This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

Promoting Objective Knowledge Transfer: A Cascaded Fuzzy System for Solving Dynamic Multiobjective Optimization Problems

Han Li, Zidong Wang, *Fellow*, *IEEE*, Nianyin Zeng, Peishu Wu, and Yurong Li

***Abstract*—In this paper, a novel dynamic multiobjective optimization algorithm (DMOA) with a cascaded fuzzy system (CFS) is developed, which aims to promote objective knowledge transfer from an innovative perspective of comprehensive information characterization. This development seeks to overcome the bottleneck of negative transfer in evolutionary transfer optimization (ETO)-based algorithms. Specifically, previous Pareto solutions, center- and knee-points of multi-subpopulation are adaptively selected to establish the source domain, which are then assigned soft labels through the designed CFS, based on a thorough evaluation of both convergence and diversity. A target domain is constructed by centroid feed-forward of multi-subpopulation, enabling further estimations on learning samples with the assistance of the kernel mean matching (KMM) method. By doing so, the property of non-independently identically distributed data is considered to enhance efficient knowledge transfer. Extensive evaluation results demonstrate the reliability and superiority of the proposed CFS-DMOA in solving dynamic multiobjective optimization problems (DMOPs), showing significant competitiveness in terms of mitigating negative transfer as compared to other state-of-the-art ETO-based DMOAs. Moreover, the effectiveness of the soft labels provided by CFS in breaking the "either/or" limitation of hard labels is validated, facilitating a more flexible and comprehensive characterization of historical information, thereby promoting objective and effective knowledge transfer.**

***Index Terms*—Evolutionary transfer optimization (ETO); cascaded fuzzy system; dynamic multiobjective optimization algorithm (DMOA); information characterization; negative transfer**

I. Introduction

Dynamic multiobjective optimization problems (DMOPs), frequently encountered in many industrial fields, are well recognized as one of the most representative and important issues. These problems aptly characterize the rich dynamic properties present in real-world applications, including both the time-varying constraints and objectives [6], [19], [29], [39], [51]. The importance of responding appropriately to environmental changes when addressing DMOPs is emphasized, and the strategies for effectively managing these dynamic behaviors have attracted significant attention [1], [11], [23], [25].

In recent years, a substantial number of dynamic multiobjective optimization algorithms (DMOAs) have been introduced, including those based on diversity, memory, and prediction [3], [27], [34], [36], [49]. A notable concern is that the development of solutions specifically tailored to isolated tasks or particular scenarios can impede the widespread adoption of some advanced algorithms. This challenge is intensified by the growing complexity of engineering systems, which complicates the assurance of effectiveness and generalization in the elaborately designed response strategies [28]. Moreover, initiating optimization algorithms without prior knowledge in new circumstances is also deemed highly inefficient [38].

In response to the challenges identified above, the evolutionary transfer optimization (ETO) technique [38] has emerged as a novel trend for developing competitive DMOAs. This approach focuses on rationally utilizing accumulated experiences from previous environments to enhance optimization efficiency in new settings. For examples of successful applications of ETO-based DMOAs, one can refer to [17], [22], [26], [46]. Specifically, ETO methods are dedicated to ensuring that individuals sufficiently learn from historical information, primarily aiming to achieve high-quality population initialization in new environments. Consequently, an ensuing challenge is to effectively transfer valuable knowledge, which undoubtedly plays a crucial role in overcoming the bottleneck of negative transfer.

Regarding the issue of "how to transfer", numerous efforts have been undertaken from various perspectives. In [40], based on a linear prediction model, key points within the population have been predicted and employed as samples in the target domain $\mathcal{D}_t$. This strategy aims to guide the population towards thoroughly exploring the most valuable regions in the decision space, thereby providing a promising direction for knowledge transfer. A pre-search strategy has been designed in [18] to establish the target domain, employing rich methods including penalty boundary intersection, crossover, and Gaussian mutation operators to sufficiently enhance the diversity of $\mathcal{D}_t$. This approach helps mitigate the gathering of inferior solutions, reducing the likelihood of negative transfer. Additionally, efforts have been made to decrease the data distribution differences between the source and target domains [22], [31], [45], addressing the challenge of

This work was supported in part by the Natural Science Foundation of China under Grant 62073271, the Fundamental Research Funds for the Central Universities of China under Grant 20720220076, the Natural Science Foundation for Distinguished Young Scholars of the Fujian Province of China under Grant 2023J06010, and the National Science and Technology Major Project of China under Grant J2019-I-0013-0013.

H. Li and Y. Li are with the College of Electrical Engineering and Automation, Fuzhou University, Fujian 350108, and also with the Fujian Provincial Key Lab. of Medical Instrument and Pharmaceutical Technology, Fuzhou University, Fujian 350108, China.

Z. Wang is with the Department of Computer Science, Brunel University London, Uxbridge UB8 3PH, U.K. Email: zidong.wang@brunel.ac.uk

N. Zeng and P. Wu are with the Department of Instrumental and Electrical Engineering, Xiamen University, Fujian 361102, China.

Corresponding author: N. Zeng. Email: zny@xmu.edu.cn

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

transferring knowledge between non-independently identically distributed populations.

Notably, in many existing ETO-based DMOAs, one-hot coding is used to characterize the value of historical information based on the dominance relationship [18], [45]. As a result, individuals in the source domain $\mathcal{D}_s$ are labeled either as "good" solutions with positive labels or "bad" ones with negative labels. It should be emphasized that such a binary criterion is highly likely to overlook potentially valuable information. On one hand, dominated solutions may still offer considerable diversity in terms of solution density; on the other hand, non-dominated solutions can lead to redundant information due to their dense distribution. Therefore, there is an urgent need to address this issue in order to facilitate a more objective and comprehensive knowledge transfer, which is beneficial in alleviating the phenomenon of negative transfer.

Building on the discussions above, this paper introduces a novel cascaded fuzzy system (CFS) designed to assign soft labels to individuals in $\mathcal{D}_s$, utilizing the advantages of fuzzy logic to enable flexible characterization of historical information. Specifically, by sequentially considering factors such as crowding distance, solution density, and Pareto rank of individuals, a comprehensive evaluation of both diversity and convergence is achieved. This approach allows for a more objective reflection of the reference value of individuals compared to the use of hard labels. Moreover, in the proposed CFS-DMOA, the source domain encompasses individuals from previous Pareto solutions, center-, and knee-points of multi-subpopulation, which forms the basis for transferring valuable knowledge. A center-point-based feed-forward prediction strategy is employed to establish the target domain $\mathcal{D}_t$, offering a promising direction for knowledge transfer. Additionally, with the aid of the kernel mean matching (KMM) method, individuals in $\mathcal{D}_t$ are used to further assess the importance of samples in $\mathcal{D}_s$, thus circumventing the need to label $\mathcal{D}_t$ without prior knowledge of searching in the new environment.

The main contributions of this study are outlined as follows.

1) A novel ETO-based algorithm, CFS-DMOA, is developed, paving a new way to overcome the challenges of negative transfer from an innovative perspective of information characterization.
2) A cascaded fuzzy system is designed to provide comprehensive evaluations of historical information, thus promoting objective knowledge transfer with considerable flexibility.
3) The KMM method is utilized to avoid labeling individuals in the target domain without prior knowledge, enabling further assessments of the importance of learning samples.

The remainder of this paper is organized as follows. Preliminaries are provided in Section II, the proposed CFS-DMOA is detailed in Section III, results and discussions are presented in Section IV and, finally, conclusions are drawn in Section V.

II. PRELIMINARIES

A. Problem Formulation of DMOPs

Without loss of generality, a minimized DMOP can be formulated as [22]:

$$\begin{cases} \min \ \boldsymbol{F}(\boldsymbol{x},t) = [f_1(\boldsymbol{x},t),\ f_2(\boldsymbol{x},t),\ ...,\ f_m(\boldsymbol{x},t)] \\ s.t.\ \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{G}(\boldsymbol{x},t) = 0,\ \boldsymbol{H}(\boldsymbol{x},t) \le 0 \end{cases} \tag{1}$$

where $\boldsymbol{x}$ lies in an n-dimensional decision space, which satisfies the equality and inequality constraints $\boldsymbol{G}$ and $\boldsymbol{H}$. $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^m$ realizes the mapping to an m-dimensional objective space, and time variable t refers to the current environment.

Dominance relationship between two decision variables at time t can be described as:

$$\begin{cases} \forall\ i \in \{1,\ 2,\ ...,\ m\},\ f_i(\boldsymbol{x_1},t) \le f_i(\boldsymbol{x_2},t) \\ \exists\ j \in \{1,\ 2,\ ...,\ m\},\ f_j(\boldsymbol{x_1},t) < f_j(\boldsymbol{x_2},t) \end{cases} \tag{2}$$

where $\boldsymbol{x_1} \prec_t \boldsymbol{x_2}$ means $\boldsymbol{x_1}$ dominates $\boldsymbol{x_2}$. Based on Eq. (2), if an individual $\boldsymbol{x}$ cannot be dominated by any other individuals, then $\boldsymbol{x}$ is termed as the Pareto solution at time t. In particular, definitions of the time-varying Pareto set (PS_t) and Pareto front (PF_t) are given as:

$$\begin{aligned} PS_t &= \{\boldsymbol{x} | \neg\exists\ \boldsymbol{x}^* \in \mathbb{R}^n,\ \boldsymbol{x}^* \prec_t \boldsymbol{x}\} \\ PF_t &= \{\boldsymbol{F}(\boldsymbol{x},t) | \boldsymbol{x} \in PS_t\} \end{aligned} \tag{3}$$

where time variable t implies the dynamic behaviors in DMOPs, and as a result, how to accurately track the changing optimal solutions is one of the major concerns in developing effective DMOAs. For the optimization tasks in each individual environment, mature evolutionary algorithms [10], [20], [30], [44], [50] can be directly employed as the static optimizer.

B. Analysis of State-of-the-art DMOAs

Recently, plenty of competitive DMOAs have been emerging, where many efforts have been carried out to effectively cope with the dynamic behaviors in DMOPs. To accurately track the movement of Pareto sets, a novel cluster prediction strategy has been proposed in [43], where the trajectory of cluster center points is employed to predict the evolutionary direction of population. As compared to the direct prediction manner based on the population centroid, the proposed method in [43] has successfully captured the deviation information of adjacent individuals. Moreover, to make potential correction of the predicted evolutionary direction, small mutations have been introduced to excellent individuals, which further improves the prediction accuracy. In [41], the authors have emphasized that most DMOAs fail to associate their response strategy with the environmental changing intensity and, in this regard, an ensemble learning-based response framework has been developed, which enables to dynamically adjust three different strategies based on the measured changing intensity. Additionally, a boundary point-oriented learning method has been introduced in [41], which has the similar effect of enhancing the population diversity as introducing the individual mutations in [43].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

To keep both well population convergence and diversity, in [1], a combinational response mechanism has been developed, which provides three different groups of solutions to form the initial population, including the propagated previous Pareto solutions based on the crowding distance, the random individuals generated by differential evolution and mutation operators, and the solutions obtained via a special points-based knowledge transfer method. Particularly, it is noticeable that when making the transfer learning response, the source domain covers information from the Pareto, center-, and boundary-points. As essentially a prediction-based strategy, the knowledge transfer-based response focuses on predicting high-quality initial population, and such setting benefits a sufficient learning of the inherent predictable patterns. Consequently, the other two groups of solutions in [1] can be regarded as the supplementary to enrich the predicted initial population, thereby making it more adaptive to the new environment. Similarly, in [49], the authors have combined an elitism-based transfer learning strategy and a diversity maintenance method, where the former leverages knowledge from the elites selected from historical environments, and the latter is responsible for increasing the population diversity.

To sum up, predicting reliable solutions in the new environment is quite promising for effectively handling DMOPs, which is exactly the focus of ETO-based DMOAs, and meanwhile, above advanced work also suggests the importance of sufficient information mining. Otherwise, it is highly possible to induce the negative transfer phenomenon, and aiming at this challenging bottleneck, some state-of-the-art ETO-based advanced DMOAs are specially discussed in next subsection.

C. Alleviating the Negative Transfer

In the context of solving DMOPs, ETO-based algorithms are dedicated to effectively utilizing accumulated search experiences to accelerate evolution in new environments. Consequently, overcoming negative transfer has emerged as a particularly challenging issue. This problem refers to the misleading of search processes in incorrect directions, which can significantly hinder the progress [24].

To address negative transfer in environments with slight changes but unchanged Pareto sets (PS), the solutions before and after transfer have been merged in [35], which are then selected based on non-dominated sorting and crowding distance [2] to facilitate population initialization. Similarly, to maximize the benefits of the knowledge transfer strategy, diversity-based responses have been adopted as supplements in [23], which have been effective in mitigating negative transfer in environments that undergo drastic changes.

In [24], a clustering-based transfer strategy has been designed and demonstrated to be effective in mitigating negative transfer by facilitating knowledge transfer between similar clusters across different environments. In [45], another clustering difference-based transfer learning strategy has been developed. In this algorithm, source domain $\mathcal{D}_s$ consists of the previous Pareto solutions (denoted as PS_{t-1}) and some other dominated solutions, which are deemed as the positive and negative samples, respectively. To establish target domain, firstly the population has been divided into 5 clusters, and then the samples in $\mathcal{D}_t$ are obtained via the feed-forward of above cluster centers, where the non-dominated and dominated solutions are labeled as positive and negative, respectively. By using the clustering method, the information in population has been sufficiently explored, and moreover, the feed-forward prediction has improved the similarity between $\mathcal{D}_s$ and $\mathcal{D}_t$, thereby reducing the possibility of negative transfer. Moreover, a knowledge reconstruction-examination strategy has been proposed in [46], which incorporates all previous optimal solutions to thoroughly extract useful information. To be specific, several knowledge clusters have been obtained based on Pareto solutions in historical environment, and those cluster centers have been merged to perform the non-dominated sorting operation. For those clusters having the first Pareto rank, the individuals therein are labeled as positive; otherwise, negative labels will be given. Though above strategy is promising to maximize the utility of historical data, it still deserves further investigations on whether the individuals in clusters with poor convergence (i.e., the large Pareto rank) are totally useless.

In addition, considering the potential for low-quality solutions to cause negative transfer, special attention has been given to the information in knee-points [16]. Within an imbalanced transfer learning framework, the estimated knee-points by the proposed trend prediction model have been labeled as "1". On the contrary, those non-knee-points have been labeled as "0", and $\mathcal{D}_s$ covers the PS_{t-1} and some random solutions. This method has proven effective in generating a high-quality initial population, whereas merely referring to the knee-points possibly loses attention to other information. In [40], multiple predicted key-points including center, polar, and boundary points have been employed as the samples in $\mathcal{D}_t$, which can collectively reflect the overall population status. Besides, some individuals obtained by the feed-forward of center-point have been employed as the supplements to those screened solutions by knowledge transfer strategy, thereby compensating the prediction accuracy so as to mitigate the negative transfer. Nevertheless, similar to most algorithms, the source domain in [40] only considers Pareto solutions from previous environment, which may hinder the comprehensive knowledge transfer.

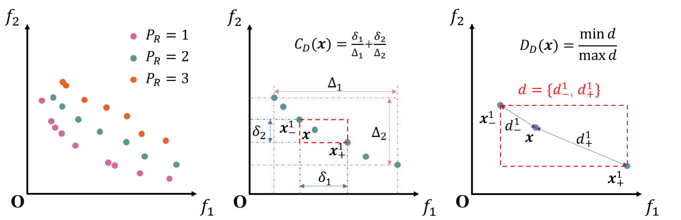
It should be pointed out that although aforementioned ETO-based algorithms have proven effective in alleviating negative knowledge transfer, it has been observed that the characterization of historical information is generally achieved through hard labels [16], [24], [45], [46]. Such "either/or" judgments on the values of previous solutions are highly likely to overlook potentially important information. This limitation has motivated the introduction of soft labels to enable a more comprehensive characterization of historical evolutionary experiences, thereby promoting a more objective knowledge transfer. Additionally, rich information has been covered in the established source domain, which aims to facilitate thorough and comprehensive knowledge transfer.

III. METHODOLOGY

In this section, the proposed CFS-DMOA is detailed, highlighting its major innovation, that is, the designed cascaded

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

fuzzy system. Particularly, the output of CFS is a soft label in $[0,1]$, which dedicates to objectively reflecting the reference values of the historical information; and the input variables of CFS take full consideration of both solution convergence and diversity, which are displayed in Fig. 1 for an intuitive illustration.



(a) Pareto rank (b) Crowding distance (c) Distribution density

Fig. 1. Visualizations of the inputs to CFS in 2-dimension case.

A. Comprehensive Historical Information Characterization

As is shown in Figs. 1(a)-1(b), the Pareto rank (denoted as P_R) and the crowding distance [2] (denoted as C_D) of previous solutions are employed as the inputs of CFS, which are commonly adopted to evaluate the convergence and diversity of individuals, respectively. In general, lower P_R means the corresponding solution is harder to be dominated, and larger C_D refers to better diversity, which can be calculated as [2]:

$$C_D(\boldsymbol{x}) = \sum_m \frac{f_m(\boldsymbol{x}_+^m) - f_m(\boldsymbol{x}_-^m)}{\max f_m - \min f_m} \tag{4}$$

where m is the number of objective functions, $\max f_m$ and $\min f_m$ stand for the maximum and minimum value of the m-th objective with regard to the Pareto front that $\boldsymbol{x}$ belongs to. $\boldsymbol{x}_+^m$ and $\boldsymbol{x}_-^m$ are two adjacent individuals to $\boldsymbol{x}$ that satisfies $f_m(\boldsymbol{x}_+^m) > f_m(\boldsymbol{x}) > f_m(\boldsymbol{x}_-^m)$, which can be obtained by sorting solutions of the same front in a descend order of $f_m(\cdot)$.

A noticeable issue is that according to Eq. (4), once a solution falls into the red dash box in Fig. 1(b), the crowding distance will be a constant, and in this regard, another distribution density (denoted as D_D) is defined in this study as a complement to C_D, which measures the uniformity of solutions on Pareto front (see Fig. 1(c)). In following Eq. (5), definition of above distribution density is given as:

$$D_D(\boldsymbol{x}) = \frac{1}{m} \sum_m \min\left\{ \frac{d_+^m}{d_-^m}, \frac{d_-^m}{d_+^m} \right\} \tag{5}$$

where d_+^m and d_-^m represent the distance between solution $\boldsymbol{x}$ to $\boldsymbol{x}_+^m$ and $\boldsymbol{x}_-^m$, respectively, which can be described as:

$$\begin{aligned} d_+^m &= \|\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{x}_+^m)\|_2 \\ d_-^m &= \|\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{x}_-^m)\|_2 \end{aligned} \tag{6}$$

where $\|\cdot\|_2$ measures the Euclidean distance, both $\boldsymbol{x}_+^m$ and $\boldsymbol{x}_-^m$ maintain the same physical meanings as in Eq. (4). Combining the illustration of Fig. 1(c), when dimension of objective space is 2, the distribution density of individual $\boldsymbol{x}$ is the ratio of shorter distance to the longer one in the red dash box, which takes value in interval $(0,1]$, and the larger D_D corresponds to more even distribution of $\boldsymbol{x}$ with the adjacent solutions.

Consequently, in total three variables will be inputted to the designed CFS and, by doing so, the diversity of an individual can be described in a more holistic manner, thereby realizing comprehensive characterization of the historical information. Moreover, it should be pointed out that in above Eqs. (4)-(6), the description of time variable t is omitted for convenience, and when calculating $C_D(\cdot,t)$ and $D_D(\cdot,t)$ in dynamic environments, fitness values at corresponding times should be matched.

B. Cascaded Fuzzy System

Based on previous introduction, in the designed CFS, the crowding distance and distribution density of an individual are firstly inputted to the first layer, whose output is deemed as the diversity score, and is then further inputted to the second layer along with the Pareto rank, which eventually outputs the soft label of the individual. Notice that the values of soft labels are mapped into interval $[0,1]$, and the larger label symbolizes the higher reference value. As a result, the designed fuzzy system is in a cascaded form, which contains two sub-systems with double inputs and single output, and the merits of such structural designing include following two aspects. On one hand, the progressive order of three inputs has rich logical hierarchy in terms of comprehensive individual evaluations, which benefits generating objective and reliable soft labels; on the other hand, it can effectively decrease the required numbers of fuzzy rules, thereby simplifying the designing difficulty. For better understanding, the diagram of CFS is shown in Fig. 2, and details are summarized in Table I.

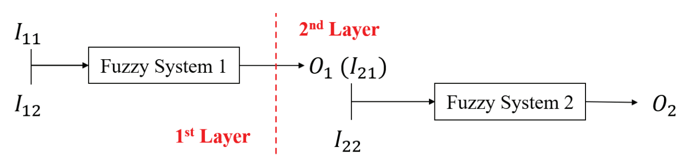


Fig. 2. Diagram of the designed CFS.

TABLE I
INPUTS AND OUTPUTS OF CFS

Layer	Variable	Domain	Membership function	Fuzzy sets
1st	C_D (I_{11})	$[\min, \max]$	Gaussian	{NB,NS,ZO,PS,PB}
	D_D (I_{12})	$[\min, \max]$	Gaussian	{NB,NS,ZO,PS,PB}
	Diversity (O_1)	$[0,1]$	Triangular	{NS,ZO,PS}
2nd	Diversity (I_{21})	$[\min, \max]$	Gaussian	{NB,NS,ZO,PS,PB}
	P_R (I_{22})	$[\max, \min]$[1]	Triangular	{NS,ZO,PS}
	Soft label (O_2)	$[0,1]$	Triangular	{NB,NS,ZO,PS,PB}

[1] Since lower Pareto rank implies better convergence, the input P_R is sorted in a descend order for fuzzification.

Remark 1: O_1 and I_{21} will be processed based on their own membership function and fuzzy sets, and the domain of I_{21} is essentially the real output range of O_1.

In Table I, $[\min, \max]$ indicates that the domains depend on the minimal and maximal values of corresponding input, and fuzzy sets are assigned evenly based on the domains, where NB, NS, ZO, PS, and PB refer to the fuzzy sets of Negative-Big, Negative-Small, Zero, Positive-Small, and Positive-Big, respectively, which generally imply the property transition

DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

from bad to good. Notice that except for the Pareto rank P_R, larger value of all variables in Table I means better property.

Moreover, the fuzzy rules applied in the designed CFS are displayed in Table II, where inference instances with the two superscript symbols are provided as follows.

- † If an individual has both large crowding distance and large distribution density, then it is deemed to own well diversity.
- ‡ If an individual locates in Pareto front with large rank, and simultaneously the diversity is poor, then it is assigned with a small soft label.

TABLE II
FUZZY RULES IN THE DESIGNED CFS ($I_{21} = O_1$)

First layer

O_1 (I_{11} \ I_{12})	NB	NS	ZO	PS	PB
NB	NS	NS	NS	ZO	ZO
NS	NS	NS	NS	ZO	ZO
ZO	NS	NS	ZO	PS	PS
PS	ZO	ZO	PS	PS	PS
PB	ZO	ZO	PS	PS	PS†

Second layer

O_2 (I_{21} \ I_{22})	PS	ZO	NS
NB	ZO	NS	NB‡
NS	ZO	NS	NS
ZO	PS	ZO	NS
PS	PS	PS	ZO
PB	PB	PS	ZO

Above inference results conform to the logic in previous discussions. Additionally, the defuzzification process in the designed CFS is realized by following centroid method [42]:

$$Out = \frac{\sum_{k=1}^{|O|} y_k \cdot \mu_k(y_k)}{\sum_{k=1}^{|O|} \mu_k(y_k)} \tag{7}$$

where $|O|$ is the number of output fuzzy sets, y_k and $\mu_k(y_k)$ refer to the fuzzy output and membership degree on the k-th set, respectively.

C. Acquisition of Source Domain

In the proposed CFS-DMOA, to realize sufficient learning from rich historical evolutionary information, the previous Pareto solutions, center- and knee-points in multiple subpopulations are employed as the samples in source domain $\mathcal{D}_s$. In this regard, the Pareto set PS_t at each environment is firstly divided into K_t subpopulations based on the K-means clustering method [14], where the number of cluster (i.e., K_t) is determined as:

$$K_t = 3 + \lfloor 3 \times \frac{|PS_t|}{N} \rfloor \tag{8}$$

where N is the population size, $\lfloor \cdot \rfloor$ is the floor function, and $|\cdot|$ is the cardinality of set.

Next, center- and knee-points of the K_t subpopulations (denoted as $P_k^{(t)}, k = 1, 2, ..., K_t$) are obtained, where the former refers to the individual closest to the sub-population centroid, and the latter is the individual whose phenotype is the farthest to the line (in 2-dimensional case, denoted as ℓ_b) determined by the two boundary points of the corresponding PF, which can be described as:

$$\boldsymbol{F}(\mathrm{kp}_k) = \arg \max_{\boldsymbol{u} \in \boldsymbol{F}(P_k)} dis(\boldsymbol{u}, \ell_b) \tag{9}$$

where kp_k and $\boldsymbol{F}(P_k)$ denote the knee-point and the Pareto front of the k-th subpopulation (time variable is omitted for simplicity), respectively, $dis(\cdot)$ measures the Euclidean distance. The obtained center- and knee-points are then further expanded in an interpolation manner to derive a center group (G_c) and a knee group (G_k), as is displayed in Algorithm. 1, where the simulated binary crossover operator [32] is adopted to enrich the diversity. Moreover, the previous optimal solutions are stored in a Pareto group (G_p), and eventually, samples in $\mathcal{D}_s$ will be selected from the above three groups according to a scoring mechanism shown in following Eq. (10).

Algorithm 1 Pseudo-code of acquiring G_c and G_k

Input:
Sub-populations $P_1^{(t)}, P_2^{(t)}, ..., P_{K_t}^{(t)}$
Output:
Center group G_c and knee group G_k

```
Obtain center-point cp and knee-point kp of each P^(t)
G_c ← cp_k,  G_k ← kp_k  (k = 1, 2, ..., K_t)
For i, j ∈ {1, 2, ..., K_t} (i ≠ j)
    Generate random numbers r_1, r_2 ∈ (0, 1)
    G_c = G_c ∪ r_1 * cp_i + (1 − r_1) * cp_j
    G_k = G_k ∪ r_2 * kp_i + (1 − r_2) * kp_j
EndFor
Further enrich G_c and G_k with simulated binary crossover
Return G_c, G_k
```

$$\begin{aligned} S_c &= \frac{|PS_{t+1|t}|}{|PS_t|} \\ S_k &= \frac{|\{\boldsymbol{x} \in PS_t | \mathrm{I}(\boldsymbol{x}) > \mathrm{I}_e\}|}{|PS_t|} \\ S_p &= 0.2 \end{aligned} \tag{10}$$

where $\mathrm{I}(\boldsymbol{x})$ refers to the impact suffered by individual $\boldsymbol{x}$, and I_e is the average impact level regarding the dynamic behaviors in environment, which are defined as [21]:

$$\begin{aligned} \mathrm{I}(\boldsymbol{x}) &= \max_{j \in \{1,2,...,m\}} \left| \frac{f_j(\boldsymbol{x}, t+1) - f_j(\boldsymbol{x}, t)}{f_j(\boldsymbol{x}, t) + \mu_0} \right| \\ \mathrm{I}_e &= \max_{j \in \{1,2,...,m\}} \frac{1}{|PS_t|} \sum_{\boldsymbol{x} \in PS_t} \left| \frac{f_j(\boldsymbol{x}, t+1) - f_j(\boldsymbol{x}, t)}{f_j(\boldsymbol{x}, t) + \mu_0} \right| \end{aligned} \tag{11}$$

where m is the number of objectives, and μ_0 avoids the denominator equaling zero.

To be specific, proportion of individuals in PS_t that still maintain non-dominated status at time $t+1$ is defined as the score of G_c, as the less impact of environmental changes on the population convergence, the more individuals in PS_t may still be Pareto solutions at time $t+1$, which implies the stability of the converged population and encourages the learning from center group. Score of G_k measures the ratio of severely affected individuals in PS_t, and the larger S_k, the more it is recommended to take full use of the diversity information in G_k, which promotes the adaptiveness to new environment. Score of G_p is fixed at a constant, which ensures that $\mathcal{D}_s$ covers information from previous Pareto solutions. Besides, the individuals in G_p are updated in a first-in-first-out manner, which avoids unnecessary memory burdens and keeps the stored solutions related to the coming environments [28].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

In the next step, the normalized scores are adopted as the sampling weights to select individuals from three groups, respectively, which are further assigned with the soft labels via the designed CFS to eventually acquire the source domain $\mathcal{D}_s$, and above procedure is summarized in Algorithm. 2, where the predefined capacity of $\mathcal{D}_s$ equals to the population size.

Algorithm 2 Acquisition of source domain

Input:
G_c, G_k, G_p, and source domain capacity n_s
Output:
Source domain $\mathcal{D}_s$
1: Initialize $\mathcal{D}_s \leftarrow \varnothing$
2: Obtain scores of three groups according to Eqs. (10)-(11)
3: Normalize above scores such that $\tilde{S}_c + \tilde{S}_k + \tilde{S}_p = 1$
4: Select $\boldsymbol{x} \in G_c$ to join $\mathcal{D}_s$ until $|\mathcal{D}_s| = \lfloor n_s \times \tilde{S}_c \rfloor$
5: Pick $\lfloor n_s \times \tilde{S}_k \rfloor$ individuals from G_k to join $\mathcal{D}_s$
6: Fill $\mathcal{D}_s$ with $\boldsymbol{x} \in G_p$ until $|\mathcal{D}_s| = n_s$
7: Assign soft label y_s for each $\boldsymbol{x_s} \in \mathcal{D}_s$ via the CFS
8: $\mathcal{D}_s \leftarrow \{(\boldsymbol{x_s}^1, y_s^1), (\boldsymbol{x_s}^2, y_s^2), ..., (\boldsymbol{x_s}^{n_s}, y_s^{n_s})\}$
9: **Return** $\mathcal{D}_s$

D. Establishment of Target Domain

In the proposed CFS-DMOA, the target domain $\mathcal{D}_t$ is established based on the centroid feed-forward of multi-subpopulation. On one hand, the movement trajectory of center-points implies the potential evolutionary tendency, which provides promising direction for knowledge transfer; and on the other hand, multiple prediction results from different subpopulations enrich the diversity of $\mathcal{D}_t$, which declines the influence of samples trapped into local optima, thereby mitigating the negative transfer phenomenon.

Take the subpopulation $P_k^{(t)}$ of PS_t as an example, it is firstly formulated as a center and a set of manifolds, which is used to describe the solution distribution in $P_k^{(t)}$ and can be calculated as [52]:

$$\tilde{X}_k^{(t)} = \{\tilde{\boldsymbol{x}} | \tilde{\boldsymbol{x}} = \boldsymbol{x} - \bar{X}_k^{(t)},\ \boldsymbol{x} \in P_k^{(t)}\} \tag{12}$$

where $\bar{X}_k^{(t)}$ is the center of $P_k^{(t)}$. By feed-forward prediction, the new position of $\bar{X}_k^{(t)}$ at next time $t+1$ can be obtained by:

$$\begin{aligned} k_0 &= \arg\min_i \|\bar{X}_k^{(t)}, \bar{X}_i^{(t-1)}\|_2 \\ \bar{X}_k^{(t+1)} &= \bar{X}_k^{(t)} + (\bar{X}_k^{(t)} - \bar{X}_{k_0}^{(t-1)}) \end{aligned} \tag{13}$$

where $\bar{X}_{k_0}^{(t-1)}$ is center of the closest subpopulation of PS_{t-1} to $P_k^{(t)}$.

Next, the new locations for $\boldsymbol{x} \in P_k^{(t)}$ can be obtained in dimension-wise as:

$$\boldsymbol{x}'_d = \bar{X}_{k,d}^{(t+1)} + \tilde{\boldsymbol{x}}_d + \varepsilon_d\ (d = 1, 2, ..., n) \tag{14}$$

where n is the dimension of decision space, $\varepsilon. \sim \mathcal{N}(0, \sigma^2)$ is the Gaussian noise used to enhance population diversity, and the variance σ^2 is calculated as [52]:

$$\sigma^2 = \frac{1}{n}\Big(\frac{1}{|\tilde{X}_k^{(t)}|}\sum_{\tilde{\boldsymbol{x}} \in \tilde{X}_k^{(t)}} \min_{\tilde{\boldsymbol{z}} \in \tilde{X}_{k_0}^{(t-1)}} \|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{z}}\|_2\Big)^2 \tag{15}$$

where k_0 keeps the same meaning as in Eq. (13), $\tilde{X}_k^{(t)}$ and $\tilde{X}_{k_0}^{(t-1)}$ are the manifolds of subpopulation $P_k^{(t)}$ and $P_{k_0}^{(t-1)}$, respectively.

At last, the target domain $\mathcal{D}_t$ can be established by merging the predicted new individuals of all subpopulations of PS_t, and above procedure is displayed in Algorithm. 3.

Algorithm 3 Establishment of target domain

Input:
Sub-populations $\{P_k^{(t)}\}_{k=1}^{K_t}$ and $\{P_k^{(t-1)}\}_{k=1}^{K_{t-1}}$
Output:
Target domain $\mathcal{D}_t$
1: Initialize $\mathcal{D}_t \leftarrow \varnothing$
2: Obtain subpopulation center of PS_{t-1} as $\mathbf{C} = \{C_k\}_{k=1}^{K_{t-1}}$
3: **For** $k = 1, 2, ..., K_t$
4: Get center and manifolds of $P_k^{(t)}$ via Eq. (12)
5: Feed-forward the $P_k^{(t)}$ center according to Eq. (13)
6: **For** $\boldsymbol{x_i} \in P_k^{(t)}$
7: Predict the new solution $\boldsymbol{x_i}'$ based on Eqs. (14)-(15)
8: $\mathcal{D}_t = \mathcal{D}_t \cup \boldsymbol{x_i}'$
9: **EndFor**
10: **EndFor**
11: **Return** $\mathcal{D}_t$

Remark 2: Different from $\mathcal{D}_s$, the label information is not contained in $\mathcal{D}_t$ in the absence of prior knowledge of searching in the new environment $t+1$.

Remark 3: Information from the previous two environments $t-1$ and t is required to establish target domain.

E. Kernel Mean Matching-based Knowledge Transfer

According to Algorithms. 2-3, label information is only accessible in source domain, and in this regard, the kernel mean matching (KMM) method [12] is applied to assist the knowledge transfer from $\mathcal{D}_s$ to $\mathcal{D}_t$, which maps samples in both domains to a high-dimensional reproducing kernel Hilbert space (RKHS), where it is expected to minimize the distribution difference of the two groups of mapped data by weighting $\boldsymbol{x_s} \in \mathcal{D}_s$ as [22]:

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} \left\| \frac{1}{n_s}\sum_{\boldsymbol{x_s} \in \mathcal{D}_s} \boldsymbol{\beta}\boldsymbol{\Phi}(\boldsymbol{x_s}) - \frac{1}{n_t}\sum_{\boldsymbol{x_t} \in \mathcal{D}_t} \boldsymbol{\Phi}(\boldsymbol{x_t}) \right\|_2 \tag{16}$$

where n_s and n_t stand for the capacity of $\mathcal{D}_s$ and $\mathcal{D}_t$, respectively, $\boldsymbol{\Phi}$ is the mapping function to generate RKHS. In particular, the finally obtained $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_{n_s}]$ contains weights of each sample in source domain, which can be deemed that the importance of $\boldsymbol{x_s} \in \mathcal{D}_s$ is further estimated with reference to $\boldsymbol{x_t} \in \mathcal{D}_t$, and larger weight corresponds to more valuable learning sample.

As a result, the acquired weighted individuals $\{(\beta_i, \boldsymbol{x_s}^i, y_s^i)\}_{i=1}^{n_s}$ can be used to train a regression model in the new environment and, by doing so, individuals in $\mathcal{D}_t$ are only adopted for initializing weights of the training samples, which avoids the labeling process without prior searching experiences. Moreover, since solving Eq. (16) is equivalent to decrease the data distribution difference between $\mathcal{D}_s$ and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

$\mathcal{D}_t$ [9], the non-independent identical distribution property of solutions at different environments [18] is well considered, which benefits the effective knowledge transfer.

In addition, the Boosting technique [33] is applied to train the regression model, which is responsible for screening the high-quality individuals in new environment to form the initial population. Above mentioned knowledge transfer process is presented in Algorithm. 4, where the support vector machine is employed as the base learner, and the selection threshold is set as $\sigma_r = 0.7$.

Algorithm 4 Knowledge transfer-based population initialization

Input:
$\mathcal{D}_s$, $\mathcal{D}_t$, the number of base learner N_b, population size N, and selection threshold σ_r
Output:
Initial population Pop_{ini}

Obtain weights $\boldsymbol{\beta}$ for $\boldsymbol{x_s} \in \mathcal{D}_s$ by Eq. (16)
For $k = 1, 2, ..., N_b$
  Train a weak model $R_k(\cdot)$ by $\{(\beta_i, \boldsymbol{x}_{\boldsymbol{s}}^i, y_s^i)\}_{i=1}^{n_s}$
  Calculate regression error ratio ϵ_k of R_k
  If $\epsilon_k \geq 0.5$
    break
  EndIf
  Obtain confidence coefficient as $\alpha_k = \frac{\epsilon_k}{1-\epsilon_k}$
  Update training weight β_i of sample $(\boldsymbol{x}_{\boldsymbol{s}}^i, y_s^i)$
EndFor
Calculate weight for each R_k as $\phi_k = ln(\frac{1}{\alpha_k})$
Form strong model $\mathcal{R}(\cdot)$ with weighted median of R_k
Screen individuals in the new environment as:

$$Pop_{ini} \leftarrow \{\boldsymbol{x} | \mathcal{R}(\boldsymbol{x}) \geq \sigma_r\}$$

until $|Pop_{ini}| = N$
Return Pop_{ini}

F. Complexity Analysis of CFS-DMOA

To estimate the computational complexity of CFS-DMOA, following aspects are mainly considered:

(1) Using K-means clustering method to obtain K_t subpopulations costs complexity of $O(Nn)$, where N is the population size and n is the dimension of decision space.
(2) When deriving the G_c and G_k based on K_t subpopulations, the interpolation operation costs complexity of $O(K_t^2) < O(N)$ and the crossover operator costs complexity of $O(Nn)$.
(3) The adaptive acquisition of $\mathcal{D}_s$ based on the scoring mechanism in Eq. (10) costs complexity of $O(N) + O(Nm)$, and when assigning soft labels via the CFS, it costs the largest complexity of $O(N^2m)$ for the non-dominated sorting operation to calculate the Pareto rank, where m is the dimension of objective space.
(4) The complexity of establishing $\mathcal{D}_t$ via centroid feed-forward of multi-subpopulation is less than $O[K_t(K_{t-1} + Nn)]$. As the numbers of subpopulations K_t and K_{t-1} are constants, above complexity can be approximated as $O(Nn)$.
(5) Lastly, the complexity of using Boosting technique to train the support vector regression model is $O(N^2n)$.

Therefore, it can be deemed that regardless of the static optimizer, the computational complexity of the proposed CFS-DMOA is $O(N^2m) + O(N^2n) = O(N^2n)$ as generally there is $n > m$. Finally, the overall framework of the proposed CFS-DMOA is displayed in Algorithm. 5 for a clear view.

Algorithm 5 Framework of CFS-DMOA

Input:
Static optimizer $\mathcal{S}_{opt}$, objective $\boldsymbol{F}$, environment indicator $\{t = 1, 2, ..., T\}$, and population size N
Output:
Pareto sets of all environments $PS = \{PS_t\}_{t=1}^T$

Initialize $PS \leftarrow \varnothing, G_p \leftarrow \varnothing$
For $t = 1, 2, ..., T$
  If $t = 1$ —— $t = 2$
    $PS_t = \mathcal{S}_{opt}(\boldsymbol{F}, t)$
  else
    Derive G_c and G_k according to Algorithm. 1
    Acquire source domain $\mathcal{D}_s$ based on Algorithm. 2
    Assign soft labels for samples in $\mathcal{D}_s$ by CFS
    Establish target domain $\mathcal{D}_t$ based on Algorithm. 3
    Obtain initial population Pop_{ini} via Algorithm. 4
    $PS_t = \mathcal{S}_{opt}(\boldsymbol{F}, t, Pop_{ini})$
  EndIf
  Get subpopulations of PS_t by K-means algorithm
  Update $G_p = G_p \cup PS_t$
  While $|G_p| > 1.5 \times N$
    Remove solutions from G_p in first-in-first-out manner
  EndWhile
EndFor
Return PS

Remark 4: At the first two environments, the population is directly initialized by the static optimizer.

IV. Results and Discussions

In this section, the proposed CFS-DMOA is comprehensively evaluated to verify the effectiveness and reliability in solving DMOPs.

A. Experimental Environments

1) Benchmark Functions and Comparison Algorithms: To comprehensively evaluate the performance of CFS-DMOA, nine well-known DF series DMOPs are adopted as the test functions in this study, including 5 bi-objective (DF1-DF4, DF8) and 4 tri-objective problems (DF11-DF14) [15]. It is noticeable that these selected functions cover rich dynamic behaviors, for example, both DF1 and DF3 have changing PF geometry, which requires the algorithm to well track the concavity-convexity variations; on the DF2 problem, the population will suffer severe diversity loss, which implies a thorough search in the whole decision space. Moreover, the time-varying PS of DF8 has a stationary centroid, and the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

changing PF contains the knee regions; on the complex tri-objective problem DF13, partial PF geometry will be disconnected; and on DF14, the number of knee regions can also be time-varying. Consequently, employing above test problems benefits a thorough validation on the proposed CFS-DMOA in terms of effectively handling various DMOPs.

To further show the competitiveness of CFS-DMOA as an algorithm designed with the idea of ETO, other 4 state-of-the-art ETO-based DMOAs are employed as the comparison algorithms, where the response strategies to the changing environments include knee-point-based imbalanced transfer learning (KITL) [16], knowledge guided Bayesian classification (KGBC) [46], clustering difference-based transfer learning (CDTL) [45], and a hybridization of key-point-based transfer and center-point-based feed-forward prediction (HKTCFP) [40]. Above comparison algorithms are named as KITL-DMOA, KGBC-DMOA, CDTL-DMOA, and HKTCFP-DMOA, respectively, and more information of these algorithms can be found in Section II-C.

2) Experimental Settings and Evaluation Metrics: For fair comparison, all the algorithms utilize the same static optimizer MOEA/D [48], where the population size is set to 100 and 120 for bi- and tri-objective problems, respectively. Dimension of the decision space for all benchmarks is fixed at 10, and on each test function, three groups of dynamic parameter setting of $(n_t, \tau_t) \in \{(1, 10), (5, 10), (10, 10)\}$ are deployed, where n_t and τ_t represent the severity and frequency of environmental changes, respectively. The maximum generation is set as $\tau = 20 \times \tau_t$, which indicates that the algorithms will search for Pareto solutions in 20 different environments, and in each individual environment, the static optimizer performs 10 epochs of optimization. In addition, to alleviate the influence of randomness, 20 independent experiments are performed for each test, and the results in average level are reported, including the Wilcoxon rank sum test [5] with the significance level of 0.05.

In this study, the mean inverted generational distance (MIGD) and mean hyper volume difference (MHVD) [21], [40] are adopted as the main evaluation metrics, which can be obtained by:

$$\begin{aligned} \text{MIGD} &= \frac{1}{T}\sum_{t=1}^{T} IGD(PF_t) \\ \text{MHVD} &= \frac{1}{T}\sum_{t=1}^{T} \big(HV(PF_t^*) - HV(PF_t)\big) \end{aligned} \tag{17}$$

where T is the number of environments, PF_t and PF_t^* refer to the obtained Pareto front and the true PF, respectively. $IGD(\cdot)$ and $HV(\cdot)$ are calculated as [21]:

$$\begin{aligned} IGD(PF_t) &= \frac{\sum_{\boldsymbol{x}\in PF_t^*} d(\boldsymbol{x}, PF_t)}{|PF_t^*|} \\ HV(PF_t) &= \bigcup_{\boldsymbol{x}\in PF_t} \nu(\boldsymbol{x}, \boldsymbol{z}_{ref}) \end{aligned} \tag{18}$$

where $d(\cdot)$ stands for the least Euclidean distance, and $\nu(\cdot)$ is the Lebesgue measure of volume enclosed by PF_t and the reference point $\boldsymbol{z}_{ref}$. Both MIGD and MHVD are comprehensive indicators, and the smaller values correspond to better performance, whereas in practice, the former focuses on convergence and the latter emphasizes diversity of algorithms.

B. Benchmark Evaluations

In Table III, benchmark evaluation results of all algorithms on MIGD are presented, where "+/-" indicates the proposed CFS-DMOA performs significantly better/worse than corresponding algorithm, and "≈" suggests equivalent performance.

According to Table III, the proposed CFS-DMOA performs significantly better than KITL-DMOA, CDTL-DMOA, KGBC-DMOA, and HKTCFP-DMOA in 25, 17, 19, and 17 out of 27 cases, respectively, which shows the superiority of CFS-DMOA as an ETO-based DMOA. It is found that merely on problem DF4, CFS-DMOA has relatively poor performance, which is slightly inferior to CDTL-DMOA in the test case of $(n_t, \tau_t) = (10, 10)$ and yields comparable results to KGBC-DMOA in the other two cases. On the complex tri-objective problem DF14, CFS-DMOA obtains the optimal MIGD results in all test cases with three groups of different dynamic parameter settings, and moreover, as can be seen from Fig. 3 that CFS-DMOA is able to provide stable convergence in most changing situations, which implies that the population can realize accurate track to the time-varying PS, thereby demonstrating the reliability of CFS-DMOA to handle the dynamic behaviors. Additionally, in Table IV, comparison on environment-level average running time is reported, which can quantify the computational expensiveness of the algorithms. It is found that the proposed CFS-DMOA yields 3 best results and ranks the second on an overall level. Combining with the above discussed convergence performance, this is also an encouraging result.

Next, the benchmark evaluation results regarding to MHVD are reported in Table V. It is found that the proposed CFS-DMOA and other four comparison algorithms obtain 10, 6, 3, 3, and 5 best results, respectively, which shows the excellent comprehensive performance of CFS-DMOA. In particular, DF2 problem poses great challenges for the DMOAs to well maintain the population diversity, as the varying PS_t will be mapped to the same PF_t in a period of time, while the proposed CFS-DMOA yields 2 out of 3 best results in related test cases. On DF8 problem, where the geometric shape of PF fluctuates from convex to concave, the proposed CFS-DMOA performs not so well as KITL-DMOA, which suggests the importance of information in knee-points. Based on above analysis, it can be concluded that the proposed CFS-DMOA is a competitive ETO-based DMOA, which is promising to overcome the bottleneck of negative transfer and exhibits both satisfactory convergence and diversity in most test cases.

Moreover, by comparing the knowledge transfer procedure in above different algorithms, it is inferred that the advantages of CFS-DMOA mainly owe to the following three aspects: 1) firstly, the source domain in CFS-DMOA covers rich information from historical Pareto solutions and the center-, knee-points of multi-subpopulation, which lays the solid foundation for comprehensive knowledge transfer; 2) the individuals in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

TABLE III
BENCHMARK EVALUATIONS OF DIFFERENT ETO-BASED DMOAS IN TERMS OF MIGD

Functions	n_t, τ_t	KITL-DMOA	CDTL-DMOA	KGBC-DMOA	HKTCFP-DMOA	CFS-DMOA
DF1	1,10	0.3314±3.19e-02(+)	0.2519±1.73e-02(-)	**0.2492±2.05e-02(-)**	0.3090±1.08e-01(≈)	0.2776±1.48e-02
	5,10	0.1520±1.78e-02(+)	0.1278±8.90e-03(+)	0.1397±1.30e-02(+)	0.1339±3.33e-02(+)	**0.1064±7.01e-03**
	10,10	0.1437±3.01e-02(≈)	0.1358±8.04e-03(+)	0.1548±1.66e-02(+)	**0.0880±1.75e-02(-)**	0.1268±6.88e-03
DF2	1,10	0.2205±1.95e-02(+)	**0.1608±1.47e-02(-)**	0.1801±1.65e-02(-)	0.1791±1.39e-02(≈)	0.1834±8.82e-03
	5,10	0.1137±1.78e-02(+)	0.0854±8.09e-03(≈)	0.1055±1.21e-02(+)	0.1089±1.29e-02(+)	**0.0817±4.24e-03**
	10,10	0.1085±1.19e-02(+)	0.0843±6.53e-03(+)	0.0979±1.12e-02(+)	0.1156±2.65e-02(+)	**0.0755±4.15e-03**
DF3	1,10	0.7937±1.58e-01(+)	0.4995±2.88e-02(+)	0.5046±3.63e-02(+)	0.6186±5.72e-02(+)	**0.4675±1.23e-02**
	5,10	0.7229±9.64e-02(+)	0.4706±3.43e-02(≈)	0.5097±1.82e-02(+)	0.5451±3.81e-02(+)	**0.4639±2.16e-02**
	10,10	1.2143±1.94e-01(+)	**0.5589±2.64e-02(-)**	0.6483±4.91e-02(+)	0.7019±6.78e-02(+)	0.5745±2.94e-02
DF4	1,10	0.7614±1.10e-01(+)	0.6072±8.50e-02(+)	**0.4982±5.31e-02(≈)**	0.6923±1.06e-01(+)	0.5191±4.35e-02
	5,10	1.4829±1.02e-01(+)	0.6294±8.47e-02(≈)	**0.5864±6.39e-02(≈)**	0.6961±1.18e-01(+)	0.6044±2.87e-02
	10,10	2.6522±8.74e-02(+)	**0.5231±5.66e-02(-)**	0.5965±8.19e-02(-)	0.6606±6.96e-02(≈)	0.6174±4.09e-02
DF8	1,10	0.8089±1.90e-02(+)	0.3329±2.85e-02(+)	**0.2605±2.48e-02(≈)**	0.3210±2.14e-02(+)	0.2642±2.47e-02
	5,10	1.0853±2.89e-02(+)	0.3712±2.26e-02(+)	0.3144±2.15e-02(≈)	0.3849±2.00e-02(+)	**0.3077±2.58e-02**
	10,10	1.1027±1.79e-02(+)	0.3569±2.07e-02(+)	0.3261±1.82e-02(+)	0.3930±1.74e-02(+)	**0.3090±2.53e-02**
DF11	1,10	0.2097±3.15e-02(+)	0.1545±7.89e-03(+)	0.1468±7.14e-03(+)	0.1643±1.05e-02(+)	**0.1363±4.38e-03**
	5,10	0.1877±2.00e-02(+)	0.1553±8.97e-03(+)	0.1478±6.79e-03(+)	0.1398±5.58e-03(≈)	**0.1378±2.90e-03**
	10,10	0.1837±1.68e-02(+)	0.1534±6.46e-03(+)	0.1465±4.16e-03(+)	**0.1362±5.47e-03(≈)**	0.1377±3.91e-03
DF12	1,10	0.6261±4.86e-02(+)	0.4887±1.98e-02(+)	0.4790±2.03e-02(+)	0.5400±7.71e-02(+)	**0.4651±1.28e-02**
	5,10	0.6572±5.78e-02(+)	**0.4994±1.44e-02(-)**	0.5192±1.22e-02(≈)	0.5117±4.31e-02(-)	0.5213±1.31e-02
	10,10	0.6501±6.77e-02(+)	**0.4418±1.79e-02(-)**	0.4863±1.61e-02(+)	0.4608±3.00e-02(≈)	0.4686±1.29e-02
DF13	1,10	1.0000±1.32e-01(+)	0.3791±2.67e-02(+)	0.3638±3.19e-02(+)	0.4213±4.50e-02(+)	**0.3320±2.30e-02**
	5,10	0.4847±8.43e-02(+)	0.3213±1.54e-02(+)	0.3427±3.04e-02(+)	0.3287±3.78e-02(+)	**0.2973±1.09e-02**
	10,10	0.3668±2.00e-02(+)	0.3190±2.34e-02(≈)	0.3350±2.81e-02(+)	**0.2892±1.71e-02(-)**	0.3066±1.45e-02
DF14	1,10	0.3485±7.76e-02(+)	0.2374±1.51e-02(+)	0.1262±2.48e-02(+)	0.2276±5.49e-02(+)	**0.1096±8.97e-03**
	5,10	0.1745±2.77e-02(+)	0.1288±1.07e-02(+)	0.1318±1.02e-02(+)	0.1764±6.78e-02(+)	**0.1152±8.23e-03**
	10,10	0.1174±8.69e-03(≈)	0.1270±9.00e-03(+)	0.1284±1.06e-02(+)	0.1354±3.72e-02(≈)	**0.1173±5.95e-03**
+ / - / ≈		25 / 0 / 2	17 / 6 / 4	19 / 3 / 5	17 / 3 / 7	-

TABLE IV
COMPARISONS ON AVERAGE RUNNING TIME PER EACH ENVIRONMENT

Functions	KITL-DMOA	CDTL-DMOA	KGBC-DMOA	HKTCFP-DMOA	CFS-DMOA
DF1	**1.87**(1)	2.15(2)	3.10(5)	2.32(4)	2.22(3)
DF2	**1.88**(1)	2.27(3)	3.26(5)	2.15(2)	2.33(4)
DF3	1.52(2)	1.60(3)	2.96(5)	**1.49**(1)	1.60(3)
DF4	**1.75**(1)	2.03(2)	4.02(5)	2.16(4)	2.12(3)
DF8	1.92(3)	**1.90**(1)	4.26(5)	1.91(2)	2.09(4)
DF11	3.95(4)	**3.52**(1)	8.38(5)	3.56(3)	3.53(2)
DF12	3.64(4)	3.11(2)	6.47(5)	3.13(3)	**3.09**(1)
DF13	3.53(4)	3.38(2)	6.23(5)	3.49(3)	**3.32**(1)
DF14	3.65(4)	3.31(2)	5.60(5)	3.37(3)	**3.24**(1)
Overall rank	3	1	5	4	2

source domain are assigned with the more flexible soft labels, which comprehensively evaluates the historical information, thereby guaranteeing the objective knowledge transfer; and 3) the individuals in target domain are used to further estimate the importance of $\boldsymbol{x_s} \in \mathcal{D}_s$, which avoids the potential negative transfer caused by the subjective sample labeling on $\mathcal{D}_t$. In this regard, three different CFS-DMOA variants are designed for comparisons in the subsequent experiments to verify the rationality of above inferences.

C. Influences of Information Diversity in $\mathcal{D}_s$

To investigate whether the rich information contained in the established source domain can promote comprehensive knowledge transfer, an algorithm variant CFS_v1-DMOA is designed, which merely learns from the previous Pareto solution. With the same basic environmental settings, comparison results under two groups of dynamic parameter settings $(n_t, \tau_t) \in \{(1, 10), (10, 10)\}$ are displayed in Table VI, where the reported indicator "MSp" refers to the mean spacing metric, which measures the algorithm diversity in terms of the solution distribution on PF, and the definition is given as:

$$
\begin{aligned}
Sp(PF_t) &= \sqrt{\frac{1}{|PF_t| - 1} \sum_{\boldsymbol{x} \in PF_t} (D(\boldsymbol{x}) - \bar{D})} \\
MSp &= \frac{1}{T} \sum_{t=1}^{T} Sp(PF_t)
\end{aligned}
\tag{19}
$$

where $D(\boldsymbol{x})$ is the distance between $\boldsymbol{x}$ and its nearest neighbor on PF_t, and $\bar{D}$ takes the average of all $D(\cdot)$. Notice that the smaller MSp, the better diversity with more evenly distributed Pareto front.

According to the results, in most test cases, as compared to the algorithm variant CFS_v1-DMOA with only one information source, the original CFS-DMOA presents better performance in both convergence and the evenness of solution distribution. On one hand, rich information can no doubts provide more selections for the knowledge transfer, which

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

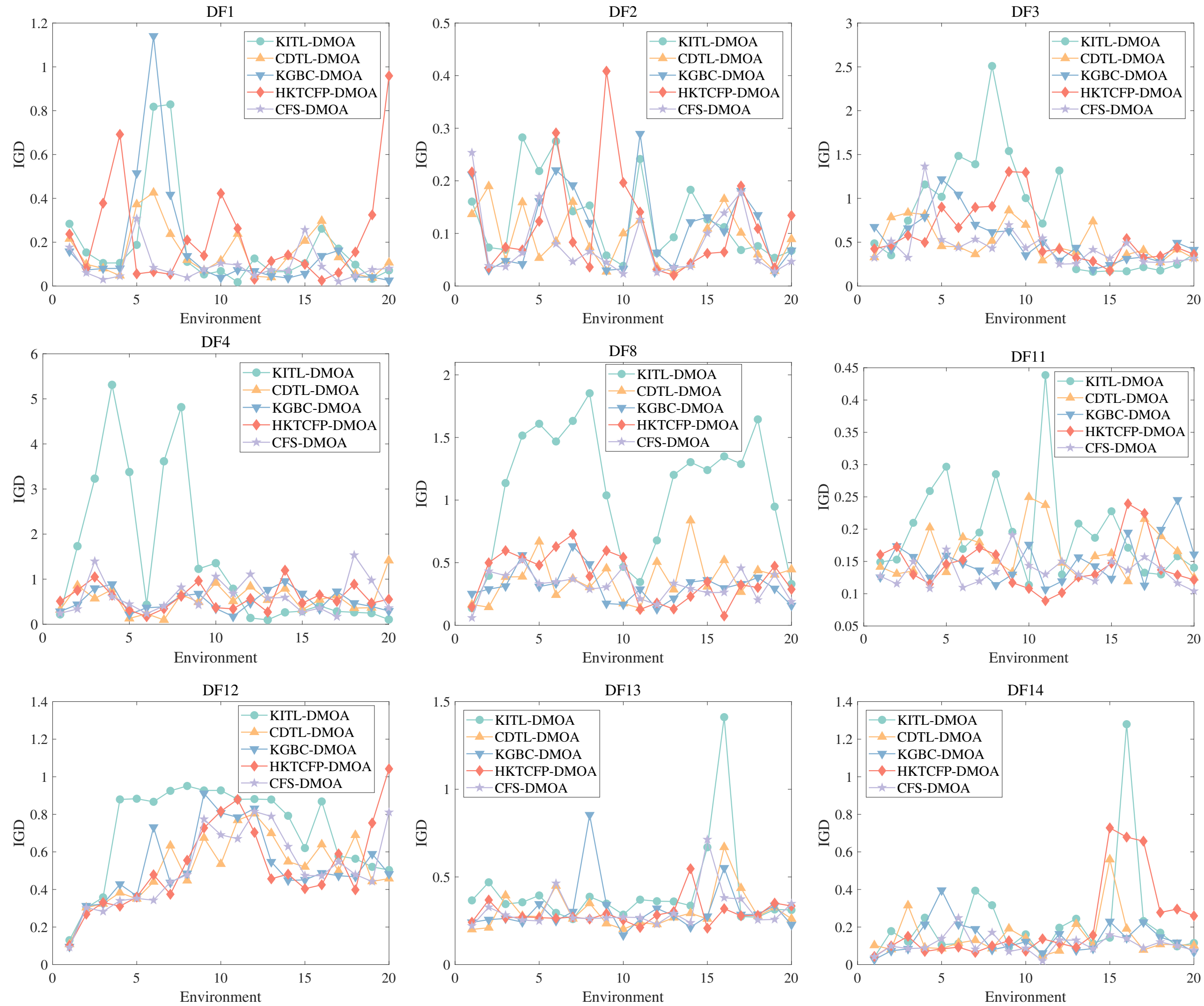


Fig. 3. Convergence comparison among different ETO-based DMOAs on 9 benchmark problems with $(n_t, \tau_t) = (5, 10)$.

benefits training a more reliable regression model to screen promising high-quality individuals in the new environment; on the other hand, the excessive reliance on the previous Pareto solutions is likely to make the population fall into the potential local optima, and the blind learning from harmful information could even lead to the negative knowledge transfer. For example, it is highly possible for CFS_v1-DMOA to suffer the negative transfer on DF3, and in Fig. 4, the obtained PF_t by above two algorithms in three different environments of DF3 is illustrated.

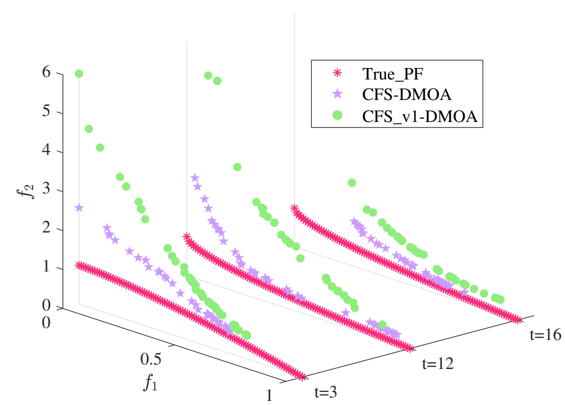


Fig. 4. PF_t in three environments on DF3 obtained by two algorithms, where $(n_t, \tau_t) = (1, 10)$.

As is shown, the original CFS-DMOA can better trace the time-varying PF, which indicates the advantages of learning from rich samples. Simultaneously, notice that the source domain is established adaptively according to the designed scoring mechanism in Eqs. (10)-(11), which associates different learning samples with the changing environment, thereby promoting more effective knowledge transfer [21].

D. Advantages of Applying Soft Labels

In the proposed CFS-DMOA, the samples in source domain is comprehensively evaluated via the cascaded fuzzy system,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

TABLE V
BENCHMARK EVALUATIONS OF DIFFERENT ETO-BASED DMOAS IN TERMS OF MHVD

Functions	n_t, τ_t	KITL-DMOA	CDTL-DMOA	KGBC-DMOA	HKTCFP-DMOA	CFS-DMOA
DF1	1,10	0.3132±2.42e-02(+)	0.2534±1.37e-02(-)	**0.2469±1.49e-02(-)**	0.2857±1.93e-02(≈)	0.2759±1.21e-02
	5,10	0.1650±1.57e-02(+)	0.1408±1.06e-02(+)	0.1493±1.14e-02(+)	0.1390±2.69e-02(+)	**0.1235±8.63e-03**
	10,10	0.1645±2.37e-02(≈)	0.1543±8.58e-03(+)	0.1684±1.52e-02(+)	**0.1035±1.81e-02(-)**	0.1492±7.45e-03
DF2	1,10	0.2193±1.78e-02(+)	**0.1639±1.32e-02(-)**	0.1821±1.55e-02(≈)	0.1802±1.39e-02(≈)	0.1880±1.08e-02
	5,10	0.1188±1.47e-02(+)	0.0926±8.89e-03(≈)	0.1106±1.15e-02(+)	0.1170±1.06e-02(+)	**0.0887±6.20e-03**
	10,10	0.1098±9.65e-03(+)	0.0934±7.99e-03(+)	0.1021±1.20e-02(+)	0.1113±1.92e-02(+)	**0.0841±6.53e-03**
DF3	1,10	0.4622±2.64e-02(+)	0.4516±1.80e-02(+)	0.4334±1.51e-02(≈)	0.4823±2.01e-02(+)	**0.4247±1.06e-02**
	5,10	0.4574±2.80e-02(+)	0.4407±2.33e-02(+)	0.4358±1.79e-02(+)	0.4434±2.94e-02(+)	**0.4178±1.55e-02**
	10,10	0.5633±4.33e-02(+)	**0.4752±1.31e-02(≈)**	0.4943±2.07e-02(+)	0.5137±3.13e-02(+)	0.4786±2.08e-02
DF4	1,10	**0.2675±4.35e-02(-)**	0.3177±2.65e-02(+)	0.2747±2.95e-02(-)	0.3311±1.96e-02(+)	0.2984±1.80e-02
	5,10	**0.2467±4.28e-02(-)**	0.3123±3.07e-02(≈)	0.2885±2.75e-02(-)	0.3259±3.39e-02(≈)	0.3108±1.51e-02
	10,10	0.3232±3.77e-02(+)	**0.2175±2.55e-02(-)**	0.2374±2.97e-02(-)	0.2713±2.71e-02(≈)	0.2543±1.79e-02
DF8	1,10	**0.0281±7.56e-03(-)**	0.0912±9.48e-03(+)	0.0704±1.00e-02(-)	0.0938±1.50e-02(+)	0.0833±8.57e-03
	5,10	**0.0499±1.00e-02(-)**	0.1702±1.62e-02(≈)	0.1605±2.01e-02(≈)	0.1897±1.45e-02(+)	0.1707±1.53e-02
	10,10	**0.0432±7.72e-03(-)**	0.1819±1.23e-02(≈)	0.1772±2.43e-02(≈)	0.1961±1.80e-02(+)	0.1810±2.19e-02
DF11	1,10	0.0732±1.20e-02(+)	0.0681±8.08e-03(+)	0.0609±1.07e-02(≈)	0.0765±1.06e-02(+)	**0.0567±5.96e-03**
	5,10	0.0670±7.50e-03(+)	0.0621±8.93e-03(+)	0.0588±7.52e-03(+)	**0.0495±5.77e-03(≈)**	0.0510±5.87e-03
	10,10	0.0680±8.31e-03(+)	0.0619±5.54e-03(+)	0.0581±5.37e-03(+)	**0.0485±6.41e-03(≈)**	0.0524±6.28e-03
DF12	1,10	0.1660±2.56e-02(≈)	0.1896±1.56e-02(+)	0.1651±1.86e-02(≈)	0.2272±3.81e-02(+)	**0.1644±1.64e-02**
	5,10	0.1133±2.08e-02(≈)	0.1053±1.05e-02(≈)	**0.0948±7.02e-03(-)**	0.1427±2.74e-02(+)	0.1048±1.32e-02
	10,10	0.1043±1.71e-02(+)	0.0938±1.42e-02(≈)	**0.0873±7.39e-03(≈)**	0.1336±2.69e-02(+)	0.0926±8.00e-03
DF13	1,10	0.3849±2.18e-02(+)	0.2417±1.50e-02(+)	0.2335±1.81e-02(+)	0.2606±3.44e-02(+)	**0.2209±1.95e-02**
	5,10	0.2809±4.35e-02(+)	0.1890±1.46e-02(+)	0.2073±1.64e-02(+)	**0.1662±2.42e-02(-)**	0.1796±1.27e-02
	10,10	0.1939±2.41e-02(≈)	0.1853±2.19e-02(≈)	0.2040±2.03e-02(+)	**0.1398±1.75e-02(-)**	0.1900±1.62e-02
DF14	1,10	0.2838±3.98e-02(+)	0.2450±1.44e-02(+)	0.1127±2.36e-02(≈)	0.1891±3.78e-02(+)	**0.1006±1.41e-02**
	5,10	0.1551±2.94e-02(+)	0.1253±1.95e-02(+)	0.1124±1.40e-02(+)	0.1361±5.91e-02(+)	**0.0914±1.32e-02**
	10,10	**0.0740±1.13e-02(-)**	0.1247±3.15e-02(+)	0.1019±1.27e-02(+)	0.0819±2.20e-02(≈)	0.0912±7.96e-03
+ / - / ≈		17 / 6 / 4	16 / 3 / 8	13 / 6 / 8	17 / 3 / 7	-

TABLE VI
INFLUENCES OF INFORMATION DIVERSITY IN SOURCE DOMAIN ON ALGORITHM PERFORMANCE

Functions	n_t, τ_t	MIGD		MSp	
		CFS_v1-DMOA	CFS-DMOA	CFS_v1-DMOA	CFS-DMOA
DF1	1,10	**0.2404±1.81e-02**	0.2776±1.48e-02	0.0356±1.81e-02	**0.0342±1.70e-02**
	10,10	0.1434±1.54e-02	**0.1268±6.88e-03**	0.0166±8.82e-03	**0.0133±5.34e-03**
DF2	1,10	**0.1692±9.15e-03**	0.1834±8.82e-03	**0.0557±1.82e-02**	0.0561±1.73e-02
	10,10	0.0964±6.97e-03	**0.0755±4.15e-03**	**0.0265±9.37e-03**	0.0332±1.47e-02
DF3	1,10	4.5740±1.06e+01	**0.4675±1.23e-02**	**0.2541±7.01e-02**	0.2798±1.04e-01
	10,10	4.6246±6.12e+00	**0.5745±2.94e-02**	0.1623±7.96e-02	**0.1143±5.84e-02**
DF4	1,10	0.5594±6.45e-02	**0.5191±4.35e-02**	0.7710±2.80e-01	**0.7295±2.17e-01**
	10,10	0.6494±5.42e-02	**0.6174±4.09e-02**	0.9059±3.28e-01	**0.8872±2.92e-01**
DF8	1,10	0.2716±1.74e-02	**0.2642±2.47e-02**	0.1257±4.20e-02	**0.1224±3.56e-02**
	10,10	0.3150±2.40e-02	**0.3090±2.53e-02**	0.0887±2.68e-02	**0.0754±2.73e-02**
DF11	1,10	0.1523±8.05e-03	**0.1363±4.38e-03**	**0.0435±1.72e-03**	0.0449±3.89e-03
	10,10	0.1506±5.90e-03	**0.1377±3.91e-03**	0.0469±2.42e-03	**0.0468±9.50e-04**
DF12	1,10	0.4788±1.76e-02	**0.4651±1.28e-02**	**0.1299±3.86e-02**	0.1368±4.52e-02
	10,10	0.4907±2.53e-02	**0.4686±1.29e-02**	**0.0932±3.70e-02**	0.0974±3.27e-02
DF13	1,10	0.3580±3.70e-02	**0.3320±2.30e-02**	0.3738±1.30e-01	**0.3204±1.18e-01**
	10,10	0.3380±3.16e-02	**0.3066±1.45e-02**	0.3583±7.12e-02	**0.3057±1.06e-01**
DF14	1,10	0.1272±1.44e-02	**0.1096±8.97e-03**	0.1379±1.12e-01	**0.1035±3.60e-02**
	10,10	0.1322±1.77e-02	**0.1173±5.95e-03**	0.1069±4.64e-02	**0.0907±3.59e-02**

which are then assigned with the soft labels in $(0,1)$. To validate the advantages of such strategy in comparison to the conventional one-hot coding, another algorithm variant CFS_v2-DMOA is designed, where the labels of source domain samples are based on the dominance relationship. That is, in CFS_v2-DMOA, the dominated and non-dominated individuals in $\mathcal{D}_s$ are labeled as "0" and "1", respectively, whereas the acquisition of $\mathcal{D}_s$ is the same as CFS-DMOA. Under the dynamic parameter settings of $(n_t, \tau_t) \in \{(5,10),(10,10)\}$, evaluation results of above two algorithms are displayed in Table VII.

TABLE VII
PERFORMANCE COMPARISONS BETWEEN ALGORITHMS USING SOFT AND HARD LABELS

Functions	n_t, τ_t	MIGD		MHVD	
		CFS_v2-DMOA	CFS-DMOA	CFS_v2-DMOA	CFS-DMOA
DF1	5,10	0.1317±1.49e-02	**0.1064±7.01e-03**	0.1474±1.18e-02	**0.1235±8.63e-03**
	10,10	0.1563±2.15e-02	**0.1268±6.88e-03**	0.1737±2.04e-02	**0.1492±7.45e-03**
DF2	5,10	0.0975±7.21e-03	**0.0817±4.24e-03**	0.1053±6.50e-03	**0.0887±6.20e-03**
	10,10	0.0911±1.08e-02	**0.0755±4.15e-03**	0.0972±1.02e-02	**0.0841±6.53e-03**
DF3	5,10	1.3660±3.87e+00	**0.4639±2.16e-02**	0.4372±2.15e-02	**0.4178±1.55e-02**
	10,10	1.2853±1.93e+00	**0.5745±2.94e-02**	0.5129±2.23e-02	**0.4786±2.08e-02**
DF4	5,10	0.6991±1.02e-01	**0.6044±2.87e-02**	0.3329±3.40e-02	**0.3108±1.51e-02**
	10,10	0.6623±6.21e-02	**0.6174±4.09e-02**	0.2657±2.04e-02	**0.2543±1.79e-02**
DF8	5,10	0.3318±2.45e-02	**0.3077±2.58e-02**	**0.1696±2.03e-02**	0.1707±1.53e-02
	10,10	0.3208±2.44e-02	**0.3090±2.53e-02**	**0.1738±1.66e-02**	0.1810±2.19e-02
DF11	5,10	0.1496±5.31e-03	**0.1378±2.90e-03**	0.0609±7.28e-03	**0.0510±5.87e-03**
	10,10	0.1471±6.24e-03	**0.1377±3.91e-03**	0.0597±5.73e-03	**0.0524±6.28e-03**
DF12	5,10	0.5338±2.46e-02	**0.5213±1.31e-02**	0.1060±1.07e-02	**0.1048±1.32e-02**
	10,10	0.4889±2.27e-02	**0.4686±1.29e-02**	**0.0925±1.26e-02**	0.0926±8.00e-03
DF13	5,10	0.3361±2.44e-02	**0.2973±1.09e-02**	0.2057±1.88e-02	**0.1796±1.27e-02**
	10,10	0.3358±3.50e-02	**0.3066±1.45e-02**	0.2075±2.11e-02	**0.1900±1.62e-02**
DF14	5,10	0.1358±1.47e-02	**0.1152±8.23e-03**	0.1156±1.53e-02	**0.0914±1.32e-02**
	10,10	0.1376±1.46e-02	**0.1173±5.95e-03**	0.1115±1.56e-02	**0.0912±7.96e-03**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

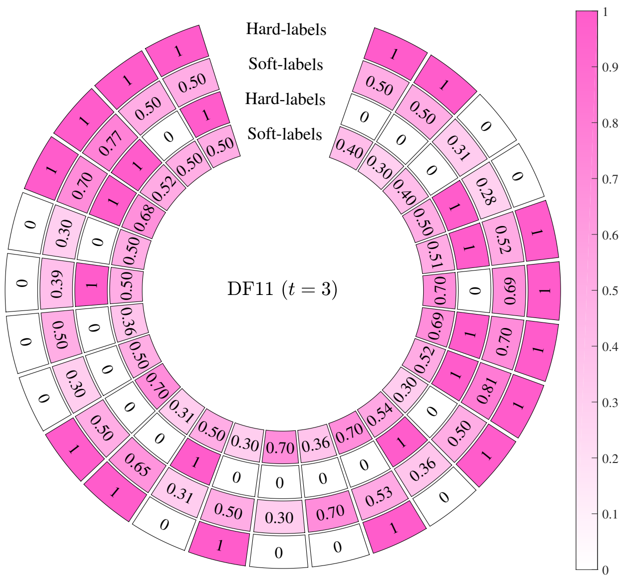


Fig. 5. Visualization of soft/hard labels on problem DF11, where $(n_t, \tau_t) = (10, 10)$.

TABLE VIII
EFFECTIVENESS VALIDATION ON APPLYING KMM

Functions	n_t, τ_t	CFS_v3-DMOA	CFS-DMOA
DF1	1,10	0.2961±2.80e-02	**0.2776±1.48e-02**
	5,10	**0.1003±9.24e-03**	0.1064±7.01e-03
DF2	1,10	0.2189±1.54e-02	**0.1834±8.82e-03**
	5,10	0.0835±1.01e-02	**0.0817±4.24e-03**
DF3	1,10	0.6303±6.99e-02	**0.4675±1.23e-02**
	5,10	**0.4542±5.80e-02**	0.4639±2.16e-02
DF4	1,10	0.8516±1.61e-01	**0.5191±4.35e-02**
	5,10	0.7436±1.68e-01	**0.6044±2.87e-02**
DF8	1,10	0.3052±2.41e-02	**0.2642±2.47e-02**
	5,10	0.3766±2.96e-02	**0.3077±2.58e-02**
DF11	1,10	0.1576±8.25e-03	**0.1363±4.38e-03**
	5,10	0.1521±7.45e-03	**0.1378±2.90e-03**
DF12	1,10	0.5523±3.85e-02	**0.4651±1.28e-02**
	5,10	**0.4152±3.66e-02**	0.5213±1.31e-02
DF13	1,10	0.3612±3.50e-02	**0.3320±2.30e-02**
	5,10	0.3506±2.68e-02	**0.2973±1.09e-02**
DF14	1,10	0.1209±1.04e-02	**0.1096±8.97e-03**
	5,10	0.1322±2.40e-02	**0.1152±8.23e-03**

As is shown from Table VII, the original CFS-DMOA wins all tests in terms of MIGD, and is slightly inferior to the variant CFS_v2-DMOA in merely three test cases with regard to MHVD, which exhibits the great advantages of using fuzzy logic to label the source domain samples as compared to the binary hard labels.

Moreover, on tri-objective problem DF11, the hard and soft labels in the same environment $t = 3$ (where the knowledge transfer strategy works for the first time, see Algorithm. 5) are visualized in Fig. 5, and it can be observed that: a) some positive samples (labeled as “1”) are deemed to have the reference value in a medium level, and b) some dominated negative learning samples (labeled as “0”) are deemed valuable by CFS, which are assigned with soft labels even larger than 0.6. Therefore, it can be concluded that the designed CFS realizes comprehensive estimations on the previous evolutionary experiences from multiple views, which enables the objective characterization of the historical information rather than simply referring to the dominance relationship, thereby effectively alleviating the negative transfer phenomenon.

E. Effectiveness of KMM Method

KMM method is applied in the proposed CFS-DMOA to map the individuals in both $\mathcal{D}_s$ and $\mathcal{D}_t$ to a high-dimensional RKHS, where the data distribution difference between the two domains is decreased so as to promote effective knowledge transfer. To verify the effectiveness of applying KMM method, an algorithm variant CFS_v3-DMOA is designed in this part, where the samples in target domain directly participate in training the regression model, whose labels are provided by the designed CFS. Comparison results between above two algorithms are displayed in Table VIII, where the reported metric is MIGD and the dynamic parameter settings include $(n_t, \tau_t) \in \{(1, 10), (5, 10)\}$.

According to Table VIII, in most test cases, the original CFS-DMOA outperforms the variant CFS_v3-DMOA, which suggests the effectiveness of applying KMM method to alleviate the negative knowledge transfer. On one hand, the weighting process on the learning sample (i.e., the individuals in $\mathcal{D}_s$) in KMM can be regarded as a further estimation on the importance of historical information, and as a result, it promisingly enables highly effective knowledge transfer with the guidance of individuals in $\mathcal{D}_t$. On the other hand, as it lacks prior searching experiences in the new environment, labeling individuals in $\mathcal{D}_t$ is more or less subjective, and in this regard, the application of KMM successfully avoids the potential negative transfer caused by the improper manual intervention.

F. Case study of Dynamic Industrial Flow-shop Scheduling

In this subsection, a case study of dynamic flow-shop scheduling is carried out, where some batches of work-piece need repairing on the flow line, and the maintenance time spent in each station (in total 6 stations are placed) depends on the situation of work-piece. By selecting every maintenance batch as the research object, a dynamic scheduling optimization model is accordingly established, whose objectives contain the minimization of 1) the makespan regarding to the whole batch, and 2) the summed idle time of all stations. In this case study, the baseline algorithm employed for comparison directly applies the static optimizer to obtain the optimal solutions of each maintenance batch, which treats the dynamic problems as a series of individual task. By contrast, our algorithm encourages the static optimizer to absorb the previous searching experiences to improve the optimization efficiency.

In the experiment, static optimizer in CFS-DMOA employs the NSGA-II [2], and other settings remain the same. In total 30 batches of maintenance tasks are used for algorithm evaluation (i.e., the experiment contains 30 environments), and average performance is compared based on following two defined evaluation metrics. Additionally, both CFS-DMOA

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

and the baseline algorithm are run 20 times to alleviate the influence of randomness.

1) Competence. Suppose P_a and P_b denote the Pareto sets obtained by two algorithms "a" and "b", respectively, and let P_{ab} denote the Pareto solutions in set $P_a \cup P_b$. The ratio of individuals in P_{ab} contributed by P_a (or P_b) is adopted to reflect the competence of algorithm "a" (or "b").
2) Competitiveness. Let n_b denote the number of individuals in P_b that can be dominated by solutions in P_a, then the proportion of "$n_b/|P_b|$" is deemed as the competitiveness of algorithm "a", and vice versa. Notice that "$n_a/|P_a| + n_b/|P_b|$" may not equal 1 due to the existence of non-dominated Pareto solutions.

TABLE IX
COMPETENCE AND COMPETITIVENESS COMPARISON IN THE DYNAMIC FLOW-SHOP SCHEDULING TASK

Experiments	Competence		Competitiveness	
	CFS-DMOA	Baseline	CFS-DMOA	Baseline
# 1	**54.27%**	45.73%	**50.03%**	40.89%
# 2	**58.17%**	41.83%	**52.82%**	34.57%
# 3	**52.75%**	47.25%	**47.40%**	44.05%
# 4	46.86%	**53.14%**	42.60%	**49.69%**
# 5	**52.96%**	47.04%	**46.48%**	37.92%
# 6	**55.27%**	44.73%	**50.23%**	37.76%
# 7	49.19%	**50.81%**	**44.32%**	42.74%
# 8	**55.92%**	44.08%	**50.93%**	37.83%
# 9	49.47%	**50.53%**	39.89%	**43.17%**
# 10	**61.91%**	38.09%	**56.83%**	34.87%
# 11	**60.39%**	39.61%	**50.89%**	37.08%
# 12	**55.17%**	44.83%	**45.28%**	37.67%
# 13	42.88%	**57.12%**	34.49%	**53.21%**
# 14	**53.27%**	46.73%	**45.45%**	39.72%
# 15	**53.88%**	46.12%	**45.23%**	34.77%
# 16	**64.33%**	35.67%	**58.37%**	27.70%
# 17	44.25%	**55.75%**	36.43%	**50.68%**
# 18	**51.44%**	48.56%	**46.45%**	43.31%
# 19	**64.15%**	35.85%	**59.50%**	31.18%
# 20	**52.70%**	47.30%	**50.22%**	43.13%
Average	**53.96%**	46.04%	**47.69%**	40.10%

According to Table IX, the proposed CFS-DMOA, which follows the idea of evolutionary transfer optimization, outperforms the baseline algorithm that individually processes each scheduling task in most cases, which can contribute 7.92% more Pareto solutions by average in 20 experiments. Notice that in the 7-th experiment, the baseline algorithm can averagely provide more than half Pareto solutions in the 30 environment, but its competitiveness score is 1.58% lower than that of CFS-DMOA. Simultaneously, 47.69% of Pareto solutions obtained by the baseline will be dominated by the PS of CFS-DMOA in an average level, and these results indicate the reliability and practicality of our algorithm in real-world optimization scenes. According to the scatter plots shown in Fig. 6, the proposed CFS-DMOA can well adapt to the changing environments, which is able to provide more Pareto solutions with better quality for the decision-makers.

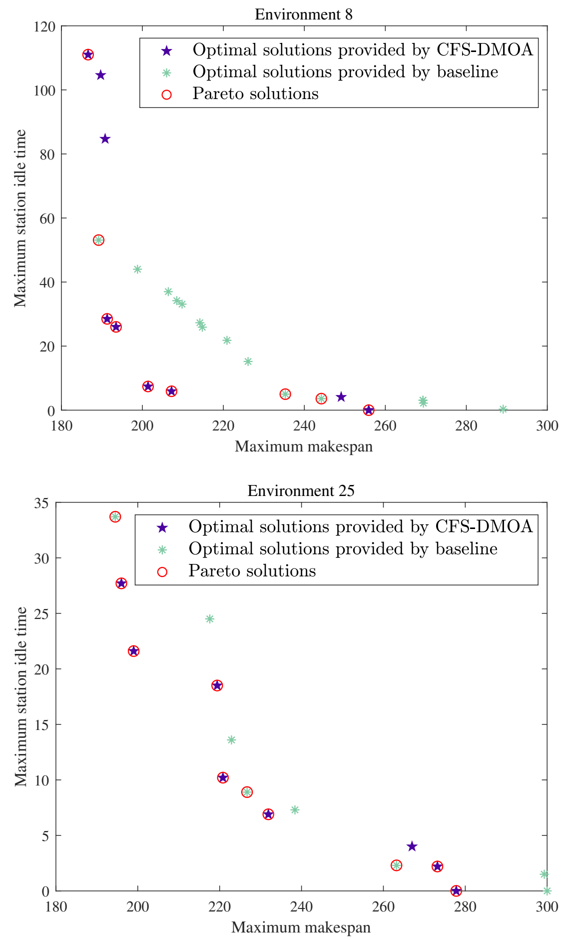


Fig. 6. Scheduling results of two maintenance tasks.

G. Outlook for Future Work

Based on above results and discussions, the proposed CFS-DMOA has achieved several promising progresses, whereas there are still some spaces for further improvement. Firstly, developing more powerful evolutionary algorithms to serve as the static optimizer no doubts benefits enhancing the overall performance [7], [13], [37]; secondly, investigating other mechanisms to quantify the dynamic behaviors could provide more in-depth insights on handling DMOPs [21], [41]; lastly, it is also promising to apply the proposed CFS-DMOA to more real-world scenes [4], [8], [47], which encourages the popularization of ETO-based algorithms in various optimization tasks, where the transfer of previous valuable experiences can promote better working efficacy.

V. CONCLUSION

In this paper, a novel CFS-DMOA has been proposed to solve DMOPs, where a cascaded fuzzy system is designed to objectively characterize the value of historical information. Based on a comprehensive evaluation in terms of both convergence and diversity, samples in $\mathcal{D}_s$ have been assigned with soft labels, which breaks the limitations of one-hot coding and benefits promoting objective knowledge transfer. In addition, KMM method has been applied to avoid labeling samples in

$\mathcal{D}_t$, which decreases the data distribution difference between above two domains, thereby benefiting efficient knowledge transfer.

Experimental results have shown the superiority of CFS-DMOA in effectively handling various dynamic behaviors, which outperforms other four state-of-the-art ETO-based DMOAs in most benchmark evaluations. Moreover, the great competitiveness of CFS-DMOA with regard to alleviating the negative transfer has been sufficiently validated, where the rich information contained in source domain, the introduction of soft labels, and the application of KMM method have collaboratively endowed our algorithm with the excellent performance. By promoting objective and comprehensive knowledge transfer, the proposed CFS-DMOA has addressed the issue of "how to transfer" from an innovative perspective of information characterization, which paves a new way to break through the bottleneck of negative transfer.

REFERENCES

[1] Z. Aliniya and S. Hossein Khasteh, A novel combinational response mechanism for dynamic multi-objective optimization, *Expert Systems with Applications*, vol. 233, art no. 120951, 2023.

[2] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182-197, 2002.

[3] K. Deb, U. B. Rao and S. Karthik, Dynamic multi-objective optimization and decision-making using modified NSGA-II: a case study on hydro-thermal power scheduling, *International Conference on Evolutionary Multi-criterion Optimization*, pp. 803-817, 2007.

[4] J. Deng, M. Tipaldi, L. Glielmo, P. Massenio and L. Del Re, A dynamic programming approach for energy management in hybrid electric vehicles under uncertain driving conditions, *International Journal of Systems Science*, vol. 55, no. 7, pp. 1304–1325, 2024.

[5] J. Derrac, S. Garcia, D. Molina and F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 3-18, 2011.

[6] J. Eaton, S. Yang and M. Gongora, Ant colony optimization for simulated dynamic multi-objective railway junction rescheduling, *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 2980-2992, 2017.

[7] J. Fang, W. Liu, L. Chen, S. Lauria, A. Miron and X. Liu, A survey of algorithms, applications and trends for particle swarm optimization, *International Journal of Network Dynamics and Intelligence*, vol. 2, no. 1, pp. 24-50, 2023.

[8] B. Freek, J. Alireza, O. Bas and D. Groen, Analysing the effect of a dynamic physical environment network on the travel dynamics of forcibly displaced persons in Mali, *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 1, art no. 100003, 2024.

[9] A. Gretton, K. Borgwardt, M. Rasch, B. Schoelkopf and A. Smola, A kernel two-sample test, *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 723-773, 2012.

[10] L. Hu, Y. Yang, Z. Tang, Y. He and X. Luo, FCAN-MOPSO: An improved fuzzy-based graph clustering algorithm for complex networks with multiobjective particle swarm optimization, *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 10, pp. 3470-3484, 2023.

[11] Y. Hu, J. Zheng, S. Jiang, S. Yang, J. Zou and R. Wang, A Mahalanobis distance-based approach for dynamic multiobjective optimization with stochastic changes, *IEEE Transactions on Evolutionary Computation*, vol. 28, no. 1, pp. 238-251, 2024.

[12] J. Huang, A. Smola, A. Gretton, K. Borgwardt and B. Schölkopf, Correcting sample selection bias by unlabeled data, *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*, vol. 19, pp. 601-608, 2007.

[13] Z. Huang, D. Zhu, Y. Liu and X. Wang, Multi-strategy sparrow search algorithm with non-uniform mutation, *Systems Science & Control Engineering*, vol. 10, no. 1, pp. 936-954, 2022.

[14] A. Ikotun, A. Ezugwu, L. Abualigah, B. Abuhaija and H. Jia, K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data, *Information Sciences*, vol. 622, pp. 178-210, 2023.

[15] S. Jiang, S. Yang, X. Yao, K. Tan, M. Kaiser and N. Krasnogor, Benchmark problems for CEC2018 competition on dynamic multiobjevtive optimisation, *Proceedings of 2018 CEC Competition*, pp. 1-18, 2018.

[16] M. Jiang, Z. Wang, H. Hong and G. Yen, Knee point-based imbalanced transfer learning for dynamic multiobjective optimization, *IEEE Transactions on Evolutionary Computation*, vol. 25, no. 1, pp. 117-129, 2020.

[17] M. Jiang, Z. Wang, L. Qiu, S. Guo, X. Gao and K. Tan, A fast dynamic evolutionary multiobjective algorithm via manifold transfer learning, *IEEE Transactions on Cybernetics*, vol. 51, no. 7, pp. 3417-3428, 2021.

[18] M. Jiang, Z. Wang, S. Guo, X. Gao and K. Tan, Individual-based transfer learning for dynamic multiobjective optimization, *IEEE Transactions on Cybernetics*, vol. 51, no. 10, pp. 4968-4981, 2021.

[19] H. Kalantari, A. Badiee, A. Dezhboro, H. Mohammadi and E. Tirkolaee, A fuzzy profit maximization model using communities viable leaders for information diffusion in dynamic drivers collaboration networks, *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 2, pp. 370-379, 2023.

[20] H. Li, H. Liu, C. Lan, Y. Yin, P. Wu, C. Yan and N. Zeng, SMWO/D: A decomposition-based switching multi-objective whale optimiser for structural optimisation of turbine disk in aero-engines, *International Journal of Systems Science*, vol. 54, no. 8, pp. 1713-1728, 2023.

[21] H. Li, Z. Fang, L. Hu, H. Liu, P. Wu and N. Zeng. A novel population robustness-based switching response framework for solving dynamic multi-objective problems, *Neurocomputing*, vol. 583, art no. 127601, 2024.

[22] H. Li, Z. Wang, C. Lan, P. Wu and N. Zeng, A novel dynamic multiobjective optimization algorithm with non-inductive transfer learning based on multi-strategy adaptive selection, *IEEE Transactions on Neural Networks and Learning Systems*, 2023. doi:10.1109/TNNLS.2023.3295461

[23] H. Li, Z. Wang, C. Lan, P. Wu and N. Zeng, A novel dynamic multiobjective optimization algorithm with hierarchical response system, *IEEE Transactions on Computational Social Systems*, 2023. doi:10.1109/TCSS.2023.3293331

[24] J. Li, T. Sun, Q. Lin, M. Jiang and K. Tan, Reducing negative transfer learning via clustering for dynamic multiobjective optimization, *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 5, pp. 1102-1116, 2022.

[25] Z. Liang, T. Wu, X. Ma, Z. Zhu and S. Yang, A dynamic multiobjective evolutionary algorithm based on decision variable classification, *IEEE Transactions on Cybernetics*, vol. 52, no. 3, pp. 1602-1615, 2022.

[26] Z. Liang, X. Xu, L. Liu, Y. Tu and Z. Zhu, Evolutionary many-task optimization based on multisource knowledge transfer, *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 2, pp. 319-333, 2022.

[27] R. Liu, P. Yang and J. Liu, A dynamic multi-objective optimization evolutionary algorithm for complex environmental changes, *Knowledge-Based Systems*, vol. 216, article no. 106612, 2021.

[28] S. Liu, Q. Lin, J. Li and K. Tan, A survey on learnable evolutionary algorithms for scalable multiobjective optimization, *IEEE Transactions on Evolutionary Computation*, 2023. doi:10.1109/TEVC.2023.3250350

[29] S. Lyshevski, Analytic design of constrained control laws for nonlinear dynamic systems with symmetric and asymmetric limits, *International Journal of Systems Science*, vol. 55, no. 3, pp. 453-466, 2024.

[30] G. Ma, Z. Wang, W. Liu, J. Fang, Y. Zhang, H. Ding and Y. Yuan, Estimating the state of health for Lithium-ion batteries: a particle swarm optimization-assisted deep domain adaptation approach, *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 7, pp. 1530-1543, 2023.

[31] X. Ma, J. Yang, H. Sun, Z. Hu and L. Wei, Feature information prediction algorithm for dynamic multi-objective optimization problems, *European Journal of Operational Research*, vol. 295, no. 3, pp. 965-981, 2021.

[32] L. Pan, W. Xu, L. Li, C. He and R. Cheng, Adaptive simulated binary crossover for rotated multi-objective optimization, *Swarm and Evolutionary Computation*, vol. 60, art no. 100759, 2021.

[33] D. Pardoe and P. Stone, Boosting for regression transfer, *Proceedings of the 27th International Conference on Machine Learning*, pp. 863-870, 2010.

[34] M. Rong, D. Gong, Y. Zhang, Y. Jin and P. Witold, Multidirectional prediction approach for synamic multiobjective optimization problems, *IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp. 3362-3374, 2019.

[35] G. Ruan, L. Minku, S. Menzel, B. Sendhoff and X. Yao, When and how to transfer knowledge in dynamic multi-objective optimization, *2019 IEEE Symposium Series on Computational Intelligence*, pp. 2034-2041, 2019.

[36] S. Sahmoud and H. Topcuoglu, A memory-based NSGA-II algorithm for dynamic multi-objective optimization problems, *European Conference on the Applications of Evolutionary Computation*, pp. 296-310, 2016.

[37] F. Tan, Z. Yuan, Y. Zhang, S. Tang, F. Guo and S. Zhang, Improved genetic algorithm based on rule optimization strategy for fibre allocation,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/TCYB.2024.3470648 IEEE Transactions on Cybernetics

Systems Science & Control Engineering, vol. 12, no. 1, art no. 2347887, 2024.

[38] K. Tan, L. Feng and M. Jiang, Evolutionary transfer optimization-a new frontier in evolutionary computation research, *IEEE Computational Intelligence Magazine*, vol. 16, no. 1, pp. 22-33, 2021.

[39] X. Wang, Y. Sun and D. Ding, Adaptive dynamic programming for networked control systems under communication constraints: a survey of trends and techniques, *International Journal of Network Dynamics and Intelligence*, vol. 1, no. 1, pp. 85-98, 2022.

[40] Y. Wang, K. Li and G. Wang, Combining key-points-based transfer learning and hybrid prediction strategies for dynamic multi-objective optimization, *Mathematics*, vol. 10, no. 12, pp. 2117, 2022.

[41] Z. Wang, L. Xue, Y. Guo, M. Han and S. Liang, Solving dynamic multi-objective optimization problems via quantifying intensity of environment changes and ensemble learning-based prediction strategies, *Applied Soft Computing*, vol. 154, art no. 111317, 2024.

[42] X. Xia, H. Song, Y. Zhang, L. Gui, X. Xu, K. Li and Y. Li, A particle swarm optimization with adaptive learning weights tuned by a multiple-input multiple-output fuzzy logic controller, *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 7, pp. 2464-2478, 2023.

[43] K. Xu, Y. Xia, J. Zou, Z. Hou, S. Yang, Y. Hu and Y. Liu, A cluster prediction strategy with the induced mutation for dynamic multi-objective optimization, *Information Sciences*, vol. 661, art no. 120193, 2024.

[44] C. Yang, H. Yang, D. Zhu, Y. Hu, Y. Zhang, H. Ma and Z. Huang, Chaotic sparrow search algorithm with manta ray spiral foraging for engineering optimization, *Systems Science & Control Engineering*, vol. 11, no. 1, 2023.

[45] F. Yao and G. Wang, Transfer learning based on clustering difference for dynamic multi-objective optimization, *Applied Science*, vol. 13, no. 8, art no. 4795, 2023.

[46] Y. Ye, L. Li, Q. Lin, K. Wong, J. Li and Z. Ming, Knowledge guided Bayesian classification for dynamic multi-objective optimization, *Knowledge-Based Systems*, vol. 250, art no. 109173, 2022.

[47] Y. Yuan, Q. Song and B. Zhou, Emergency medical service location problem based on physical bounds using chance-constrained programming approach, *International Journal of Systems Science*, vol. 54, no. 3, pp. 704-716, 2022.

[48] Q. Zhang and H. Li, MOEA/D: a multiobjective evolutionary algorithm based on decomposition, *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712-731, 2007.

[49] X. Zhang, G. Yu, Y. Jin and F. Qian, Elitism-based transfer learning and diversity maintenance for dynamic multi-objective optimization, *Information Sciences*, vol. 636, art no. 118927, 2023.

[50] Y. Zhang, P. Wang, H. Yang and Q. Cui, Optimal dispatching of microgrid based on improved moth-flame optimization algorithm based on sine mapping and Gaussian mutation, *Systems Science & Control Engineering*, vol. 10, no. 1, pp. 115-125, 2022.

[51] T. Zheng, Y. Zhou, M. Hu and J. Zhang, Dynamic scheduling for large-scale flexible job shop based on noisy DDQN. *International Journal of Network Dynamics and Intelligence*, vol. 2, no. 4, article no. 100015, 2023.

[52] A. Zhou, Y. Jin and Q. Zhang, A population prediction strategy for evolutionary dynamic multiobjective optimization, *IEEE Transactions on Cybernetics*, vol. 44, no. 1, pp. 40-53, 2014.