Novel Algorithm for Identifying Kinematics Characteristic Data under Road Conditions in Xi'an

Wensheng Wang[a], Xindong Hu[b], Shengdun Zhao[c], Liying Jin*[a,b,c], Yangfeng Cao[c], Yang Xie[b]
[a]AVIC AEROSPACE LIFE-SUPPORT INDUSTRIES,LTD. 441003,China
[b]Xiangyang Key Laboratory of Intelligent Manufacturing and Machine Vision, Hubei University of Arts and Science, Xiangyang, Hubei Province,441053, China
[c]School of Mechanical Engineering, Xi'an Jiaotong University, Shannxi Xi'an 710049
* Corresponding author: JLY@hbuas.edu.cn

ABSTRACT

Aiming at the problem that the kinematics feature dataset of different traffic flow, average speed, acceleration, and traffic time are affected by high-dimensional, irrelevant, and redundant factors, and the kinematics feature dataset is a multi-objective, multi-constrained and complex nonlinear optimization system, the improved multi-objective evolutionary soft subspace clustering algorithm (iMOSSC) is proposed to mine the micro-stroke segments with different kinematic characteristics and realize data classification. The algorithm uses iNSGA-II as the base algorithm and performs local search operator and repair operator operation in the feature space to accelerate convergence and improve the accuracy of the solution. The feasibility and effectiveness of the algorithm are verified by 12 sets of UCI standard dataset. The classified kinematics characteristic data is used to construct the Xi'an urban road trajectory database. Compared with the iMWK-HD algorithm in the collected kinematics feature data of circulation condition, the feature importance degree of the iMOSSC algorithm is more reasonable, the stability is better, the accuracy is higher, and the classification effectiveness is more obvious than the iMWK-HD algorithm. The excavated kinematics data is imported into the Optimumlap simulation software to construct the actual road circulation condition trajectory database. Based on the ADVISOR commercial software platform for the simulation module. The simulation results show that the fuel economy, acceleration time, and gradeability of the novel dual-mode&dual-motor hybrid drive system are better than those of gasoline vehicles and Prius vehicles when operating under actual road conditions.

Keywords: Kinematics feature dataset; iMOSSC; iNSGA-II algorithm; Optimumlap simulation software; ADVISOR commercial software

1. INTRODUCTION

The kinematic features dataset of the dual-motor & dual-mode power-split hybrid electric vehicle is a multi-objective, multi-constraint, and complex nonlinear optimization system. Hence, each sample in the kinematic features dataset is high-dimensional, irrelevant, and redundant in the feature space. The direct use of these kinematic features data to simulate the dynamic performance of the target vehicle will result in large randomness, which cannot truly reflect the dynamic performance of the target vehicle. Therefore, it is necessary to use the dimensionality reduction technology to process the circulation condition characteristic data to eliminate irrelevant and redundant features, and to mine the micro-stroke segments with different kinematic features to realize the classification of data.

As some kinematic features of driving conditions can better reflect a certain type of working conditions but not sensitive to another type of working conditions, this characteristic is called subspace. If the subspace is not considered, the kinematic information of different working conditions is classified under the same feature subset, which will affect the dimensionality reduction and also affect the accuracy of the classification. Even if certain features are effective for the recognition of multiple working conditions, there will be cases where each feature is sensitive to different working conditions. Therefore, a more appropriate weighting method is used to express the relationship between the features and various working conditions. The subspace under the weight representation is also called the soft subspace [1].

The soft subspace clustering technology can realize pattern recognition of high-dimensional data. The kinematic features data is different under the driving conditions, the characteristics of different driving conditions, and the distribution of each driving condition. In the high-dimensional data space, the irrelevant, redundant, and noise features make the

Third International Symposium on Computer Engineering and Intelligent Communications (ISCEIC 2022), edited by Xianye Ben, Proc. of SPIE Vol. 12462, 124620V · doi: 10.1117/12.2661176

distribution characteristics of each driving condition covered up. In addition to different feature values of each driving condition, different driving conditions also have a different number of features. Therefore, each driving condition is generally in its own unique subspace. The soft subspace clustering algorithm can measure each feature in the high-dimensional space according to the importance degree of each driving condition, i.e. the weight value significantly removes the irrelevant and redundant features, restore the kinematic features distribution law, and effectively improve the distance metrics affected by the sparsity of samples in high-dimensional space.

The soft subspace clustering technology can not only effectively deal with the dimensionality reduction of high-dimensional data, but also explore the subspace corresponding to each working condition and give sensitive features of different working conditions. However, the existing soft subspace clustering technique has the disadvantages of being easy to fall into local extremum and slow convergence, while the evolutionary soft subspace clustering algorithm has a good global optimization ability and does not require the differentiability and continuity of the objective function, and therefore, it is used by many scholars for data clustering [2-4]. In the fields of artificial intelligence, pattern recognition, and statistical analysis, research scholars attach great importance to and extensively study cluster analysis [5-7].

In practical engineering application, the kinematic features data face multiple problems: Firstly, the data is affected by irrelevant and redundant features, the sample distribution is not regular enough, and it is difficult to divide it reasonably, thereby affecting the recognition rate; Secondly, the time cost caused by directly processing high-dimensional data is too high, which affects the operational efficiency of the algorithm; Thirdly, the external environment is changeable, and the test signal is disturbed, which increases the workload of data collection. Therefore, the single-objective evolutionary soft subspace clustering algorithm is not suitable for data recognition in complex situations. The multi-objective evolutionary soft subspace clustering algorithm can effectively solve this problem [8,9], because it optimizes multiple objective functions at the same time, taking into account the optimization direction of each objective function, and evaluating the feature subset and clustering quality from different angles. A multi-objective optimal solution is a set of solutions that are weighed against each other, which is also called a Pareto optimal solution set [10,11], such as improved multi-objective particle swarm algorithm [12], multi-objective evolutionary soft subspace clustering algorithm [13], intelligent Minkowski metric feature weights soft subspace clustering algorithm through hybrid dissimilarity measure (iMWK-HD) [14], improved multi-objective sine and cosine optimization algorithm [15], adaptive soft subspace clustering combining within-cluster and between-cluster information [16], and so on.

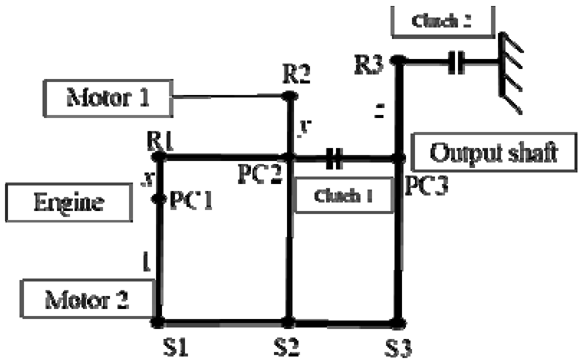
However, these algorithms are limited by design mechanisms, resulting in poor performance when processing large-scale high-dimensional data. In view of this, the improved multi-objective evolutionary soft subspace clustering algorithm (iMOSSC) was proposed. This algorithm inherits the excellent performance of the original iMWK-HD algorithm [14], adds an objective function, can accurately mine the micro-stroke segments with different kinematic features, and realize data classification. During the algorithm execution, real number coding is used. A local search of motion information is performed and operator operations are repaired in the feature space to accelerate convergence and improve the accuracy of the solution. The iNSGA-II as the base algorithm [17], the redundancy is used as the target during mutation, and features having a high correlation with selected features are excluded.

2. RELATED WORKS

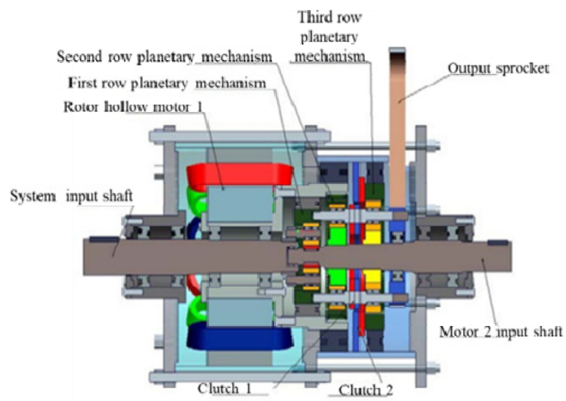
Tsinghua University, Jilin University, Shanghai Jiao Tong University, Harbin Institute of Technology, and other scientific research institutions [18-21], as well as China's FAW Group Corporation, Dong Feng Motor Corporation, Chery Automobile Co., Ltd., and Changan Automobile Co., Ltd. [22], are conducting research on hybrid vehicles, and have achieved fruitful results in their research, besides, some related results have also been applied to urban buses, passenger cars, and other vehicles. The research and development of hybrid power systems provide a new platform for the comprehensive development of automotive technology, which is of great significance to promote electric drive in China's automotive industry [23-26].

Hybrid Electric Vehicle (HEV) provides vehicle power requirements by distributing engine or motor power or torque, which can ensure fuel economy without considering the problem of "mileage anxiety" [27,28]. A hybrid electric vehicle has two power sources, an internal combustion engine and an electric motor. It is an energy management strategy to improve the fuel economy of the hybrid system to achieve energy conservation and emission reduction [29, 30]. The hybrid type adopts the planetary row structure to realize the output power coupling of the engine and the two motors, to achieve the complete decoupling of the torque-speed and the wheels, promote the engine to operate with high efficiency, and effectively simplify the power system structure [31, 32].

Reference [17] analyzed the structural parameter optimization of key components of the novel dual-mode&dual-motor hybrid drive system and proposed an improved NSGA-II algorithm (iNSGA-II). The combination of ADVISOR and Simulink software was used to simulate the performance of the novel system. The simulation results showed that the battery is basically maintained in the expected working range, which meets the vehicle's dynamic performance requirements. The novel dual-mode&dual-motor hybrid drive system rod model is shown in Fig. 1(a), and the cross-sectional view is shown in Fig. 1(b).



(**a**) Rod model



(**b**) The cross-sectional view

Figure 1. Schematic diagram of the novel dual-mode&dual-motor hybrid drive system

For the construction of the Xi'an urban road trajectory database, the kinematic information of the automobile roads in the urban area of Xi'an city was collected by GPS and CAN-Analyzer equipment recorder. As the collected kinematic information such as different traffic flow, average speed, acceleration, and traffic time contain high-dimensional, irrelevant, and redundant data, the resulting kinematic features data of driving conditions is multi-variable, multi-constraint, complex, and nonlinear. Using this kinematic data directly to simulate the vehicle's dynamic performance of the target model will cause it to be very random, and thus cannot truly reflect the dynamic vehicle performance of the target model. In practical engineering applications, decision-makers must simultaneously consider the low cost, high performance, and high efficiency of components. However, there is usually a conflict between low cost, high performance, and high efficiency. Promoting one of them may cause a reduction or degradation of the other two. Problems like this are called multi-objective optimization problems [33]. The multi-objective evolutionary soft subspace clustering algorithm is an integration of multi-objective optimization problems and a soft subspace clustering algorithm. Based on the iMWK-HD algorithm [14], to solve the dimensionality reduction problem of high-dimensional data of cyclic kinematic features, an objective function was added to construct a multi-objective evolutionary soft subspace clustering algorithm (iMOSSC) to solve the kinematic features data of road driving conditions in urban areas of Xi'an city in Shaanxi Province, and the corresponding software module which can effectively realize the dimensionality reduction of the kinematic information and significantly remove redundant data in the kinematic information was developed.

3. IMOSSC ALGORITHM

The intelligent identification of the cycle conditions of the novel dual-mode&dual-motor hybrid drive system is built on data mining based on multi-objective evolutionary soft subspace clustering model analysis. It aims to find potentially searchable rules in the disordered sample data and extract different features.

3.1 Multi-objective mathematical model

$$
\begin{gathered}
\min J_1(U,V,W)=\sum_{k=1}^{C}\sum_{i=1}^{N}u_{ik}\times\left(\sum_{j=1}^{M}w_{kj}^{\beta}\times\left[(x_{ij}-v_{kj})^{\beta}-\eta\times\left(\frac{v_{kj}\times x_{ij}}{\|v_k\|\times\|x_i\|}-\frac{1}{M}\right)\right]\right)^{1/\beta}\\
\min J_2(U,V,W)=\gamma\times\sum_{i=1}^{N}\sum_{j=1}^{M}w_{kj}\times\log w_{kj}-\eta\times\sum_{k=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{M}u_{ik}^{m}\times w_{kj}\times\left(1-\exp\left(-\frac{(v_{kj}-v_{oj})^2}{\sigma^2}\right)\right)\\
s.t.\quad u_{ik}\in\{0,1\},\quad \sum_{k=1}^{C}u_{ik}=1,\quad 1\le i\le N\\
w_{kj}\in[0,1],\quad \sum_{j=1}^{M}w_{kj}=1,\quad 1\le k\le C
\end{gathered} \tag{1}
$$

where C ---- Number of clusters; N ---- Total number of samples; M ---- Total number of features; $U = [u_{ik}]$ ---- The degree of membership of the k-th class on the i-th sample; $V = [v_{kj}]$ ---- The center value of the k-th class on the j-th dimension; $W = [w_{kj}]$ ---- Feature weight value of the k-th class on the j-th dimension feature; x_{ij} ---- The value of the i-th sample on the j-th feature; $k_\sigma = \exp(-(v_{kj} - v_{oj})^2 / \sigma^2)$ ---- Gaussian kernel function, which changes the original measurement method to a certain extent; σ ---- Parameter, the value of this article is 2, (σ = [2, 5, 10]) [34]; γ ---- Information entropy coefficient, which is used to coordinate the influence of entropy on the clustering results, the value range is (0,1); η ---- Reciprocal of sample data variance, such as $\eta_j = 1 / s_j$, $s_j = \Sigma_{i=1}^N (x_{ij} - \dot{x}_j)^2 / N$, $\dot{x}_j = \Sigma_{i=1}^N x_{ij} / N$; v_{oj} ---- The average value of all sample points in the class center of the j-th feature dataset.

In Matlab software, the real number coding method is used to mix the class centers and feature weights. For example, a dataset with a sample number of N and a dimension of M is divided into C classes, and its coding method is $\{v_{11}, \ldots, v_{1M}, v_{21}, \ldots, v_{2M}, \ldots, v_{C1}, \ldots, v_{CM}, w_{11}, \ldots, w_{1M}, w_{21}, \ldots, w_{2M}, \ldots, w_{C1}, \ldots, w_{CM}\}$, with a length equal to $2 \times C \times M$. The initial value of each gene locus of the class center in the coding is randomly selected in the dataset, and the initial value of each gene locus of feature weight is equal to the reciprocal of the dimension. Because the dataset has been normalized before running, their the value range is [0, 1].

3.2 Search operator and repair operator

According to the iterative idea of the clustering algorithm, a local search operator is used to improve the class center and feature weight accuracy. Their calculation formulas are shown in formula (2) and formula (3).

$$v_{kj} = \sum_{i=1}^{N} u_{ik} \times x_{ij} \times \left(1 + \frac{\eta}{2 \times \|v_k\| \times \|x_i\|}\right) / \sum_{i=1}^{N} u_{ik}$$
$$\begin{cases} u_{ik} = 1, & \sum_{j=1}^{M} w_{kj}^{\beta} \times D_{kj} \le \sum_{j=1}^{M} w_{kj}^{\beta} \times D_{lj} \\ u_{ik} = 0, & 1 \le l \le C, \quad l \ne k \end{cases} \tag{2}$$

$$w_{kj} = \exp\left(\sum_{p=1}^{C} |v_{pj} - v_{kj}|\right) / \sum_{q=1}^{M} \exp\left(\sum_{k=1}^{C} v_{kq} - v_{kj}\right) \tag{3}$$

where $D_{kj} = (x_{ij} - v_{kj})^\beta - \eta \times [(v_{kj} \times x_{ij}) / (\|v_k\| \times \|x_i\|) - (1 / M)]$, When $u_{ik} = 1$, it means that the i-th sample x_i is assigned to the k-th category v_k. When the value of $\Sigma_{p=1}^C |v_{pj} - v_{kj}|$ is larger, it means that the k-th category and the remaining categories are easier to distinguish in the j-th dimensional feature, and then brought into equation (3), the feature weight value will also increase, and vice versa.

In the clustering process, there are two kinds of random clustering division. The first type of division is that the number of classes represented in the encoding is greater than the actual number of classes; the second type of division is that the samples with different encodings refer to the same division after decoding. Therefore, the results of these two situations need to be repaired. The operation flow of the first repair operator is: If there is no sample in one or more classes after each class is divided, then randomly select a sample from the class with the largest number of samples as the class center of the first empty class, then, under the constraint condition that the feature weights satisfy $\Sigma_{j=1}^{M} w_{kj} = 1$, the feature weights of this class are randomly generated within the range of [0, 1]. Then randomly select a sample different from the selected sample as the class center of the second empty class. The feature weight value is also randomly generated in the range of [0, 1]; Repeat the operation by analogy until there is no empty class repair. In the case of repair operators, each class contains at least one sample. The second repair operator is solved based on the crowding distance calculation formula in environment selection. In order to maintain the diversity of the population, priority is given to giving up samples with small crowding distances.

3.3 iMOSSC algorithm flow

Step 1: Normalize all sample values in the dataset [35], and set the required parameter values. The initial value of evolution algebra *ite* is set to 1, and the maximum algebra MaxGen is 500;

Step 2: Initialize class center V_0 and feature weight W_0. V_0 is randomly selected from the normalized sample, and W_0 is the reciprocal of the individual's total dimension. Then bring V_0 and W_0 into equation (1) to calculate the objective function value;

Step 3: Select the parent individual from the parent population P according to the binary league rule, and perform the crossover operation with probability P_c to obtain the child population Q_1. The comparison principle is: Pareto dominates the individual first, and the two do not dominate each other individuals with a large congestion distance are preferred, and they are randomly selected when the two have the same congestion distance;

Step 4: Select progeny individuals from progeny population Q_1 with probability P_s, and perform mutation operation with probability P_m to obtain progeny population Q_2;

Step 5: According to the formulas of (2) and (3), perform local search and repair operations on all descendants of Q_2 in turn to obtain the Q_3 population;

Step 6: Calculate the objective function values of all individuals in Q_3, and merge the parent populations P and Q_3. Then, according to the non-inferior sorting method in NSGA-II, the environment selection operation is performed, and the highest-level $|P|$ individuals are selected to form a new parent population P;

Step 7: If the maximum evolution algebra MaxGen is reached, the evolution will stop, otherwise $ite = ite + 1$, and proceed to step 3 to continue execution.

4. IMOSSC ALGORITHM PERFORMANCE TEST AND ANALYSIS

In this paper, 12 sets of UCI standard dataset [36] are used to test the performance of the iMOSSC algorithm, and the Entropy Weighting Fuzzy Clustering in Composite Kernel Space (CKS-EWFC) [37], entropy weight Soft subspace clustering algorithm (Entropy Weighting K-Means algorithm, EWKM) [38] and fuzzy subspace clustering algorithm (Fuzzy Weighted Soft Subspace Clustering Algorithm, FSC) [39, 40] Three excellent soft subspace clustering algorithms Compared. Table 1 lists the relevant information of the UCI standard dataset. The dataset is represented by the data matrix of “number of samples × features”.

Table 1. UCI dataset related information

Name	Number of samples	Feature dimension	Class number
Australian	690	15	2
Breast tissue	106	9	6
Bupa	345	6	2
Pima Indians diabetes	768	8	2
Vehicle	846	18	4
Wdbc	569	30	2
Letter-ABCD	3096	16	4
Letter-UVWX	3116	16	4
Heart	270	13	2
Iris	150	4	3
Parkinsons	195	23	2
Wine	178	13	3

Each of the four algorithms is executed 10 times using different initial parameters. Each algorithm runs independently 10 times under different parameter combinations, and then uses the average of 10 times *RI* index [41] and normalized mutual information index (NMI) [42] to evaluate the quality of the clustering results, The value range of the two index is [0,1], and the value is proportional to the cluster quality. The calculation formulas of *NMI* and *RI* are equations (4) and (5), respectively. Table 2 shows the values of various parameters of the algorithm during the execution of the experiment.

$$NMI = \sum_{k=1}^{C}\sum_{i=1}^{C} d_{ik} \times \log\frac{N \times d_{ik}}{d_i \times d_k} / \sqrt{(\sum_{k=1}^{C} d_k \times \log\frac{d_k}{N})/(\sum_{i=1}^{C} d_i \times \log\frac{d_i}{N})} \tag{4}$$

Where C ---- The number of sample classifications; N ---- The number of samples; d_{ik} ---- The number of samples belonging to the i-th category, and also the number of samples belonging to the k-th category; d_i ---- The number of samples in the i-th category only; d_k ---- The number of samples in class k only.

$$RI = \frac{d_{00} + d_{11}}{N \times (N-1) \times 0.5} \tag{5}$$

Where d_{00} ---- After the algorithm is run, the samples are divided into different classes, and the real class markers are also divided into the number of sample pairs of different classes; d_{11} ---- After the algorithm is run, the samples are divided into the same class, and the real class labels are also divided into the logarithm of the samples of the same class; N ---- the total number of samples.

Table 2 Parameter setting of each algorithm in the experiment

Algorithm	Parameter setting
iMOSSC	γ = [0, 1], σ = [2, 5 ,10]
CKS-EWFC [37]	m = [1.05, 1.2]; η = [1, 5, 10, 100, 1000], γ = [1, 5, 10, 100, 1000, 10000]
EWKM [38]	γ = [1e-3, 1e-2, 1e-1, 1e0, 1e1, 1e2, 1e3, 1e4, 1e5, 1e6]
FSC [39]	β = [1.05~4] γ = [0.0001, 0.001, 0.01, 0.1]

4.1 UCI dataset test results and analysis

The best results achieved by the four algorithms of iMOSSC, CKS-EWFC, EWKM, and FSC are shown in Table 3, and the bold results indicate the best results of the four algorithms.

It can be seen from Table 3 that most of the solutions obtained by the iMOSSC algorithm proposed in this study are better than or close to the best results of the three traditional soft subspace clustering algorithms of CKS-EWFC, EWKM, and FSC. The iMOSSC algorithm has a more stable performance than the CKS-EWFC, EWKM, and FSC on various datasets. However, on the Heart and Pima Indians diabetes dataset, the iMOSSC algorithm has the best *RI* evaluation index, and the *NMI* evaluation index is almost the same as the CKS-EWFC algorithm, which is superior to the EWKM algorithm and FSC algorithm. On the Parkinsons dataset, the *RI* evaluation index and *NMI* evaluation index of the iMOSSC algorithm are slightly lower than the CKS-EWFC algorithm but higher than the EWKM algorithm and FSC algorithm. This shows that the iMOSSC algorithm has strong robustness, and also shows that no algorithm has better clustering results for all the datasets than other algorithms of the same type. In addition, when the *RI* evaluation index achieves the best result in a set of data, the *NMI* evaluation index may not necessarily achieve the optimal result in the same set of data. In the same set of data, these two evaluation indicators cannot achieve the optimal clustering result at the same time. Therefore, it is necessary to evaluate the performance of the same clustering algorithm under different evaluation indexes. Finally, the performance of the clustering algorithm was comprehensively evaluated. As shown in the variance in Table 3, the iMOSSC algorithm shows better robustness on each dataset than the other three algorithms.

4.2 Cluster Analysis of Kinematics Data of Driving Conditions in Xi'an City

While performing kinematic features processing on the collected data, the data can be divided into segments of driving conditions that are independent of each other, which are called kinematic micro-travel segments. The kinematic micro-travel segments describe the vehicle's driving process, starting from the idle speed, after accelerating, cruising, and decelerating to a complete single-stroke kinematic process. According to the kinematic features and parameter indexes of driving conditions, and based on the analysis of the influence of kinematic features on fuel economy and power of hybrid vehicles, the kinematic features parameters of driving conditions used in this study are shown in Table 4.

Table 3 The best results of the four algorithms on 12 UCI dataset

Dataset	iMOSSC		CKS-EWFC[37]		EWKM[38]		FSC[39]	
	RI	*NMI*	*RI*	*NMI*	*RI*	*NMI*	*RI*	*NMI*
Australian	m = 1, η = 3		m=1.2,η=1, γ =10		γ = 50		β = 1.05, γ = 5	
Mean	**0.7563**	**0.4394**	0.7204	0.3686	0.7122	0.3621	0.7518	0.4279
Variance	0	0	0.0888	0.1345	0	0	0	0
Wine	m = 1, η = 4				γ = 2		β = 2.5, γ = 10^{-5}	
Mean	**0.9400**	**0.8648**	0.8964	0.8464	0.9331	0.8336	0.7241	0.3621
Variance	0	0	0.0649	0	0	0	0	0
Heart	m = 1, η = 0				γ = 100		β = 1.05, γ = 0	
Mean	**0.6941**	0.3061	0.6816	**0.3062**	0.6788	0.2795	0.6788	0.2795
Variance	0.0001	0.2100	0.0088	0	0	0	0	0
Vehicle	m = 1, η = 8				γ = 100		β = 1.5, γ = 10^{-3}	
Mean	**0.6840**	**0.200**	0.6641	0.1848	0.6378	0.1354	0.6505	0.0943
Variance	0.0001	0	0.0019	0.0097	0.0100	0.0233	0	0
Breast tissue	m = 1, η = 7				γ = 10		β = 2.3, γ = 10^{-3}	
Mean	**0.8001**	**0.5013**	0.7457	0.4856	0.7242	0.3189	0.7353	0.3179
Variance	0. 0120	0.0021	0.0258	0.0218	0.0240	0.0173	0.0235	0.0118
Bupa	m = 1, η = 1				γ = 1000		β =1.05, γ = 0.1	
Mean	**0.5512**	**0.0381**	0.5047	0.0196	0.5107	0.0105	0.5052	0.0102
Variance	0.1001	0	0.0008	0	0.0004	0	0.0017	0.0011
Iris	m = 1, η = 4				γ = 2		β = 4, γ = 10^{-2}	
Mean	**0.9514**	**0.8525**	0.8737	0.8513	0.8667	0.7416	0.9381	**0.8525**
Variance	0.0101	0	0	0	0.0112	0.0291	0.0241	0.0246
Parkinsons	m = 1, η = 2				γ = 1		β = 2.4, γ = 0.1	
Mean	0.9688	0.9859	**0.9799**	**1**	0.6606	0.3206	0.6280	0.3059
Variance	0	0	0.0636	0	0.1193	0.0142	0.0022	0
Pima Indians diabetes	m = 1, η = 7				γ = 50		β = 2, γ = 10^{-2}	
Mean	**0.6120**	0.1251	0.5574	**0.1306**	0.5507	0.0297	0.5390	0.0204
Variance	0	0	0.0050	0.0021	0	0	0	0
Wdbc	m = 1, η = 8				γ = 1000		β = 3.1, γ = 10^{-3}	
Mean	**0.9121**	**0.6840**	0.8605	0.6833	0.8365	0.5944	0.7515	0.3932
Variance	0	0	0.0031	0	0	0.0031	0.1170	0.2006
Letter-ABCD	m = 1, η = 4				γ = 1000		β = 1.05, γ =0.1	
Mean	**0.8001**	**0.4703**	0.7623	0.4656	0.7707	0.4480	0.7287	0.3897
Variance	0.0100	0.0214	0.0188	0.1638	0.0374	0.0785	0.0187	0.0830
Letter-UVWX	m = 1, η = 5				γ = 1000		β = 2.1, γ = 10^{-3}	
Mean	**0.3104**	**0.3220**	0.2613	0.3111	0.2836	0.2836	0.2567	0.2567
Variance	0	0	0.0282	0.0471	0.0318	0.0318	0.0833	0.0833

Accuracy and effectiveness are the multi-objective optimization soft subspace clustering algorithms for hybrid electric passenger vehicle driving conditions, which can well identify the purpose and requirements of the kinematic features of the urban area in Xi'an. The importance of kinematic features of the multi-objective optimization soft subspace clustering algorithm in the high-dimensional space in each driving condition is measured by the weight value. This method can significantly remove irrelevant and redundant features, restore the distribution of kinematic features, and effectively reduce the effect of the sample's sparseness in a high-dimensional space on a distance metric. It can be used for the construction of the city circular track database in Xi'an. However, in the test under a complex environment, the kinematic parameters are easily interfered with by the outside world, and the dimensions of these parameters are different. Therefore, the original test data collected usually contains noise, and its accuracy and effectiveness will be relatively poor. It needs to be standardized before analyzing the data. Mapped to the range of [0, 1] and converted into the corresponding dimensionless form, the formula for non-quantization is:

$$data(i,j)^* = \frac{data(i,j) - data(:,j)_{\min}}{data(:,j)_{\max} - data(:,j)_{\min}} \tag{6}$$

where i ---- The i-th sample data; j ---- The j-th dimension feature; $data(:,j)_{\min}$ ---- The minimum value of all samples on the j-th dimension feature; $data(:,j)_{\max}$ ---- The maximum value of all samples on the j-th dimension feature.

Table 4 Kinematic parameters of driving conditions

Kinematic characteristic parameters	Symbol	Unit
Total distance traveled	L	km
Total travel time	T	s
Maximum speed	V_{max}	km/h
Maximum/Min acceleration	a	m/s^2
Maximum/Min deceleration	d	m/s^2

The iMOSSC algorithm was used to perform cluster analysis on the kinematic features (12000 samples, 4000 samples of smooth, congested, and slow movement) in driving conditions, and compared with the single-objective iMWK-HD algorithm [14]. The feature importance ranking method is to count the number of occurrences of each feature in the Pareto solution after running on the single-objective iMWK-HD and multi-objective iMOSSC algorithms. The importance of the features is positively correlated with the number of occurrences. The serial numbers corresponding to the important features in the kinematic features of driving conditions are indicated in bold black font, as shown in Table 5. The clustering performance of the two algorithms was evaluated, and the optimal results are indicated in bold, as shown in Table 6. The feature subsets of the two algorithms are shown in Fig. 2.

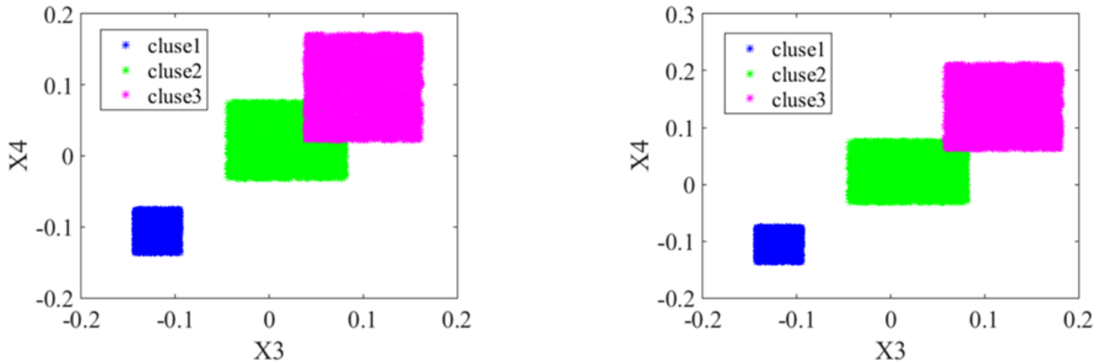
Table 5 The two algorithms are important to the kinematics characteristics of driving conditions in Xi'an

Dataset	iMWK-HD algorithm	iMOSSC algorithm
Kinematics data of driving conditions in Xi'an	**4, 1**, 3, 2	**1, 4**, 3, 2

Table 6 Clustering results of the two algorithms on kinematic feature data

Algorithm	*ACC* (%)	*RI* (%)	Time (s)
iMWK-HD	95.8140	93.4139	**2.1896**
iMOSSC	**98.0122**	**95.9530**	2.2406

It can be seen from Table 5 that both algorithms can correctly select features. It can be seen from Table 6 that the multi-objective iMOSSC algorithm proposed in this study is superior to the single-objective iMWK-HD algorithm in the *ACC* evaluation index and *RI* evaluation index in the 12000 driving conditions kinematic dataset. Compared with the single-objective iMWK-HD algorithm, the multi-objective iMOSSC algorithm's *ACC* evaluation index results increased by 2.1982%, and the evaluation index *RI* value increased by 2.5391%. However, the running time of the multi-objective iMOSSC algorithm only increased by 0.051 seconds.



(a) iMWK-HD algorithm (b) iMOSSC algorithm

Fig. 2 Comparison between the two algorithms for distinguishing kinematics data

It can be seen from Fig. 2 that the feature subset corresponding to the iMOSSC algorithm can basically separate the overlapped kinematic features data, so as to better distinguish the data collected in congestion, slow movement, and unobstructed driving conditions, respectively. This means that the introduction of the objective function between classes is very important for the ranking of feature importance, which can lengthen the distance between feature subsets. Therefore, the robustness of the iMOSSC algorithm and the quality of the clustering results are superior to the single-objective iMWK-HD algorithm.

5. ANALYSIS OF VEHICLE FUEL PERFORMANCE AND POWER PERFORMANCE UNDER ROAD CONDITIONS IN XI'AN

The excavated kinematic features data was imported into Optimumlap simulation software to construct a trajectory database of vehicles operating in the cycling conditions in Xi'an's urban area. Based on the ADVISOR commercial software platform and the above-mentioned Xi'an urban road trajectory database, a co-simulation model including the novel dual-mode&dual-motor hybrid drive system, a gasoline-powered vehicle of the same specifications and size, and the Japanese Toyota Prius 2012 hybrid car on the urban roads of Xi'an was established. The results are shown in Table 7.

Table 7 Comparison of different driving methods under roads in Xi'an urban area

	Performance	Road	Gasoline vehicle	Prius	dual-mode&dual-motor hybrid system
1.8-liter gasoline engine	Fuel consumption (100km)	Urban	8.82-liter	4.06-liter	3.64-liter
		High	7.628-liter	4.271-liter	3.50-liter
		Comprehensive	7.56-liter	4.102-liter	**3.57**-liter
	Dynamic performance	Acceleration (s)	10.3	10.15	**7.55**
		Climbing (%)	63.5	64.5	**73.7**

It can be seen from Table 7 that under the urban roads of Xi'an, the new hybrid electric vehicle system has more advantages than the traditional 1.8-liter standard gasoline vehicle and the Prius 2012 hybrid electric vehicle produced by Toyota Japan. With good fuel economy and power, fuel consumption per 100 kilometers is reduced by 52.8% and 13.0%, acceleration time per 100 kilometers is reduced by 26.7% and 25.6%, and the maximum gradeability is increased by 13.8% and 12.5%, respectively.

6. CONCLUSIONS

Considering that the collected kinematic data is subject to multi-objective, multi-constraint, and complex nonlinear characteristics due to the changeable external environment and test signal interference, and based on the iMWK-HD algorithm to solve the problem of dimensionality reduction of high-dimensional data in the kinematic information under cyclic conditions, a new multi-objective evolution soft subspace clustering algorithm(iMOSSC) was proposed to dig out micro-stroke segments of driving conditions with different kinematics, achieving data clustering and classification. The algorithm used iNSGA-II as the base algorithm and performed local search operator and repair operator operation in the feature space to accelerate convergence and improve the accuracy of the solution. The feasibility and effectiveness of the algorithm were verified by 12 sets of UCI standard datasets.

The classified kinematic features data was used to construct the Xi'an urban road trajectory database. Based on the ADVISOR commercial software platform and the Xi'an urban road trajectory database, a co-simulation model, including the novel dual-mode&dual-motor hybrid drive system, a gasoline-powered vehicle of the same specifications and size, and the Japanese Toyota Prius 2012 hybrid car on the urban roads of Xi'an, was developed. The simulation results showed that, compared with the standard 1.8 liters gasoline-powered vehicle and the hybrid system of the Japanese Toyota Prius 2012, the new hybrid-electric system has better fuel economy, acceleration performance, and gradability when running under the urban road conditions of Xi'an. The fuel consumption per 100 kilometers is reduced by 52.8%、13.0 %, respectively, the acceleration time per 100 kilometers is reduced by 26.7%、25.6% respectively, and the maximum grade is increased by 13.8%、12.5% respectively..

FUNDING

This work was supported by the Key Research and Development Program of Hubei Province (Grant no. 2021BID001) and the Research on Multiple Regression and Fitting Technology of Simulation Data for Dynamic Umbrella Opening of Lifesaving Umbrella (Grant No. HX2021157).

REFERENCES

[1] Hu Xia. Research on Processing High-Dimensional Dataset Based on an Intelligent Diagnosis Approach with Multi-objective Evolutionary Optimization and Soft Subspace Clustering [D]. Xi'an: Xi'an Jiaotong University, 2013.

[2] Hu Xia, Jian Zhuang, Dehong Yu. Novel soft subspace clustering with multi-objective evolutionary approach for high-dimensional data [J], Pattern Recognition, 46 (2013): 2562-2575.

[3] Weimin Zhong, Dayu Tan, Xin Peng, et al. Fuzzy high-order hybrid clustering algorithm for swarm intelligence sets [J]. Neurocomputing, 314(2018):347-359.

[4] Chao Liu, Jing Xie, Qi Zhao, et al. Novel evolutionary multi-objective soft subspace clustering algorithm for credit risk assessment [J]. Expert Systems With Applications, 138 (2019):1-11.

[5] Erzhou Zhu, Yuanxiang Zhang, Peng Wen, et al. Fast and stable clustering analysis based on Grid-mapping K-means algorithm and new clustering validity index [J]. Neurocomputing, 363(2019),149-170.

[6] Luque-Casado Antonio, Fogelson Noa, Iglesias-Soler Eliseo, et al. Exploring the effects of Transcranial Direct Current Stimulation over the prefrontal cortex on working memory: A cluster analysis approach [J]. Behavioural brain research, 375 (2019):112144.

[7] Yongping Tang; Zizhen Peng. A Customer Group Mining Method Based on Cluster Analysis [J]. Advanced Multimedia and Ubiquitous Engineering. MUE/FutureTech 2019, (2020) 351-357.

[8] Deb K. Multi-objective optimization. In: Search methodologies, Springer, 2014: 403–449.

[9] Guanci Yang, Ansi Zhang, Shaobo Li, et al. Multi-objective evolutionary algorithm based on decision space partition and its application in hybrid power system optimization [J]. Appl Intell (2017) 46:827-844. DOI 10.1007/s10489-016-0864-1.

[10] Cruz-Ramírez M, Hervás-Martínez C, Sánchez-Monedero J, et al. Metrics to guide a multi-objective evolutionary algorithmfor ordinal classification [J]. Neurocomputing, 135(2014): 21-31.

[11] Zhaomin Chen, KiatYeo Chai, BuSung Lee, et al. Evolutionary multi-objective optimization based ensemble autoencoders for image outlier detection [J]. Neurocomputing, 309(2018): 192-200.

[12] Jiaxing You, Jili Chen, Minggang Dong. An improved multistage multi-objective particle swarm optimization algorithm [J]. Journal of Chinese Computer Systems, 2015, 36(4):792-796.

[13] Shengdun Zhao, Liying Jin, Yuehui Wang, et al. Soft Subspace Clustering with a Multi-objective Evolutionary Approach [J], Association for Computing Machinery, 2018, DOI: 10.1145/3271553.3271610. (EI: 20185106273289).

[14] Liying Jin, Xiaobin Zhi, Shengdun Zhao. Enhanced subspace clustering through combining Minkowski distance and Cosine dissimilarity [J]. Journal of Intelligent & Fuzzy Systems, 2018, 35(5): 5541-5556.

[15] Wanliang Wang, Weikun Li, Yule Wang, et al. Improved Multi-objective Sine Cosine Optimization Algorithm [J]. Journal of Chinese Computer Systems, 2019, 40(10):2102-2108.

[16] Liying Jin, Shengdun Zhao*, Congcong Zhang, etc. Adaptive Soft Subspace Clustering Combining Within-cluster and Between-cluster Information [J]. Journal of Intelligent & Fuzzy Systems, 2020, 38(3): 3319-3330.
[17] Liying Jin, Shengdun Zhao, Wei Du, et al. Research on Multi-objective Optimization of Switched Flux Motor Based on Improved NSGA-II Algorithm [J]. PART E-JOURNAL OF PROCESS MECHANICAL ENGINEERING, 2019, 233(6): 1268-1279.
[18] Huiyong Liu, Yeping Xiong and Qing Zhao. Research Status and Development Trends of Petrol-Electric Hybrid Construction Machinery [J]. MACHINE TOOL & HYDRAULICS, 2017, 45(15): 166-171.
[19] Fei Jiang. Research on Design and Control Strategy of the Hydraulic Hybrid Excavator Slewing [D]. Harbin: Harbin Institute of Technology, 2015.
[20] Hong Wang. Research on Dynamic Modeling and Control Strategy for Hybrid Tracked Bulldozer [D]. Bei Jing: Beijing Institute of Technology, 2015.
[21] Peng Li. Research on Control Strategy and Simulation of Hybrid in Crawler Bulldozer [D]. Xi'an: Chang'an University, 2015.
[22] Miao Xu. Research on parameter design and control strategy of hybrid excavator power assembly [D]. Changchun: Jilin University, 2017..
[23] Kanghui LÜ, Ping Liao, Chen Wang, et al. Research on fuel consumption of vehicle's power split hybrid powertrain at high-speed operation [J]. Modern Electronies Technique, 2019, 42(9): 72-75, 80.
[24] Hamut HS, Dincer I, Naterer GF. Analysis and optimization of hybrid electric vehicle thermal management systems [J]. Journal of power sources, 2014, 247: 643-654.
[25] Kaur J, Gaur P, Saxena, et al. Speed control of hybrid electric vehicle using artificial intelligence techniques [J]. International journal of computing and network technology, 2014, 2(1): 33-39.
[26] Bing Han. Single-mode Compound Power Split Hybrid Power System Development and Heat Balance Technology Research [D]. Zhenjiang: Jiangsu University, 2011.
[27] Li Feng. Simulate Investigation of the Power Matching for the Parallel Hybrid Electric Vehicle [D]. Xi'an: Xi'an University of Technology, 2006.
[28] Weimin Li. Research of Hybrid Electric Vehicle Control System and Energy Management Strategy [D]. Shanghai: Shanghai Jiao Tong University, 2008.
[29] Yong Wang, Jianyong Li, Xin Zhang, et, al. Research on Control Strategy of Hybrid Electric Vehicle [J]. Beijing Auto, 2004 (2): 11-14 .
[30] Xiaoqi Tang, Fawen Liu, Guoqing Zhao, et, al. Research on Control Strategy of Hybrid Electric Vehicle [J]. Automotive Technology, 2004 (2): 21-24.
[31] Shaohua Wang, Jiaxin Li, Dehua Shi, et, al. Control strategy of power-split HEV based on optimal transmission efficiency [J]. Journal of Jiangsu University, 2018, 39(6): 621-627.
[32] Andaloro L, Arista A, Agnello G, et al. Study and design of a hybrid electric (lithium batteries-PEMFC) [J]. International Journal of Hydrogen Energy, 2017, 42(5):3166-3184.
[33] Xia H, Zhuang J, Yu DH. Novel Soft Subspace Clustering with Multi-objective Evolutionary Approach for High-dimensional Data[J]. Pattern Recognition, 2013,46(9):2562-2575.
[34] Yajun Xu. Research on Subspace Clustering Algorithms and its Applications [D]. Zhenjiang: Jiangnan University, 2016.
[35] Amorim RC, Mirkin B. Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering[J]. Pattern Recognition, 2012, 45(3): 1061-1075.
[36] MUR PP, AHA DW. UCI repository of machine learning databases [EB/OL]. [2014-09-06]. http: / /mlearn. ics. uci.edu/MLRepository.html.
[37] Wang J, Deng Z, Choi KS, et al. Distance metric learning for soft subspace clustering in composite kernel space[J], Pattern Recognition, 2016, 52 : 113-134.
[38] Jing L, Ng MK, Huang JZ. An Entropy Weighting K-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data[J], IEEE Transactions on Knowledge & Data Engineering, 2007, 19: 1026-1041.
[39] Gan GJ, Wu JH, Yang ZJ. Advanced Data Mining and Applications[M]. Berlin: Springer, 2006:271-278.
[40] Gan G, Wu J. A convergence theorem for the fuzzy subspace clustering (FSC) algorithm [J]. Pattern Recognition, 2008, 41: 1939-1947.
[41] Rand WM. Objective Criteria for the Evaluation of Clustering Methods[J], Journal of the American Statistical Association, 1971, 66 (336) 846-850.
[42] Strehl A, Ghosh J. Cluster ensembles: a knowledge reuse framework for combining multiple partitions[J]. Journal of Machine Learning Research, 2008, 3(3): 583-617.