\|}{\|\boldsymbol{\lambda}_i\|},$$
$$d_2 = \|\mathcal{F}(\mathbf{b}_i) - (\mathbf{z}^* + d_1 \boldsymbol{\lambda}_i)\|. \tag{4}$$

$\theta > 0$ is a preset penalty parameter, and $\boldsymbol{\Omega}$ is the range. Minimizing Eq. (3) is actually the sub-problem for individual $\mathbf{b}_i$. The current population get close to $\mathbf{z}^*$ based on an evolution strategy according to the evaluation results by Eq. (3) and Eq. (4).

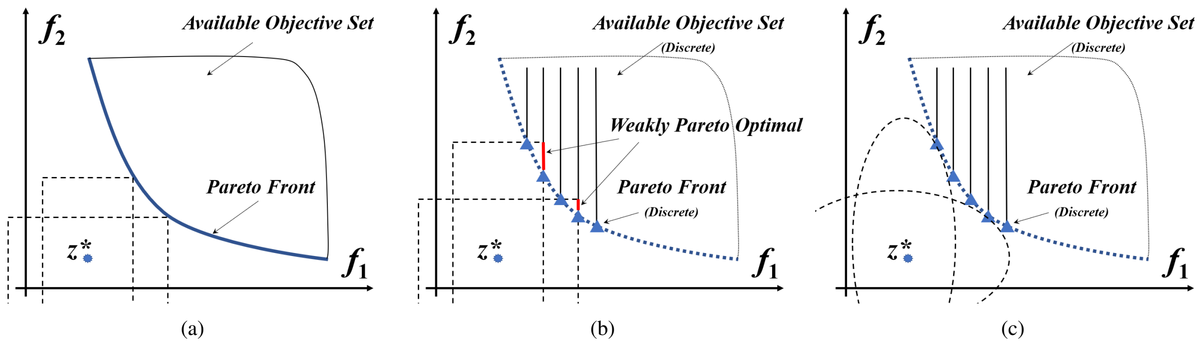

Fig. 1: Illustrations about the final Pareto fronts by different methods and conditions. (a) Weighted Tchebycheff decomposition in continuous range. (b) Weighted Tchebycheff decomposition in discrete range. (c) The proposed method in discrete range.

The evolution strategy used in BiMOBS is random flipping, which is the same as [5, 6]. Here we do not put it into details. After sufficient iterations the population will stabilize in the marginal part of the range, which is called Pareto (optimal) front.

2.3. Analysis and Discussion

In this section, we analyze and discuss why BiMOBS could overcome the weakly Pareto optimal problem. From the definition of weighted Tchebycheff we can conclude that the contour lines are concentric boxes, each of which corresponds to a certain $\boldsymbol{\lambda}_i$. The points under the same box have the same distance to the ideal point, thus they are all optimal. Fig. 1(a) is an illustration for the condition of using weighted Tchebycheff in continuous range. We can see that each box has a unique intersection with the range, *i.e.*, for a certain $\boldsymbol{\lambda}_i$ there is only one solution which is optimal. Collecting all the intersection points we can get the Pareto front.

However, the situation is different when the range is discrete, as is shown in Fig. 1(b). According to Eq. (1), the range of band selection problem should be many parallel rays. In this case, it is almost impossible for the Tchebycheff boxes to just exactly intersect on the edge (shown by triangles in Fig. 1(b)) of the range. More probably, there would be a coincident line with the range, as is shown by red lines in Fig. 1(b). All the available points under the red lines will be selected as the final solutions, *i.e.*, Pareto front. This situation is called weakly Pareto optimal. Theoretically, solutions in the same red line are equivalently optimal. However, this result cannot satisfy the uniqueness of the solution in band selection problem.

By comparison, the weakly Pareto optimal problem can be solved in BiMOBS, as is shown in Fig. 1(c). Geometrically, Eq. (3) and Eq. (4) have actually modified the shape of contour lines in weighted Tchebycheff methods. The proposed method is able to overcome this problem by forcing the contour lines to concentric ellipses. This improvement has little influence when the range is continuous, however, it will lead to a unique intersection in the specific application of band selection.

3. EXPERIMENTS

3.1. Setups

In this section, we evaluate the performance of BiMOBS on two public HSI data sets, Indian Pines and Pavia University[1]. They are both very popular data sets [9], with 200 available bands in Indian Pines and 103 in Pavia University. The parameters p and θ are set as 100 and 0.5 respectively.

Because BiMOBS is improved based on MOBS [5] which adopted weighted Tchebycheff decomposition approach, in this paper we compare their results. Another MO-based band selection method, RMOBS [6], focused on solutions concentration and thus used different objectives. Therefore the comparison with RMOBS is not necessary.

Both BiMOBS and MOBS are general unsupervised band selection methods which do not aim at specific applications. So only the band selection results are shown. As a conference paper, we will show the most meaningful experimental results. We will further show the classification results using the selected bands in our extended journal paper. Comparison with [2, 3, 6] and parameters discussion will also be supplemented.

3.2. Experimental Results

Results on the two data sets are shown in Fig. 2. Since $f_1(\cdot)$ and $f_2(\cdot)$ are error values of Eq. (1), there are no units in this figure. All the results are obtained after 100 iterations. The circles in Fig. 2 denote current solutions. Because the solutions tend stable in these cases, current solution sets could be considered as Pareto optimal front.

However, it can be observed that in Fig. 2(a)(c) there are apparent weak Pareto sets. Take Fig. 2(a) for example. There

[1] Available online: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

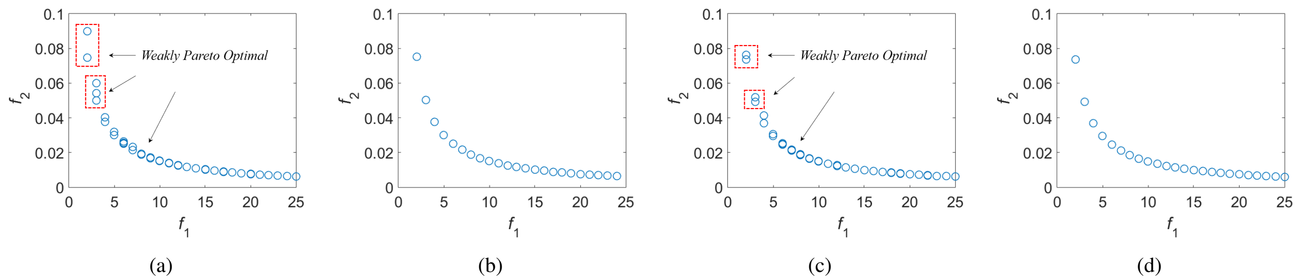

Fig. 2: Band selection results by MOBS on (a) Indian Pines, (c) Pavia University, and BiMOBS on (b) Indian Pines, (d) Pavia University. The circles denote the solutions, and the red boxes are the so-called weakly Pareto optimal.

are 2 and 3 individuals when $f_1(\cdot) = 2$ and $f_1(\cdot) = 3$, respectively. These individuals are all equivalently optimal [4]. This is the so-called weakly Pareto optimal. Fig. 2(a)(c) are also the experimental verification for the assumption and analysis in Fig. 1(b). Specially, some nearly coincident points can be observed in Fig. 2(a)(c), such as $f_1(\cdot) =$ 8-12, 15, 20 in Fig. 2(a) and $f_1(\cdot) =$ 7-10, 12, 18 in Fig. 2(c). Because one solution is a series of selected bands, although these coincident points look like very similar, they may correspond to completely different results. In this case how to determine a unique solution from this weak Pareto set is an open problem.

By comparison, BiMOBS can avoid the weakly Pareto optimal problem. From Fig. 2(b)(d) we can see that there are no redundant individuals. The obtained Pareto front is smooth which will help users find the required solutions. Results in Fig. 2(b)(d) indicate that BiMOBS is able to overcome the weakly Pareto optimal problem in the two test data sets.

4. CONCLUSION

In this paper, a new unsupervised HSI band selection method, BiMOBS, is proposed based on the idea of multi-objective optimization. BiMOBS is inspired by the fact that MO-based methods could generate a series of solutions in a single running. However, weighted Tchebycheff decomposition based methods may suffer from weakly Pareto optimal in band selection, which will lead to the non-uniqueness of the solutions. BiMOBS deals with this problem by introducing a boundary intersection approach to the decomposition process in MOEA/D. Our improvement could force the shape of contour lines around the ideal point from rectangle to ellipse, so as to generate a smooth Pareto front in the band selection problem whose range is discrete. Experiments on two popular HSI data sets have demonstrated the effectiveness of the proposed method.

5. REFERENCES

[1] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1279, 2016.

[2] W. Sun, L. Tian, Y. Xu, D. Zhang, and Q. Du, "Fast and robust self-representation method for hyperspectral band selection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 11, pp. 5087–5098, Nov 2017.

[3] X. Geng, K. Sun, L. Ji, and Y. Zhao, "A fast volume-gradient-based band selection method for hyperspectral image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7111–7119, Nov 2014.

[4] Kaisa Miettinen, *Nonlinear Multiobjective Optimization*, vol. 12, Kluwer Academic Publishers, Dordrecht, 1999.

[5] Maoguo Gong, Mingyang Zhang, and Yuan Yuan, "Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 544–557, 2016.

[6] Xia Xu, Zhenwei Shi, and Bin Pan, "A new unsupervised hyperspectral band selection method based on multiobjective optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2112–2116, 2017.

[7] Qingfu Zhang and Hui Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on evolutionary computation*, vol. 11, no. 6, pp. 712–731, 2007.

[8] Xia Xu and Zhenwei Shi, "Multi-objective based spectral unmixing for hyperspectral images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 124, pp. 54–69, 2017.

[9] Bin Pan, Zhenwei Shi, and Xia Xu, "MugNet: Deep learning for hyperspectral image classification using limited samples," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017.