Local Landscape Patterns for Fitness Landscape Analysis

Shinichi Shirakawa[1] and Tomoharu Nagao[2]

[1] College of Science and Engineering, Aoyama Gakuin University
Sagamihara, Kanagawa, Japan
shirakawa@it.aoyama.ac.jp

[2] Graduate School of Environment and Information Sciences,
Yokohama National University, Yokohama, Kanagawa, Japan
nagao@ynu.ac.jp

Abstract. Almost all problems targeted by evolutionary computation are black-box or heavily complex, and their fitness landscapes usually are unknown. Selection of the appropriate search algorithm and parameters is a crucial topic when the landscape of a given target problem could be unknown in advance. Although several landscape features have been proposed in this context, examining a variety of landscape features is useful for problem understanding. In this paper, we propose a novel feature vector for characterizing the fitness landscape using the local landscape patterns (LLP). The proposed feature vector is composed by the histogram of the fitness patterns of the local candidate solutions. We extract the proposed LLP feature vector from well-known continuous optimization benchmark functions and BBOB 2013 benchmark set to investigate the properties of the proposed landscape feature and discuss about its effectiveness.

Keywords: Fitness Landscape Analysis, Local Feature, Problem Understanding, Continuous Optimization Problem

1 Introduction

The main target problems of evolutionary computation are black-box or heavily complex, and the landscape and characteristics of a target search space are usually unknown. Therefore, the users of the evolutionary computation method have to decide which algorithm is most suitable for their target problem by trial and error or heuristic knowledge. If we are able to know the characteristics of the landscape for the target problem in advance or during the search, it is helpful for the selection of the appropriate algorithm or parameter setting. In this context, the field of fitness landscape analysis [7, 19] have been developed, and several features for characterizing the landscape are proposed such as ruggedness [24], fitness distance correlation (FDC) [3], evolvability [1, 21], fitness cloud [17], neutrality [20], and dispersion metric [6].

FDC is a global feature which examines the correlation between fitness values and the distance to the global optimum. Let x_g be a global optimum, $d(\cdot, \cdot)$

be an appropriate distance function, $d_i = d(x_i, x_g)$, and λ candidate solutions $X = \{x_1, \cdots, x_\lambda\}$ are given, then FDC is defined as

$$\mathrm{FDC} = \frac{\frac{1}{\lambda}\sum_{i=1}^{\lambda}(f(x_i) - \bar{f})(d_i - \bar{d})}{\sigma_F \sigma_D}, \tag{1}$$

where $\bar{f}$ and $\bar{d}$ indicate the average values of fitness and distance, respectively, and σ_F and σ_D denote the standard deviation of fitness and distance, respectively. Note that the global optimum is usually unknown, although it can be given in the original definition of FDC. Therefore the global optimum is usually approximated by using finite candidate solutions such as $x_g = \mathrm{argmin}_{x \in X} f(x)$[1]. FDC is one of the useful landscape features and applied various problems such as combinatorial [10] and continuous [14] ones.

Dispersion metric [6] is defined by average pairwise distance between the q best points.

$$\mathrm{DISP} = \frac{1}{q(q-1)} \sum_{i=1}^{q} \sum_{j=1, j \neq i}^{q} d(x_{\mathrm{rk}_i}, x_{\mathrm{rk}_j}), \tag{2}$$

where x_{rk_i} denotes the i-th best point among λ samples. Dispersion metric is mainly applied to analyze the continuous optimization problems [6, 13, 15]. In order to compare the values between different search space scales, the candidate solutions are normalized to the $[0, 1]^n$ when the dispersion is computed in practice [15], where n is the problem dimension.

Fitness cloud [17] is represented as a scatter plot of the fitness values of parents against those of their offsprings (or neighbors), and negative slope coefficient (nsc) [23] is a measure of the problem hardness which is extracted from a plot of fitness cloud. Motif difficulty (MD) is introduced in [4] as a predictive difficulty measure for evolutionary algorithms by extracting motif properties from directed fitness landscape networks (FLNs). Concretely, the subgraphs of FLNs are classified into three types of classes, neutral, guide and deceptive, and then the predictive difficulty measure is calculated based on the number of these motifs in FLN. Recently, Morgan and Gallagher [11] have proposed a landscape feature called length scale and applied it to the analysis of BBOB 2010 benchmark functions. Length scale is defined by dividing the difference of the fitness of two candidate solutions by their distance. They conclude length scale is one of the promising features for the continuous optimization problems. Mersmann et al. propose an approach called exploratory landscape analysis which cheaply and automatically extracts problem properties from a concrete problem instance [9], and they investigate the relationship between low-level features and expert knowledge for the benchmark problems [8]. Smith-Miles and Tan [22] extract various features from traveling salesman problems (TSP) and investigate the relationship between the performance of several search algorithms and the problems. Muñoz et al. [13] present a neural network model for predicting the performance measure of the search algorithm. The model is input the landscape features and the algorithm parameters and outputs the predicted number of function evaluations for solving a given problem.

[1] This is the case of the minimization problem.

We can combine several landscape features to investigate the characteristic of a target problem. We, however, believe that it is an attractive approach to directly extract the landscape feature as a vector form which represents various characteristics of the problem. In this paper, we propose a novel feature vector for characterizing the fitness landscape, which uses the fitness patterns of the local candidate solutions. Then we extract the proposed feature from the well-known benchmark functions and BBOB 2013 benchmark set for continuous optimization problem to confirm its properties and effectiveness.

The next section of this paper presents our proposed feature, local landscape patterns (LLP), for characterizing the fitness landscape. Then, in Section 3, we extract the LLP feature vector from several continuous optimization problems and discuss about the experimental results. Finally, in Section 4, we describe our conclusions and future work.

2 Landscape Feature Using Local Landscape Patterns

Definition of Fitness Landscape In this paper, we refer to the definition of the fitness landscape in [19]. Let $\mathscr{S}$ be a set of all candidate solutions (or a search space), $f : \mathscr{S} \to \mathbb{R}$ be a fitness function, and $d : \mathscr{S} \times \mathscr{S} \to \mathbb{R}$ be a distance function between the candidate solutions. Then the fitness landscape $\mathscr{F}$ is defined as a pair of functions of f and d in the candidate solutions $\mathscr{S}$, namely $\mathscr{F} = (\mathscr{S}, f, d)$. Let consider the continuous optimization without constraints, then the candidate solutions are $\mathscr{S} = \mathbb{R}^n$ where n is a dimension of the search space, and the Euclidean distance is usually used as the distance function d. In the context of evolutionary computation, it is impossible to get the analytical form of f or check for the fitness values and distance of all candidate solutions. We therefore consider to estimate the fitness landscape or its characteristics by only using a given finite sample. Let X be a set of sampled candidate solutions from $\mathscr{S}$ by the specific sampling method, then the approximated fitness landscape can be represented as $\bar{\mathscr{F}} = (X, f, d)$.

Note that $\bar{\mathscr{F}}$ heavily depends on the sampling method of X and may be quite different from $\mathscr{F}$ if the biased sampling is used. For example, the fitness landscape is viewed as a unimodal function when all candidate solutions are sampled from one basin area even if f is a multi-modal function. In other words, $\bar{\mathscr{F}}$ represents the fitness landscape from the viewpoint of the finite samples. In practice, we have to estimate the landscape using sampled candidate solutions and exploit it for selection of algorithm or parameters. Morgan and Gallagher discuss about the sampling issue in dispersion metric in [12].

Local Landscape Patterns Most conventional landscape features such as FDC and dispersion metric are summarized as a scalar value. It may be difficult to represent the exhaustive features of the fitness landscape, and the multiple different landscape features are used to characterize it [8, 13, 22]. It is, therefore, attractive to directly extract the landscape feature as a vector which represents various characteristics of the problem. We consider to characterize the fitness

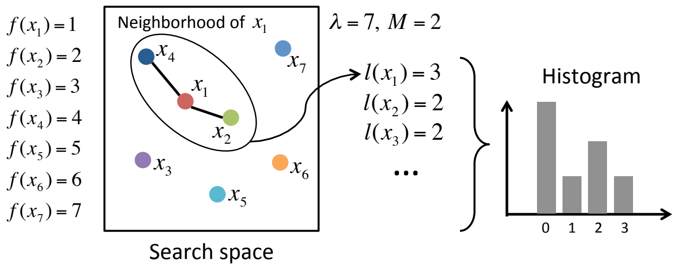


Fig. 1: Conceptual image of the procedure for calculating the LLP feature. λ and M indicate the size of candidate solution sample X and the neighborhood size, respectively.

landscape in a form of a feature vector by using the finite samples. The relationship between the fitness values and the distance of the candidate solutions is the most important characteristic, and the neighborhood structure is also an important property for almost all evolutionary algorithms. From this observation, our proposed feature focusses on the fitness patterns of the local candidate solutions.

Let $X = \{x_1, \cdots, x_\lambda\}$, $x_i \in \mathscr{S}$ be a set of the λ sampled candidate solutions, and $x_{i:j} (1 \leq j \leq \lambda - 1)$ denotes the j-th nearest neighbor candidate solution of the i-th candidate solution x_i measured by the distance function d. Then the local landscape pattern (LLP) $l(x_i)$ around x_i is defined by using the fitness values of the $M (> 0)$ nearest candidate solutions from x_i as

$$l(x_i) = \sum_{j=1}^{M} \delta_{ij} 2^{j-1}, \tag{3}$$

$$\delta_{ij} = \begin{cases} 1 & (\text{if } f(x_i) \text{ better than } f(x_{i:j})) \\ 0 & (\text{otherwise}). \end{cases} \tag{4}$$

where $f(x_{i:j})$ means the fitness value of the j-th nearest neighbor candidate solution of x_i. The value of $l(x_i)$ is an integer within the range of $[0, 2^M]$ and indicates the pattern number of x_i. The pattern number corresponds to the binary number of the fitness pattern. Let $\mathcal{L} = \{l(x_i) | 1 \leq i \leq \lambda\}$ be a set of the local landscape patterns. To summarize these local landscape patterns, we construct the histogram of $\mathcal{L}$. As we employ the histogram of $\mathcal{L}$ as a feature vector, the value of $l(x_i)$ is irrelevant for the eventual feature vector. Consequently, the LLP feature vector $\mathrm{LLP}(X)$ is given by the normalized histogram of $\mathcal{L}$, and the k-th element of $\mathrm{LLP}(X)$ is computed by $H_k(\mathcal{L})/\lambda$, where $H_k(\mathcal{L})$ denotes the k-th element of the histogram of $\mathcal{L}$. The elements of $\mathrm{LLP}(X)$ are divided by λ to normalize the scale of the bin values caused by the different sample size. Note that the number of the bins in the histogram is 2^M, namely the LLP feature is not a scalar but a vector form for representing the landscape features.

The conceptual image of the calculation of the LLP feature is shown in Fig. 1. In this figure, the size λ of candidate solution sample X and the neighborhood size M are set to 7 and 2, respectively. The two nearest neighbors of x_1 are x_2 and x_4, namely $x_{1:1} = x_2$, $x_{1:2} = x_4$, and $f(x_{1:1}) = f(x_2) > f(x_1)$, $f(x_{1:2}) = f(x_4) > f(x_1)$. Then the local landscape pattern of x_1 is computed by $l(x_1) = 1 \cdot 2^0 + 1 \cdot 2^1 = 3$, and it is added to the histogram bin of 3. Analogously, other local landscape patterns $l(x_2) \cdots l(x_7)$ are calculated, then the histogram feature can be obtained.

In order to extract the LLP feature vector from the given finite candidate solutions, it is only necessary to define the fitness function f and the distance function d among the candidate solutions. It means the LLP feature can be extracted from various problem domains such as continuous or combinatorial optimization problems. One of the advantages of the LLP feature is its invariant property under monotonicity-preserving transformations of f because the LLP feature only uses the magnitude relation of the fitness values. For example, the LLP feature vectors on such as $f(x) = x^{\mathrm{T}}x$ and $f(x) = \exp(x^{\mathrm{T}}x)$ are same without any normalization of the fitness function values.

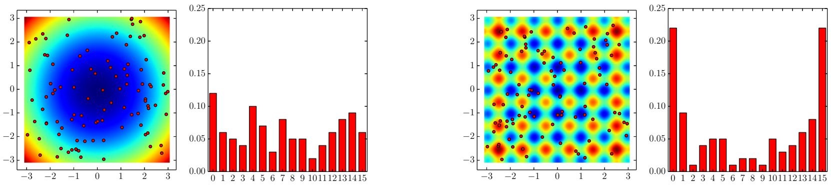
The concept of the proposed method which uses the histogram of the local patterns has succeeded in the computer vision community, e.g. bag-of-features using SIFT descriptors [2]. The LLP feature shares the common concept with the image feature of the local binary patterns (LBP) [18] with respect to focus on the local relationship between the fitness values. LBP can be, however, applicable only for the two dimensional pixel spaces but the fitness landscape is always high dimensional and it may not even be Euclidean space.

3 Experiments and Results

To investigate the properties and the effectiveness of the LLP feature, we apply it to the fitness landscape of the continuous minimization problems defined as $\mathrm{argmin}_{x \in \mathbb{R}^n} f(x)$ and use the Euclidean distance as the distance function between the candidate solutions. Namely, the search space $\mathscr{S}$ described in Section 2 is $\mathscr{S} = \mathbb{R}^n$ and the distance function d is the Euclidean distance in the experiments.

3.1 Basic Properties of LLP

First of all, we consider two dimensional problems whose landscape can be easily visualized. We generate $\lambda = 100$ candidate solutions by drawing random numbers uniformly from $[-3, 3]^2$ and then extract the LLP feature by the setting of the neighborhood size $M = 4$. Figure 2 (a) and (b) show the visualized two dimensional landscapes, the points of the candidate solutions, and the extracted histogram of the LLP on the sphere and Rastrigin functions, respectively. From these figures, we can see the shapes of the histogram are different between the unimodal (sphere) and the multi-modal (Rastrigin) functions. The pattern 0 or 15 indicates all fitness values of the neighborhood solutions are worse or better than that of the focused candidate solution. In Rastrigin function, the patterns of 0 and 15 (which corresponds to the binary number 0000 and 1111, respectively) have a large value because the points on the basin and ridge tend to



(a) Sphere function

(b) Rastrigin function

Fig. 2: Visualized fitness landscapes and the sampled candidate solutions for calculating the LLP feature (left), and the histogram of the LLP (right) on (a) Sphere and (b) Rastrigin functions, respectively.

become the best and the worst point in the local region, respectively. On the other hand, the patterns which have intermediate value such as 7 (which corresponds to the binary number 0101) mean that there are both better and worse neighborhood candidate solutions than the focused one. As a lot of the local regions in the sphere function are the monotonous landscape (i.e. the fitness value monotonically decreases or increases toward a certain direction), the patters which have intermediate value are increased. The shape of the histogram of the sphere function becomes like in Fig. 2 (a) in the end.

Next, we extract the LLP feature vectors from well-known six continuous benchmark functions, Sphere, Ellipsoid, Rosenbrock, Ackley, Schaffer, and Rastrigin functions. In this experiment, the problem dimension of each function n is 20, and the 100 different sets of the candidate solutions are generated for calculating the proposed feature. The candidate solution set is sampled by drawing random numbers uniformly from $[-3, 3]^n$ for all benchmark functions. And the neighborhood size of $M = 4$ is employed in this experiment. Since our proposed feature is a form of vector, principal component analysis (PCA) is applied to reduce the dimension from $2^M = 16$ to 2 which is easy to be visualized.

Figure 3 illustrates the two dimensional plots using first and second principal components with the varying sample size $\lambda = 100, 500, 1000, 2000$. When the sample size is 100, it is difficult to discriminate each function. We can, however, observe each function is clustered when the sample size becomes larger. Both the sphere and ellipsoid functions are unimodal but the ellipsoid is ill-conditioning function. Therefore, we can regard that the difference between these functions is the range of the landscape because the LLP feature uses only the rank of fitness values, and the size of locally monotonous (or basin) regions becomes different if the same sampling region is used. This is the reason why the sphere and ellipsoid functions become different plots. From this result, we can conclude the LLP feature vector has a potential to discriminate the fitness landscape.

A large number of the candidate solutions cause the increase of computational cost. The LLP feature uses the distance between the candidate solutions, and their computational order is $O(\lambda^2)$. Of course, it requires the fitness evaluations for λ candidate solutions. Obviously, there is a tradeoff between the

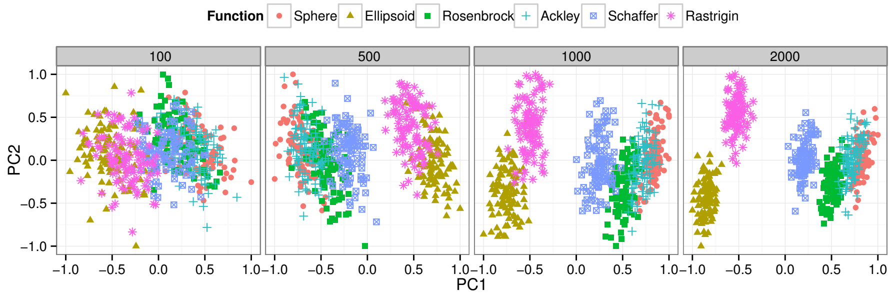


Fig. 3: Two dimensional plots using first and second principal components of the LLP feature vectors extracted from the 20 dimensional benchmark functions with the varying sample size $\lambda = 100, 500, 1000, 2000$. The 100 different sets of candidate solutions with the same size are generated for each function.

computational costs and the knowledge gain. We can observe that it seems to be sufficient to discriminate the landscapes by the 2000 samples in this case. Note that it may have additional information in the hidden dimension because Fig. 3 shows the reduced two dimensional feature by PCA.

3.2 Comparison with Conventional Landscape Features

To confirm the effectiveness of the LLP feature, we compare it with the conventional landscape features, fitness distance correlation (FDC) [3] and dispersion metric [6]. The parameter of dispersion metric is set to $q = \lfloor 0.1\lambda \rfloor$. The sampling method of the candidate solution sample X is same as the previous experiment. Figure 4 illustrates the plots of FDC and dispersion metric extracted from the 20 dimensional benchmark functions with the varying sample size $\lambda = 100, 500, 1000, 2000$. The tendency of the plots is similar to that of the LLP feature. When the sample size is small, the boundary of each function is not clear. The ellipsoid and Rastrigin functions are clustered far away from the other functions. Further the relative position of each plotted function in the two dimensional space is very similar between Fig. 3 and 4. From this result, at least the LLP feature vector has the same ability of characterizing and discriminating the fitness landscape as that of FDC and dispersion metric.

We then conduct the experiment of clustering to quantitatively evaluate the quality of each feature. We use k-means as the clustering method and 2^M dimensional feature vector is used in the proposed LLP feature. The number of the clusters is set to 6 which is the same number of the benchmark functions. To evaluate the clustering quality, we compute purity and adjusted rand index (ARI) [16] as the performance measure by referring to an ideal clustering result. Purity focuses on the frequency of the most common category in each cluster, and ARI is based on the similarity between two data clustering results. The

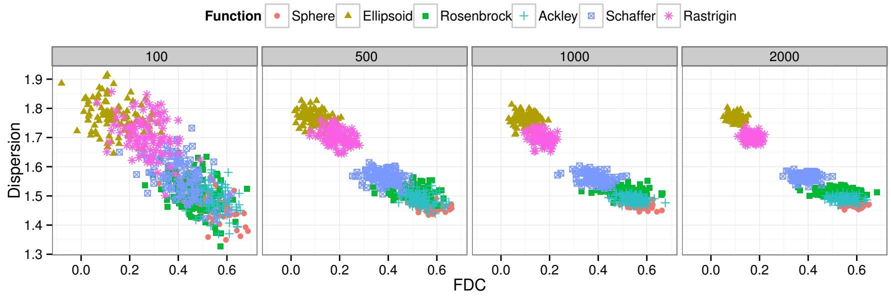


Fig. 4: Two dimensional plots using FDC and dispersion metric extracted from the 20 dimensional functions with the varying sample size $\lambda = 100, 500, 1000, 2000$. The 100 different sets of candidate solutions with the same size are generated for each function.

range of both measures is $[0, 1]$ and the higher value of them indicates the better clustering performance. Table 1 shows the purity and ARI using the LLP and conventional features with the varying problem dimension of $n = 10, 20, 30$, the sample size of $\lambda = 500, 1000, 2000$, and the neighborhood size of $M = 3, 4, 5, 6$. In all settings, the LLP feature outperforms the conventional features, FDC and dispersion metric, in terms of purity and ARI. From this table, we can see the most stable and best setting is $M = 4$ and $\lambda = 2000$. The performance basically improves as the sample size becomes larger. This result is intuitive understandable because a large number of samples are useful to grasp the detailed landscape structure. Consequently, we can consider the LLP feature vector is one of the promising landscape features for discriminating the continuous optimization problems.

3.3 Fitness Landscape Analysis of BBOB 2013 Functions

Finally, we apply our LLP feature to analyze one of the major benchmark function sets, BBOB 2013[2], which contains the 24 noiseless benchmark functions. In this experiment, the problem dimension of each function is 20 and the candidate solutions are sampled by drawing random numbers uniformly from $[-5, 5]^n$. The range of the search space is based on the definition of BBOB 2013 benchmark functions. The neighborhood size $M = 4$ and the sample size $\lambda = 2000$ are used for extracting the LLP features.

Table 2 shows the clustering result by using k-means clustering, which the number of the clusters is set to 6. The cluster 1 and 2 contain the unimodal functions except Schwefel function (f_{20}). The reason why the multi-modal Schwefel function is assigned to the cluster 1 is that its landscape is macroscopically similar to the sphere function with the exception of near the optimum. The cluster

[2] `http://coco.gforge.inria.fr/doku.php?id=bbob-2013`

Table 1: Comparison of purity and ARI for the clustering results. Each feature is extracted from the 10, 20, and 30 dimensional benchmark functions with the varying sample size $\lambda = 500, 1000, 2000$. The neighborhood size varies as $M = 3, 4, 5, 6$ for the LLP feature, and the 100 different sets of the candidate solutions with the same setting are generated for each function. The bold value indicates best performance among the varying parameters.

	LLP (M=3)			LLP (M=4)			LLP (M=5)			LLP (M=6)			FDC & DISP		
λ	500	1000	2000	500	1000	2000	500	1000	2000	500	1000	2000	500	1000	2000
Problem dimension = 10															
Purity	0.74	0.75	0.84	0.75	0.76	0.87	0.77	0.76	0.84	0.75	0.78	**0.88**	0.60	0.65	0.68
ARI	0.59	0.64	0.74	0.61	0.64	**0.77**	0.62	0.64	0.72	0.58	0.66	**0.77**	0.39	0.43	0.45
Problem dimension = 20															
Purity	0.70	0.81	0.86	0.67	0.82	**0.88**	0.70	0.82	**0.88**	0.71	0.80	**0.88**	0.65	0.72	0.75
ARI	0.50	0.66	0.75	0.46	0.70	**0.79**	0.51	0.68	**0.79**	0.52	0.66	**0.79**	0.46	0.53	0.61
Problem dimension = 30															
Purity	0.74	0.80	0.91	0.76	0.83	**0.93**	0.72	0.83	0.92	0.73	0.83	0.91	0.67	0.72	0.74
ARI	0.54	0.66	0.82	0.58	0.70	**0.85**	0.53	0.69	0.84	0.53	0.68	0.83	0.45	0.55	0.59

3 consists of the multi-modal functions such as Rastrigin and Schaffers variants and highly conditioned functions (f_6, f_7, and f_{12}), and the cluster 5 is composed by the multi-modal functions and Rosenbrock family. These two clusters contain both the unimodal and multi-modal functions. All the functions in the cluster 4 and 6 are highly rugged multi-modal function. Original Rosenbrock (f_8) and its rotated version (f_9) are assigned different clusters because the sampling region is not rotated along with the rotation of the fitness function and then the local landscales of the samples become different. This clustering result is similar with the qualitative grouping by the human observation.

Table 3 shows the expected running time (ERT) of BIPOP-aCMA [5] for reaching precision $\Delta f = 10^{-7}$ in the BBOB 2013 result[3] for 20 dimension in BBOB 2013. BIPOP-aCMA has only succeeded to reach the target precision at least once all benchmark functions. From Table 2 and 3, the functions in the cluster 2 are relatively easy to solve for BIPOP-aCMA because the ERT of these functions is less than 2.0×10^4. However, other clusters are not so easily understandable with respect to the ERT.

At the end of this section, we investigate whether our proposed LLP feature has the ability to predict the ERT in the manner of the supervised learning. We construct a linear regression model and predict the values which are taken the common logarithm of the ERT ($\log_{10}$ ERT). Each feature is calculated using the 2000 candidate solutions, and the neighborhood size of the LLP is 4. The LLP feature is reduced to two dimensions by PCA. To evaluate the generalized error, we adopt leave-one-out cross-validation. The generalization error of the

[3] `http://coco.gforge.inria.fr/doku.php?id=bbob-2013-results`

Table 2: Clustering result for the BBOB 2013 benchmark functions by k-means clustering using the LLP feature. Each feature is extracted from the 20 dimensional benchmark functions with the sample size $\lambda = 2000$. The neighborhood size M is set to 4.

Cluster 1	Cluster 2	Cluster 3
Sphere (f_1)	Ellipsoid (f_2)	Rastrigin (f_3)
Original Rosenbrock (f_8)	Linear Slope (f_5)	Attractive Sector (f_6)
Sharp Ridge (f_{13})	Rotated ellipsoid (f_{10})	Step Ellipsoid (f_7)
Different Powers (f_{14})	Discus (f_{11})	Bent Cigar (f_{12})
Schwefel (f_{20})		Non-separable Rastrigin (f_{15})
		Schaffers F7 (f_{17})
		Ill-conditioned Schaffers (f_{18})
Cluster 4	Cluster 5	Cluster 6
Büche-Rastrigin (f_4)	Rotated Rosenbrock (f_9)	Weierstrass (f_{16})
Gallagher's Gaussian	Composite	Katsuura (f_{23})
101-me Peaks (f_{21})	Griewank-Rosenbrock (f_{19})	
Gallagher's Gaussian	Lunacek bi-Rastrigin (f_{24})	
21-hi Peaks (f_{22})		

Table 3: Expected running time (ERT) of BIPOP-aCMA [5] for reaching the precision $\Delta f = 10^{-7}$ in the BBOB 2013 result. The problem dimension is 20.

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}
ERT	6.9e2	1.1e3	4.1e3	1.5e4	4.1	9.3e3	2.5e4	4.9e3	1.7e4	1.3e4	7.7e3	2.1e4
	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{18}	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	f_{24}
ERT	3.3e4	1.1e4	3.5e5	2.6e5	8.9e4	2.0e5	6.3e6	5.3e6	7.6e5	6.5e6	1.3e6	1.4e8

root mean squared error (RMSE) of the LLP feature is 0.961, while that of the conventional landscape features (FDC and dispersion metric) is 1.13. Although the LLP feature outperforms the conventional feature, it is not so large improvement. We, therefore, conclude at least the LLP feature has the comparable ability for predicting the problem difficulty to the conventional features.

4 Conclusion and Future Work

In this paper, we propose a novel landscape feature vector, LLP, which focusses on the local fitness patterns. The proposed method constructs the histogram of the local landscape patterns and extracts the feature as a vector form. The advantage of the LLP feature is that it is possible to extract landscape features by unified procedure, i.e. it has sufficient characterizing performance of the fitness landscape without combining the multiple landscape features. We extract the proposed LLP feature vector from the well-known benchmark functions, and show the effectiveness of it through the comparison with existing landscape features. Then we extract the LLP feature vector from the BBOB 2013 benchmark functions and present the results of the clustering and the prediction of ERT.

The clustering results are understandable from the viewpoint of the qualitative properties of each function, and predicting the performance of the ERT outperforms the conventional landscape features.

In this paper, we compared the proposed LLP feature with FDC and dispersion metric. The comparison with other landscape features such as length scale [11], evolvability [1, 21], and fitness cloud [17] should be conducted to verify the effectiveness of the LLP feature. Further, overall the results depend on the data analysis and the sampling methods used. It may be interesting to investigate the compatibility between the LLP feature and the data analysis techniques. We should attempt to use more sample size because the sample sizes λ used up to 2000 in the experiments might be insufficient for the 20 dimensional problems. As we noted in Section 2, the sampling method and the sample size have a big impact to the landscape features. One possible work is to generate the concatenated LLP feature using different sample sizes or sampling methods to extract the landscape characteristics from various viewpoints.

In the future, we plan to apply the proposed LLP feature to another type of problems such as combinatorial problems and real-world problems, and investigate the properties and effectiveness of it. We are able to use and investigate the several standard statistics as a local feature such as the number of the improving solutions and the probability to improve instead of the fitness patterns of the local candidate solutions. Furthermore, we will develop an efficient algorithm which switches the strategy parameters based on the LLP feature.

References

1. Altenberg, L.: The evolution of evolvability in genetic programming. In: Kinnear, Jr., K.E. (ed.) Advances in Genetic Programming, pp. 47–74. MIT Press, Cambridge, MA, USA (1994)
2. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV. pp. 1–22 (2004)
3. Jones, T., Forrest, S.: Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In: Proceedings of the Sixth International Conference on Genetic Algorithms. pp. 184–192. Morgan Kaufmann (1995)
4. Liu, J., Abbass, H.A., Green, D.G., Zhong, W.: Motif difficulty (MD): a predictive measure of problem difficulty for evolutionary algorithms using network motifs. Evolutionary computation 20(3), 321–47 (2012)
5. Loshchilov, I., Schoenauer, M., Sèbag, M.: Bi-population CMA-ES agorithms with surrogate models and line searches. In: Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO '13) Companion. pp. 1177–1184. ACM, New York, NY, USA (2013)
6. Lunacek, M., Whitley, D.: The dispersion metric and the CMA evolution strategy. In: Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation (GECCO '06). pp. 477–484. ACM, New York, NY, USA (2006)
7. McClymont, K.: Recent advances in problem understanding: Changes in the landscape a year on. In: Proceeding of the 15th Annual Conference Companion on Genetic and Evolutionary Computation Conference Companion (GECCO '13) Companion. pp. 1071–1078. ACM, New York, NY, USA (2013)

8. Mersmann, O., Bischl, B., Trautmann, H., Preuss, M., Weihs, C., Rudolph, G.: Exploratory landscape analysis. In: Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation (GECCO '11). pp. 829–836. ACM, New York, NY, USA (2011)
9. Mersmann, O., Preuss, M., Trautmann, H.: Benchmarking evolutionary algorithms: Towards exploratory landscape analysis. In: Parallel Problem Solving from Nature - PPSN XI. Lecture Notes in Computer Science, vol. 6238, pp. 73–82. Springer Berlin Heidelberg (2010)
10. Merz, P.: Advanced fitness landscape analysis and the performance of memetic algorithms. Evolutionary computation 12(3), 303–325 (2004)
11. Morgan, R., Gallagher, M.: Length scale for characterising continuous optimization problems. In: Parallel Problem Solving from Nature - PPSN XII. Lecture Notes in Computer Science, vol. 7491, pp. 407–416. Springer Berlin Heidelberg (2012)
12. Morgan, R. and Gallagher, M.: Sampling techniques and distance metrics in high dimensional continuous landscape analysis: limitations and improvements, IEEE Trans. Evol. Comput., 18(3), 456–461 (2014)
13. Muñoz, M., Kirley, M., Halgamuge, S.: A meta-learning prediction model of algorithm performance for continuous optimization problems. In: Parallel Problem Solving from Nature - PPSN XII. Lecture Notes in Computer Science, vol. 7491, pp. 226–235. Springer Berlin Heidelberg (2012)
14. Müller, C.L., Sbalzarini, I.F.: Global Characterization of the CEC 2005 Fitness Landscapes Using Fitness-Distance Analysis. In: Applications of Evolutionary Computation. Lecture Notes in Computer Science, vol. 6624, pp. 294–303. Springer Berlin Heidelberg (2011)
15. Müller, C., Baumgartner, B., Sbalzarini, I.: Particle swarm cma evolution strategy for the optimization of multi-funnel landscapes. In: 2009 IEEE Congress on Evolutionary Computation (CEC '09). pp. 2685–2692 (2009)
16. Murphy, K.P.: Machine Learning: A Probabilistic Perspective. The MIT Press (2012)
17. Philippe, C., Vérel, S., Manuel, C.: Local search heuristics : Fitness Cloud versus Fitness Landscape. In: Proceedings of the 16th Eureopean Conference on Artificial Intelligence (ECAI 2004). pp. 973–974. IOS Press (2004)
18. Pietikäinen, M., Zhao, G., Hadid, A., Ahonen, T.: Computer Vision Using Local Binary Patterns. No. 40 in Computational Imaging and Vision, Springer (2011)
19. Pitzer, E., Affenzeller, M.: A comprehensive survey on fitness landscape analysis. In: Fodor, J., Klempous, R., Suárez Araujo, C.P. (eds.) Recent Advances in Intelligent Engineering Systems, Studies in Computational Intelligence, vol. 378, pp. 161–191. Springer Berlin Heidelberg (2012)
20. Reidys, C.M., Stadler, P.F.: Neutrality in fitness landscapes. Applied Mathematics and Computation 117(2-3), 321–350 (2001)
21. Smith, T., Husbands, P., Layzell, P., O'Shea, M.: Fitness landscapes and evolvability. Evolutionary Computation 10(1), 1–34 (2002)
22. Smith-Miles, K., Tan, T.: Measuring algorithm footprints in instance space. In: 2012 IEEE Congress on Evolutionary Computation (CEC '12). pp. 1–8 (2012)
23. Vanneschi, L., Tomassini, M., Collard, P., Vérel, S.: Negative Slope Coefficient: A Measure to Characterize Genetic Programming Fitness Landscapes. In: Proceedings of the 9th European Conference on Genetic Programming (EuroGP '06). Lecture Notes in Computer Science, vol. 3905, pp. 178–189. Springer Berlin Heidelberg (2006)
24. Weinberger, E.: Correlated and uncorrelated fitness landscapes and how to tell the difference. Biological Cybernetics 63(5), 325–336 (1990)