11-2014

Band selection for hyperspectral images using probabilistic memetic algorithm

Liang FENG

Ah-hwee TAN
Singapore Management University, ahtan@smu.edu.sg

Meng-Hiot LIM

Si Wei JIANG

Soft Comput (2016) 20:4685–4693
DOI 10.1007/s00500-014-1508-1

FOCUS

Band selection for hyperspectral images using probabilistic memetic algorithm

Liang Feng · Ah-Hwee Tan · Meng-Hiot Lim · Si Wei Jiang

Published online: 6 November 2014

Abstract Band selection plays an important role in identifying the most useful and valuable information contained in the hyperspectral images for further data analysis such as classification, clustering, etc. Memetic algorithm (MA), among other metaheuristic search methods, has been shown to achieve competitive performances in solving the NP-hard band selection problem. In this paper, we propose a formal probabilistic memetic algorithm for band selection, which is able to adaptively control the degree of global exploration against local exploitation as the search progresses. To verify the effectiveness of the proposed probabilistic mechanism, empirical studies conducted on five well-known hyperspectral images against two recently proposed state-of-the-art MAs for band selection are presented.

Keywords Hyperspectral image · Band selection · Memetic algorithm

Communicated by Y.-S. Ong.

L. Feng (✉) · A.-H. Tan
Center for Computational Intelligence, School of Computer Engineering, Nanyang Technological University, Singapore, Singapore
e-mail: feng0039@ntu.edu.sg

A.-H. Tan
e-mail: asahtan@ntu.edu.sg

S. W. Jiang
Institute of Manufacturing Technology (SIMTech), Singapore 638075, Singapore
e-mail: jiangsw@simtech.astar.edu.sg

M.-H. Lim
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore
e-mail: emhlim@ntu.edu.sg

1 Introduction

Today, due to the rapid development of sensor technologies, high-dimensional spectral images have attracted significant attentions in a tremendous range of real world applications, from satellite remote sensing imaging and surveillance monitoring systems to industrial product inspections and medical imaging examinations. It provides vast spectral information for data analysis, i.e., classification, clustering, etc. Among spectral images, a hyperspectral image denotes a set of data which measure the spectrum of solar radiation reflected by the earth's surface (Lillesand 2006). It is obtained from hyperspectral sensors by scanning the objects and collecting imagery simultaneously in tens and thousands of narrow and contiguously spaced spectral bands, with wavelength ranging from the visible spectrum to the infrared region (Chang 2003). The information contained in hyperspectral images allows the accurate characterization, identification, and classification of the land covers. However, hyperspectral images are always of abundant high correlated spectral bands of images, which require expensive computational resources and are full of redundancies that do not contribute to the data analysis accuracy. Consequently, determining the most useful and valuable information in hyperspectral images has become essential.

Band selection for hyperspectral image is the process to reduce the band size and identify the most informative bands for further analysis on the hyperspectral image data. Like the feature selection problem, band selection is NP-hard with only explication enumeration approaches known to solve it optimally. However, large scale problems are generally computationally intractable due to the poor scalability of most enumeration methods. From a survey of the literature, many heuristic approaches have played an important role in algorithms capable of providing good solutions within tractable computational time. In Chang et al. (2009), Chang et al. proposed a parallel particle swarm optimization scheme for band selection of hyperspectral images, while Wu et al. employed genetic algorithm to identify the most informative bands of hyperspectral images (Wu et al. 2010). Further, Yin used immune clonal strategy, which is an improved evolutionary strategy algorithm, to select effective bands for hyperspectral images (Yin et al. 2012). Zhu et al. proposed a memetic algorithm framework to identify the essential hyperspectral images bands (Zhu et al. 2010). It is worth noting here that, among these proposed heuristic approaches, memetic algorithm is able to achieve superior performance since it has an individual learning process which is incorporated with domain-specific knowledge, to local fine tune the found solution for band selection.

Memetic algorithms (MA) (Moscato 1999) are population-based metaheuristic search methods inspired by the Darwinian principles of natural evolution and the Dawkins' notion of a meme. It has materialized as a form of population-based search with lifetime learning as a separate process capable of local refinement for accelerating search. Studies on MAs have demonstrated that they converge to high-quality solutions more efficiently than their conventional counterparts (Ong and Keane 2004; Le et al. 2009; Zhu et al. 2007; Tang et al. 2009; Chen et al. 2011; Neri et al. 2011) on many real world applications. Recently, Nguyen et al. proposed a theoretic probabilistic memetic framework (PrMF) that unifies the local search frequency, intensity and selection of solutions undergoing local search under a single theme (Nguyen et al. 2009). The proposed algorithm was demonstrated to exhibit superior performances on a set of continuous benchmark problems. To extend the framework for combinatorial optimization, we have previously proposed a probabilistic memetic algorithm for capacitated arc routing problem (Feng et al. 2010). In this paper, due to the combinatorial and NP-hard nature of band selection, we further extend our proposed probabilistic memetic algorithm to identify the informative bands for hyperspectral images. In particular, two recent published memetic solvers (Zhu et al. 2010) for band selection have been employed to exhibit the true efficacy on theoretic rigor of the probabilistic memetic approach. Further, five well-known hyperspectral images data sets, namely "Indian Pines", "Salinas", "Salinas-A", "Pavia Centre", "Pavia University", which are differing in number of class, class balance, number of bands, etc., are used in the empirical study to verify the effectiveness of the probabilistic memetic approach.

The rest of this paper is organized as follows. Section 2 presents an introduction of band selection problem for hyperspectral images and a brief review of memetic algorithm. The proposed probabilistic memetic algorithm for hyperspectral image band selection is then given in Sect. 3. Section 4 subsequently provides the empirical studies on five well-known hyperspectral images data sets. Lastly, the brief conclusive remarks of this paper are drawn in Sect. 5.

2 Preliminary

In this section, we first give an introduction to the band selection problem for hyperspectral images. Secondly, a brief review of memetic algorithm is provided.

2.1 Band selection for hyperspectral images

Usually, the hyperspectral images, each collected from a spectral band, are combined together to form a three-dimensional image cube (as shown in Fig. 1) for further data analysis. Band selection denotes the process to identify the most informative bands. It can not only reduce the band size of the hyperspectral image data, but also reduces the embedded redundancies, which would improve the accuracy of data analysis. Take classification as an example of data analysis for hyperspectral images, which deals with assigning pre-

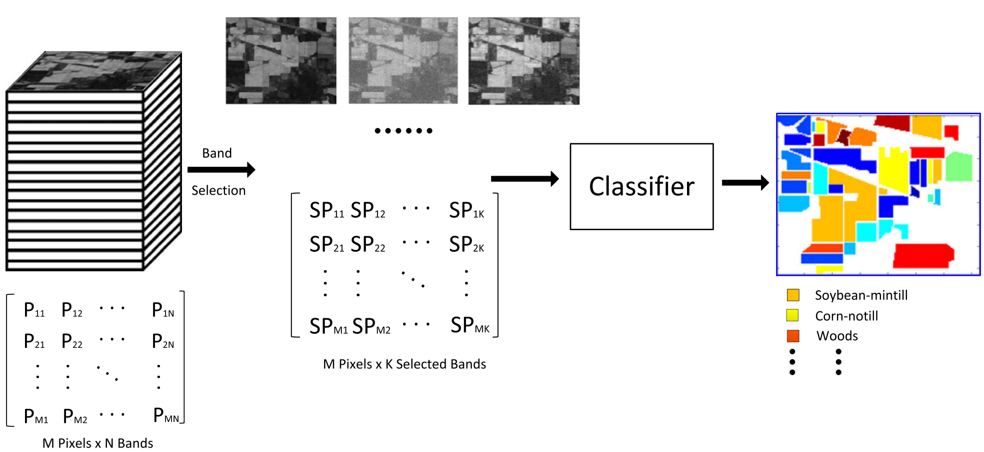

Fig. 1 An illustration of band selection for hyperspectral images

defined classes, such as grass, trees, houses, crops, water and so on, to the imaged terrains or objects. As depicted in Fig. 1, for each pixel i on the ground truth image, there is a vector $\mathbf{P}_i$ of N image measurements from each of the spectral band, where N is the total number of spectral bands. In such a way, the three-dimensional image cube is represented by a $M \times N$ matrix, where M is the total number of pixels. Before the classification analysis process, the $M \times N$ matrix undergoes band selection to produce a $M \times K$ matrix, where K denotes the number of selected bands. Please note that K is usually much smaller than N. Through band selection, a more effective and efficient classification is expected on the given hyperspectral images. In this paper, classification is considered as the data analysis for verifying the proposed band selection algorithm for hyperspectral images.

2.2 Memetic algorithm

To date, many dedicated MAs have been crafted to solve domain-specific problems more efficiently. In a recent special issue dedicated to MA research (Ong et al. 2007), several new design methodologies of memetic algorithms (Lim and Xu 2005; Liu et al. 2007; Hasan et al. 2008; Wang et al. 2009; Bansal et al. 2013), and specialized memetic algorithms designed for tackling the permutation flow shop scheduling (Liu et al. 2007), optimal control systems of permanent magnet synchronous motor (Caponio et al. 2007), VLSI floor planning (Tang and Yao 2007), quadratic assignment problem (Tang et al. 2006, 2007), gene/feature selection (Zhu et al. 2007), have been introduced. From a survey of the area, it is now well established that potential algorithmic improvement can be achieved by considering some important issues of MA (Nguyen et al. 2009; Ong et al. 2006; Lozano et al. 2004):

1. Local search frequency, hereby denoted as f_{il}: defines how often should local learning be applied. f_{il} can be represented as a percentage of the population, i.e., the percentage of individuals in the population that undergoes local learning, or the ratio of evolutionary to local search, i.e., in how many generations of global search should local learning be conducted. Alternatively, f_{il} can be replaced with the local search probability, P_{il}, which defines the probability at which each individual in the population should undergo local learning.
2. Local search intensity, t_{il}: defines how much computational budget should be allocated to each local learning process. t_{il} may be represented in terms of number of the objective function evaluations or time budget.
3. Subset of solution undergoing local search, Ω_{il}: represents the subset of the solution population that undergoes local learning.
4. Local search method: which among a given set of available local learning strategies should be employed on a given problem at hand.

While the above issues have been studied extensively in the literature, for examples, Hart (1994) and Ku et al. (2000) on the local search frequency, Land (1998) on selecting appropriate individuals among the EA population that should undergo local search, Goldberg and Voessner (1999) on local search intensity, Ong and Keane (2004) and Kendall et al. (2002) on the selection of local search; it is worth noting that the works only consider the design issues separately. On the other hand, the theoretic probabilistic memetic framework (PrMF) (Nguyen et al. 2009) unifies the local search frequency, intensity and selection of solutions undergoing local search under a single theme. It has demonstrated superior performances on both continuous benchmark problems (Nguyen et al. 2009) and NP-hard combinatorial optimization problems (Feng et al. 2010). In this paper, we further extend this probabilistic memetic approach for hyperspectral images band selections, which is described in the next section.

3 Probabilistic memetic algorithm for hyperspectral image band selection

In this section, we present the details of the proposed probabilistic memetic algorithm (PMA) for hyperspectral image band selection.

3.1 Encoding representation

In this proposed probabilistic memetic algorithm, we use a binary string chromosome to represent a candidate band subset. The length of the binary string chromosome equals to the total number of bands N, with each bit in the chromosome encoding a single band. A bit of "1" ("0") indicates the corresponding band is selected (excluded). An illustration of our encoding representation for band selection is depicted in Fig. 2.

3.2 Fitness function

The objectives of band selection are to maximize the classification accuracy and meanwhile minimize the number of selected bands. One way to handle this problem is to use the aggregating function method, where the objective function is defined as a linear combination of the classification error and the number of selected bands. The respective weights can be tuned empirically. However, the weights would be significantly different for various datasets, and the corresponding

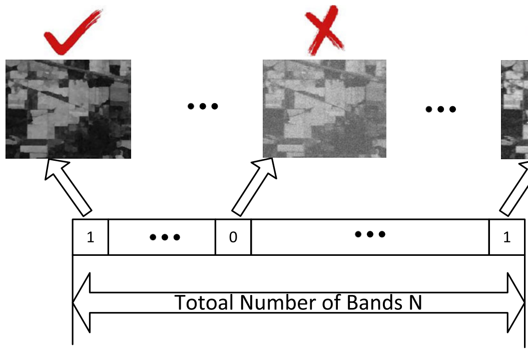

Fig. 2 An illustration of encoding representation for band selection

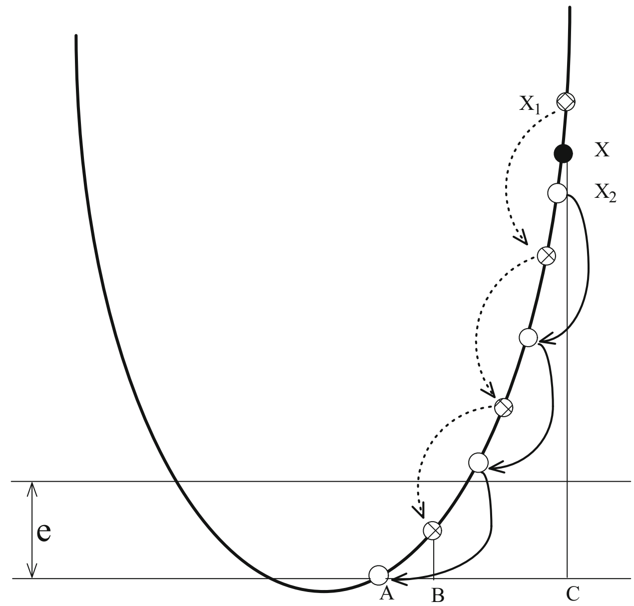

Fig. 3 A depiction on the estimation of t_{upper} in the probabilistic memetic framework for non-linear programming

fitness function tuned on one data set usually cannot be reused on another. To avoid this problem, according to Zhu et al. (2010), here we consider that the classification accuracy is more important than the number of selected bands. Thus the fitness function is defined as the classification error of the selected classifier.

Further, when two chromosomes happen to have the same fitness, the one with a smaller number of selected bands will be given a higher chance for surviving to the next generation.

3.3 Proposed PMA for band selection

In this subsection, we first present the theoretical upper bound of local search derived for memetic searches. Subsequently, how this upper bound is used that leads to a probabilistic memetic algorithm for solving band selection is introduced.

3.3.1 Theoretical upper bound on local search

The theoretical upper bound on local search intensity in memetic algorithm derived in Nguyen et al. (2009) is described by:

$$t_{\text{upper}} = \frac{t_g}{n} \frac{ln(1 - p_2^{(k)})}{ln(1 - p_1^{(k)})} \tag{1}$$

where t_g denotes the function evaluations incurred in a generation and n is the population size. $p_1^{(k)}$ gives the probabilities of an individual, in generation k, hitting the global optimum. $p_2^{(k)}$ are the probabilities of an individual falling within the basin of attraction of the global optimum in generation k. Based on Taylor series expansion, the above equation was simplified to the following when t_g is configured to n:

$$t_{\text{upper}} = \frac{p_2^{(k)}}{p_1^{(k)}} \tag{2}$$

By figuring out $p_1^{(k)}$ and $p_2^{(k)}$ in the search process, this upper bound can be used to determine whether the current individual should undergo local search and/or how much computational budget should be allocated to the local search phase.

Figure 3 presents a depiction on the estimation of t_{upper} in the probabilistic memetic framework for non-linear programming. In the figure, e denotes the precision accuracy for convergence to global optimum and X is the current chromosome or solution. X_1 and X_2 are the nearest neighbors of X (based on simple Euclidean distance) selected from the database of solution vectors obtained by previous local search processes. Note that the local search traces for X_1 and X_2 are also depicted in the figure. From here, the best solution A found in the neighborhood of X and the furthest search point B within the range of e from A are then used to approximate $p_1 = \frac{|AB|}{\text{volume of search space}}$, and $p_2 = \frac{|AC|}{\text{volume of search space}}$. The upper bound for local search intensity, t_{upper}, on X is subsequently determined based on Eq. 2. On the other hand, the expected local search intensity, t_{expected}, required to reach the local optimum of X is then defined as the estimated average length of local search traces of X_1 and X_2. Finally, local search is performed on the current individual only when $t_{\text{expected}} \leq t_{\text{upper}}$.

On the other hand, for combinatorial optimization problem, due to its discrete nature, Euclidean distance is no long suitable for measuring the closeness or similarity between solutions. To approximate the upper bound given by Eq. 2 combinatoric problem context, distance measure such as *Hamming Distance*, *Minkowski-r-distance* (Sörensen 2007), *exact match distance* (Ronald 1998), *deviation distance* (Ronald 1998), *edit distance* (Sörensen 2007) and the *Jaccard's similarity coefficient* (Najera and Bullinaria 2009) can be used. However, based on our previous experiences in solving capacitated arc routing problem (Feng et al. 2010), a modified *Jaccard's similarity coefficient* is considered here as the similarity measure between the two solutions, which is given by:

$$\mathrm{Dis}(S_a, S_b) = |S_a \cup S_b| - |S_a \cap S_b| \tag{3}$$

where S_a and S_b denote the two index sets of bands selected for the given hyperspectral image. $|\cdot|$ gives the number of selected bands in the respective solution. The distance defined in Eq. 3 gives us the number of different selected bands between two solutions, which lies in the range of $[0, |S_a \cup S_b|]$.

Subsequently, the values of p_1 and p_2 in the local search upper bound (i.e., Eq. 2) can be estimated with distance defined in Eq. 3 for search on combinatorial problems. First of all, since the definition of "nearest" relationship is generally loose in the combinatoric context, a single nearest neighbor of the current individual is found to be sufficient for accurate estimation. Let the current solution be X. The nearest neighbor of X, denoted as X_{nber}, is then identified from database Ω which archives all previous local search traces. The local optimum reached, starting from X_{nber}, is then labeled here as X_A, while the individual solution found along the search trace before converging to X_A is labeled as X_B, as depicted in Fig. 4.

In Fig. 4, a "step" denotes a solution jump or transition to a higher quality solution from the initial point. The dashed line represents the trace of X_{nber} generated by the local search process. Probability p_1 is then derived as:

$$p_1 = \frac{\mathrm{Dis}(X_B, X_A)}{\text{volume of search space}} \tag{4}$$

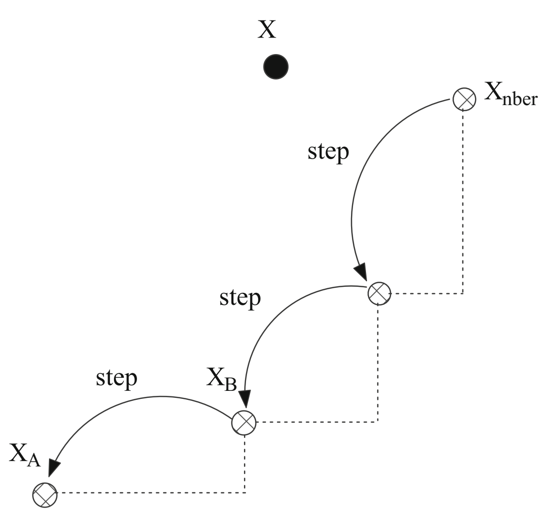

Fig. 4 A depiction on estimation of t_{upper} in probabilistic memetic algorithm for combinatoric problem

Probability p_2, on the other hand, is derived as:

$$p_2 = \frac{\mathrm{Dis}(X, X_A)}{\text{volume of search space}} \tag{5}$$

Subsequently, the upper bound is then derived as:

$$t_{\mathrm{upper}} = \frac{p_2}{p_1} = \frac{\mathrm{Dis}(X, X_A)}{\mathrm{Dis}(X_B, X_A)} \tag{6}$$

The expected local search intensity t_{expected} on the other hand denotes the number of local search steps needed for X to reach X_A, which is approximated by the number of steps from X_{nber} to X_A. Here, we summarize the basic steps of the probabilistic memetic approach in Fig. 5. Note that the upper bound is considered only if the current solution X is sufficiently close to its nearest neighbor; otherwise the search will proceed with local search process.

3.3.2 *Memetic solver for band selection*

In this paper, we consider two recently published memetic solvers as the baselines for band selection, namely *MAFR* and *MAAMB* in the original published work Zhu et al. (2010). *MAFR* and *MAAMB* use the same global exploration evolutionary operators, and are mainly differing in the local search methods used.

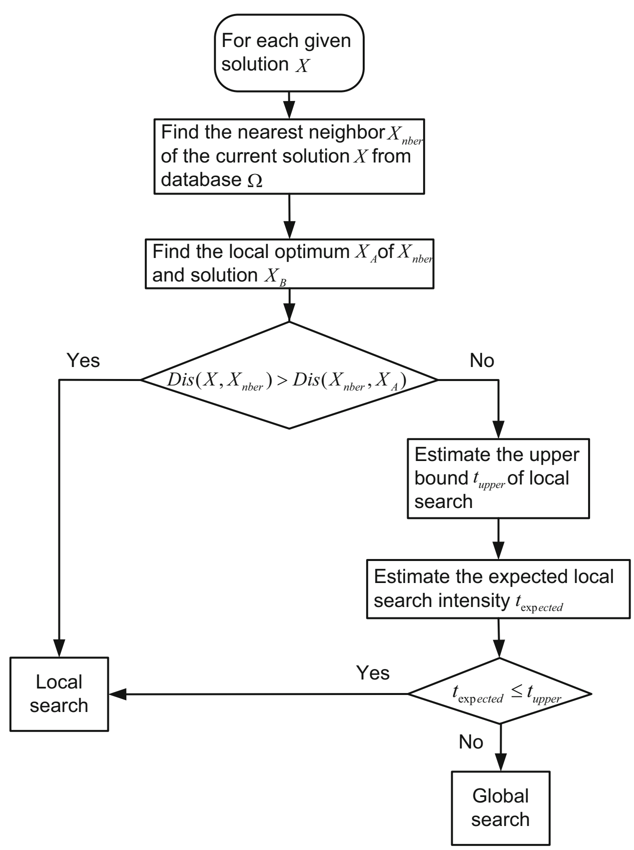

Fig. 5 Outline of the proposed probabilistic memetic approach

In particular, *MAFR* uses the *filter ranking* as the local search operator. It provides the relevance value of each band based on an intuitive idea that a band is more relevant if it distinguishes between a data instance and its nearest neighbor instances from different classes, and less relevant if it distinguishes between an instance and its nearest neighbors from the same class. On the other hand, *MAAMB* employs the *Approximate Markov Blanket* as the local search operator. In this method, the relevance measure is defined as C-correlation (Yu and Liu 2004), which evaluates the correlation between a band and the class label vector. In addition, redundancy measure between bands is also defined in this method. For more details about these two memetic solvers, readers can refer to Zhu et al. (2010).

3.3.3 PMA for band selection

By equipping the approximated upper bound of local search intensity for band selection into the considered memetic solver, the resultant probabilistic memetic algorithm is outlined in Algorithm 1. As the initial population is randomly generated, for the first m generations, the memetic solver proceeds conventionally to collect the local search traces for the approximation of t_{upper} in Eq. 6. Subsequently, the probabilistic memetic approach kicks in to determine whether the current solution should undergo local search or not based on the derived theoretical upper bound for local search. The whole search process terminates when the stopping condition is satisfied.

Algorithm 1: Outline of the memetic solver equipped with the proposed probabilistic framework.

Begin:
Initialization: Generate the initial population randomly
for $j = 1 : m$ *generations* **do**
  while *the termination criteria are not met* **do**
    Select two chromosomes from the current population
    Perform crossover operator to generate offspring
    Apply the local search process with tracking capability on the generated offspring with a certain probability
    Update the current population with the newly generated offspring
for $g = m + 1 : MaxGen$ *generations* **do**
  while *stopping conditions are not satisfied* **do**
    Select two chromosomes from the current population
    Perform crossover operator to generate offspring
    Apply probabilistic memetic framework to the offspring /*see Fig. 5*/
    Update the current population with the newly generated offspring
End

4 Empirical study

In this section, an empirical study on five well-known real world hyperspectral images is conducted using the proposed probabilistic memetic algorithm (PMA). The performance efficacy of PMA is subsequently compared to the two recently proposed memetic solvers for hyperspectral images band selection as aforementioned in Sect. 3.3.2, which form the baseline for comparison.

4.1 Detailed setting

4.1.1 Data set

Five well-known hyperspectral images are used in the present experimental study, which are "Indian Pines", "Salinas", "Salinas-A", "Pavia Centre", and "Pavia University". These images consist of diverse data that differ in terms of the number of class, number of bands, and land cover, etc. The detailed property of the hyperspectral images is summarized in Table 1. In particular, $|Class|$, $|Bands|$, and $Pixels$ denote the number of class, the number of collected band, and the pixel size of each band, respectively. Further, the sampled bands and the respective ground truths of the hyperspectral images are depicted in Fig. 6. In the present paper, for each image, we randomly sample ten bands from each class to form the empirical data for every considered hyperspectral image data set.

4.1.2 Algorithm setup

First of all, to calculate the fitness, we adopt the K-Nearest Neighborhood (KNN), which is one of the most fundamental and simple classification methods in the literature. In our study, k of KNN is configured to 1. Secondly, for the two memetic solvers, i.e., *MAFR* and *MAAMB*, the evolutionary operators and parameter settings are set according to the original paper (Zhu et al. 2010). In particular, the restrictive crossover and mutation operations (Zhu et al. 2007) are employed as exploration operators. The population size, crossover probability, and mutation rate are set to 50, 0.6, and 0.1, respectively. The stopping criteria are defined by a convergence to global optimal or a maximum computational budget of 60,000 fitness functional calls or a maximum generation of 200 is reached. It is worth noting that the fitness function calls made in the local search are also included as part of the total fitness function calls for fair comparison.

Table 1 Detailed properties of the hyperspectral images used in our experiment

Data	$\|Class\|$	$\|Bands\|$	$Pixels$	Description
1. Indian pines	16	224	145 × 145	Gathered by AVIRIS sensor over the Indian pines test site in North-western Indiana
2. Salinas	16	224	512 × 217	Collected over Salinas valley, California, characterized by high spatial resolution
3. Salinas-A	6	224	83 × 86	Small subscene of Salinas image
4. Pavia centre	9	102	1,096 × 715	Acquired by the ROSIS sensor during a flight campaign over Pavia, nothern Italy
5. Pavia University	9	103	610 × 340	Acquired by the ROSIS sensor over Pavia University

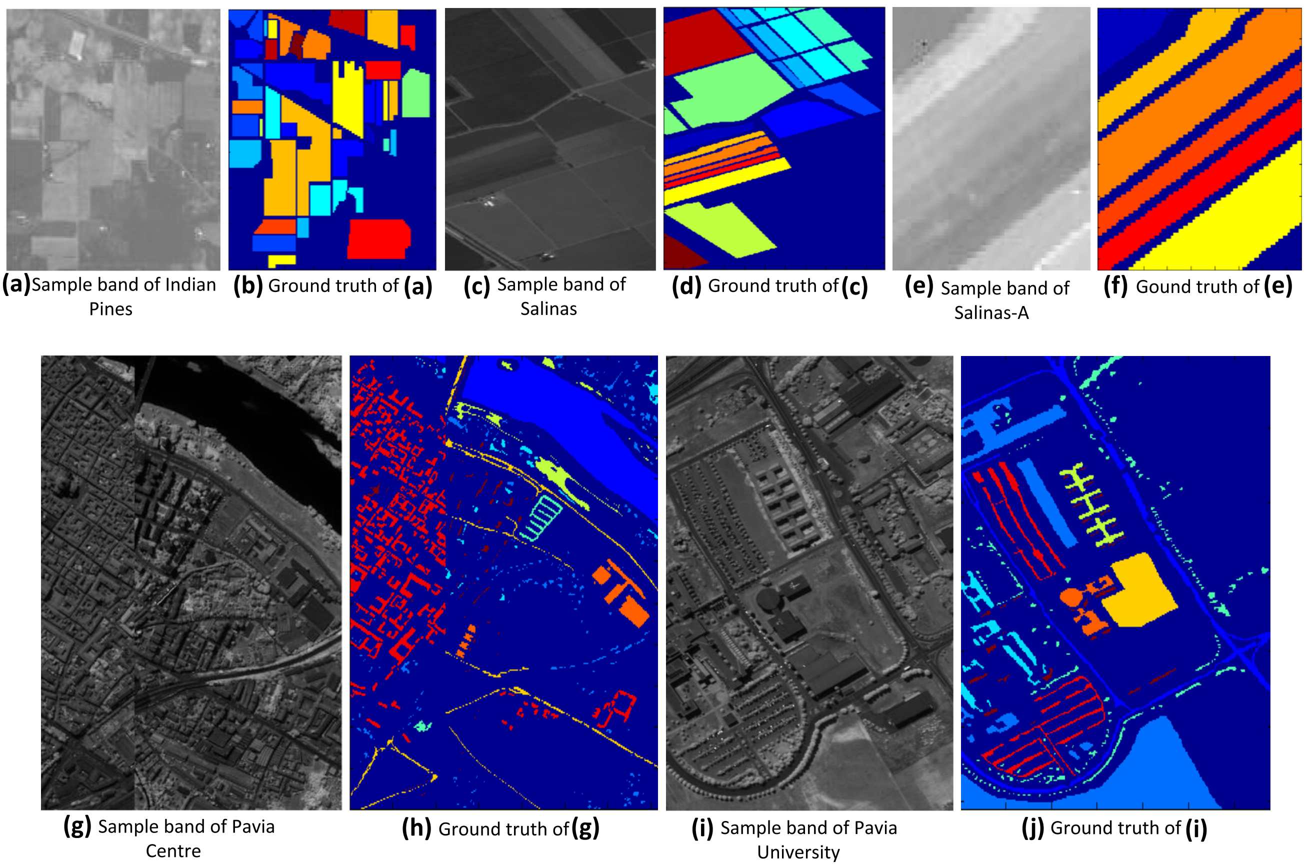

Fig. 6 Sampled bands of the hyperspectral images and their ground truths

On the other hand, both *MAFR* and *MAAMB* use the best 5 individuals in each generation to undergo local search process. For *MAFR* and *MAAMB* equipped with the proposed probabilistic approach, labeled as *MAFR-PMA* and *MAAMB-PMA*, individual undergoes local search or global exploration based on the proposed probabilistic approach as depicted in Fig. 5. Further, to generate local search traces for estimating the upper bound in Eq. 6, in the first 30 generations, *MAFR-PMA* and *MAAMB-PMA* perform exactly the same as *MAFR* and *MAAMB*, respectively.

4.2 Results and Discussion

The performance of *MAFR-PMA* against *MAFR* and *MAAMB-PMA* versus *MAAMB* is summarized in Tables 2 and 3, respectively. Due to the stochastic nature of evolutionary algorithm, the averaged results for ten runs are reported. In particular, the training accuracy is obtained by leave-one-out cross validation and the test accuracy is obtained by fivefold cross validation on the hyperspectral images. In the tables, $Train.Acc$, $Test.Acc$ and $|SF|$ denote averaged training accuracy, testing accuracy, and number of selected bands, respectively. The value inside the brackets gives the corresponding standard deviation. Further, to obtain the statistically comparison, Wilcoxon rank sum test with 95 % confidence level has been conducted on the experimental results.

As can be observed, both *MAFR-PMA* and *MAAMB-PMA* achieved competitive or better performance against their counterparts in terms of training accuracy and testing accuracy on all the considered hyperspectral image data sets.

Table 2 Performance of *MAFR-PMA* and *MAFR* on hyperspectral datasets.

Test	*MAFR-PMA*			*MAFR*		
image	*Train.Acc*	*Test.Acc*	$\|SF\|$	*Train.Acc*	*Test.Acc*	$\|SF\|$
Indian pines	69.84(±2.32)	61.88(±7.13)+	27.00(±2.74)	69.84(±3.20)	57.50(±10.50)	14.20(±6.87)
Salinas	97.97(±1.05)	93.50(±3.12)≈	24.60(±4.51)	97.81(±1.34)	92.50(±3.56)	12.20(±3.03)
Salinas-A	98.33(±0.93)	95.00(±7.45)≈	8.4(±1.67)	98.33(±0.93)	93.34(±9.13)	2.8(±0.84)
Pavia centre	91.67(±2.59)	86.67(±6.33)≈	13.40(±2.07)	92.50(±2.11)	86.67(±8.43)	9.40(±5.32)
Pavia University	78.89(±3.54)	70.00(±8.29)≈	20.00(±4.17)	78.89(±3.85)	70.00(±8.42)	8.60(±4.22)

≈, + and − denote *MAFR-PMA* statistically significant similar, better, and worse than *MAFR*, respectively

Table 3 Performance of *MAAMB-PMA* and *MAAMB* on hyperspectral datasets.

Test	*MAAMB-PMA*			*MAAMB*		
image	*Train.Acc*	*Test.Acc*	$\|SF\|$	*Train.Acc*	*Test.Acc*	$\|SF\|$
Indian Pines	70.31(±1.91)	60.00(±6.77)≈	18.00(±5.70)	68.75(±2.98)	58.75(±9.47)	21.8(±9.34)
Salinas	97.82(±0.86)	89.50(±2.05)+	20.60(±4.03)	97.34(±1.18)	87.50(±4.94)	23.60(±4.97)
Salinas-A	98.75(±1.14)	96.67(±4.56)+	4.80(±2.77)	98.33(±0.93)	93.34(±6.97)	2.00(±0.00)
Pavia Centre	92.22(±2.10)	86.67(±9.10)+	9.8(±3.42)	92.22(±2.11)	83.34(±11.78)	8.2(±2.77)
Pavia University	79.45(±3.59)	70.01(±8.42)≈	11.00(±4.00)	79.17(±4.17)	68.89(±11.52)	9.80(±6.53)

≈, + and − denote *MAAMB-PMA* statistically significant similar, better, and worse than *MAAMB*, respectively

In particular, *MAFR-PMA* obtained better testing accuracy than *MAFR* in 3 out of 5 data sets, while *MAAMB-PMA* demonstrated superior testing accuracy than *MAAMB* in all the data sets. It is worth noting that, on images "Indian Pines" and "Salinas", which have 16 classes and contain relatively more complex information than the others, both *MAFR-PMA* and *MAAMB-PMA* achieved superior performance. This is because the effectiveness of the proposed PMA is more obvious when the search space is more complicated and guidance of the balance between local search and global search is more necessary.

In summary, since the only difference between *MAFR-PMA* (or *MAAMB-PMA*) and *MAFR* (or *MAAMB*) lies in the probabilistic approach introduced in the control of global and local search balance of the former, the superior performance obtained by the former verifies the effectiveness of the proposed probabilistic memetic algorithm.

5 Conclusion

In this paper, we have presented a probabilistic memetic algorithm (PMA) for band selection to identify the most informative bands for hyperspectral image data analysis. The proposed probabilistic approach can adaptively balance the exploration and exploitation of the evolutionary search by a theoretical upper bound of local search while the search progresses online. By equipping the probabilistic approach on two recent memetic solvers for band selection, empirical studies on five well-known hyperspectral image data sets highlighted the efficacy of PMA in converging to competitive or improved band selection for more accurate classification when compared to the baseline memetic solvers.

Acknowledgments The authors would like to thank Dr. Zhu zexuan for his fruitful discussions and valuable suggestions on the empirical study.

References

Bansal J, Sharma H, Arya KV, Nagar A (2013) Memetic search in artificial bee colony algorithm. Soft Comput 17(10):1911–1928

Caponio A, Cascella GL, Neri F, Salvatore N, Sumne M (2007) A fast adaptive memetic algorithm for online and offline control design of pmsm drives. IEEE Trans Syst Man Cybern Part B 37(1): 28–41

Chang C-I (2003) Hyperspectral imaging: techniques for spectral detection and classification. Plenum Publishing Co.

Chang YL, Fang JP, Benediktsson JA, Chang L, Ren H, Chen KS (2009) Band selection for hyperspectral images based on parallel particle swarm optimization schemes. IEEE Int Geosci Remote Sens Symp (IGARSS) 5:84–87

Chen XS, Ong YS, Lim MH, Tan KC (2011) A multi-facet survey on memetic computation. IEEE Trans Evolut Comput 15(5):591–607

Feng L, Ong YS, Nguyen QH, Tan AH (2010) Towards probabilistic memetic algorithm: an initial study on capacitated arc routing problem. IEEE Congr Evolut Comput 2010:18–23

Goldberg DE, Voessner S (1999) Optimizing global-local search hybrids. Genet Evolut Comput Conf 1:220–228

W. E. Hart. Adaptive global optimization with local search. PhD thesis, University of California, San Diego, 1994.

Hasan SMK, Sarker R, Essam D, Cornforth D (2008) Memetic algorithms for solving job-shop scheduling problems. Memet Comput 1(1):69–83

Kendall G, Cowling P, Soubeiga E (2002) Choice function and rrandom hyperheuristics. Fourth Asia-Pacific conference on simulated evolution and learning, pp 667–671

Ku KWC, Mak MW, Siu WC (2000) A study of the lamarckian evolution of recurrent neural networks. IEEE Trans Evolut Comput 4(1):31–42

Land MWS (1998) Evolutionary algorithms with local search for combinatorial optimization. Ph. D. Thesis, University of California, San Diego

Le MN, Ong YS, Jin YC, Sendhoff B (2009) Lamarckian memetic algorithms: local optimum and connectivity structure analysis. Memet Comput 1(3):175–190

Lillesand TM (2006) Remote sensing and image interpretation. Wiley

Lim MH, Xu YL (2005) Application of hybrid genetic algorithm in supply chain management. Int J Comput Syst Sig, 6(1)

Liu D, Tan KC, Goh CK, Ho WK (2007) A multiobjective memetic algorithm based on particle swarm optimization. IEEE Trans Syst Man Cybern Part B 37(1):42–50

Liu B, Wang L, Jin YH (2007) An effective pso-based memetic algorithm for flow shop scheduling. IEEE Trans Syst Man Cybern Part B 37(1):18–27

Lozano M, Herrera F, Krasnogor N, Molina D (2004) Real-coded memetic algorithms with crossover hill-climbing. Evolut Comput 12(3):273–302

Moscato P (1999) New ideas in optimization. chapter memetic algorithms: a short introduction, McGraw-Hill Ltd., Maidenhead, pp 219–234

Najera AG, Bullinaria JA (2009) Bi-objective optimization for the vehicle routing problem with time windows: Using route similarity to enhance performance. In: Proceedings of the 5th international conference on evolutionary multi-criterion optimization, pp 275–289

Neri F, Cotta C, Moscato P (2011) Handbook of memetic algorithms. Springer, Incorporated

Nguyen QH, Ong YS, Lim MH (2009) A probabilistic memetic framework. IEEE Trans Evolut Comput 13(3):604–623

Ong YS, Keane AJ (2004) Meta-lamarckian learning in memetic algorithms. IEEE Trans Evolut Comput 8(2):99–110

Ong YS, Lim MH, Zhu N, Wong KW (2006) Classification of adaptive memetic algorithms: a comparative study. IEEE Trans Syst Man Cybern Part B Cybern 36(1):141–152

Ong YS, Krasnogor N, Ishibuchi H (2007) Special issue on memetic algorithm. IEEE Trans Syst Man Cybern Part B 37(1):2–5

Ronald S (1998) More distance functions for order-based encodings. IEEE Cong Evolut Comput, pp 558–563

Sörensen K (2007) Distance measures based on the edit distance for permutation-type representations. J Heuristics 13(1):35–47

Tang J, Lim MH, Ong YS (2006) Parallel memetic algorithm with delective local search for large scale quadratic assignment problems. Int J Innov Comput Inf Control 2(6):1399–1416

Tang M, Yao X (2007) A memetic algorithm for vlsi floorplanning. IEEE Trans Syst Man Cybern Part B 37(1):62–69

Tang J, Lim MH, Ong YS (2007) Diversity-adaptive parallel memetic algorithm for solving large scale combinatorial optimization problems. Soft Comput J 11(9):873–888

Tang K, Mei Y, Yao X (2009) Memetic algorithm with extended neighborhood search for capacitated arc routing problems. IEEE Trans Evolut Comput 13(5):1159–1166

Wang HF, Wang DW, Yang SX (2009) A memetic algorithm with adaptive hill climbing strategy for dynamic optimization problems. Soft Comput 13(8–9):763–780

Wu H, Zhu JL, Li SJ, Wan DS, Lin L (2010) A hybrid evolutionary approach to band selection for hyperspectral image classification. In: Zeng Zhigang, Wang Jun (eds) Advances in neural network research and applications, vol 67., Lecture notes in electrical engineeringSpringer, Berlin, pp 329–336

Yin JH, Wang YF, Hu JK (2012) A new dimensionality reduction algorithm for hyperspectral image using evolutionary strategy. IEEE Trans Ind Inform 8(4):935–943

Yu L, Liu H (2004) Efficient feature selection via analysis of relevance and redundancy. J Mach Learn Res 5:1205–1224

Zhu ZX, Ong YS, Dash M (2007) Wrapper-filter feature selection algorithm using a memetic framework. IEEE Trans Syst Man Cybern Part B 37(1):70–76

Zhu ZX, Jia S, Ji Z (2010) Towards a memetic feature selection raradigm. IEEE Comput Intell Mag 5(2):41–53