

Article

Time Series Feature Selection Method Based on Mutual Information

Lin Huang¹, Xingqiang Zhou², Lianhui Shi¹ and Li Gong^{1,*}¹ Ship Comprehensive Test and Training Base, Naval University of Engineering, Wuhan 430033, China² 91251 Army of PLA, Shanghai 200940, China

* Correspondence: 13971276201@139.com

Abstract: Time series data have characteristics such as high dimensionality, excessive noise, data imbalance, etc. In the data preprocessing process, feature selection plays an important role in the quantitative analysis of multidimensional time series data. Aiming at the problem of feature selection of multidimensional time series data, a feature selection method for time series based on mutual information (MI) is proposed. One of the difficulties of traditional MI methods is in searching for a suitable target variable. To address this issue, the main innovation of this paper is the hybridization of principal component analysis (PCA) and kernel regression (KR) methods based on MI. Firstly, based on historical operational data, quantifiable system operability is constructed using PCA and KR. The next step is to use the constructed system operability as the target variable for MI analysis to extract the most useful features for the system data analysis. In order to verify the effectiveness of the method, an experiment is conducted on the CMAPSS engine dataset, and the effectiveness of condition recognition is tested based on the extracted features. The results indicate that the proposed method can effectively achieve feature extraction of high-dimensional monitoring data.

Keywords: time series; feature extraction; mutual information; system operability; condition identification



Citation: Huang, L.; Zhou, X.; Shi, L.; Gong, L. Time Series Feature Selection Method Based on Mutual Information. *Appl. Sci.* **2024**, *14*, 1960. <https://doi.org/10.3390/app14051960>

Academic Editors: Roque A. Osornio-Rios, Athanasios Karlis and Andres Bustillo Iglesias

Received: 4 August 2023

Revised: 12 February 2024

Accepted: 18 February 2024

Published: 28 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Time series data mining has extensive and important applications in the field of machine learning, such as data classification, clustering, or prediction, which can help explore the potential patterns in time series data that are useful for subsequent research. Currently, most researchers mainly focus on the processing of univariate time series. However, with the development of data collection technology, multidimensional time series data have become increasingly common and contain a large amount of potentially valuable information. Feature selection is an important step in data processing of high-dimensional data (such as regression and classification). Its main function is to reduce computational complexity, avoid the “curse of dimensionality” problem [1], reduce training time, and improve the performance of the predictor [2]. Therefore, how to effectively extract system features is one of the key issues in the field of time series analysis [3], which has been widely used in the following fields: image recognition [4,5], natural language processing [6,7], data mining [8,9], fault diagnosis [10–12], remaining useful life prediction [13,14], microbes classification [15], fatigue detection [16], image classification [17], intrusion detection [18–21], etc.

Feature selection methods can be roughly divided into three categories [22] including filter, wrapper, and embedded, according to their relationship with learning methods [23,24]. On the other hand, according to the utilized training data, feature selection methods can be divided into supervised [25], unsupervised, and semi-supervised models [26]. In this paper, we focus on the feature selection of time series. At present, the feature selection of time series, in most cases, is mainly conducted by observing the variable trends, amplitude, noise, and other characteristics [27]. This kind of method has problems such as strong subjectivity, inability to conduct quantitative analysis, and inaccuracy.

To solve these problems, we propose a time series feature extraction method based on mutual information. Mutual information is a measure of the interdependence between two variables, which indicates how much information is shared between two variables [28]. The greater the mutual information between variables, the stronger the correlation. Firstly, we determine the correlation between features and then select features with a strong correlation with the target variable. This is an unsupervised feature selection method and has become an important feature selection method [29]. In combination with other methods, MI-based feature selection has derived many other methods, such as MI with a correlation coefficient [30], variance impact factor [31], fisher score [32], binary butterfly optimization algorithm [33], conditional mutual information [34], deep neural network [35,36], etc. Among them, the authors of [37] proposed a deep generative network model for feature extraction of multivariate time series and introduced mutual information into the loss function to improve the expression capability and accuracy of the model. In the study [31], the authors proposed a variable selection method based on mutual information and the variance inflation factor (MI-VIF), which eliminated the variables based on MI and VIF, respectively, and showed good prediction performance. A feature selection algorithm based on MI and particle swarm optimization, which optimized the traditional particle update mechanism and swarm initialization strategy, was proposed in the work [23]. The study [38] proposed a novel fuzzy multi-information-based multi-label feature selection approach that was suitable for multi-label learning. The authors of [39] proposed a conditional multiple information-based feature selection algorithm for maximum reliability and minimum redundancy.

The above research mainly focuses on the improvement of the feature extraction algorithm based on mutual information. However, for the feature extraction analysis of time series, the main problem of the mutual information-based method is in finding the most appropriate target variables, that is, the key to the mutual information-based method is to find the target variables. In most cases, there is no intuitive or directly available target information. PCA is a multivariate statistical analysis technique for data compression and feature extraction, which can effectively remove linear correlations between data [40]. The purpose of PCA is to discard a small portion of information using linear transformation and replace the original variable with a few new comprehensive variables while ensuring minimal data loss. Therefore, it is required that the principal components fully reflect the information of the original variables, while PCA can eliminate the linear correlation between variables and suppress noise by fusing multiple variables [41,42]. Based on the above problems and the technical advantages of the PCA method, this paper uses the PCA method to reduce dimensions and fuse sensor data to obtain the target variables that can represent the system state.

Given the problem, this paper proposes a target information extraction method based on PCA dimension reduction, and on this basis, mutual information is used to distinguish and extract the suitable sensor signal. To validate the effectiveness of the method, analysis and validation are conducted on NASA's publicly available CMAPSS aircraft engine dataset, and the effectiveness of operating condition recognition based on the extracted features is tested. The specific content is as follows:

- (1) We propose a time series feature selection method based on PCA dimension reduction and MI. MI is used to quantify the correlation between two variables, but for time series, it is difficult to obtain effective target variables. Therefore, PCA dimension reduction is used to extract the target variables from the time series, and feature selection of the time series is conducted based on this.
- (2) We design a specific technical process for feature selection based on the above theoretical methods. In this process, we focus on the construction of target variables and the method of sensor selection and conduct experimental verification.
- (3) The effectiveness of the proposed method is verified based on a publicly available aviation engine operation dataset. The experimental results are compared with other

methods to verify the feasibility of the proposed method, and experiments are conducted on condition recognition based on the selected features.

The rest of this paper is organized as follows: In Section 2, we describe the basic principle of the algorithm in detail and propose a feature extraction method flow based on MI. In Section 3, we analyze the experimental dataset and conduct comparative experiments using the algorithms, and in Section 4, we analyze the obtained result and compare the proposed idea with existing approaches. Section 5 is the conclusion of this paper, which includes recommendations for future work.

2. Methodology

2.1. Problem Description

This paper focuses on the feature selection of multivariate time series. Each sample in the dataset is a set of time series, and the length of each time series may vary. A set of univariate time series can generally be defined as follows:

$$T = t_1, t_2, \dots, t_n \quad (1)$$

where n represents the length of the time series. Correspondingly, an m -dimensional multi-variable time series can be written as:

$$\begin{aligned} T_1 &= t_{11}, t_{12}, \dots, t_{1n_1} \\ T_2 &= t_{21}, t_{22}, \dots, t_{2n_2} \\ &\dots \\ T_m &= t_{m1}, t_{m2}, \dots, t_{mn_m} \end{aligned} \quad (2)$$

In the feature selection of time series, the dimension of the monitoring data is the number of features in the time series. When processing time series data, if the number of features in the dataset is too large, it can seriously affect the effectiveness of model training, leading to the so-called “curse of dimension” [43]. Effective feature selection can lay a good foundation for subsequent anomaly detection, fault diagnosis, condition identification, remaining useful life prediction, etc.

2.2. Feature Selection Based on Mutual Information

Mutual information is a measure of the interdependence between two variables, which mainly represents the correlation between them. Assuming that the joint probability distribution of two random variables X and Y is $p(x, y)$, and their marginal probability distribution is $p(x)$ and $p(y)$, respectively, then the mutual information $MI(X, Y)$ is the Kullback–Leibler divergence of the joint probability distribution $p(x, y)$ and the marginal probability distribution $p(x)$ and $p(y)$, which is defined as follows [28]:

$$MI(X, Y) = - \int_x \int_y f_{x,y}(x, y) \log \frac{f_{x,y}(x, y)}{f_x(x)f_y(y)} dx dy \quad (3)$$

From the above definition, when variables are completely independent or mutually independent, their mutual information is the smallest, and the result is 0. The greater the mutual information between variables, the stronger the correlation. In data processing, features with large mutual information with the target variable should be selected as much as possible to improve the prediction ability of the algorithm, and features with small mutual information should be eliminated to reduce data redundancy, that is, so-called feature selection based on maximum correlation and minimum redundancy [44].

In the specific application of mutual information, the probability density estimation function can be used to approximate the edge probability distribution to simplify the above formula and express it as follow:

$$MI(X, Y) = -\frac{1}{N} \sum_{i=1}^N \log \frac{\hat{f}_{x,y}(x_i, y_i)}{\hat{f}_x(x_i) \hat{f}_y(y_i)} \quad (4)$$

where $\hat{f}_x(x_i)$ and $\hat{f}_y(y_i)$ are the probability density estimations of X and Y , $\hat{f}_{x,y}(x_i, y_i)$ is the joint probability density estimation function, and N is the number of data samples.

According to Equation (4), the key to mutual information is the estimation of the probability density function. For the univariate edge probability distribution, the kernel probability density estimation method can be used for its estimation. In this way, without prior knowledge of the data distribution, the characteristics of the data distribution can be learned based on the data samples themselves. This is a non-parametric method for estimating probability density functions. For example, for variable X , suppose it has n sample points x_1, x_2, \dots, x_n , which are independent and identically distributed random variables. Then, its kernel probability density estimation can be expressed as:

$$\hat{f}_x(x) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (5)$$

where $K(\cdot)$ is the kernel function and h is the scale parameter of the kernel function, which defines the similarity between samples. Usually, Gaussian kernel functions are used, which can be expressed as:

$$K\left(\frac{x - x_i}{h}\right) = \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{(x - x_i)^2}{2h^2}\right) \quad (6)$$

Therefore, the probability density function of sample points can be estimated as:

$$\hat{f}_x(x) = \frac{1}{Nh_w} \sum_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x - x_j)^2}{2\sigma_x^2 h^2}\right) \quad (7)$$

where σ_x is the variance in variable X , and the scale parameter h can be determined according to the following equation:

$$h_w = \left(\frac{4}{d+2}\right)^{1/(d+4)} N^{-1/(d+4)} \quad (8)$$

where d is the dimension of variable X , and for one-dimensional random variables, d is equal to 1.

For the joint probability distribution $\hat{f}_{x,y}(x_i, y_i)$, assuming $z = [x \ y]^T$, the estimation of the joint probability distribution can be expressed as:

$$\hat{f}_z(z) = \frac{1}{Nh} \sum_{i=1}^N K(u') \quad (9)$$

where:

$$u' = \frac{(z - z_i)^T C_z^{-1} (z - z_i)}{h^2} \quad (10)$$

$$K(u') = \frac{1}{(2\pi)^{d/2} \det(C_z)^{1/2}} \exp(-u'/2) \quad (11)$$

where C_z is the covariance matrix of z and $\det(C_z)$ is the value of the determinant of C_z .

2.3. Construction of the Time Series Target Variable Based on PCA

As mentioned before, for the feature selection problem of time series based on mutual information, one of the difficulties is finding appropriate target variables. In response to this issue, this paper proposes a time series target variable construction method based on PCA. By performing dimensionality reduction processing on multidimensional time series variables using PCA, variables that can characterize changes in system lifetime performance, which can be referred to as system operability, are obtained. On this basis, mutual information analysis is carried out between it and the monitoring variables of the system to conduct feature selection, so as to ensure that the selected features can maximize the prediction accuracy.

Assuming that the dataset $X = (x_1, x_2, \dots, x_L)$ contains L samples and x is an N -dimensional variable, then the empirical mean of x can be expressed as:

$$u = (\mu_1, \mu_2, \dots, \mu_N)^T \quad (12)$$

$$\mu_n = \frac{1}{L} \sum_{l=1}^L X[n, l] \quad (13)$$

The covariance matrix of the samples can be expressed as:

$$C = \frac{1}{L-1} (X - uh) \cdot (X - uh)^T \quad (14)$$

where h is an N -dimensional vector that is all 1. The covariance matrix C is decomposed into eigenvectors, and the first M eigenvectors are selected. The selection of the number of eigenvectors M depends on the desired data variance. In this paper, in order to obtain a one-dimensional objective variable that can characterize the system operability, $M = 1$ is chosen.

2.4. Workflow of the Proposed Feature Selection Method for Time Series

Figure 1 shows the proposed time series feature selection workflow. After obtaining the system time series monitoring data, the sensor signals are initially screened and then normalized. Afterward, in order to obtain the system target variables, PCA dimensionality reduction is performed on the normalized data to obtain one-dimensional time variables that can characterize the trend in the system operability. Then, the mutual information between each time series features with the target variable, as well as between each time series feature, is calculated. According to the value of mutual information, the time series features with large correlations and small redundancy with the target variables are selected.

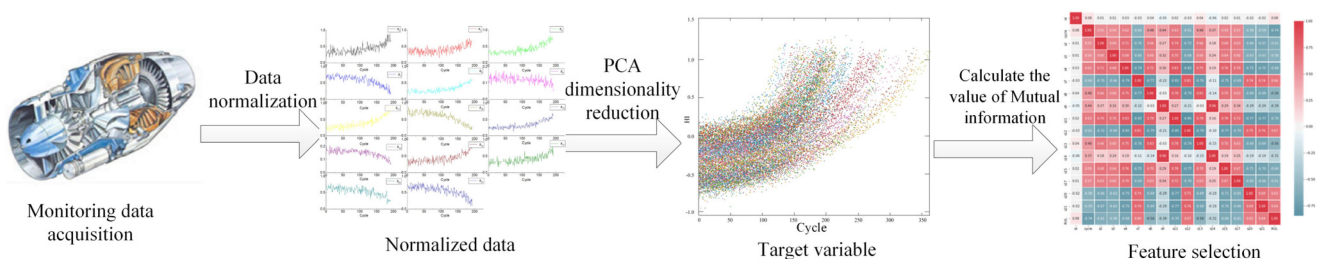


Figure 1. Workflow of the proposed feature selection method.

3. Experiments

To verify the effectiveness of the method proposed in this paper, the CMAPSS aircraft engine simulation time series dataset published by NASA [45] is used to validate the proposed method, which is then compared with different methods. All simulation experiments are carried out in the Pycharm 2021 (Community Edition) environment on a PC with an AMD Ryzen 7 3700X eight-core CPU and 16 GB memory.

3.1. C-MAPPS Datasets

As shown in Figure 2, the structure diagram of the aero-engine based on CMAPSS contains several components such as a fan, combustion chamber, high-/low-pressure booster, turbine, nozzle, etc.

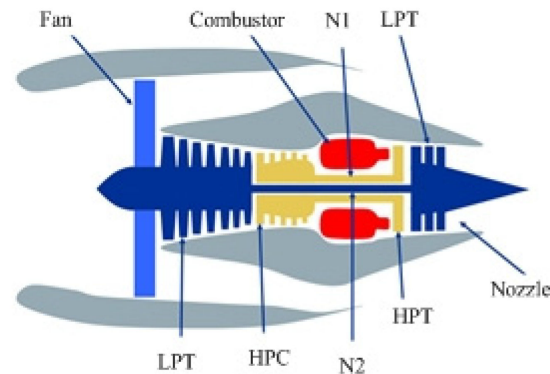


Figure 2. Structure diagram of the C-MAPPS aero-propulsion system.

Table 1 shows the relevant parameters of the C-MAPSS dataset and the monitoring data for each work cycle. Gaussian white noise is added to the monitoring data to simulate the actual sensor noise. This dataset can realistically simulate engine systems with high reliability.

Table 1. Engine sensor data description [45].

No.	Symbol	Description	Unit
1	T2	Total temperature at the fan inlet	°R
2	T24	Total temperature at the low-pressure booster outlet	°R
3	T30	Total temperature at the high-pressure booster outlet	°R
4	T50	Total temperature at the low-pressure turbine outlet	°R
5	P2	Pressure at the fan inlet	psia
6	P15	Total pressure of the external bypass	psia
7	P30	Total pressure at the high-pressure booster outlet	psia
8	Nf	Physical speed of the fan	rpm
9	Nc	Physical speed of the core engine	rpm
10	epr	Engine pressure ratio (P50/P2)	—
11	Ps30	Static pressure at the high-pressure booster outlet	psia
12	Phi	Ratio of fuel flow and P30	pps/psi
13	NRf	Corrected speed of the fan	ppm
14	NRc	Corrected speed of the core engine	rpm
15	BPR	Bypass ratio	—
16	farB	Gas–air ratio of the combustion chamber	—
17	htBleed	Bleed air entropy	—
18	Nf_dmd	Set speed of the fan	rpm
19	PCNfR_dmd	Set corrected speed of the core engine	rpm
20	W31	Cooling bleed air flow of the high-pressure turbine	lbm/s
21	W32	Cooling bleed air flow of the low-pressure turbine	lbm/s

In the CMAPSS dataset, some sensor signals are constant and do not vary with engine operation. After removing these data, the CMAPSS dataset has 14 sensor signals in total, that is, the data dimension of the original data space is reduced from 21 dimensions to 14 dimensions. Further, in order to reduce the redundancy between variables, based on the idea of the feature selection of maximum correlation and minimum redundancy, before feature selection, the variables with high redundancy are removed according to the mutual information value between the variables. Figure 3 shows the mutual information

thermogram of 14 sensor signals in the C-MAPSS dataset. The diagonal value of the thermogram is 1, and it is symmetrical with the diagonal.



Figure 3. Mutual information matrix among variables (figure drawn by seaborn [46]).

From Figure 3, it can be seen that some sensor signals are highly correlated. For example, the value of mutual information between s9 and s14 is 1, indicating that there is a high degree of redundancy between them. From Table 1, it can be seen that s9 represents the physical speed of the core engine, and s14 represents the corrected speed of the core engine. The measured values of the two are highly correlated, and the calculated mutual information is consistent in a physical sense. Based on the mutual information matrix, the system features are further filtered. With 0.95 as the threshold (manually set), some redundant features are deleted, and a total of nine sensor signals in the system can be obtained, that is, from 21 dimensions of the original data space to 9 dimensions after mutual information filtering among variables.

3.2. Data Normalization

Due to the different physical quantities measured by different sensors, the data need to be normalized to make the variable range consistent, so as to achieve better classification and prediction effects. This paper uses Z-score standardization to normalize the input data, which can be expressed as:

$$X' = \frac{X - \bar{X}}{\sigma} \quad (15)$$

where X is the time series data, \bar{X} and σ are the average and variance of X , respectively, and X' is the normalized monitoring data. Figure 4 shows the trend in nine variables after normalization.

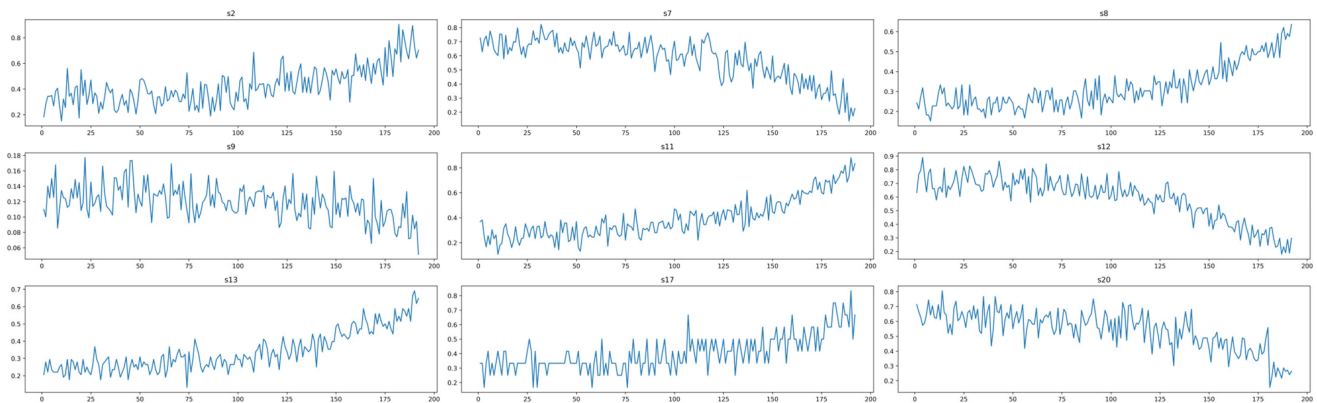


Figure 4. The trends in the 9 variables after normalization (figure drawn by matplotlib [47]).

3.3. Target Variable Construction

Two issues need to be considered in the construction of the target variable of the system. One is that the variable needs to be able to reflect the change in system operability, and the other is that the variable needs to ensure monotonicity and irreversibility in the whole life cycle. For monitoring and control systems, in most cases, the monotonicity, predictability, and irreversibility of a single sensor signal cannot be guaranteed, nor can it fully reflect the system operability. Therefore, this paper uses the sensor information fusion method based on PCA dimension reduction and kernel smoothing to extract the system target variables, thus ensuring the monotonicity and irreversibility of the degradation trajectory [48].

It should be noted that the degradation trajectory extracted using PCA and KR is the system operability mentioned above, which reflects the overall operational status of the system. The specific approach is to first fuse the multidimensional monitoring variables of the system using PCA to extract the principal component that contains the most system information. At the same time, in order to ensure the monotonicity and irreversibility of the system operability, KR is used to smooth the extracted principal components, so as to obtain the curve of system operability, as shown in Figure 5.

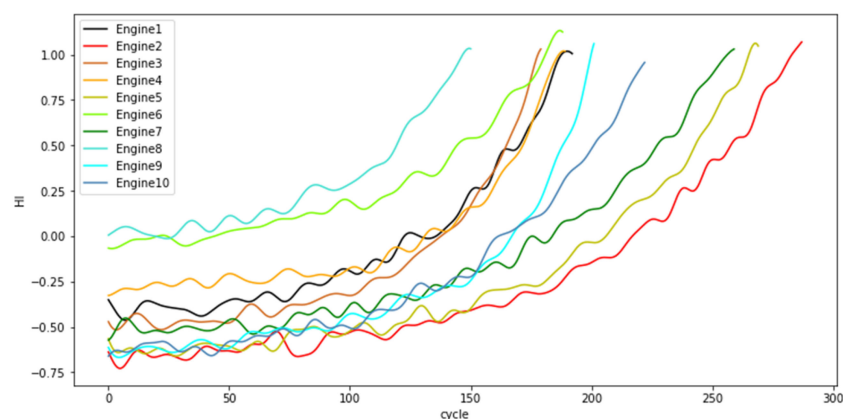


Figure 5. Target variables for engines 1 to 10 based on PCA dimensionality reduction (figure drawn by matplotlib).

Figure 5 shows the system operability of engines from numbers 1 to 10 in the dataset. It can be seen from the figure that by extracting the system degradation curve as the target variable, the monotonicity and irreversibility of the variable can be guaranteed, and the system operability can reflect the health status of the system to a certain extent.

3.4. Calculation of the Mutual Information of Sensor Signals

The first engine data in the CMAPSS dataset is taken as an example to analyze the mutual information values of different sensor signals with the target variable. It should be noted that among the 21 sensor signals in Table 1, some sensor signals are constant values, which do not vary with engine operation, such as s1, s5, s6, s10, s16, s18, and s19. After removing these data, a total of 14 sensor signals are obtained in the C-MAPSS dataset. Table 2 shows the mutual information values of the nine different sensor signals with the target variable, and Figure 6 shows the histogram of the mutual information values of sensor signals with the target variable.

Table 2. Value of mutual information between the sensor signal and the target variable.

Sensor Number	Value of Mutual Information
2	0.570
7	0.697
8	0.597
9	1.000
11	0.951
12	0.705
13	0.682
17	0.485
20	0.506

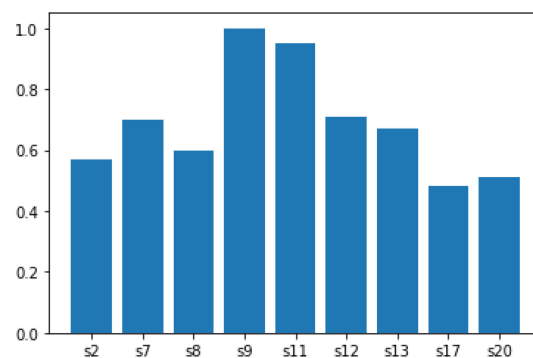


Figure 6. Mutual information histogram of sensor signals with the target variable (figure drawn by matplotlib).

It can be seen from Table 1 and Figure 6 that the five sensor signals with high mutual information with the target variable are s7, s9, s11, s12, and s13, corresponding to the total pressure at the high-pressure booster outlet, the physical speed of the core engine, the static pressure at the high-pressure booster outlet, the ratio of fuel flow and P30, and the corrected speed of the fan.

4. Result Analysis

Using the calculation in Section 3, we obtained the values of MI for each sensor signal based on the target variable. To verify the effectiveness of the proposed method, we compared and analyzed it with the scores obtained from the F-test score to clarify that the proposed idea is better than the existing approach. On the other hand, in order to measure the performance of the extracted features, we use the extracted features to identify six different operating conditions of the engine in the CMAPSS dataset and verify them with visual data clusters.

4.1. Comparative Analysis of the Selected Features

In order to verify the effect of the selected features, the correlation between the nine sensor signals with the target variable is visualized, and the F-test scores between sensor

signals and target variables are given, as shown in Figure 7. The F-test score [49] is used to extract the linear relationship between the nine sensor signals with the target variable, which can be mutually verified with feature selection based on mutual information to a certain extent.

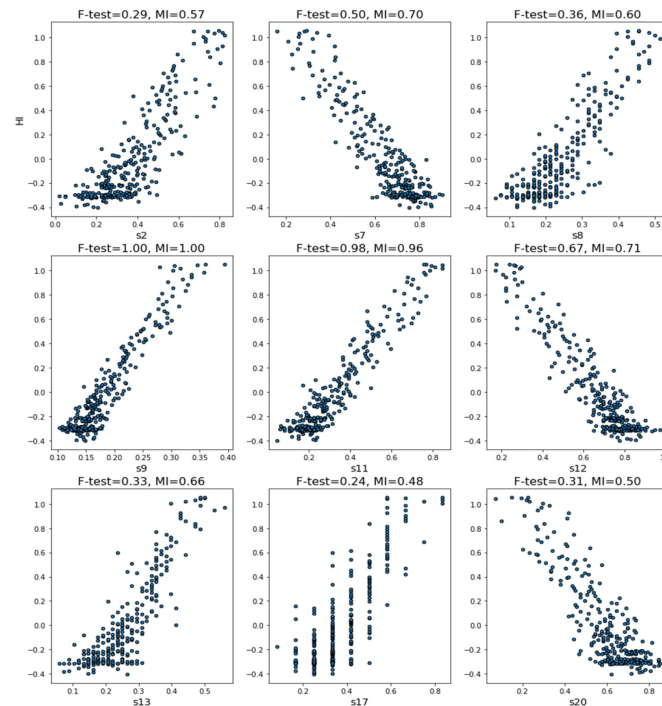


Figure 7. F-test and mutual information comparison of the nine sensor signals (figure drawn by matplotlib).

As can be seen from Figure 6, the features selected based on the F-test are basically consistent with those selected based on mutual information. For example, for sensor signal s9, both give scores of 1 and for sensor signal s11, the scores given by both are 0.98 and 0.96, respectively. Obviously, this indicates a strong linear relationship between s9 and s11 with the target variable. However, the difference is that feature extraction based on mutual information can better analyze and extract the nonlinear relationship between features and target variables. For example, when s13, s17, and s20 do not show an obvious linear relationship with the target variables, the value of mutual information is higher than the F-test score. Based on the above analysis, the five time series characteristics that can best represent the state of the system are obtained, and their value of mutual information is s9, s11, s12, s7 and s13 from high to low.

4.2. Effect Analysis of Working Condition Recognition Based on the Selected Features

Figure 7 shows the effect of working condition recognition based on the selected system time series features. In the CMAPSS dataset, there are a total of six different operating conditions for the engines, and they continuously alternate between the six different operating conditions throughout their entire lifetime. Accurately identifying the system working conditions based on the selected sensor signals is an important foundation for subsequent research such as system operability analysis, etc.

Figure 8a shows the effect of recognition based solely on sensor signal s9, Figure 8b shows the effect of working condition recognition based on s9 and s11, and Figure 8c shows working condition recognition based on s9, s11, and s12. From Figure 8, it can be seen that Figure 8a,b can achieve working condition recognition to a certain extent. However, there are some data points that overlap, which means that the system's working condition identification cannot be effectively and completely achieved. By selecting the first three

features, the recognition of complex working conditions of the system can be effectively realized, which shows the effectiveness of selecting time series features of the system based on mutual information.

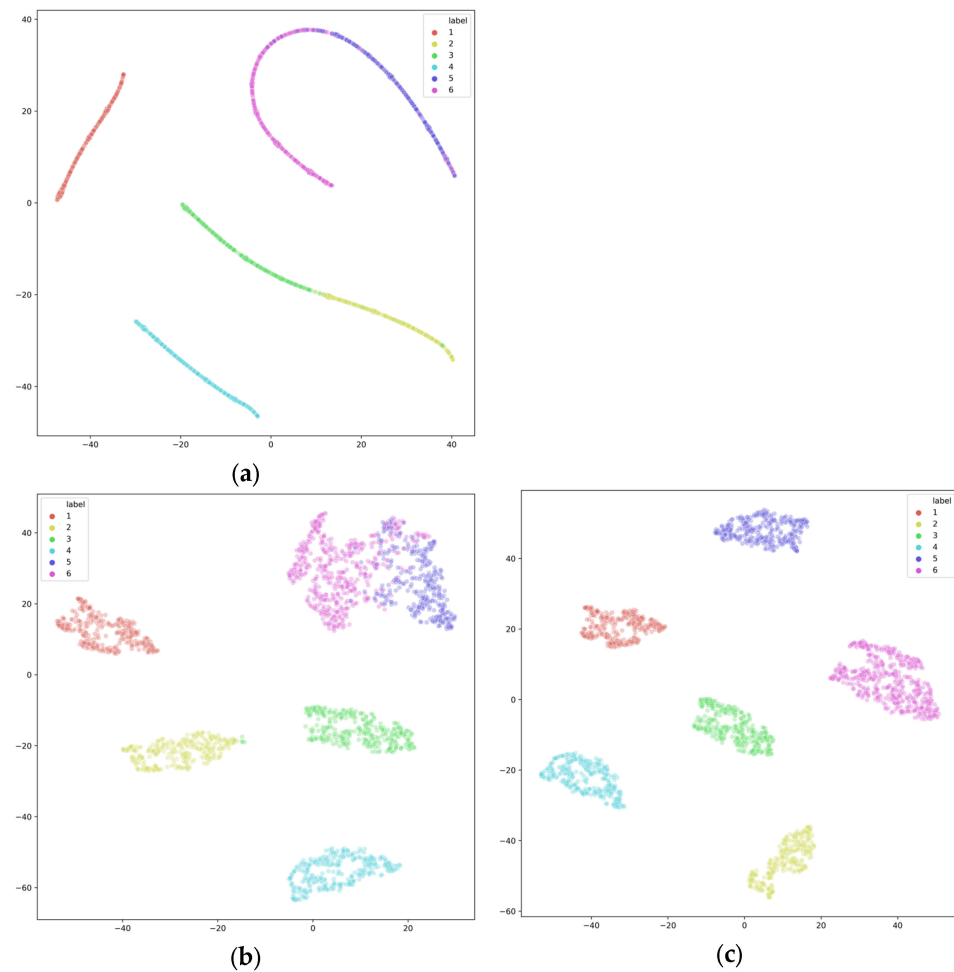


Figure 8. Recognition of working conditions based on the selected features (figure drawn by seaborn): (a) working condition recognition based on s9; (b) working condition recognition based on s9 and s11; and (c) working condition recognition based on s9, s11, and s12.

5. Conclusions

Extracting effective features of a system based on mutual information values is a commonly used feature extraction method in the field of machine learning. However, one of the most challenging issues with this method is finding suitable target variables. Theoretically, these target variables must include the main performance indicators of system operation to ensure that the extracted features can effectively reflect the multiple characteristics of the system. Especially for time series, the historical data of system operation contains a large amount of information, but at the same time, there is significant redundancy and noise within this information. To solve this problem, this paper proposes a method for feature selection based on MI, in which system operability is used as a target variable. Firstly, the system operability is extracted based on PCA dimension reduction and kernel smoothing, which is monotonic and irreversible. Using it as the target variable for MI analysis is more credible compared to using individual sensor signals as the target variable. On this basis, features are filtered based on the value of mutual information between features and the target variable. In order to verify the effectiveness of the method, comparisons and analyses were conducted with F-test scores. The results show that feature extraction based on mutual information can better analyze and extract the non-linear relationship between

features and target variables. Finally, to further validate the effectiveness of the method, we conducted tests on condition recognition based on the CMPASS dataset. The CMPASS dataset contains continuous operational data of the engine throughout its entire life cycle under six different operating conditions. Effectively partitioning these data between the six different operating conditions is crucial for subsequent analysis. Therefore, based on the selected feature variables, the system condition recognition was carried out. The results indicate that by using the first three extracted feature variables, the complex condition recognition of the system can be achieved. From the above analysis, we identified that the proposed method can effectively achieve system feature extraction and is suitable for feature extraction problems of time series data.

However, it should be noted that one of the most critical steps of the method proposed in this paper is constructing the system operability based on sufficient historical operational data. Therefore, the method is mainly applicable to devices with a large amount of historical operating data. If the data are insufficient, the constructed system operability may not reflect the overall characteristics of the system, leading to inaccurate estimation of MI values. For example, the CMAPSS dataset FD001 used in this paper contains monitoring parameter information for the full life cycle status of 200 engines.

Regarding future work, we plan to study the construction of system operability based on online data and the analysis of MI values, in order to reduce the dependence on historical operating data and improve the applicability of this method.

Author Contributions: Conceptualization, L.H.; methodology, X.Z.; software, L.H.; validation, L.S.; formal analysis, X.Z.; investigation, L.G.; resources, L.G.; data curation, L.S.; writing—original draft preparation, L.H.; writing—review and editing, L.G.; supervision, L.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in the ‘Turbofan Engine Degradation Simulation Data Set’ at <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/> (accessed on 17 February 2024).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aremu, O.O.; Hyland-Wood, D.; McAree, P.R. A machine learning approach to circumventing the curse of dimensionality in discontinuous time series machine data. *Reliab. Eng. Syst. Saf.* **2020**, *195*, 106706. [\[CrossRef\]](#)
2. Khaire, U.M.; Dhanalakshmi, R. Stability of feature selection algorithm: A review. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 1060–1073. [\[CrossRef\]](#)
3. Zebari, R.; Abdulazeez, A.; Zeebaree, D.; Zebari, D.; Saeed, J. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *J. Appl. Sci. Technol. Trends* **2020**, *1*, 56–70. [\[CrossRef\]](#)
4. Jiao, W.; Cheng, X.; Hu, Y.; Hao, Q.; Bi, H. Image Recognition Based on Compressive Imaging and Optimal Feature Selection. *IEEE Photonics J.* **2022**, *14*, 1–12. [\[CrossRef\]](#)
5. Afza, F.; Khan, M.A.; Sharif, M.; Kadry, S.; Manogaran, G.; Saba, T.; Ashraf, I.; Damaševičius, R. A framework of human action recognition using length control features fusion and weighted entropy-variances based feature selection. *Image Vis. Comput.* **2021**, *106*, 104090. [\[CrossRef\]](#)
6. Zhao, H.; Liu, Z.; Yao, X.; Yang, Q. A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. *Inf. Process. Manag.* **2021**, *58*, 102656. [\[CrossRef\]](#)
7. Sharma, M.; Kaur, P. A Comprehensive Analysis of Nature-Inspired Meta-Heuristic Techniques for Feature Selection Problem. *Arch. Comput. Methods Eng.* **2021**, *28*, 1103–1127. [\[CrossRef\]](#)
8. Abualigah, L.; Dulaimi, A.J. A novel feature selection method for data mining tasks using hybrid Sine Cosine Algorithm and Genetic Algorithm. *Clust. Comput.* **2021**, *24*, 2161–2176. [\[CrossRef\]](#)
9. Wang, L.; Jiang, S.; Jiang, S. A feature selection method via analysis of relevance, redundancy, and interaction. *Expert Syst. Appl.* **2021**, *183*, 115365. [\[CrossRef\]](#)

10. Zheng, J.; Pan, H.; Tong, J.; Liu, Q. Generalized refined composite multiscale fuzzy entropy and multi-cluster feature selection based intelligent fault diagnosis of rolling bearing. *ISA Trans.* **2022**, *123*, 136–151. [\[CrossRef\]](#)
11. Cao, Y.; Sun, Y.; Xie, G.; Li, P. A Sound-Based Fault Diagnosis Method for Railway Point Machines Based on Two-Stage Feature Selection Strategy and Ensemble Classifier. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 12074–12083. [\[CrossRef\]](#)
12. Buchaiah, S.; Shakya, P. Bearing fault diagnosis and prognosis using data fusion based feature extraction and feature selection. *Measurement* **2022**, *188*, 110506. [\[CrossRef\]](#)
13. Wang, Y.; Zhao, Y. Three-stage feature selection approach for deep learning-based RUL prediction methods. *Qual. Reliab. Eng. Int.* **2023**, *39*, 1223–1247. [\[CrossRef\]](#)
14. Abushark, Y.B. An intelligent feature selection approach with systolic tree structures for efficient association rules in big data environment. *Comput. Electr. Eng.* **2022**, *101*, 108080. [\[CrossRef\]](#)
15. Dhindsa, A.; Bhatia, S.; Agrawal, S.; Sohi, B.S. An Improvised Machine Learning Model Based on Mutual Information Feature Selection Approach for Microbes Classification. *Entropy* **2021**, *23*, 257. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Halomoan, J.; Ramli, K.; Sudiana, D.; Gunawan, T.S.; Salman, M. ECG-Based Driving Fatigue Detection Using Heart Rate Variability Analysis with Mutual Information. *Information* **2023**, *14*, 539. [\[CrossRef\]](#)
17. Islam, M.R.; Ahmed, B.; Hossain, M.A.; Uddin, M.P. Mutual Information-Driven Feature Reduction for Hyperspectral Image Classification. *Sensors* **2023**, *23*, 657. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Thakkar, A.; Lohiya, R. A survey on intrusion detection system: Feature selection, model, performance measures, application perspective, challenges, and future research directions. *Artif. Intell. Rev.* **2022**, *55*, 453–563. [\[CrossRef\]](#)
19. Halim, Z.; Yousaf, M.N.; Waqas, M.; Sulaiman, M.; Abbas, G.; Hussain, M.; Ahmad, I.; Hanif, M. An effective genetic algorithm-based feature selection method for intrusion detection systems. *Comput. Secur.* **2021**, *110*, 102448. [\[CrossRef\]](#)
20. Nimbalkar, P.; Kshirsagar, D. Feature selection for intrusion detection system in Internet-of-Things (IoT). *ICT Express* **2021**, *7*, 177–181. [\[CrossRef\]](#)
21. Alalhareth, M.; Hong, S. An Improved Mutual Information Feature Selection Technique for Intrusion Detection Systems in the Internet of Medical Things. *Sensors* **2023**, *23*, 4971. [\[CrossRef\]](#)
22. Maldonado, J.; Riff, M.C.; Neveu, B. A review of recent approaches on wrapper feature selection for intrusion detection. *Expert Syst. Appl.* **2022**, *198*, 116822. [\[CrossRef\]](#)
23. Song, X.; Zhang, Y.; Gong, D.; Sun, X. Feature selection using bare-bones particle swarm optimization with mutual information. *Pattern Recognit.* **2021**, *112*, 107804. [\[CrossRef\]](#)
24. Sánchez-Maróño, N.; Alonso-Betanzos, A.; Tombilla-Sanromán, M. Filter Methods for Feature Selection—A Comparative Study. In *Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, Birmingham, UK, 16–19 December 2007*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 178–187.
25. Di Mauro, M.; Galatro, G.; Fortino, G.; Liotta, A. Supervised feature selection techniques in network intrusion detection: A critical review. *Eng. Appl. Artif. Intell.* **2021**, *101*, 104216. [\[CrossRef\]](#)
26. Cai, J.; Luo, J.; Wang, S.; Yang, S. Feature selection in machine learning: A new perspective. *Neurocomputing* **2018**, *300*, 70–79. [\[CrossRef\]](#)
27. González-Vidal, A.; Jiménez, F.; Gómez-Skarmeta, A.F. A methodology for energy multivariate time series forecasting in smart buildings based on feature selection. *Energy Build.* **2019**, *196*, 71–82. [\[CrossRef\]](#)
28. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Pascoal, C.; Oliveira, M.R.; Pacheco, A.; Valadas, R. Theoretical evaluation of feature selection methods based on mutual information. *Neurocomputing* **2017**, *226*, 168–181. [\[CrossRef\]](#)
30. Zhou, H.; Wang, X.; Zhu, R. Feature selection based on mutual information with correlation coefficient. *Appl. Intell.* **2022**, *52*, 5457–5474. [\[CrossRef\]](#)
31. Cheng, J.; Sun, J.; Yao, K.; Xu, M.; Cao, Y. A variable selection method based on mutual information and variance inflation factor. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2022**, *268*, 120652. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Sun, L.; Wang, T.; Ding, W.; Xu, J.; Lin, Y. Feature selection using Fisher score and multilabel neighborhood rough sets for multilabel classification. *Inf. Sci.* **2021**, *578*, 887–912. [\[CrossRef\]](#)
33. Sadeghian, Z.; Akbari, E.; Nematzadeh, H. A hybrid feature selection method based on information theory and binary butterfly optimization algorithm. *Eng. Appl. Artif. Intell.* **2021**, *97*, 104079. [\[CrossRef\]](#)
34. Liang, J.; Hou, L.; Luan, Z.; Huang, W. Feature Selection with Conditional Mutual Information Considering Feature Interaction. *Symmetry* **2019**, *11*, 858. [\[CrossRef\]](#)
35. Hammad, M.; Chelloug, S.A.; Alayed, W.; El-Latif, A.A.A. Optimizing Multimodal Scene Recognition through Mutual Information-Based Feature Selection in Deep Learning Models. *Appl. Sci.* **2023**, *13*, 11829. [\[CrossRef\]](#)
36. Li, K.; Fard, N. A Novel Nonparametric Feature Selection Approach Based on Mutual Information Transfer Network. *Entropy* **2022**, *24*, 1255. [\[CrossRef\]](#)
37. Li, J.; Ren, W.; Han, M. Mutual Information Variational Autoencoders and Its Application to Feature Extraction of Multivariate Time Series. *Int. J. Pattern Recognit. Artif. Intell.* **2022**, *36*, 2255005. [\[CrossRef\]](#)
38. Liu, J.; Lin, Y.; Ding, W.; Zhang, H.; Du, J. Fuzzy Mutual Information-Based Multilabel Feature Selection with Label Dependency and Streaming Labels. *IEEE Trans. Fuzzy Syst.* **2023**, *31*, 77–91. [\[CrossRef\]](#)

39. Gu, X.; Guo, J.; Xiao, L.; Li, C. Conditional mutual information-based feature selection algorithm for maximal relevance minimal redundancy. *Appl. Intell.* **2022**, *52*, 1436–1447. [[CrossRef](#)]
40. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **2008**, *26*, 303–304. [[CrossRef](#)] [[PubMed](#)]
41. Hasan, B.M.S.; Abdulazeez, A.M. A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *J. Soft Comput. Data Min.* **2021**, *2*, 20–30.
42. Schreiber, J.B. Issues and recommendations for exploratory factor analysis and principal component analysis. *Res. Soc. Adm. Pharm.* **2021**, *17*, 1004–1011. [[CrossRef](#)] [[PubMed](#)]
43. Verleysen, M.; François, D. The Curse of Dimensionality in Data Mining and Time Series Prediction. In *Proceedings of the International Work-Conference on Artificial Neural Networks, Barcelona, Spain, 8–10 June 2005*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 758–770.
44. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [[CrossRef](#)] [[PubMed](#)]
45. Saxena, A.; Goebel, K.; Simon, D.; Eklund, N. Damage propagation modeling for aircraft engine run-to-failure simulation. In *Proceedings of the 2008 International Conference on Prognostics and Health Management, Denver, CO, USA, 6–9 October 2008*; pp. 1–9.
46. Seaborn: Statistical Data Visualization. Available online: <https://seaborn.pydata.org/> (accessed on 17 February 2024).
47. Matplotlib: Visualization with Python. Available online: <https://matplotlib.org/> (accessed on 17 February 2024).
48. Huang, L.; Gong, L.; Chen, Y.; Li, D.; Zhu, G.; Ma, J. Trajectory Similarity Matching and Remaining Useful Life Prediction Based on Dynamic Time Warping. *Math. Probl. Eng.* **2022**, *2022*, 5344461. [[CrossRef](#)]
49. Hui, F.; Müller, S.; Welsh, A.H. Testing random effects in linear mixed models: Another look at the F-test (with discussion). *Aust. New Zealand J. Stat.* **2019**, *61*, 61–84. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.