Letter

An Information-Based Elite-Guided Evolutionary Algorithm for Multi-Objective Feature Selection

Ziqian Wang ⓘ, Shangce Gao ⓘ, *Senior Member, IEEE*,
Zhenyu Lei ⓘ, and Masaaki Omura ⓘ

Dear Editor,

This letter is concerned with the evolution strategy for addressing multi-objective feature selection problems in classification. Previous methods suffer from limitations such as being trapped in local optima and lacking stability. To overcome them, we propose a novel elite-guided mechanism based on information theory. Firstly, an elite solution is generated through a dimension reduction strategy and incorporated to the initialization population. Then, a symmetrical uncertainty-based mutation operator is developed to implement local search after the crossover operator. Finally, a special crowding distance is utilized to analyze duplicates in the environmental selection. The effectiveness and superiority of the proposed method are verified on 20 datasets, including high-dimensional ones.

Introduction: With the continuous advancements in data collection technologies, the dimensionality of datasets has significantly increased. High-dimensional data often contains irrelevant and redundant features, leading to performance degradation in classification tasks due to the "curse of dimensionality". Consequently, effective data mining techniques are essential to address this challenge.

Feature selection (FS) is a widely adopted data mining technique that tackles this issue by selecting a subset of relevant features. FS methods can be broadly categorized into two main types: filter and wrapper-based methods [1]. Filter-based methods evaluate the importance of features using specific criteria and select top-n features accordingly. Wrapper-based methods employ a learning algorithm to iteratively evaluate candidate feature subsets and directly output an optimal feature subset. Generally, wrapper-based methods outperform filter methods and provide superior performance [1], [2]. However, existing approaches still face challenges, including computational complexity and being prone to local optima, due to the exponentially growing search space as the number of features increases.

Evolutionary algorithms (EAs) have recently gained significant attention and been applied in various domains [3]. EAs are particularly suitable for wrapper-based FS methods as they can obtain multiple solutions within only one population. Current state-of-the-art EAs treat FS as a multi-objective problem with two conflicting objectives: 1) reducing selected features and 2) classification error rate [1]. For instance, Xue *et al.* [4] introduced nondominated sorting into a multi-objective particle swarm optimization-based FS algorithm. More recently, Cheng *et al.* [5] proposed a steering matrix to guide the evolution of the population in multi-objective evolutionary algorithms (MOEAs). Despite the advancements in MOEA-based FS algorithms, challenges such as complex feature interactions and irregular Pareto fronts, still persist [6].

Motivated by the aforementioned discussions, this letter presents a

Corresponding author: Shangce Gao.

The authors are with the Faculty of Engineering, University of Toyama, Toyama-shi 930-8555, Japan (e-mail: d2172002@ems.u-toyama.ac.jp; gaosc@eng.u-toyama.ac.jp; leizg@eng.u-toyama.ac.jp; momura@eng.u-toyama.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

novel multi-objective evolutionary algorithm (termed IEMOEA) for FS. The algorithm is based on the elite-guided solution generation strategy and the information-based mutation operator. The elite-guided solution generation strategy facilitates the learning of a feature subspace and accelerates population convergence. The main contributions of this work are as follows: 1) Introducing a novel elite-guided solution generation strategy to enhance convergence speed and solution quality. 2) Incorporating symmetrical uncertainty (SU) into the mutation operator to further improve solution quality by removing redundant features or adding relevant features. 3) Utilizing a specialized crowding distance measure to eliminate duplicate solutions and ensure population diversity. Experiments are conducted on 20 datasets, encompassing both low and high-dimensional scenarios, to verify the effectiveness and superiority of the proposed IEMOEA method.

Methodology: Our proposed method is realized in Algorithm 1. It can be introduced from three aspects in details:

1) Generating elite solutions (Line 2 in Algorithm 1): Fig. 1 shows the process of generating the elite solution. Consider a problem with $D = 5$ features, the problem can be formulated as maximizing the classification accuracy of the solution E, s.t. $E = (F_1, \ldots, F_i, \ldots, F_5)$, where F_i indicates the i-th feature. Firstly, our proposed method calculates SU values between the features and the label. The SU value between the i-th feature F_i and the label vector C is given by

$$U(F_i, C) = 2 \times \frac{I(F_i, C)}{H(F_i) + H(C)} \tag{1}$$

where $H(C) = -\sum_{c \in C} p(c) \log p(c)$ is the entropy of C. $I(F_i, C)$ is the mutual information between F_i and C, given by

$$I(F_i, C) = \sum_{c \in C} \sum_{f_i \in F_i} p(f_i, c) \log \left(\frac{p(f_i, c)}{p(f_i) p(c)} \right) \tag{2}$$

where $p(f_i, c)$ is the joint probability distribution function of F_i and C, and $p(f_i)$ and $p(c)$ are the marginal probability distribution functions of F_i and C, respectively. Then, the method employs tournament selection to initialize individuals. As shown in Fig. 1, two features are randomly selected, and the feature with the larger U value is chosen. This process is repeated $D = 5$ times. The corresponding positions in the individual are encoded as "1", while the other positions are set to "0". Then the proposed method applies a dimension reduction operator to the individual. The number of features to be reduced is set as $\lceil R \times D \rceil$, where R represents the reduction rate. Tournament selection is also used to reduce features, where two features are randomly selected, and the feature with the smaller U value is reduced. This process is repeated $\lceil R \times D \rceil$ times. Inspired by [7], a linear decrease of R is applied to balance exploration and exploitation, given as

$$R = (0.3 - 0.001) \times ((\hat{t}_e - t_e)/t_e) + 0.001 \tag{3}$$

Algorithm 1 IEMOEA($N, D, T_D, \hat{t}$)

Input: Population size N, decision space dimensionality D, training data T_D; termination criterion $\hat{t}$;

Output: Final population P;

1: $U \leftarrow$ Calculate SU by (1);
2: $E \leftarrow EliteGeneration(T_D, U)$;
3: $P \leftarrow Initialize(N - 1)$;
4: $P \leftarrow P \cup E$;
5: **while** t $< \hat{t} - \hat{t}_e$ **do**
6:    $P' \leftarrow$ Randomly select N parents;
7:    $O \leftarrow Variation(P, P')$;
8:    $P \leftarrow Duplication analysis(P \cup O)$;
9:    $P \leftarrow Nondominated sorting(P)$;
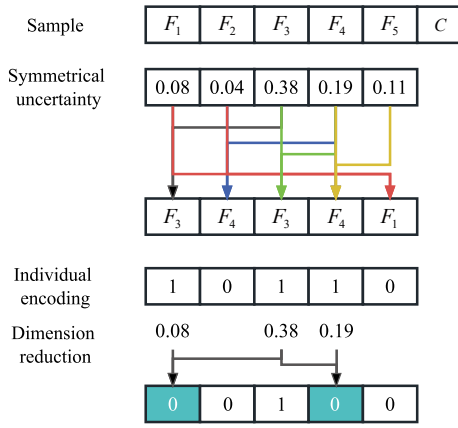10: $P \leftarrow$ the first N solutions in P;
11:   $t = t + 1$;
12: **end**

Fig. 1. Illustration of the elite solution formulation.

where $\hat{t}_e = \begin{cases} 0.1 \times \hat{t}, & if \quad D \geq 200 \\ 0.3 \times \hat{t}, & if \quad D < 200 \end{cases}$ is the termination criterion in elite generation operator and $\hat{t} = \begin{cases} 20\,000, & if \quad D \geq 200 \\ 100 \times D, & if \quad D < 200 \end{cases}$ is the maximum number of iterations. It should be noticed that if all the features are reduced after several steps, the features in the individual will not undergo further reduction. Conversely, one non-selected feature will be randomly selected.

2) The SU-based variation operator (Line 7 in Algorithm 1): The proposed IEMOEA adopts existing single-point crossover and SU-based mutation for binary variation. Specifically, the SU-based mutation operator is designed to flip one of the selected or non-selected variables in the individual with a probability of 0.5. If $rand > 0.5$, two non-selected features are randomly chosen and the corresponding position of the feature with larger U value is flipped to "1". Conversely, if $rand < 0.5$, two selected features are randomly chosen and the corresponding position of the feature with smaller U value is flipped to "0". This mutation operator facilitates a local search mechanism and ensures the quality of the generated offsprings.

3) The special crowding distance-based duplication analysis (Line 8 in Algorithm 1): An issue of FS is the duplicate solutions in both decision and objective space, which can lead to population diversity degradation. Therefore, as part of the environmental selection process in IEMOEA, we initially eliminate all duplicate solutions, retaining only a single unique solution for each set of duplicates. Subsequently, we employ the special crowding distance technique introduced by [8] to handle solutions that share the same objective values but differ in the decision space. The solution with the maximum special crowding distance in the decision space is preserved. Following the duplication analysis, we utilize the non-dominated sorting [9] to rank the population and retain the top-N individuals.

Experiments: The implementation of all the MOEAs was performed using the open-source platform PlatEMO [10]. To ensure consistency, we adopted the specific parameter settings outlined in their original papers [5], [9], [11], [12].

During the experiments, the datasets were separated into two subsets: a training set comprising approximately 70%, and a testing set comprising approximately 30% of the samples. It is worth mentioning that, in the experiments, the SU values were calculated only based on the training data since the testing data is typically unknown in real-world applications. All the algorithms were executed within a wrapper-based approach, and K-nearest neighbors (KNN) (K = 5) was chosen as the wrapped classifier because of its efficiency. Furthermore, to address the issue of FS bias, we employed a 10-fold cross-validation strategy during the training process.

Table 1 presents the 20 datasets utilized in our experiments, encompassing diverse fields such as life sciences, physics, video processing, image analysis, and bioinformatics. These datasets were obtained from the UCI machine learning repository [13], and their

Table 1. Benchmark Datasets

No.	Name	Features	Instances	Classes
1	Australian	14	690	2
2	Zoo	16	101	7
3	Vehicle	18	946	4
4	Wine	22	267	2
5	Breast	30	569	2
6	Ionosphere	34	351	2
7	KrVsKpEW	36	3196	2
8	Waveform	40	5000	3
9	Spect	44	267	2
10	Spambase	57	4601	2
11	Sonar	60	208	2
12	Libras movement	90	360	15
13	Hillvalley	100	1212	2
14	Urban land cover	147	675	9
15	Semeion handwritten digit	256	1593	10
16	LSVT voice rehabilitation	310	126	2
17	Madelon	500	2600	2
18	Multiplefeatures	649	2000	10
19	Colon	2000	62	2
20	Leukemia	7070	72	2

specific descriptions can be found there.

In order to assess the performance of the proposed algorithm, we employ the hypervolume (HV) indicator [3], a well-established measure for MOEAs. The two objectives optimized in this letter, namely the selected features ratio and classification error rate, are already scaled to the range $(0, 1)$ in each objective direction. As a result, we set the reference point for HV as $(1, 1)$ to evaluate the performance of FS algorithms. A higher HV value indicates better algorithm performance.

Table 2 presents the mean and standard deviation (std) values of the HV for all the MOEAs. The highest mean HV value is highlighted in bold. As shown in Table 2, IEMOEA achieves the best performance in 10 out of the 20 datasets. Additionally, we conducted the Wilcoxon signed-rank test to determine whether the performance difference between IEMOEA and the other algorithms is statistically significant, using a significance level of 0.05. The p-values obtained from the test are reported at the bottom of Table 2. Notably, all these p-values are smaller than 0.05, indicating a significant difference between the performance of IEMOEA and its competitors.

To provide a more intuitive assessment of the performance of the proposed IEMOEA, we present the Pareto fronts obtained by all the competitors on the testing set in Fig. 2. It is important to note that the dominated solutions in the final population have been removed. The results of four datasets are depicted in Fig. 2, including two datasets with a relatively small number of features (< 500), namely Spect and Madelon, as well as two high-dimensional datasets (with > 500 features), namely Multiplefeatures and Colon. Fig. 2 clearly shows that in all the displayed datasets, IEMOEA successfully obtains solutions with lower classification error rates. Specifically, in the small-dimensional datasets, solutions generated by IEMOEA can dominate those acquired by its competitors. In the large-scale datasets, IEMOEA can find solutions with lower classification error rates compared to DAEA, SparseEA, and MOEA/PSL. Meanwhile, solutions generated by IEMOEA can dominate those acquired by SMMOEA. The visualization analysis of the Pareto fronts confirms the conclusion that IEMOEA outperforms its peers.

Conclusion: This letter addresses the problem of multiobjective FS and presents a novel information-based elite-guide MOEA (named IEMOEA) for FS. The proposed algorithm introduces SU to generate an elite solution, which facilitates rapid convergence of the population. Furthermore, the algorithm leverages feature information in the mutation process to conduct local search and enhance the quality of solutions. Additionally, duplication analysis is implemented to refine the population and maintain population diversity. Experimen-

Table 2. Mean and Std HV Values Obtained by All Compared Algorithms

| No.	IEMOEA (Mean ± std)	DAEA [9]	SparseEA [11]	MOEA/PSL [12]	SMMOEA [5]
1	**8.4677E−01** ± 2.3490E−03	8.4602E−01 ± 2.1310E−03	8.4570E−01 ± 1.5153E−03	8.4515E−01 ± 1.1240E−03	8.0203E−01 ± 4.7164E−02
2	7.2624E−01 ± 1.0499E−02	7.2245E−01 ± 3.4312E−16	**7.2755E−01** ± 1.3262E−02	7.2361E−01 ± 8.0405E−03	7.1697E−01 ± 2.7151E−02
3	6.9324E−01 ± 7.0587E−03	**6.9568E−01** ± 7.8244E−03	6.8030E−01 ± 1.0024E−02	6.8370E−01 ± 1.1509E−02	6.5498E−01 ± 1.9859E−02
4	**8.8242E−01** ± 4.5590E−16	8.8242E−01 ± 2.3506E−16	8.8242E−01 ± 4.5168E−16	8.8242E−01 ± 4.7238E−16	8.6721E−01 ± 2.3637E−02
5	**9.3575E−01** ± 4.5667E−03	9.1498E−01 ± 2.3052E−03	9.1324E−01 ± 2.3981E−03	9.1377E−01 ± 2.3796E−03	8.9492E−01 ± 2.8058E−02
6	9.0189E−01 ± 4.5427E−03	9.0316E−01 ± 2.4175E−03	**9.0320E−01** ± 2.3357E−03	8.9479E−01 ± 1.2509E−02	8.6420E−01 ± 1.9124E−02
7	9.3393E−01 ± 1.1436E−03	**9.3522E−01** ± 1.3062E−03	9.1022E−01 ± 3.1237E−02	9.3299E−01 ± 1.9249E−03	8.3559E−01 ± 2.7546E−02
8	8.2261E−01 ± 3.7823E−03	**8.2567E−01** ± 2.6582E−03	8.2489E−01 ± 2.4688E−03	8.2220E−01 ± 3.5282E−03	7.9641E−01 ± 1.3525E−02
9	**8.7135E−01** ± 1.0828E−02	8.6917E−01 ± 1.3400E−02	8.6461E−01 ± 1.4366E−02	8.5776E−01 ± 1.4514E−02	8.4581E−01 ± 3.9790E−02
10	**9.1167E−01** ± 2.3529E−03	9.0992E−01 ± 1.7182E−03	9.0947E−01 ± 2.7002E−03	9.0760E−01 ± 1.9082E−03	7.0462E−01 ± 2.1556E−02
11	8.5558E−01 ± 2.1958E−02	8.5347E−01 ± 2.1441E−02	**8.5726E−01** ± 2.6387E−02	8.3753E−01 ± 3.0787E−02	8.1980E−01 ± 2.6059E−02
12	4.4766E−01 ± 2.5663E−02	4.4003E−01 ± 2.6348E−02	4.1028E−01 ± 3.3935E−02	4.2146E−01 ± 3.4786E−02	**4.6443E−01** ± 1.9710E−02
13	6.2791E−01 ± 8.2577E−03	6.2755E−01 ± 7.2754E−03	**6.2880E−01** ± 7.2730E−03	6.2244E−01 ± 9.2373E−03	6.1457E−01 ± 1.1980E−02
14	**8.6550E−01** ± 1.2784E−02	8.3643E−01 ± 6.5004E−03	8.2313E−01 ± 9.9165E−03	8.3068E−01 ± 1.0708E−02	7.8322E−01 ± 1.6446E−02
15	**8.7498E−01** ± 1.1117E−02	8.6333E−01 ± 7.9474E−03	8.6765E−01 ± 5.2192E−03	8.7287E−01 ± 5.1543E−03	8.5655E−01 ± 4.0966E−03
16	**9.3865E−01** ± 1.9349E−02	9.1934E−01 ± 1.1290E−02	8.9719E−01 ± 1.6819E−02	8.8575E−01 ± 2.1952E−02	8.8521E−01 ± 3.3281E−02
17	**9.1745E−01** ± 2.6874E−03	8.8373E−01 ± 5.4245E−03	8.9963E−01 ± 9.9456E−03	8.9500E−01 ± 6.7642E−03	8.0226E−01 ± 1.2905E−02
18	**9.7808E−01** ± 1.4267E−02	9.6719E−01 ± 2.7814E−03	9.6075E−01 ± 3.3129E−03	9.6184E−01 ± 5.8221E−03	9.5261E−01 ± 3.5350E−03
19	9.5096E−01 ± 3.6284E−02	9.5409E−01 ± 2.4256E−02	**9.5571E−01** ± 1.7326E−02	9.3205E−01 ± 3.7994E−02	9.4544E−01 ± 3.9691E−02
20	9.8531E−01 ± 1.9264E−02	9.9697E−01 ± 1.0980E−02	**9.9985E−01** ± 1.1392E−05	9.5428E−01 ± 7.1374E−02	9.8511E−01 ± 1.7729E−02
p-value		3.07E−02	1.41E−02	7.15E−05	1.18E−04
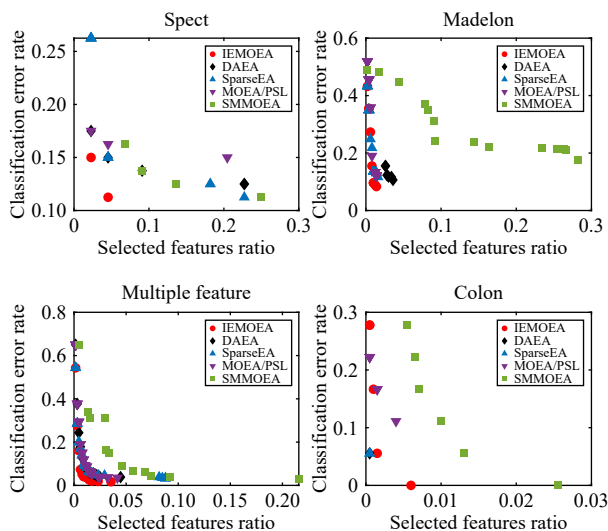
Fig. 2. The rank one Pareto fronts obtained by all compared algorithms on the testing set.

tal results demonstrate the superior performance of IEMOEA compared to other state-of-the-art algorithms. The elite-guided search approach introduced in IEMOEA can also be applied in future research to enhance the search capabilities of other evolutionary FS methods.

References

[1] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Trans. Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, 2015.

[2] K. Dunne, P. Cunningham, and F. Azuaje, "Solutions to instability problems with sequential wrapper-based approaches to feature selection," *J. Machine Learning Research*, vol. 1, pp. 1–22, 2002.

[3] Y. Hua, Q. Liu, K. Hao, and Y. Jin, "A survey of evolutionary algorithms for multi-objective optimization problems with irregular pareto fronts," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 2, pp. 303–318, 2021.

[4] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimization for feature selection in classification: A multi-objective approach," *IEEE Trans. Cyber.*, vol. 43, no. 6, pp. 1656–1671, 2012.

[5] F. Cheng, F. Chu, Y. Xu, and L. Zhang, "A steering-matrix-based multiobjective evolutionary algorithm for high-dimensional feature selection," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9695–9708, 2021.

[6] Z. Wang, S. Gao, M. Zhou, S. Sato, J. Cheng, and J. Wang, "Information-theory-based nondominated sorting ant colony optimization for multiobjective feature selection in classification," *IEEE Trans. Cyber.*, vol. 53, no. 8, pp. 5276–5289, 2023.

[7] B. Ahadzadeh, M. Abdar, F. Safara, A. Khosravi, M. B. Menhaj, and P. N. Suganthan, "SFE: A simple, fast and efficient feature selection algorithm for high-dimensional data," *IEEE Trans. Evolutionary Computation*, pp. 1–15, 2023. DOI: 10.1109/TEVC.2023.3238420

[8] C. Yue, B. Qu, and J. Liang, "A multiobjective particle swarm optimizer using ring topology for solving multimodal multiobjective problems," *IEEE Trans. Evolutionary Computation*, vol. 22, no. 5, pp. 805–817, 2017.

[9] H. Xu, B. Xue, and M. Zhang, "A duplication analysis-based evolutionary algorithm for biobjective feature selection," *IEEE Trans. Evolutionary Computation*, vol. 25, no. 2, pp. 205–218, 2020.

[10] Y. Tian, R. Cheng, X. Zhang, and Y. Jin, "PlatEMO: A MATLAB platform for evolutionary multi-objective optimization," *IEEE Computational Intelligence Magazine*, vol. 12, no. 4, pp. 73–87, 2017.

[11] Y. Tian, X. Zhang, C. Wang, and Y. Jin, "An evolutionary algorithm for large-scale sparse multiobjective optimization problems," *IEEE Trans. Evolutionary Computation*, vol. 24, no. 2, pp. 380–393, 2019.

[12] Y. Tian, C. Lu, X. Zhang, K. C. Tan, and Y. Jin, "Solving large-scale multiobjective optimization problems with sparse optimal solutions via unsupervised neural networks," *IEEE Trans. Cyber.*, vol. 51, no. 6, pp. 3115–3128, 2020.

[13] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml.