International Journal of Innovative
Computing, Information and Control
Volume **19**, Number **4**, August **2023**

ICIC International ©2023 ISSN 1349-4198
pp. 1007–1021

DEMAND FORECASTING FOR RUSH REPAIR SPARE PARTS OF POWER EQUIPMENT USING FUZZY C-MEANS CLUSTERING AND THE FUZZY DECISION TREE

Yuanyuan Hao[1,2], Mingheng Tian[1,2], Yanxin Wang[3,*] and Minfang Huang[1,2]

[1]School of Economics and Management
[2]Beijing Key Laboratory of New Energy and Low-Carbon Development
North China Electric Power University
No. 2, Beinong Road, Beijing 102206, P. R. China
{ haoyy; mingheng_tian; huangmf }@ncepu.edu.cn

[3]Business College
Beijing Union University
No. 3, Yanjing Dongli, Beijing 100025, P. R. China
[*]Corresponding author: yanxin.wang@buu.edu.cn

Received November 2022; revised March 2023

Abstract. *Rush repair spare parts of power equipment are necessary for ensuring the normal operation of the critical equipment of power grid. Once the power equipment is damaged due to disasters and the rush repair spare parts are out of stock, the repercussions of the accident will be worsened, causing further harm to the power grid and potentially large losses to production and life. On the other hand, the excessive inventory of rush repair spare parts will cause a waste of cost. Therefore, accurate prediction of rush repair spare parts demand is very important. It is particularly difficult to forecast the demand for rush repair spare parts of power equipment due to the nature of their demand, which is usually highly uncertain, random, and with a small amount of historic data. Aiming to improve the scientificity and practicality, this paper proposes a demand forecasting method of rush repair spare parts by using the Fuzzy C-Means (FCM) clustering and the Fuzzy Decision Tree (FDT). At first, we analyze the characteristics of emergency events causing spare parts demand and the attributes of spare parts' demand data. Secondly, FCM is applied to dividing the historic demand data into clusters. Then, FDT is used to mine the correlation between the spare parts demand and the emergency events and forecast the demand for rush repair spare parts according to the best cluster of data, and thus realize to predict the demand even with a small data set. Finally, we verify the proposed method by a data set of emergency demand for rush repair spare parts during the last three years from a power company.*

Keywords: Rush repair spare parts, Demand forecasting, Fuzzy clustering, Fuzzy decision tree

1. **Introduction.** In recent years, the large-scale interruption of power grid caused by emergencies has resulted in a chain of social problems. If they are not properly responded to, the power grid is likely to experience a series of failures. Rush repair for power equipment is an important work in power emergency response management. Rush repair spare parts are those spare parts or components only used for rush repair, temporary accident defects and quick treatment of power equipment. If the rush repair spare parts of power equipment are out of stock, it will aggravate the occurrence of accidents, and even lead to the collapse of the power grid, which will have a significant impact on production and life. In order to make a reasonable allocation plan of rush repair spare parts in time

DOI: 10.24507/ijicic.19.04.1007

and improve the stability of the power system after the occurrence of power equipment emergencies, decision-makers need to accurately predict the demand for rush repair spare parts under different situations of emergencies. Especially with the rapid development of intelligent interconnection of power grid, there are various application scenarios of rush repair spare parts of power equipment. This requires that rush repair spare parts can respond to emergency requirements more quickly and accurately. Consequently, the demand forecasting for rush repair spare parts has become an important issue in advancing the efficiency and agility of rush repair spare parts management. It is beneficial to improve the ability of power enterprises to deal with accidents and save social costs.

The selection of rush repair spare parts of power equipment needs to comprehensively consider the scale and operation mode of the power grid, the inherent quality and operation defects of power equipment, the severity and harm of accidents involved, the compatibility of different types of equipment, climatic conditions and other external factors. Therefore, the features of rush repair spare parts of power equipment, such as the variety, big difference, scattered storage locations and difficult demand forecasting, bring great difficulties to the problem decision-making. In recent years, some power enterprises have put forward the spare parts storage strategy of centralized procurement and regional warehousing, which improves the spare parts utilization rate and reduces the capital occupation of regional inventory. However, there is still a lack of effective means to scientifically determine the inventory indicators such as the types of spare parts and inventory capacity in different areas. Based on the historical inventory data, some scholars have studied the prediction of spare parts demand and spare quantity by using big data technology and achieved certain achievements. However, due to the fact that the demand for rush repair spare parts occurs in the rush repair scenario, the amount of historical data is small, and the trained model is often not as good as expected. It is difficult to improve the accuracy between emergency and the demand for rush repair spare parts from two dimensions of time and space by forecasting only from the perspective of historical spare parts inventory data. Since many of the decision factors – natural, social, and human – involved in the pre-event, mid-event and post-event stages of an emergency are fuzzy and cannot be measured quantitatively, the resulting demand for rush repair spare parts and its influencing factors are also fuzzy.

In summary, there are two difficulties existing in the demand forecasting for rush repair spare parts of power equipment, small data and fuzzy demand information. Hence, some traditional spare parts demand forecasting methods are not suitable for forecasting the demand for the rush repair spare parts of power equipment. For example, traditional exponential smoothing or moving averages primarily rely on the latest regular data points, while rush repair spare parts show erratic and intermittent demand patterns. Croston's research [1] points out this problem and proposes an alternative, using exponential smoothing to predict intermittent demand time and demand size, respectively. However, Croston's research (called Croston method) and its modifications still rely on a certain amount of historical data. When the historical data of rush repair spare parts is few and the information is fuzzy, the prediction result is not satisfied. In recent years, many studies focus on improving the forecast by demand classification. Demand classification schemes categorize data and recommend the best-performing forecasting method for each demand type [2]. The study draws on the idea of demand classification, classifies the demand for rush repair spare parts, and mines the association rules between different categories and demand values, thus to realize the prediction under a small scale of data. Since there is no clear classification label for the data, the clustering method is adopted to classify the demand value of rush repair spare parts first, and then the association between the category and the demand value is mined using classification.

In recent years, fuzzy logic has been widely recognized as a successful approach for dealing with data uncertainty. Therefore, we use clustering and classification methods based on fuzzy logic to cope with the difficulty of fuzzy information. Aiming at the incomplete and fuzzy data of the demand for rush repair spare parts of power equipment, this paper presents a method based on fuzzy theory to predict the demand for rush repair spare parts of power equipment. First, the paper analyzes the factors affecting the demand for spare parts from the perspective of emergencies, and adopts Fuzzy C-Means (FCM) to cluster the demand data. According to the clustering results, we use the Fuzzy Decision Tree (FDT) to classify the demand for rush repair spare parts and obtain the corresponding fuzzy membership degree. In FCM method, the sample may belong to all the clusters with a certain fuzzy membership degree [3]. The same is true in FDT, where fuzzy membership can indicate the extent that a sample belongs to a certain class. Thus, we can predict the demand for rush repair spare parts by using a small amount of historical data. The rationale behind this method is that fuzzy logic and classification are used to solve the small amount and fuzziness of the demand data for rush repair spare parts, which should lead to better demand forecasts.

The rest of this paper is structured as follows. In Section 2, we review the literature on spare parts demand forecasting as well as the application of fuzzy decision methods. We propose a fuzzy-based methodology for demand forecasting in Section 3. In Section 4, we present the experimental results from spare parts data of power enterprises. Concluding remarks and future work are described in Section 5.

2. **Literature Review.** Forecasting spare parts demand is an important part of spare parts inventory management. A scientific spare parts inventory strategy will depend heavily on accurate demand forecasting. However, the demand for spare parts is commonly random and uncertain. Especially for rush repair spare parts, demand prediction is extremely difficult. Hence, the research results of traditional periodical sequence prediction [4] and spare parts prediction [5] in general industries cannot be directly applied to rush repair spare parts demand prediction. They prefer to incorporate probability and statistics theory into the demand prediction, which describes the demand distribution regularity of spare parts by using common probability distribution functions, such as negative exponential distribution, normal distribution and Poisson distribution [6].

For stochastic demand prediction of spare parts, the main research results are as follows. Croston [1] previously proposed the research of intermittent demand forecasting problem. On this basis, Nikolaos [7] proposed an improved neural network model to solve this problem. According to a recent study by Kalaya et al. [8], they compared the Croston method with TSB (the Teunter, Syntetos, and Babai's) method, and predicted unstable demand by combining average demand interval and average demand. In addition, for intermittent demand with a large number of 0 values, some researchers estimated the probability distribution of demand using the bootstrapping method [9]. Dombi et al. [10] used knowledge engineering-related theories to conduct a long-term prediction of the typical demand for electronic spare parts from the perspective of the procurement life cycle of electronic spare parts. Some scholars have proposed data-centered research ideas to improve the accuracy and scientificity of demand forecasting. For example, Jiang et al. [11] proposed an adaptive autoregressive Support Vector Machine (SVM) model to predict unstable spare parts demand, automatically generated relevant attributes from historical demand time series data, and used the data-driven method to identify attribute dimensions, nonlinear kernel functions, and key model parameters.

The preceding findings are primarily concerned with the intermittent demand for spare parts. From Croston method to data-centered method, prediction accuracy has been improved while the dependence on the historical data has increased. However, predictions on the rush repair spare parts of power equipment suffer from two difficulties. First, due to the special nature of rush repair spare parts of power equipment, there are few historical data with many 0 values, making it challenging to estimate with such a limited quantity of historical data. Second, the demand for rush repair spare parts of power equipment is highly uncertain, which brings difficulty to rule mining and demand prediction. To better solve these difficulties, fuzzy theory has been utilized in demand forecasting. Tsaur and Kuo [12] believed that fuzzy time series is capable of dealing with very small data sets and does not require linear assumptions. Fuzzy theory can cope with uncertainty effectively. Cheng et al. [13] introduced a new multi-attribute fuzzy time series method based on fuzzy clustering that integrated fuzzy clustering into the fuzzy time series process and enabled the processing of multiple attributes. Therefore, we use the advantages of fuzzy theory to forecast the demand for rush repair spare parts based on the fuzzy clustering prediction model. Given that the data in the same cluster show similar demand patterns, the cluster-based prediction model outperforms the prediction model based on the complete dataset [14]. Thus, FCM is utilized to cluster data of rush repair spare parts demand. After clustering the data, the classification method is selected to predict the classification. Because of its ease of interpretation, the decision tree is often used in demand forecasting [15]. To integrate the decision tree and fuzzy theory, we apply FDT to predicting the classification and corresponding fuzzy membership degree of spare parts to realizing the demand forecasting for rush repair spare parts of power equipment.

This study incorporates the fuzzy theory into the demand forecasting for rush repair spare parts. We use FCM and FDT to develop a prediction model, which realizes the prediction with a small scale of data set. To the best of our knowledge, previously no results have been proposed that utilize FCM and FDT to forecast the demand for rush repair spare parts of power equipment, as we do in this study.

3. **Proposed Methodology.**

3.1. **The procedure of demand forecasting for rush repair spare parts.**

According to the real data of emergency resources scheduling of power enterprises and the characteristics of power emergencies, the paper presents a method based on FCM and FDT to forecast the demand for rush repair spare parts. It considers the impact of power emergencies on the demand for rush repair spare parts of power equipment and uses the random forest model to analyze the influencing factors. To solve the two difficulties of less data and fuzziness in the demand for rush repair spare parts, the clustering and classification method based on fuzzy logic is adopted. FCM clustering is used first and then FDT classification is used to achieve higher accuracy under less data. The flow chart of the proposed forecasting is shown in Figure 1, which mainly consists of the following five steps.

Step 1: Collect historical spare parts data. Gather and sort out the historical demand data of rush repair spare parts of power equipment caused by power emergencies. Each data includes a description of the power emergency and the resulting demand for rush repair spare parts.

Step 2: Determine the factors affecting the demand for rush repair spare parts of power equipment. Adopt the random forest model to analyze the correlation of influencing factors, and identify six related factors (including population, equipment state, weather, wind level, emergency level and inventory status) associated with power emergencies as attributes of data clustering of rush repair spare parts demand.

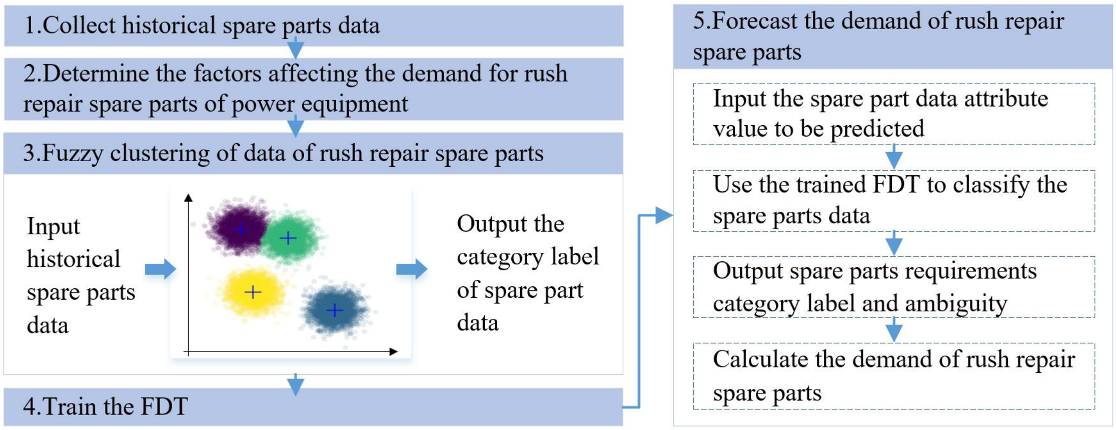

FIGURE 1. The flow chart of the proposed demand forecasting method using FCM and FDT

Step 3: Fuzzy clustering of data of rush repair spare parts. According to six data attributes, use the fuzzy clustering method based on FCM to cluster the data of rush repair spare parts of power equipment, and add a category label for each data.

Step 4: Train the FDT. With the demand data and cluster label as input, the FDT model is built and trained.

Step 5: Forecast the demand for rush repair spare parts. Input the data attribute value of rush repair spare parts caused by power emergency, classify it into a spare parts demand category and its corresponding fuzzy membership degree by the FDT, so as to forecast the corresponding demand.

In the following, Steps 2-5 in Figure 1 will be elaborated in detail.

3.2. **Factors affecting the demand for rush repair spare parts of power equipment.** Power emergencies include natural disasters, social emergencies, energy shortages, equipment defects, engineering construction accidents, etc. All of them may impact the power system infrastructure, leading to equipment failures or operational accidents, and then generate the demand for rush repair spare parts. These emergencies are usually sudden, urgent, with serious implications, non-procedural decision-making, etc. Ineffective response will result in the effect of hazard diffusion. Through interviewing spare parts manager in grid enterprises and referring to the literature of power spare parts management [16], the following factors affecting the demand for rush repair spare parts of power equipment are summarized from the perspective of power emergencies.

1) Category and extent of emergencies. Power system emergencies can be divided into natural emergencies and man-made emergencies. Depending on the kind of emergency, various rush repair spare parts will be required. In this paper, the categories of power emergencies mainly include flood, typhoon, snow disaster, explosion accident, emergency rush repair, debris flow, etc. In addition, the severity of emergencies will also have an impact on the demand for spare parts. We choose the population affected by emergencies and the emergency level to measure the severity of different emergencies, so as to mine the correlation between emergencies and the demand for rush repair spare parts of power equipment.

2) The environment of emergencies. The environment of emergencies has meteorological environment, geological environment, geomorphic environment, power equipment environment, etc. We primarily consider the meteorological environment and select the characteristics of weather state and wind level to analyze the influence of meteorological

environment on the demand for spare parts. Another significant factor influencing the demand for spare parts is the geographical characteristic of emergencies. For example, similar emergencies occuring in different locations such as downtown area, residential area, suburb and field, will result in different demands for rush repair spare parts. Therefore, it is necessary to comprehensively consider the geographical dimension of emergencies. In our study, the geographical environment of emergencies is represented by the location characteristics of emergencies.

3) Power equipment status. Not all emergencies will generate the demand for spare parts. The state of power equipment and the inventory status of spare parts also have a great impact on the production of spare parts demand. For example, compared with the power equipment in normal operation, the power equipment scheduled for maintenance is more susceptible to the impact of emergencies, and more rush repair spare parts are needed. When spare parts inventory is insufficient to cope with emergencies, the demand for rush repair spare parts will also increase.

When forecasting the demand for rush repair spare parts of power equipment, it is necessary to consider not only the state of power equipment and the inventory status of rush repair spare parts, but also the related factors of emergencies. In order to forecast future demand for rush repair spare parts, we first choose the initial relevant influencing factors, which include the emergency events, the population affected by emergencies, the emergency level, the location of emergency, the weather state, the wind level, the state of power equipment and the inventory status of spare parts.

To prevent factors with weak correlation from increasing the complexity of model computation, it is also necessary to analyze the above characteristics in practice. Identify which characteristics have a stronger correlation with the demand for rush repair spare parts of power equipment, that is, so as to further analyze the factors affecting the demand for rush repair spare parts and establish a forecasting model that is more accurate. In this paper, we use the random forest model to compare the correlation degree of spare parts data characteristics. Random forest is an ensemble learning algorithm based on the decision tree [17], which can be used for both characteristic selection and prediction problems. The random forest model will measure the importance of variables on data attributes after data fitting. In the sklearn toolkit, the feature_importances_parameter of the random forest model can return an array object. The elements in the array represent the importance of the training attribute column obtained after random forest model fitting. The higher the value of the attribute column, the more important it is to the accuracy of the prediction.

Figure 2 shows the importance of factors obtained by random forest model fitting. As can be seen from the figure, weather state, population affected by emergencies, emergency level, wind scale, inventory status of spare parts and power equipment state are strongly correlated with spare parts demand. Therefore, we select these 6 data attributes for subsequent clustering and forecasting.

3.3. **Fuzzy clustering of data of rush repair spare parts.** The purpose of this step is to perform fuzzy clustering according to the demand data and distinguish different demand categories of rush repair spare parts. The clustering algorithm is a kind of unsupervised learning. The common clustering methods include K-means clustering and FCM clustering. Traditional clustering division is hard clustering, which strictly divides the data to be processed into a certain class. For example, K-means algorithm is a classical hard clustering algorithm, and many kinds of literature use K-means to cluster and predict the demand [18]. However, most samples in real life do not have strict attribute division, and the hard clustering algorithm cannot accurately capture the actual relationship

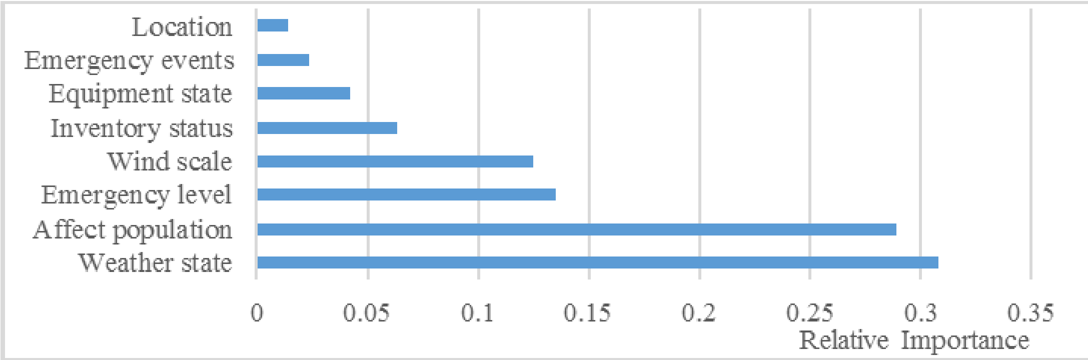

FIGURE 2. The importance of characteristics by random forest

between samples and categories, which is not suitable for the prediction of rush repair spare parts with randomness and less data. Fuzzy clustering combines the principle of fuzzy mathematics with soft cluster analysis [19], which describes the uncertainty for the sample category and also can objectively reflect the relationship in the real world. Since the demand for rush repair spare parts of power equipment is random and there is only a small scale of historical data, the fuzzy clustering method is adopted in this paper to soft divide different spare parts demand categories and add category labels to the data based on the input historical spare parts data.

Among the fuzzy clustering algorithms, the most widely used classical algorithm is FCM algorithm. FCM is a soft clustering algorithm based on the objective function. Its clustering result is a membership matrix, which can not only cluster data samples but also record the fuzzy membership degree of each sample belonging to each category. FCM provides a better description of covered data, and the distribution and processing effect of boundary points between clusters is close to the real relationship between samples and categories. Besides, its clustering result is more stable than that of C-means. The main principles of FCM algorithm are as follows.

Assume the input data have n samples described by x_j $(j = 1, 2, \ldots, n)$, respectively. We classify the data into c classes. And c category centers are generated, which are represented as v_i $(i = 1, 2, 3, \ldots, c)$, respectively. The degree that each sample j belongs to cluster i is called degree of membership, represented as u_{ij}. The objective function and the constraints of FCM are defined as follows:

$$J(u, v) = \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^m \|x_j - v_i\|^2 \tag{1}$$

$$\sum_{i=1}^{c} u_{ij} = 1, \quad j = 1, 2, \ldots, n \tag{2}$$

Equation (1) is the objective function of FCM. It is the sum of the product of the membership u_{ij} of the sample and the distance between the sample and the center of each class. Herein m is the fuzzy weight index. The smaller the objective function J is, the better the clustering algorithm will be. Equation (2) is the constraint condition, indicating that the sum of membership degrees of the samples belonging to various classes is 1, in which $u_{ij} \in (0, 1)$. FCM algorithm is transformed into the optimization problem of finding the membership degree u_{ij} and clustering center v_i which minimize J under the constraints. We use the Lagrange multiplier method to solve the problem and fuse the constraints into the objective function. To determine the values of the two parameters, calculate the partial derivatives of the membership degree u_{ij} and the clustering center v_i, respectively. When the objective function is the minimum value, the values of membership degree u_{ij} and clustering center v_i are derived as follows:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{x_j - v_i}{x_j - v_k} \right)^{2/(m-1)}}, \quad (i = 1, 2, \ldots, c;\ j = 1, 2, \ldots, n) \tag{3}$$

$$v_i = \frac{\sum_{j=1}^{n} \left(u_{ij}^m x_j \right)}{\sum_{j=1}^{n} u_{ij}^m} = \sum_{j=1}^{n} \frac{u_{ij}^m}{\sum_{j=1}^{n} u_{ij}^m} x_j, \quad (i = 1, 2, \ldots, c) \tag{4}$$

It can be seen from Equations (3) and (4) that the calculation of u_{ij} is related to the value of v_i, and the calculation of v_i is related to the value of u_{ij}; the two are related to each other. Therefore, we can first assign a random value to u_{ij}; on this basis v_i can be obtained. From there, we can calculate a new u_{ij}, etc. Iteratively updating, the objective function J moves closer to the minimum value. When J is no longer numbered or the specified number of iterations is reached, the algorithm ends. The specific algorithm steps are as follows:

1) Determine the number of categories c, the number of iterations and the value of fuzzy weight index m;

2) Initialize the membership matrix U;

3) Calculate clustering center v_i according to the membership matrix U;

4) Calculate the objective function J;

5) Calculate the membership matrix U according to the clustering center v_i obtained in (3), and then the obtained U is used to calculate v_i, and so on, until the number of iterations is reached;

6) After the end of the iteration, we have a final membership matrix U. Each sample has a membership degree u_{ij} for each category, and the category corresponding to the maximum u_{ij} is the class to which the sample belongs. Add category labels to the samples according to the maximum membership degree.

As can be seen from the above steps, how to determine the number of clusters is the key of FCM algorithm. The silhouette coefficients provide a way to assess the initial value of the number of clusters, and how well a sample fits into its own cluster. The silhouette coefficients will be used to evaluate the effect of clustering and determine the suitable number of clusters. The silhouette coefficient $s(j)$ of sample x_j $(j = 1, 2, \ldots, n)$ is calculated as follows [20]:

$$s(j) = \frac{b(j) - a(j)}{\max\{a(j), b(j)\}} = \begin{cases} 1 - a(j)/b(j), & \text{if } a(j) < b(j) \\ 0, & \text{if } a(j) = b(j) \\ b(j)/a(j) - 1, & \text{if } a(j) > b(j) \end{cases} \tag{5}$$

Herein $a(j)$ is the average dissimilarity distance between the sample j and its assigned cluster, and $b(j)$ is the minimum of the average dissimilarity of the sample j to all members of other clusters except for its assigned cluster. The silhouette coefficient ranges from -1 to 1. The closer its value is to 1, the more suitable sample j is for this class; Conversely, the closer it is to -1, the less suitable sample j is for this class and the more it should be assigned to other clusters. When $s(j)$ approaches 0, sample j is on or very close to the boundary between its own and the neighbouring cluster. The average silhouette coefficient for all samples can be used to compare the clustering effect for different clustering numbers.

3.4. **Training the FDT.** The main purpose of this part is to establish a classification model, match the spare parts demand category label for the input data, and then obtain the corresponding demand. The common classification algorithms include the decision tree method, neural network method, Bayesian method, K-nearest neighbor method, support vector machine method, etc. The decision tree method is widely used because of its strong interpretation and ease of use. The traditional decision tree method mainly includes the ID3 algorithm, which selects the attribute with maximum information divergence as the extended attribute based on information entropy. The classification nodes of the ID3 algorithm are traditional subsets, but their exact description features enable them unsuitable for fuzzy knowledge acquisition in an uncertain environment. In fuzzy theory, the FDT is an extension of the traditional decision tree. Each classification node of FDT is a fuzzy subset, which can better reflect the uncertainty in real world.

Due to the small amount of historical demand data and strong uncertainty of emergencies in this paper, it is necessary to combine fuzzy set theory and decision tree. In order to analyze the relationship between emergencies and spare parts demand based on the aforementioned FCM, we adopt FDT to establish a classification model. We input the data attributes of sample spare part to the classification model, forecast its demand category and the corresponding fuzzy membership degree, and then the demand of the sample spare part is calculated. The main steps of FDT adopted in this paper are as follows.

1) Data fuzzification. In order to fuzzify the attribute data with a continuous value, first divide each input and output variable of the given sample data into fuzzy sets (also known as fuzzy regions), and then define a function for each fuzzy set to describe the degree of the elements in the data set belonging to the data set (called membership degree), which is called membership degree function. The commonly used membership functions are the Gaussian function, ladder function, trigonometric function, etc. For example, divide the input factors of population affected by emergencies into three fuzzy sets, A_1, A_2 and A_3, which respectively represent the three categories of low, medium and high. Table 1 lists the values of membership degrees of different fuzzy sets after using the Gaussian function as the membership function to complete the fuzzification.

Table 1. Examples of fuzzy set partitioning

Variable number	The population affected by emergencies (ten thousand people)	The membership degree of A_1	The membership degree of A_2	The membership degree of A_3
x_1	175.71	0.022619	0.923857	0.147402
x_2	77.50	0.480024	0.520201	0.002202
x_3	242.40	0.000733	0.352246	0.660831

2) Extract the classification knowledge from the training set and establish FDT. Based on the information entropy, select the attribute that minimizes the degree of class confusion as the extended attribute. According to the input training set, FDT model can be trained. The FDT algorithm regards decision tree nodes as fuzzy subsets in space, so it is necessary to adjust the formula of information entropy and information divergence [21].

3) Transform FDT path into fuzzy rules. Each path from the root to the leaf of traditional decision tree can be transformed into a clear rule, while the path of FDT can be transformed into a fuzzy rule. The partition of data space by fuzzy rules is fuzzy, and its boundary is determined by the fuzzy inference mechanism. According to FDT formed by training the test set, we can obtain the corresponding fuzzy rules. After receiving new data to forecast, the model classifies it according to the corresponding rules and outputs its membership degree based on the attribute value of the input data.

3.5. **Forecasting the demand for rush repair spare parts.** The purpose of this part is to forecast the final demand for rush repair spare parts of power equipment. Based on the classification model established in the preceding steps, input the attribute value of spare parts data to be forecasted (including six attributes including weather state, the population affected by emergencies, emergency level, wind level, spare parts inventory state and power equipment state). Use the trained FDT to classify the spare parts data. According to the fuzzy rules in the model, divide the spare parts demand category for each data, and the category label of rush repair spare parts of power equipment and its corresponding fuzzy membership degree are output. Finally, calculate the demand D_k of rush repair spare parts of power equipment according to the following formula:

$$D_k = \sum_{i=1}^{c} P_{ki} \cdot C_i \tag{6}$$

In Equation (6), c is the number of categories, C_i is the corresponding demand value of different category centers, and P_{ki} represents the membership degree of the article k data divided into the i class.

4. Computational Experiments.

4.1. **Data description.** According to the allocation data of rush repair spare parts of a power grid company in the past three years, there are 87 pieces of data in total. Combining the weather, influencing population and other information, a numerical example is constructed to verify the efficiency of the forecasting method proposed in this paper. Overhead insulated wire is a crucial spare part for the maintenance and safety of power grid lines, and it also plays a significant role in power grid emergency repairs. Therefore, this study takes the commonly used spare parts of overhead insulated wire as an example to demonstrate the feasibility of the forecasting method. Figure 3 shows the demand distribution curve of overhead wires. It can be seen that due to the small amount of data and strong randomness of demand, it is difficult to establish a model by traditional forecasting method. In this paper, we divide the data into the training data set and the testing data set to measure the performance of the forecasting scheme, 60% of which is used as the training set (52 pieces of data), and the rest is used as the testing set.

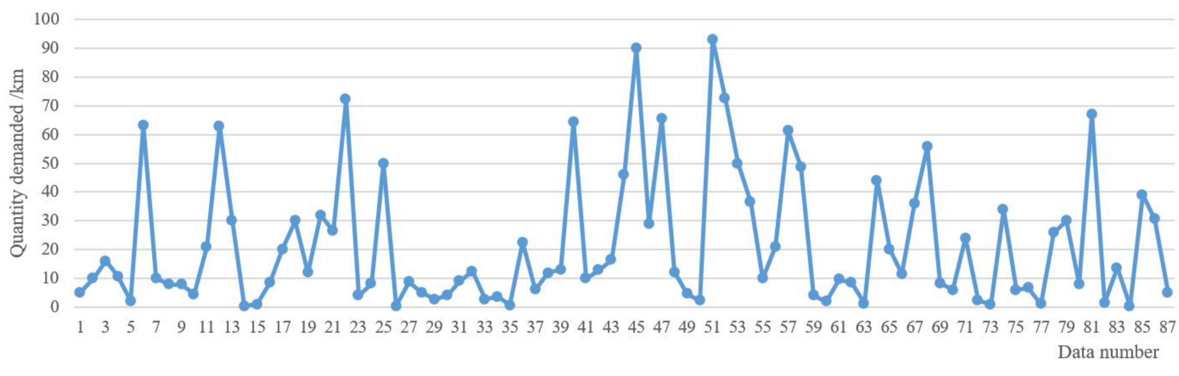

FIGURE 3. Demand curve of rush repair spare parts

The input attributes of the dataset include population affected by emergencies, power equipment state, weather state, wind scale, emergency level and inventory status of spare parts. The specific description of each attribute is shown in Table 2.

This paper uses Python to realize the demand forecasting method and it is run on a notebook with Intel(R) Core(TM) i5-10400 CPU @ 2.90GHz and 8G memory. Since the

Table 2. Attribute description

Attribute	The numerical description	Unit	Value range
Population affected by emergencies	Represent the number of people affected by emergencies.	Ten thousand people	$[0.3, 300.4]$
Power equipment state	1 indicates that the equipment is in good condition. 0 indicates that the equipment is in poor condition.	–	$\{0, 1\}$
Weather state	Represent the weather when the core scenario occurs.	–	Category-type data
Wind scale	Represent the wind level of the day	Level	$[1, 6]$
Emergency level	Represent the level of emergency response that causes the demand for spare parts	–	Category-type data
Inventory status of spare parts	1 indicates that the spare parts are in sufficient stock. 0 indicates that the spare parts are in insufficient stock.	–	$\{0, 1\}$

purpose of this paper is to enhance the forecasting accuracy, we use the root mean square error (RMSE) as the evaluation criterion of the forecasting performance, and we determine the number of clustering classifications with the minimum RMSE as the objective. RMSE represents the deviation between the forecasting value and the real value, which is defined as Formula (7).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - d_i)^2} \tag{7}$$

4.2. **Experimental results.** Firstly, we use the average silhouette coefficient to determine the initial clustering number. The silhouette coefficients of clustering numbers from 2 to 10 are shown in Table 3.

Table 3. The silhouette coefficients of different clustering numbers

	The number of clusters (c)								
	2	3	4	5	6	7	8	9	10
Silhouette coefficients	0.69	0.69	0.69	0.62	0.59	0.63	0.57	0.51	0.53

The value of silhouette coefficient near 1 indicates that the sample accordingly fits well into its own cluster. However, too little classification will easily lead to a large gap in demand prediction. Thus, we select the first six clustering numbers with large silhouette coefficients and comprehensively judge the optimal clustering numbers by comparing the prediction errors. We use FCM to divide the training data into 2, 3, 4, 5, 6 and 7 classes, respectively, which correspond to different categories of spare parts demand. Input the class data into FDT to train the classification model. Finally, input the test data into the established FDT model to obtain the category label and membership degree of spare parts, and then calculate the demand for rush repair spare parts of power equipment. In the power industry, the Support Vector Regression (SVR) is commonly used to predict power demand [22]. In this paper, SVR is used as a comparison model, and establish the SVR model and the fuzzy decision model proposed in this paper respectively based on the same training set. The comparison of the effects of different categories of the fuzzy decision model and SVR model is shown in Table 4.

TABLE 4. Comparison of prediction effects of different models

Error index \ Prediction method	SVR	Fuzzy decision method					
		2 classes	3 classes	4 classes	5 classes	6 classes	7 classes
Precision	0.11	0.96	0.97	0.97	0.98	0.93	0.96
RMSE	20.04	20.14	17.88	16.82	15.30	10.27	12.38

As shown in Table 4, due to the small amount of data, the fuzzy decision method outperforms the SVR method in most categories, with higher accuracy, smaller RMSE and relatively higher prediction accuracy (except 2 classes). For the fuzzy decision method proposed in this paper, the prediction effect is slightly different when the number of classes is different. When the number of classes is 5, the classification effect of FDT is the best, with an accuracy of 0.98 but a significant RMSE. When the number of classes is 6, FDT has a good classification effect, with an accuracy of 0.93 and a minimum RMSE. In this study, the RMSE is taken as the evaluation criterion of the comprehensive prediction model, and the FCM cluster number is determined to be 6. The FDT is then used to forecast the classification, and Figure 4 shows a comparison of the predicted demand based on the cluster number of 6 classes, the predicted demand of the SVR method, and the actual demand. When there is a limited amount of data, the red line corresponding to the fuzzy decision method is relatively closer to the actual demand, which indicates that the demand forecasting method proposed in this paper can predict more effectively.

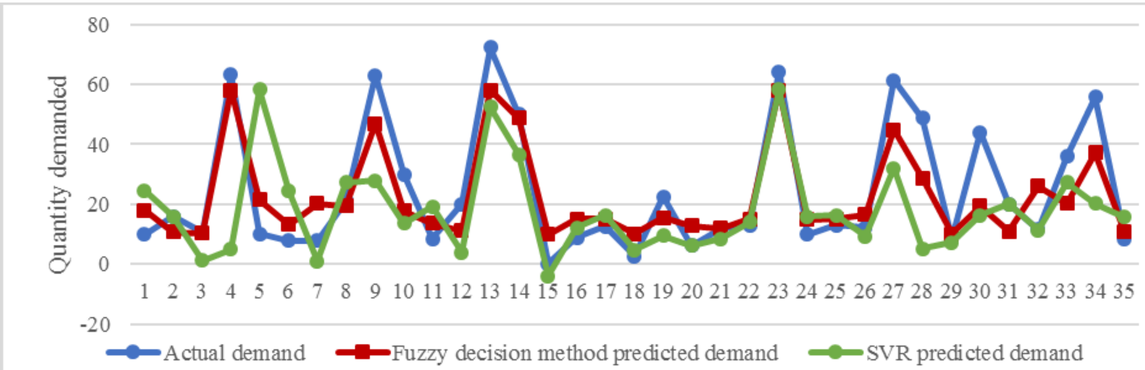

FIGURE 4. The proposed forecasting scheme based on fuzzy decision method

5. **Conclusions.** With the objective of overcoming the nature of uncertainty, randomness and a small scale of data, the paper establishes a demand predicting method for rush repair spare parts of power equipment by using the Fuzzy C-Means (FCM) clustering and the Fuzzy Decision Tree (FDT). It introduces the fuzzy theory into rush repair spare parts demand forecasting, which provides a new idea for uncertain demand forecasting and improves the scientificity and practicability of demand forecasting model. It is beneficial for decision makers to determine the corresponding spare parts demand in time from different emergencies. Furthermore, it facilitates the emergency response from a predictive-response mode to a scenario-response mode. The achievements and innovations achieved can be summarized as follows. 1) For the uncertainty and fuzziness of the demand for rush repair spare parts, fuzzy theory was introduced to establish the relationship between the emergency events and the spare parts demand, which could convert the uncertainty of spare parts demand from qualitative to quantitative and enhance the practicality of the established demand prediction model. 2) Using the Fuzzy C-Means with the Fuzzy

Decision Tree, the characteristics of different emergency events are considered, and the number of clusters is determined which decides the prediction accuracy of the model. The fuzzy clustering and classification directly contribute to the prediction effect of the model, thus realizing the prediction for rush repair spare parts of power equipment even with a small amount of historic data.

Future studies will include the following: 1) According to the feature of a small amount of data, establish an online prediction model, and optimize the prediction model constantly with the real-time input data flow; 2) Consider demand forecasting of spare parts based on the inventory of spare parts and real-time operating data of power equipment, and further study the distribution and storage for rush repair spare parts of power equipment; 3) The emergency events will be further studied and incorporated to the demand forecasting for spare parts. The simulation drill and feedback evaluation of the emergency events will be conducted to build the emergency response decision-making system and to improve the ability of rapid recovery from the crisis brought by the emergency events.

Acknowledgment. This work was supported by the program of the National Natural Science Foundation of China (72071078). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] J. D. Croston, Forecasting and stock control for intermittent demands, *Operational Research Quarterly*, vol.23, no.3, pp.289-303, 1972.

[2] E. Pine, L. Turrini and J. Meissner, Intermittent demand forecasting for spare parts: A critical review, *Omega*, vol.105, no.1, 102513, 2021.

[3] M. Ferreira and F. Carvalho, Kernel fuzzy C-means with automatic variable weighting, *Fuzzy Sets & Systems*, vol.237, pp.1-46, 2014.

[4] F. T. Liljana, M. Blaz and T. Aleš, Demand forecasting with four-parameter exponential smoothing, *International Journal of Production Economics*, vol.181, pp.162-173, 2016.

[5] M. Akpinar and N. Yumusak, Day-ahead natural gas forecasting using nonseasonal exponential smoothing methods, *2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, pp.1-4, 2017.

[6] A. J. Jiang, K. L. Tam, X. Guo and Y. Zhang, A new approach to forecasting intermittent demand based on the mixed zero-truncated Poisson model, *Journal of Forecasting*, vol.39, no.1, pp.69-83, 2020.

[7] K. Nikolaos, Intermittent demand forecasts with neural networks, *International Journal of Production Economics*, vol.143, no.1, pp.198-206, 2013.

[8] P. Kalaya, P. Termsuksawad and T. Wasusri, Forecasting lumpy demand of planning inventory: The case of community hospitals in Thailand, *2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp.686-690, 2019.

[9] C. Zhou and S. Viswanathan, Comparison of a new bootstrapping method with parametric approaches for safety stock determination in service parts inventory systems, *International Journal of Production Economics*, vol.133, no.1, pp.481-485, 2011.

[10] J. Dombi, T. Jónás and Z. E. Tóth, Modeling and long-term forecasting demand in spare parts logistics businesses, *International Journal of Production Economics*, vol.201, pp.1-17, 2018.

[11] P. Jiang, X. Liu and Y. Huang, An adaptive data-driven model for erratic demand forecasting, *IFAC PapersOnLine*, vol.50, no.1, pp.9020-9025, 2017.

[12] R. C. Tsaur and T. C. Kuo, The adaptive fuzzy time series model with an application to Taiwan's tourism demand, *Expert Systems with Applications*, vol.38, no.8, pp.9164-9171, 2011.

[13] C. H. Cheng, G. W. Cheng and J. W. Wang, Multi-attribute fuzzy time series method based on fuzzy clustering, *Expert Systems with Applications*, vol.34, no.2, pp.1235-1242, 2008.

[14] C.-J. Lu and L.-J. Kao, A clustering-based sales forecasting scheme by using extreme learning machine and ensembling linkage methods with applications to computer server, *Engineering Applications of Artificial Intelligence: The International Journal of Intelligent Real-Time Automation*, vol.55, pp.231-238, 2016.

[15] P. K. Bala, Decision tree based demand forecasts for improving inventory performance, *2010 IEEE International Conference on Industrial Engineering and Engineering Management*, Macao, China, pp.1926-1930, 2010.
[16] Y. Wang, C. Chen, J. Wang and R. Baldick, Research on resilience of power systems under natural disasters – A review, *IEEE Transactions on Power Systems*, vol.31, no.2, pp.1604-1613, 2016.
[17] A. Liaw and M. Wiener, Classification and regression by randomForest, *R. News*, vol.12, no.2, pp.18-22, 2002.
[18] L. Sun, X. Xing, Y. Zhou and X. Hu, Demand forecasting for petrol products in gas stations using clustering and decision tree, *Journal of Advanced Computatioanl Intelligence & Intelligent Informatics*, vol.22, no.3, pp.387-393, 2018.
[19] W. Wei, X. Zhang and L. Yang, Full-cycle state evaluation of S700K switch machine based on residual network and fuzzy clustering, *International Journal of Innovative Computing, Information and Control*, vol.18, no.4, pp.1203-1216, 2022.
[20] C. Barth and C. Becker, Machine learning classification for field distributions of photonic modes, *Communications Physics*, vol.1, no.58, 2018.
[21] B. Jeng, Y.-M. Jeng and T.-P. Liang, FILM: A fuzzy inductive learning method for automated knowledge acquisition, *Decision Support Systems*, vol.21, no.2, pp.61-73, 1997.
[22] Q. Guo, Y. Feng, X. Sun and L. Zhang, Power demand forecasting and application based on SVR, *Procedia Computer Science*, vol.122, pp.269-275, 2017.

Author Biography

Yuanyuan Hao received the B.Eng. degree in electronic commerce from the Tianjin University of Commerce, Tianjin, China, in 2020. She is currently pursuing the Master's degree in North China Electric Power University, Beijing, China. Her main research interests include e-commerce logistics management and emergency rush spare parts management.

Mingheng Tian received the B.Eng. degree in management science (engineering project management) from the Henan Agricultural University, Zhengzhou, China, in 2021. He is currently pursuing the Master's degree in North China Electric Power University, Beijing, China. His main research interests include e-commerce logistics management and emergency rush spare parts management.

Yanxin Wang received the B.Sc. degree in technological economics, the M.Sc. degree in technological economics and management and the Ph.D. degree in management science and engineering from Harbin Institute of Technology, China in 1999, 2001 and 2011, separately.

Dr. Wang is currently a full-time associate professor at the Business College of Beijing Union University, China. His research interests include data analysis and intelligent decision. He has published over 30 papers in journals and conferences.

Minfang Huang received the B.Sc. degree and the M.Sc. degree in system engineering from Harbin Institute of Technology, China in 1999 and 2001, separately; the Ph.D. degree in e-commerce and logistics management from Dalian University of Technology, China, 2009.

Dr. Huang is currently a full-time associate professor at the School of Economics and Management, North China Electric Power University, China. Her research interests include logistics management and data-driven optimization method. She has published over 40 papers in journals and conferences.