

Evaluate the Effectiveness of Multiobjective Evolutionary Algorithms by Box Plots and Fuzzy TOPSIS

Xiaobing Yu^{1,2,*}, Chenliang Li^{1,2}, Hong Chen^{1,2}, Xianrui Yu^{1,2}

¹School of Management Science and Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China

²Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science & Technology, Nanjing 210044, China

ARTICLE INFO

Article History

Received 09 Oct 2018

Accepted 17 Apr 2019

Keywords

Multiobjective problems

MOEAs

Box plots

Fuzzy TOPSIS

ABSTRACT

Now, there are a lot of multiobjective evolutionary algorithms (MOEAs) available and these MOEAs argue that they are competitive. In fact, these results are generally unfair and unfaithful. In order to make fair comparison, comprehensive performance index system is established. The weights among the performance index system are solved by an adaptive differential evolution (ADE) algorithm. An approach is proposed to estimate MOEAs based on box plots and fuzzy TOPSIS. Box plots are employed to depict features of performance indicators and fuzzy TOPSIS is used to make evaluation. Experiments have been tested on IEEE CEC2009. The experiment results have revealed that the evaluation approach is effective, fair, and faithful when evaluating MOEAs.

© 2019 The Authors. Published by Atlantis Press SARL.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

Evolutionary algorithm (EA) is a useful approach to solve multiobjective problems (MOPs). multiobjective evolutionary algorithms (MOEAs) try to find a good approximation [1,2]. Nowadays, there are a lot of MOEAs available. They are widely used to solve the real applications.

Many performance indicators are designed to evaluate MOEAs. NSGAI employed convergence and diversity indicators to compare with SPEA and PAES [3]. The set convergence and inverted generation distance (IGD) were used to evaluate the performance of MOEAD [4]. Convergence measurement, spread, hyper volume (HV), and computational time (CT) were selected as performance indicators in epsilon-MOEA [5]. To validate the proposed MOPSO, four quantitative performance indicators (success counting IGD, set coverage, two set difference HVs) and qualitative performance indicators (plotting the Pareto fronts) were adopted [6]. Three indicators: spread, additive unary epsilon, and HV were considered in SMPSO [7]. Spacing and binary C were used in GDE3 [8]. Three performance indicators: spread, generation distance (GD), and HV were used to estimate ABYSS [9]. GD, diversity, CT, and box plot were considered as measurements in MOSOS [10]. GD and diversity indicators were adopted in MOEDA [11].

In fact, every indicator can't demonstrate the whole performance of MOEAs. It can just depict some specific qualifications of performance. It can't estimate the whole performance of MOEAs. Therefore, it is unfaithfully to measure MOEAs performance by one or

two indicators. Besides, it is known that we can't get any useful conclusions based on a single run. Generally speaking, statistical measurement is often used, such as mean values. However, the mean values can't give more information of indicators, such as the max value, min value, and distribution.

In order to deal with these problems, a comprehensive performance index and a novel evaluation method are proposed. Evaluations of MOEAs involve many indicators, it can be considered as a multiple criterion decision-making (MCDE) problem. MCDE techniques can be used to make evaluations. Comprehensive performance index system (HV, GD, IGD, ϵ_+ -Indicator, spacing, and CT) is proposed. For MOEAs, it is very important to accelerate convergence of MOEAs, while keep diversity of population. Among the proposed performance index system, HV, IGD, and ϵ_+ -Indicator measure the diversity and convergence, GD is utilized to measure convergence of MOEAs, and spacing is utilized to measure diversity of population. Therefore, they can evaluate the whole performance of MOEAs. However, the relative weights among six indicators are not equal. The problem of acquiring the weights can be converted to a typical constraint optimization problem (COP). In order to deal with the COP, an adaptive algorithm based on differential evolution (DE) algorithm is presented.

An approach is proposed by box plots and fuzzy TOPSIS to make fair comparisons. Box plot is commonly employed to make comparisons [3,10], which is a useful approach to depict groups of data. It is used to depict the results of performance indicators. As it is hard to compare box plots directly, fuzzy method is used to evaluate them. The efforts can give more fair and faithful comparisons than single indicator and mean values. To validate the effectiveness

* Corresponding author. Email: yuxb111@163.com

of the evaluation method, 14 functions from CEC 2009 are used to perform testing and make comparisons.

The main advantage of the proposed method can be listed as follows:

1. The performance index system is proposed, which includes HV, GD, IGD, ϵ_+ -Indicator, spacing, and computation time. They can give more fair and faithful comparisons than single indicator and mean values.
2. A novel method on how to solve the weights among the performance index system is proposed, which is different from the traditional method.
3. Box plots and fuzzy TOPSIS are combined to evaluate MOEAs. It is very convenient to depict features of performance indicators by Box plots. Fuzzy TOPSIS is used to make evaluation.

Section 2 presents the previous work and Section 3 introduces performance index system. The related methodologies are introduced in Section 4. In Section 5, proposed approach is presented. In Section 6, experiments are presented. In Section 7, conclusions are illustrated.

2. PREVIOUS WORK

Over the past decades, research on MOEAs has produced a large number of algorithms. These MOEAs are compared with very few, often a decade older MOEAs using some performance indicators. In fact, a common baseline for comparison should be established. Otherwise, the state of the art in MOEAs is a disputed topic [12]. Four different performance indicators: IGD, spacing, maximum spread (MS), and HV are adopted to make evaluations among different MOEAs [13]. Five indicators Pareto dominance indicator, IGD, spacing, MS, and HV are chosen to evaluate MOEAs [1]. Epitropakis *et al.* adopted average percentage of fault detection (APFD) indicator to evaluate seven MOEAs [14]. Error rate (ER), generational distance (GD), and Pareto subset (PS) are used to evaluate five MOEAs [15].

In the real application, AMALGAM outperformed NSGA-II, NSHS, and NSDE for the evaluation of rehabilitation of urban storm water drainage systems in terms of convergence and diversity indicators [16]. The comparative performance of MOEAs has been investigated according to their overall end-of-run results within pre-specified computational budgets when solving water resource optimization problems [17].

However, these comparisons are generally unfair and unfaithful. On the one hand, they are based on few performance indicators. Systematic, comprehensive evaluations are not implemented. On the other hand, some new proposed MOEAs are compared with very older MOEAs. In fact, evaluations of MOEAs can be considered as MCDE problems. Therefore, comprehensive performance index system should be established.

3. PERFORMANCE INDICATORS

For different performance indicators, different MOEAs tend to perform different [18]. However, it is very important to accelerate

convergence of MOEAs, while keep diversity of population [19]. Thus, convergence and diversity should be given more attentions. The following outlines six of the commonly used performance indicators which can be used to evaluate MOEAs [20].

3.1. Hyper Volume

HV is demonstrated in Figure 1. The indicator computes the volume of the space which is dominated by these nondominated solutions. It is bounded by a reference point, which is plus some fixed increment. This fixed increment is necessary to allow the external points to contribute to the HV. HV can reflect the degree of outperformance between two approximation sets [21].

3.2. Generational Distance

GD is the average distance. It is from every solution in the non-dominated solutions to the nearest solutions in the reference set in Figure 2.

$$GD = \frac{\sqrt{\sum_{i=1}^n d_i^2}}{n} \quad (1)$$

The performance indicator measures convergence to the reference set. GD by itself can be misleading, as an approximation set containing a single solution in close convergence to the reference set produces low GD measurements, and is often combined with diversity measures in practice [22].

3.3. Inverted Generational Distance

IGD is the inverse of GD. It is also the average distance. But, it is from every solution to the nearest solution in the nondominated solutions in Figure 3.

$$IGD = \frac{\sum_{v \in RS} d(v, P_A)}{|RS|} \quad (2)$$

The distance is Euclidean distance. IGD measures diversity and convergence since an approximation set is required to have solutions near each reference set point in order to achieve low IGD measurements [20].

3.4. ϵ_+ - Indicator

Good convergence and good diversity both result in low ϵ values, as the distance that the approximation needs to be translated is reduced. However, if there is a region of the reference set that is poorly approximated by the solutions in the approximation set, a large ϵ is required. Therefore, the ϵ_+ -indicator measures the consistency of an approximation set [23].

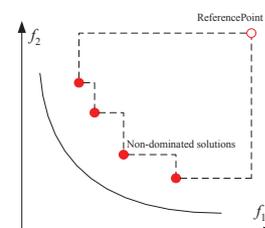


Figure 1 | Hyper volume measurement.

3.5. Spacing

Spacing, shown in Figure 4, measures the uniformity of the spacing between solutions in an approximation set [20]. An approximation set that is well spaced will not contain dense clusters of solutions separated by large empty expanses. Note that, since spacing does not involve a reference set in its calculation, an approximation can register good spacing while having poor convergence to the reference set. It is therefore recommended to use spacing in conjunction with a performance indicator for convergence.

3.6. Computational Time

The CT can reflect the space and time complexity of an algorithm. The less the CT is, the better the MOEA is. However, with the development of computer technology, the speed of computer becomes faster and faster.

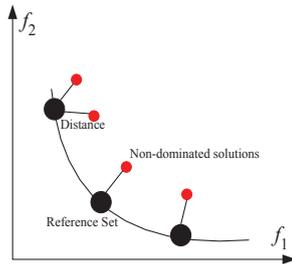


Figure 2 | Generational distance.

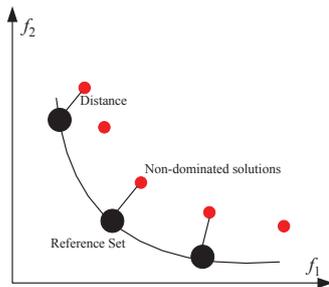


Figure 3 | Inverted generational distance measurement.

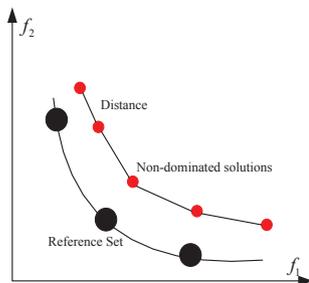


Figure 4 | Spacing measurement.

4. RELATED METHODOLOGIES

In this section, some related methodologies are briefly introduced. They are used to formulate the proposed approach in the next section.

4.1. Box Plot

It is well known that we can't get any conclusion according to a single run. Generally speaking, statistical measurement is often employed, such as box plot. The approach is a kind of box-and-whisker diagram. It is very convenient. It is a useful method to depict groups of data by five parameters, which include the smallest observation, largest observation, lower quartile (Q1), median (Q2), and upper quartile (Q3) [24]. They are presented in Figure 5.

The approach is useful to demonstrate differences between populations. The approach can be drawn either vertically or horizontally.

The approach is widely used to evaluate MOEAs [3,10]. However, it is a graphical method. It is still difficult to compare the difference of box plots directly because it uses five numbers from the results of MOEAs. In order to solve the issue, the fuzzy theory is applied.

4.2. Fuzzy TOPSIS

Fuzzy method is widely used in uncertain environments [25,26]. Fuzzy TOPSIS is based on fuzzy theory. Triangular fuzzy number is widely used. A triangular fuzzy number \tilde{A} can be presented by a triplet (b_1, b_2, b_3) [27].

$$\mu_{\tilde{b}}(x) = \begin{cases} 0, & x \leq b_1 \\ \frac{x - b_1}{b_2 - b_1}, & b_1 < x \leq b_2 \\ \frac{b_3 - x}{b_3 - b_2}, & b_2 < x \leq b_3 \\ 0, & x > b_3 \end{cases} \quad (3)$$

The distance between triangular fuzzy number $\tilde{b} = (b_1, b_2, b_3)$ and $\tilde{c} = (c_1, c_2, c_3)$ can be computed by the following Equation (4) [27].

$$d(\tilde{b}, \tilde{c}) = \sqrt{\frac{1}{3} [(b_1 - c_1)^2 + (b_2 - c_2)^2 + (b_3 - c_3)^2]} \quad (4)$$

Besides, the multiplication operation between the real number k and triangular fuzzy number \tilde{b} is as following [27]:

$$k \times \tilde{b} = k \times (b_1, b_2, b_3) = (k \times b_1, k \times b_2, k \times b_3) \quad (5)$$

In the real application, fuzzy linguistic is often adopted to deal with the ambiguities of human assessments. Therefore, it is necessary to

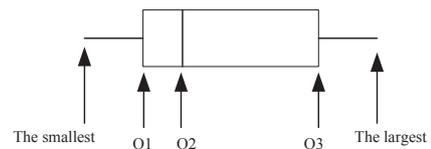


Figure 5 | Box plots diagram.

establish the relationship between fuzzy linguistic and triangular fuzzy number. The fuzzy linguistic and their corresponding fuzzy number are listed in Table 1 and Figure 6 [27].

TOPSIS is one of widely used MCDM techniques as it is useful, convenient, and practical to estimate alternatives and rate them [28–30]. TOPSIS is according to the fact that the best alternative should be the farthest from the negative-ideal point and the shortest from the positive-ideal point. The distance is often Euclidean distances. However, it is difficult to give a precise performance rating for a decision maker. Fuzzy TOPSIS is used to solve the problem by introducing fuzzy number. Fuzzy TOPSIS has the following steps [29]:

Step 1: The suitable linguistic variables should be selected for the alternatives. They can be used to evaluate alternatives, so that the fuzzy performance matrix x_{ij} can be established.

Step 2: Establish the decision matrix, which is the weighted normalized matrix. The value v_{ij} of the matrix can be computed by Equation (6).

$$v_{ij} = x_{ij} \times w_j, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n, \quad (6)$$

where w_j is the weight of the j th criterion, and $\sum_{j=1}^n w_j = 1$.

Step 3: Finding out the negative-ideal point (A^-) and positive-ideal point (A^*).

The fuzzy negative-ideal point (FNIS, A^-) and fuzzy positive-ideal point (FPIS, A^*) are presented as follows:

$$A^- = \{v_1^-, v_2^-, \dots, v_n^-\} = \{(\min_j v_{ij} | i \in I'), (\max_j v_{ij} | i \in I'')\} \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (7)$$

$$A^* = \{v_1^*, v_2^*, \dots, v_n^*\} = \{(\max_j v_{ij} | i \in I'), (\min_j v_{ij} | i \in I'')\} \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (8)$$

where I' is associated with the benefit criteria, and I'' is associated with the cost criteria.

Table 1 | The correspondence of linguistic terms and linguistic values.

Linguistic Value	Triangular Fuzzy Number
VP	(0, 0, 0.25)
P	(0, 0.25, 0.5)
FP	(0.25, 0.5, 0.75)
FG	(0.5, 0.75, 1)
G	(0.75, 1, 1)

FG, fairly good; FP, fairly poor; G, good; P, poor; VP, very poor.

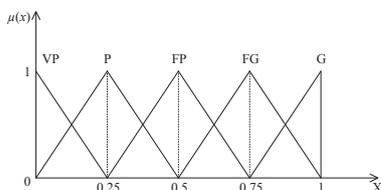


Figure 6 | The correspondence of linguistic terms and linguistic value.

Step 4: Compute the distance of each alternative from FNIS and FPIS, respectively.

$$D_i^- = \sum_{j=1}^n d(v_{ij}, v_j^-) \quad i = 1, 2 \dots m \quad (9)$$

$$D_i^* = \sum_{j=1}^n d(v_{ij}, v_j^*) \quad i = 1, 2 \dots m \quad (10)$$

The distance can be computed by Equation (4).

Step 5: Calculate the closeness coefficients by Equation (11).

$$CC_i = \frac{D_i^-}{D_i^* + D_i^-} \quad (11)$$

Step 6: Rank order.

Rank alternatives based on CC_i results in descending order.

Now, how to solve relative weights w among criteria is the next issue that should be solved. Analytic Hierarchy Process (AHP) is a good approach and it is widely used in the decision problems.

4.3. AHP

AHP is a useful approach to solve MCDM in the different kinds of real applications. In the theory, multiple pair wise comparisons have to be established from 1 ~ 9.

According to the above analysis, there are six performance indicators: HV, GD, IGD, ϵ_+ -Indicator, spacing, and computation time in Figure 7.

Let $D = \{D_j | j = 1, 2, \dots, n\}$, $n = 6$ be the set of criteria. The matrix E can be obtained. Each element e_{ij} ($i, j = 1, 2, \dots, n$), $n = 6$ in the matrix E represents the relative weights of the criteria:

$$E = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1n} \\ e_{21} & e_{22} & \dots & \dots \\ \dots & \dots & e_{ii} & \dots \\ e_{n1} & e_{n2} & \dots & e_{nn} \end{bmatrix} \quad (12)$$

When the following conditions are met, the matrix E can achieve complete consistent:

$$e_{ii} = \frac{w_i}{w_i} = 1$$

$$e_{ji} = \frac{w_j}{w_i} = \frac{1}{a_{ij}} \quad (13)$$

$$e_{ij}e_{jk} = \frac{w_i}{w_j} \times \frac{w_j}{w_k} = \frac{w_i}{w_k} = e_{ik}$$

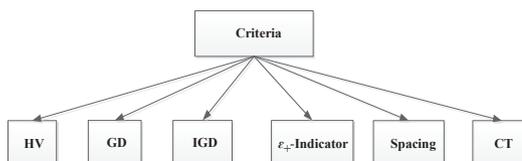


Figure 7 | Six criteria.

According to the Equation (13), the following equations can be obtained:

$$\sum_{k=1}^n (e_{ik}w_k) = \sum_{k=1}^n \left(\frac{w_i}{w_k} \right) w_k = nw_i, i = 1, 2 \quad (14)$$

$$\sum_{i=1}^n \left| \sum_{k=1}^n (e_{ik}w_k) - nw_i \right| = 0 \quad (15)$$

Thus, when the judgment matrix E meets the Equation (15), it is a complete consistent matrix. In fact, it is very hard to meet the demand. The matrix E has to satisfy the satisfactory consistency in the real applications. The Equation (15) can be transformed into Equation (16):

$$\min f(w) = \frac{\sum_{i=1}^n \left| \sum_{k=1}^n (e_{ik}w_k) - nw_i \right|}{n} \quad (16)$$

$$0 < w_k < 1, \sum_{k=1}^n w_k = 1 \quad (17)$$

Now, solving weights among performance indicators is converted to COP: the objective function is to minimize f and $0 < w_k < 1, \sum_{k=1}^n w_k = 1$ is the constraint condition.

The constraints $\sum_{k=1}^n w_k = 1$ are transformed to inequality constraints:

$$\left| \sum_{k=1}^n w_k - 1 \right| - \delta \leq 0 \quad (18)$$

Where δ is a very small positive value. The Equation (18) can be transformed into Equation (19).

$$-\delta \leq \sum_{k=1}^n w_k - 1 \leq \delta \rightarrow \begin{cases} \sum_{k=1}^n w_k - 1 \leq \delta \\ -\left(\sum_{k=1}^n w_k - 1\right) \leq \delta \end{cases} \quad (19)$$

Where k is from 1 to the number of criteria. The constraint violation (CV) can be presented as follows:

$$CV(x) = \max \left(\left(\sum_{k=1}^n w_k \right) - 1 - \delta, 0 \right) + \max \left(- \left(\left(\sum_{k=1}^n w_k \right) - 1 \right) - \delta, 0 \right) \quad (20)$$

From above discussion, it can be observed that solving weight is a typical COP. How to solve the typical COP is very important. An adaptive DE (ADE) algorithm is presented to solve the COP.

4.4. Solving COP by the Proposed ADE Algorithm

DE is one of EAs [31]. Mutation strategy is very important during evolution. The main strategies of DE are given [32]:

DE/best/1:

$$V_i = X_{best} + F \cdot (X_{r_1} - X_{r_2}) \quad (21)$$

DE/rand/1:

$$V_i = X_{r_1} + F \cdot (X_{r_2} - X_{r_3}) \quad (22)$$

Three parameters r_1^i, r_2^i , and r_3^i must be different from each other. The range of them is $[0, 1]$. At the same time, they are also of difference from the index i .

There are many mutation strategies. They have different performances. Among these strategies, DE/rand/1 was the first proposed [33]. It has been reported that the strategy is widely used and has achieved successful [34]. The strategy can keep diversity of population by random search. However, it may result in poor convergence [22]. The mutation strategy DE/best can accelerate convergence of algorithm with the help of the best information found so far during evolution process [18,19]. But, it may also have some problems. For example, diversity of population may become poor as the algorithm searches around the best information [20]. Algorithm may get into local optimum. It is very critical to deal with convergence and diversity during evolution. In order to realize the aim, DE/rand and DE/best are combined together to generate offspring shown in Table 2.

Where $rand$ is uniformly generated between 0 and 1. Parameter ϕ is the feasible solution proportion. For COPs, there are little feasible solutions at the early stage. Therefore, ϕ is small at this stage. The condition $rand < \phi$ cannot be met frequently. Thus, the mutation strategy DE/rand/1 is used more frequently. At the later stage, feasible solutions are increased and kept. So, parameter ϕ will get bigger. The condition $rand < \phi$ will be met more frequently. DE/best will be used more frequently during evolution.

The crossover operation of the proposed algorithm is similar to the conventional DE.

The crossover operation is employed to $V_i = \{v_i^1, v_i^2, \dots, v_i^D\}$ and $X_i = \{x_i^1, x_i^2, \dots, x_i^D\}$. A trial vector $U_i = \{u_i^1, u_i^2, \dots, u_i^D\}$ is generated as follows:

$$u_i^j = \begin{cases} v_i^j & \text{if } rand_j [0, 1] \leq CR \text{ or } (j = j_{rand}) \\ x_i^j & \text{others} \end{cases} \quad (23)$$

The crossover constant is $CR \in [0, 1]$. The index $j_{rand} \in [1, D]$ makes sure that the trial vector U_i will be different from X_i .

The comparison among two vectors X_i, U_i is to select one vector for next evolution. If a solution meets constraints, it is feasible. Otherwise, it is infeasible. When comparing two feasible solutions with each other, the one with lower fitness value should be maintained. When comparing infeasible solutions with infeasible solutions, feasible solutions with infeasible solutions, it should be observed that infeasible solutions are of importance to keep diversity of population. These solutions should be made good use of. The

Table 2 | The generation method of offspring.

if $rand < \phi$ then
$V_i = X_{best} + F \cdot (X_{r_1} - X_{r_2})$ //DE/best/1
else
$V_i = X_{r_1} + F \cdot (X_{r_2} - X_{r_3})$ //DE/rand/1
end

ϵ -constrained method is a useful method to cope with the problem, which uses a relaxation of the limit ϵ to regard an infeasible solution as a feasible solution [35]. Let $x_1, x_2, f(x_1), f(x_2)$,

$CV(x_1), CV(x_2)$ be the two solutions, fitness values and the CV, respectively. Then, ϵ -comparison between two solutions is presented as follows:

$$(f(x_1), CV(x_1)) <_{\epsilon} (f(x_2), CV(x_2)) \Leftrightarrow \begin{cases} f(x_1) < f(x_2) & \text{if } CV(x_1), CV(x_2) \leq \epsilon \\ f(x_1) < f(x_2) & \text{if } CV(x_1) = CV(x_2) \\ CV(x_1) < CV(x_2) & \text{otherwise} \end{cases} \quad (24)$$

Now, the main steps of the ADE is revealed as follows :

Step 1 Parameters settings. Number of Population (NP); tolerance level ϵ ; F ; crossover rate (CR).

Step 2 Generate NP solutions $P_G = \{X_{1,G}, \dots, X_{NP,G}\}$ with $X_{i,G} = \{x_{i,G}^1, \dots, x_{i,G}^D\}$ randomly, $i = 1, \dots, NP$ in the range $[X_{min}, X_{max}]$.

Step 3 Compute fitness value and CV of each solution, and feasible proportion ϕ of the current population.

Step 4 While *stopping condition is not met*

Step 4.1 if ($\epsilon \leq 0.000001$) $\epsilon = 0$; end

Step 4.2 Generate vector V_i according to procedure in Table 2.

Step 4.3 Generate vector U_i according to DE crossover and V_i according to Equation (23)

Step 4.4 If vectors U_i are not in the boundary, randomly generate them in the search space

Step 4.5 Compute the fitness value, total CVs of U_i , feasible proportion ϕ of U_i .

Step 4.6 Selection

Compare Vector X_i, U_i by Equation (24) and retain the better one

Step 4.7 Update ϵ linearly

Step 5 End while

When the proposed algorithm is employed to acquire the weights of performance indicators, the fitness value is the result of Equation (16) and the CV is the result of Equation (20).

5. PROPOSED EVALUATION APPROACH

The approach for evaluating MOEAs, composed of ADE, box plot, and fuzzy TOPSIS approaches, has following four steps. It is depicted in Figure 8.

1. Select alternative MOEAs and identify performance indicators
In this step, MOEAs and performance indicators are determined. The decision hierarchy is established. Indicators are in the first level and MOEAs are located in the second level. The performance indicators are HV, GD, IGD, ϵ_+ Indicator, spacing, and CT. The MOEAs include NSGAII [3], GDE3 [8], MOEAD [4], SMSEMOA [36], eNSGAII [37], and eMOEA [5].
2. Determine weights
In this step, the weights of six performance indicators are determined. As weights among the six performance indicators are not equal, how to solve weights is very critical. AHP and proposed ADE algorithm are combined to solve weights.
3. Select benchmark functions and draw box plots
In this step, benchmark functions are needed. As we can get any useful conclusions based on a single run, 30 runs are generally required. The results of performance indicators are gotten. Box plots can be drawn based on the results of 30 runs.
4. Evaluate MOEAs and determinate the final rank
Ranking MOEAs is determined by fuzzy TOPSIS in the fourth step. Linguistic values are employed to evaluate box plots. Ranking MOEAs is achieved based on the value of CC_i .

6. EXPERIMENTS

6.1. Solving Weights of Performance Indicators

In the above evaluation index system, HV, IGD and ϵ_+ Indicator measure the diversity and convergence, GD is utilized to measure

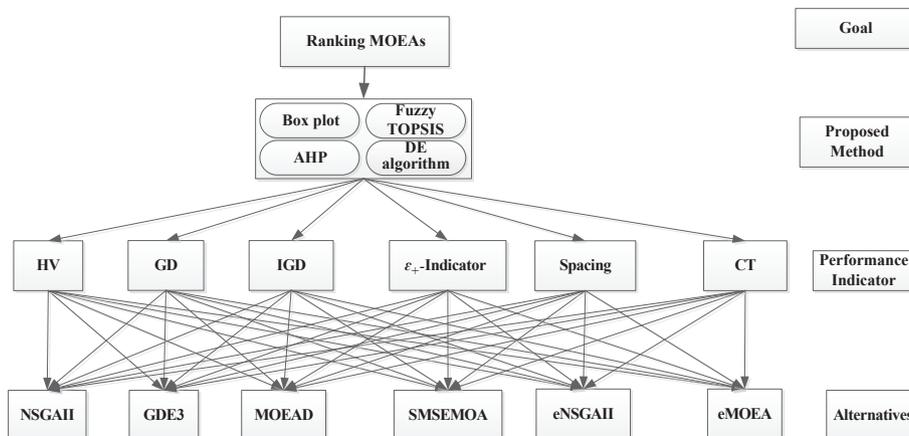


Figure 8 | The proposed method.

convergence of MOEAs and spacing are utilized to measure diversity of population. For a good MOEA, diversity and convergence are equally important. Thus, the importance of the above indicators is same. However, with the development of science and technology, computer speed becomes faster and faster. Thus, the CT is less important. Thus, we can give the following evaluation matrix according to AHP theory.

$$E = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 5 \\ 1 & 1 & 1 & 1 & 1 & 5 \\ 1 & 1 & 1 & 1 & 1 & 5 \\ 1 & 1 & 1 & 1 & 1 & 5 \\ 1 & 1 & 1 & 1 & 1 & 5 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 1 \end{bmatrix}$$

The $\min f(w)$ is as follows:

$$\min f(w) = \frac{\sum_{i=1}^6 |\sum_{k=1}^6 (E(k, i) \times w_k) - 6w_i|}{n} \quad (25)$$

The proposed algorithm ADE can be used to optimize the $f(w)$. To validate the results, weights from ADE and AHP are listed in Table 3. It can be noticed that the results are very consistent with each other. The observations have demonstrated that ADE has the capability to solve weights.

In order to validate the consistency of the judgment matrix, the consistency is tested. By following the steps of AHP, CR is 0.0451, which

Table 3 The weights comparison.

Method	w_1	w_2	w_3	w_4	w_5	w_6
ADE	0.1874	0.1875	0.1875	0.1874	0.1875	0.0627
AHP theory	0.1875	0.1875	0.1875	0.1875	0.1875	0.0625

ADE, adaptive differential evolution.

is less than 0.1 [38]. Therefore, it can be concluded that evaluation matrix is consistent.

6.2. Evaluate MOEAs

The experiments are implemented to evaluate six MOEAs including NSGAII [3], GDE3 [8], MOEAD [4], SMSEMOA [36], eNSGAII [37], and eMOEA [5]. In order to make faithful comparisons, 14 test functions from CEC 2009 are used in MOPs and employed in the experiments. They can be divided into two groups: constrained (CF1 CF7) and unconstrained (UF1 UF7). The detail information of these test functions can be obtained from the reference [39].

The parameter settings of these algorithms are same to the original paper. It is recommended that each algorithm should be run independently 30 times for each test problem in CEC 2009 [39]. The values of performance indicators are obtained. The box plots of these values are presented.

In order to elaborate the whole calculation process, the CF1 results of six performance indicators are acquired and box plots are presented in Figure 9. Linguistic values are adopted to rank six MOEAs.

Algorithms with larger HV, smaller GD, IGD, ϵ_+ -Indicator, spacing, and CT values are desirable. From the Figure 9, it is observed that NSGAII has the largest HV, whose Q1, Q2, Q3 and the largest observation are the biggest among the six MOEAs. In other words, NSGAII has achieved the best values in HVE indicator. Thus, the HV performance of NSGAII is very good (VG). SMSEMOA, eNSGAII, and eMOEA have the similar achievement. Three of them are good (G). MOEAD is less than the three algorithms. It can be considered as fairly good (FG). GDE3 is the worst one. The performance of GDE3 is very poor (VP). According to the above discussion, the linguistic evaluations of HV are listed in Table 4.

The linguistic evaluations are converted to triangular fuzzy number according to Table 1 and Figure 6. Every element belong to [0, 1] in Table 5. Therefore, there is no need to normalize them. Fuzzy

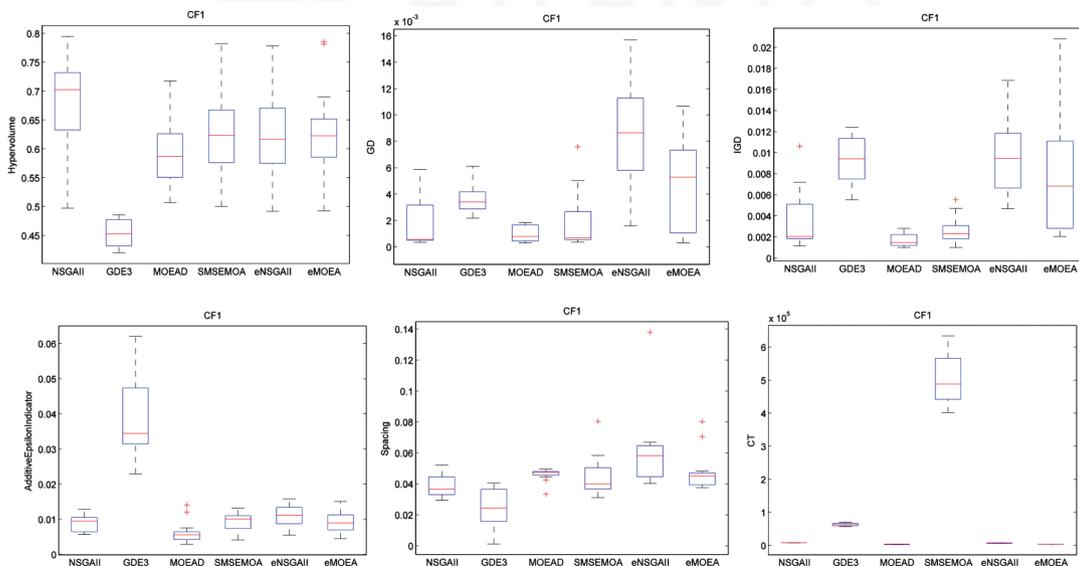


Figure 9 The box plots of six performance indicators, respectively.

positive-ideal point (FPIS, A^*) and negative-ideal point (FNIS, A^-) can be defined as following:

$$(FPIS, A^*) = \{(0.8, 1, 1) \times w_1, (0.6, 0.8, 1) \times w_2, (0.6, 0.8, 1) \times w_3, (0.6, 0.8, 1) \times w_4, (0.6, 0.8, 1) \times w_5, (0.8, 1, 1) \times w_6\};$$

$$(FNIS, A^-) = \{(0, 0, 0.2) \times w_1, (0, 0, 0.2) \times w_2, (0, 0, 0.2) \times w_3, (0, 0, 0.2) \times w_4, (0, 0.2, 0.4) \times w_5, (0, 0, 0.2) \times w_6\};$$

The distance can be computed by Equations (9) and (10). For example, D_1^+ , D_1^- , and CC_1 are taken as an example and results are listed in Table 6.

Similar procedures can be fulfilled for the other algorithms. The global performance of each algorithm is determined by CC. So, the ranking of six algorithms is: NSGAI > MOEAD > SMSEMOA > eMOEA > eNSGAI > GDE3. NSGAI is the best algorithm and GDE3 is the worst one for CF1.

Table 7 gives the complete rankings for all benchmark functions. For the 14 test functions, NSGAI wins in three problems: CF1, CF3, and UF6. NSGAI performs better for CF4, CF5, and UF3. GDE3 wins in five problems: CF4, CF5, CF7, UF1, and UF7. GDE3 provides better performance for CF2, CF6, UF2, and UF4. However, it achieves worse results for CF1 and UF5. MOEAD obtains better performance for CF1 and CF7. SMSEMOA gets the best performance for UF3 and better results for UF5. eNSGAI achieves the superior results for CF2, UF2, UF4, and UF5. It obtains better performance in four problems: CF3, UF1, UF6, and UF7. eMOEA ranks the first for CF6.

6.3. The Effectiveness of the Proposed Method

In order to make more comparisons among six MOEAs, the Pareto front and nondominated set of six MOEAs are presented in Figure 10. NSGAI can achieve many nondominated solutions. These solutions are evenly distributed along the Pareto front, while eNSGAI and GDE3 obtain nondominated solutions, in which both convergence and diversity are poor.

In order to prove the advantage of the evaluation method, the set coverage is used. The set coverage can be utilized to compare the difference between two algorithms [4]. There are two nondominated solutions A and B . $C(A, B)$ is denoted as following:

$$C(A, B) = \frac{|\{u \in B | \exists v \in A : v \text{ dominates } u\}|}{|B|} \quad (26)$$

Where $|B|$ is the number of solutions in the B . Six nondominated solutions of CF1 from above six algorithms are gotten. Pair wise comparisons are made. The results are presented in Table 8. 96.3% of the solutions from GDE3 are dominated by those from NSGAI, and only 21.4% vice versa. Thus, the performance of the NSGAI is better than that of GDE3. According to Table 8, it can be found that NSGAI is the best. GDE3 is the worst one. MOEAD is better than GDE3, SMSEMOA, eNSGAI, and eMOEA. SMSEMOA is better than eNSGAI and eMOEA, eMOEA is a little better than eNSGAI. Therefore, the ranking is NSGAI > MOEAD > SMSEMOA

Table 4 | The five performance indicators according to the box plots.

Algorithm	Hyper Volume	GD	IGD	ϵ_+ Indicator	Spacing
NSGAI	VG	G	FG	FG	G
GDE3	VP	FG	FP	VP	FG
MOEAD	FG	G	G	G	P
SMSEMOA	G	P	P	FP	P
eNSGAI	G	P	VP	P	P
eMOEA	G	VP	P	FP	FP

FG, fairly good; FP, fairly poor; G, good; GD, generational distance; IGD, inverted generation distance; P, poor; VP, very poor.

Table 5 | The triangular fuzzy number of the six indicators for the six MOEAs.

Algorithm	Hyper Volume	GD	IGD	ϵ_+ Indicator	Spacing
NSGAI	(0.8, 1, 1)	(0.6, 0.8, 1)	(0.4, 0.6, 0.8)	(0.4, 0.6, 0.8)	(0.6, 0.8, 1)
GD3	(0, 0, 0.2)	(0.4, 0.6, 0.8)	(0.2, 0.4, 0.6)	(0, 0, 0.2)	(0.4, 0.6, 0.8)
MOEAD	(0.4, 0.6, 0.8)	(0.6, 0.8, 1)	(0.6, 0.8, 1)	(0.6, 0.8, 1)	(0, 0.2, 0.4)
SMSEMOA	(0.6, 0.8, 1)	(0, 0.2, 0.4)	(0, 0.2, 0.4)	(0.2, 0.4, 0.6)	(0, 0.2, 0.4)
eNSGAI	(0.6, 0.8, 1)	(0, 0.2, 0.4)	(0, 0, 0.2)	(0, 0.2, 0.4)	(0, 0.2, 0.4)
eMOEA	(0.6, 0.8, 1)	(0, 0, 0.2)	(0, 0.2, 0.4)	(0.2, 0.4, 0.6)	(0.2, 0.4, 0.6)

GD, generational distance; IGD, inverted generation distance; MOEAs, multiobjective evolutionary algorithms.

Table 6 | The D^* , D^- , and CC of the six MOEAs.

Algorithm	D^*	D^-	CC
NSGAI	0.159	0.475	0.749
GDE3	0.447	0.234	0.344
MOEAD	0.188	0.437	0.699
SMSEMOA	0.386	0.278	0.419
eNSGAI	0.442	0.270	0.379
eMOEA	0.432	0.270	0.384

MOEAs, multiobjective evolutionary algorithms; CC, closeness coefficients.

Table 7 | The final ranking for the six MOEAs.

	NSGAII	GDE3	MOEAD	SMSEMOA	eNSGAII	eMOEA
CF1	1	6	2	4	5	3
CF2	4	2	3	6	1	5
CF3	1	4	3	6	2	5
CF4	2	1	3	4	6	5
CF5	2	1	3	4	6	5
CF6	3	2	6	4	5	1
CF7	5	1	2	3	6	4
UF1	4	1	5	6	2	3
UF2	5	2	6	4	1	3
UF3	2	4	5	1	3	6
UF4	5	2	6	4	1	3
UF5	3	6	4	2	1	5
UF6	1	3	6	5	2	4
UF7	3	1	6	4	2	5

MOEAs, multiobjective evolutionary algorithms.

The bold values are ranked the first, which indicates that the corresponding algorithm has the best performance.

Table 8 | The set coverage of pair wise comparisons among the six algorithms.

	NSGAII	GDE3	MOEAD	SMSEMOA	eNSGAII	eMOEA
NSGAII		96.3%	56.5%	79.4%	60%	73.5%
GDE3	21.4%		5.6%	4.8%	22.2%	36.7%
MOEAD	55.7%	92.6%		81%	68.9%	71.4%
SMSEMOA	60%	96.3%	27.8%		68.9%	69.8%
eNSGAII	52.9%	96.3%	38.9%	50.8%		55.1%
eMOEA	52.9%	85.2%	36.1%	67.3%	60%	

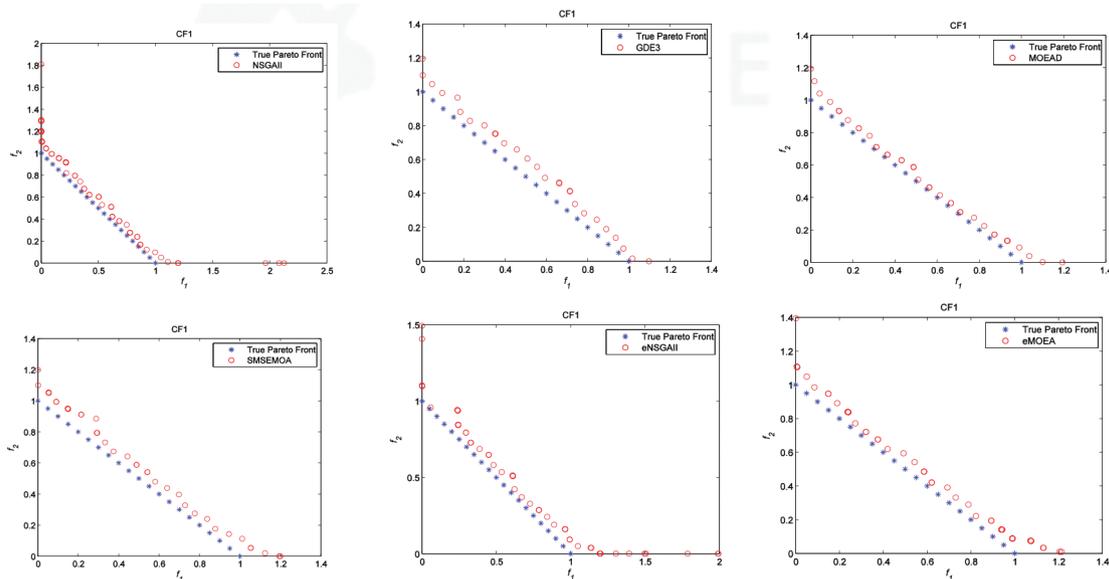


Figure 10 | The Pareto front and the nondominated solutions of six multiobjective evolutionary algorithms (MOEAs).

> eMOEA > eNSGAII > GDE3. The result is same to the proposed method.

However, the set coverage has to make pair wise comparisons. If the number of algorithms is n , $n \times (n - 1)$ comparisons have to be made. The comparison task will remarkably increase when n is large. The comparison is qualitative. The proposed evaluation method doesn't have to deal with so many tasks.

7. CONCLUSIONS

Many MOEAs are available and the experiment results often get the conclusion that these MOEAs are the best based on few performance indicators. These comparisons are unfair. In order to solve the problem, an approach is proposed to evaluate MOEAs. The

approach uses box plots and fuzzy TOPSIS. Six performance indicators HV, GD, IGD, ϵ_+ - Indicator, spacing, and CT are selected. Six MOEAs including NSGAII, GDE3, MOEAD, SMSEMOA, eNSGAII, and eMOEA are evaluated. The weights of six indicators are solved by an ADE algorithm.

The experiments have been conducted based on CEC 2009. The results from CEC 2009 have implied that the proposed evaluation method is fair and accurate. The observation of experimental results demonstrates that the capability of MOEAs to solve MOPs counts on both MOEAs and the features of MOPs.

In the following research, we will adopt Pythagorean fuzzy sets (PFSs) to evaluate MOEAs, which was proposed by Yager [40]. The approach can deal with uncertain situation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

Xiaobing Yu conceived the research, edited, and revised the manuscript. Chenliang Li and Hong Chen drafted the manuscript and designed the experiments; Xianrui Yu performed the experiments and analyzed the data.

Funding Statement

This study was funded by Major Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (2019SJZDA039), China Natural Science Foundation (71503134), Qinglan project in Jiangsu province, Key Project of National Social and Scientific Fund Program (16ZDA047).

ACKNOWLEDGMENT

This study was funded by China Natural Science Foundation (grant number. 71503134, 71871121), Qinglan project in Jiangsu province, Key Project of National Social and Scientific Fund Program (16ZDA047), Top-notch Academic Programs Project of Jiangsu Higher Education Institutions (TAPP).

REFERENCES

- [1] G.G. Yen, Z. He, Performance metric ensemble for multiobjective evolutionary algorithms, *IEEE Trans. Evol. Comput.* 18 (2014), 131–144.
- [2] A. Zhou, B.-Y. Qu, H. Li, S.-Z. Zhao, P.N. Suganthan, Q. Zhang, Multiobjective evolutionary algorithms: a survey of the state of the art, *Swarm Evol. Comput.* 1 (2011), 32–49.
- [3] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast elitist multi-objective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.* 6 (2000), 182–197.
- [4] H. Li, Q. Zhang, Multiobjective optimization problems with complicated pareto sets, MOEA/D and NSGA-II, *IEEE Trans. Evol. Comput.* 13 (2009), 284–302.
- [5] K. Deb, A fast multi-objective evolutionary algorithm for finding well-spread Pareto-optimal solutions, *KanGAL Rep. 2003002* (2003), 849–858.
- [6] S.Hr. Aghay Kaboli, J. Selvaraj, N.A. Rahim, Rain-fall optimization algorithm: a population based algorithm for solving constrained optimization problems, *J. Comput. Sci.* 19 (2017), 31–42.
- [7] A.J. Nebro, J.J. Durillo, J. Garcia-Nieto, C.A. Cello Coello, F. Luna, E. Alba, SMPSO: a new PSO-based metaheuristic for multi-objective optimization, in *2009 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making, Nashville, 2009*, pp. 66–73.
- [8] S. Kukkonen, J. Lampinen, GDE3: the third evolution step of generalized differential evolution, *KanGAL Rep. 2005013* (2005), 443–450.
- [9] A.J. Nebro, F. Luna, E. Alba, B. Dorronsoro, J.J. Durillo, A. Beham, AbYSS: adapting scatter search to multiobjective optimization, *IEEE Trans. Evol. Comput.* 12 (2008), 439–457.
- [10] A. Panda, S. Pani, A symbiotic organisms search algorithm with adaptive penalty function to solve multi-objective constrained optimization problems, *Appl. Soft Comput.* 46 (2016), 344–360.
- [11] M. Ali, P. Siarry, M. Pant, An efficient differential evolution based algorithm for solving multi-objective optimization problems, *Eur. J. Oper. Res.* 217 (2012), 404–416.
- [12] L.C.T. Bezerra, M. López-Ibáñez, T. Stützle, A large-scale experimental evaluation of high-performing multi-and many-objective evolutionary algorithms, *Evol. Comput.* 26 (2018), 621–656.
- [13] Z. He, G.G. Yen, Comparison of many-objective evolutionary algorithms using performance metrics ensemble, *Adv. Eng. Softw.* 76 (2014), 1–8.
- [14] M.G. Epitropakis, S. Yoo, M. Harman, E.K. Burke, Empirical evaluation of Pareto efficient multi-objective regression test case prioritization, in *Proceedings of the 2015 International Symposium on Software Testing and Analysis, ACM, Baltimore, 2015*, pp. 234–245.
- [15] T.P. França, T.F. de Queiroz Lafetá, L.G.A. Martins, A comparative analysis of MOEAs considering two discrete optimization problems, in *2017 Brazilian Conference on Intelligent Systems (BRACIS), IEEE, Uberlandia, 2017*, pp. 402–407.
- [16] J. Yazdi, S. Mohammadiun, R. Sadiq, S.A.A.S. Neyshabouri, A.A. Gharahbagh, Assessment of different MOEAs for rehabilitation evaluation of urban stormwater drainage systems—case study: eastern catchment of Tehran, *J. Hydro Environ. Res.* 21 (2018), 76–85.
- [17] F. Zheng, A.C. Zecchin, H.R. Maier, A.R. Simpson, Comparison of the searching behavior of NSGA-II, SAMODE, and Borg MOEAs applied to water distribution system design problems, *J. Water Res. Plann. Manage.* 142 (2016), 04016017.
- [18] P.A.N. Bosman, D. Thierens, The balance between proximity and diversity in multiobjective evolutionary algorithms, *IEEE Trans. Evol. Comput.* 7 (2003), 174–188.
- [19] K. Deb, S. Jain, Running performance metrics for evolutionary multi-objective optimization, *KanGAL Rep. 2002004* (2002), 1–18.
- [20] C.A. Coello Coello, G.B. Lamont, D.A. Van Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems*, Springer Science+Business Media LLC, New York, 2007.
- [21] J. Knowles, D. Corne, On metrics for comparing non-dominated sets, in *Congress on Evolutionary Computation, Honolulu, HI, 2002*, pp. 711–716.
- [22] D. Hadka, P. Reed, Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization, *Evol. Comput.* 20 (2012), 423–452.
- [23] D. Hadka, P. Reed, Borg: an auto-adaptive many-objective evolutionary computing framework, *Evol. Comput.* 21 (2013), 231–259.
- [24] J.W. Tukey, *Exploratory Data Analysis*. Addison-Wesley, Reading, 1977.
- [25] X. Peng, Y. Yang, Algorithms for interval-valued fuzzy soft sets in stochastic multi-criteria decision making based on regret theory and prospect theory with combined weight, *Appl. Soft Comput.* 54 (2017), 415–430.
- [26] X. Peng, H. Garg, Algorithms for interval-valued fuzzy soft sets in emergency decision making based on WDBA and CODAS with new information measure, *Comput. Ind. Eng.* 119 (2018), 439–452.
- [27] C.-J. Lin, W.-W. Wu, A causal analytical method for group decision-making under fuzzy environment, *Expert Syst. Appl.* 34 (2008), 205–213.

- [28] Y.-H. He, L.-B. Wang, Z.-Z. He, M. Xie, A fuzzy TOPSIS and rough set based approach for mechanism analysis of product infant failure, *Eng. Appl. Artif. Intell.* 47 (2016), 25–37.
- [29] R. Lourenzutti, R.A. Krohling, A generalized TOPSIS method for group decision making with heterogeneous information in a dynamic environment, *Info. Sci.* 330 (2016), 1–18.
- [30] X. Peng, J. Dai, Approaches to single-valued neutrosophic MADM based on MABAC, TOPSIS and new similarity measure with score function, *Neural Comput. Appl.* 29 (2018), 939–954.
- [31] K. Price, R.M. Storn, J.A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, Natural Computing Series, Springer-Verlag, Berlin, 2006.
- [32] H. Sharma, J.C. Bansal, K.V. Arya, Fitness based differential evolution, *Memetic Comp.* 4 (2012), 303–316.
- [33] R. Gamperle, S.D. Muller, P. Koumoutsakos, A parameter study for differential evolution, in *Proceeding of Advances Intelligent*, Vancouver, BC, Canada, 2002, pp. 293–298.
- [34] U. Pahner, K. Hameyer, Adaptive coupling of differential evolution and multiquadrics approximation for the tuning of the optimization process, *IEEE Trans. Magn.* 36 (2000), 1047–1051.
- [35] T. Takahama, S. Sakai, N. Iwane, Constrained optimization by the epsilon constrained hybrid algorithm of particle swarm optimization and genetic algorithm, in *Advances Artificial Intelligence*, Muroran, Japan, 2005, pp. 389–400.
- [36] N. Beume, B. Naujoks, M. Emmerich, SMS-EMOA: Multiobjective selection based on dominated hypervolume, *Eur. J. Oper. Res.* 181 (2007), 1653–1669.
- [37] J.B. Kollat, P.M. Reed, The value of online adaptive search: a performance comparison of NSGA-II, epsilon-NSGAII, and epsilon-MOEA, in: C.C. Coello, A.H. Aguirre, E. Zitzler (Eds.), *The Third International Conference on Evolutionary Multi-Criterion Optimization (EMO 2005)*, Lecture Notes in Computer Science, vol. 3410, Springer Verlag, Guanajuato, 2005, pp. 386–398.
- [38] X. Yu, S. Guo, J. Guo, X. Huang, Rank B2C e-commerce websites in e-alliance based on AHP and fuzzy TOPSIS, *Expert Syst. Appl.* 38 (2011), 3550–3557.
- [39] Q.F. Zhang, A.M. Zhou, S.Z. Zhao, N.S. Ponnuthurai, W.D. Liu, T. Santosh, Multiobjective optimization Test Instances, in *Special Session and Competition*, University of Essex, Colchester, England; Nanyang Technological University, Nanyang, Singapore, 2009.
- [40] R.R. Yager, Pythagorean fuzzy subsets, in *Proceeding of Joint IFSA World Congress and NAFIPS Annual Meeting*, Edmonton, Canada, 2013, pp. 57–61.