

# A recognition model of survival situations for survivable systems

Zhao Guosheng<sup>1</sup> Shao Zihao<sup>1</sup> Wang Jian<sup>2</sup> Li Yingmei<sup>1</sup>

(<sup>1</sup>College of Computer Science and Technology, Harbin Normal University, Harbin 150025, China)

(<sup>2</sup>School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China)

**Abstract:** Due to the lack of pre-recognition and post-prediction in existing survivable systems, a recognition model of survival situations for survivable systems is proposed. First, the survival situation data is clustered into several survival clusters with different service levels based on the Ward method, and then the survival clusters are classified and recognized by means of the error-eliminating decision-making method, which can realize the pre-recognition of the system's survival situation. Secondly, the differentiated survival situation data is used to generate stationary predicting sequences. The autoregressive integrated moving average (ARIMA) model is constructed, and the stability, randomness and reversibility index of the model are verified by the autocorrelation function and partial auto-correlation function. Finally, fuzzy particles and the residual correction for the support vector regression (SVR) model are applied to realize the post-prediction of the survival situation. Compared with traditional decision-making methods, the simulation experiments show that the pre-recognition module can not only cluster the survival situation data and identify the service ranks, but can also recognize the illegal users. According to the prediction of abnormal situations numbers and residual correction, the model can effectively realize the post-prediction of survival situations for survivable systems.

**Key words:** survivability; recognition; fuzzy particle; residual correction

**DOI:** 10.3969/j.issn.1003-7985.2018.03.002

Survivability research is a hot issue in the next generation of cyberspace security. According to the definition of Ref. [1], the research of survivability is divided into three aspects, which are resistance, recognition and recovery (3R attributes). In reality, systems are always being invaded and destroyed by different degrees, and the failures of systems are inevitable. How to use the current survival situation data to predict the future survival situations has become an urgent problem. Therefore, we focus

on the pre-recognition and the post-prediction of survival situations for survivable systems.

At present, most of the existing literature focuses on the recovery and resistance research. Yaghlane et al.<sup>[2]</sup> introduced the concept of system survivability under attack in analogy with system reliability. Zhao et al.<sup>[3]</sup> described the survivability of RSCN by the fault repair and anti-failure technology in network systems. Raja et al.<sup>[4]</sup> proposed a multi-dimensional measurement method for the survivability of open source software. Most research on the recognition of survivable systems often focuses on perception. It is still in the initial stage and related literature is insufficient. Zhao et al.<sup>[5]</sup> proposed an autonomous recognition unit from the survivability of autonomous recognition. They focused on the recognition of the detection parameter definition, the autonomous recognition model and the threshold variable method. The recognition monitoring mechanism can improve the ability of self-cognition and the capacity of services, but it cannot predict future survival situations. Wang et al.<sup>[6]</sup> mainly studied the hierarchical cognitive model, the self-management model of cognitive unit and the transformation process of survival state. The proposed method improves survivability by enhancing self-recovery ability, but lacks the processing of residual data and prediction of future survival situations. Dharmaraja et al.<sup>[7]</sup> applied the concept of survivability into vehicle ad hoc network. This method created a mobile network to facilitate communication between vehicles. It can ensure the safety of roads and reduce security risks.

Therefore, existing literature on the survival situations for survivable systems can be divided into two fields. The first is the definition of survivability<sup>[2-4]</sup>. It mainly focuses on the studies of resistance and recovery. Most research belongs to pre-recognition research. The second is the application of survivability<sup>[5-7]</sup>. There is little literature in the field of recognition for survivable systems. Most research focuses on the Internet of things and artificial intelligence, such as context awareness<sup>[8]</sup>, compressed sensing, etc. On the whole, there is very little literature concerning the recognition techniques, and the recognition not only includes pre-recognition of the current survival situation, but also includes the post-prediction of the future survival situation.

## 1 Pre-Recognition

The Ward method<sup>[9]</sup> can cluster the current situation da-

**Received** 2018-01-16, **Revised** 2018-04-02.

**Biography:** Zhao Guosheng (1977—), male, doctor, professor, zgswj@163.com.

**Foundation items:** The National Natural Science Foundation of China (No. 61202458, 61403109), the Natural Science Foundation of Heilongjiang Province (No. F2017021), Harbin Science and Technology Innovation Research Funds (No. 2016RAQXJ036).

**Citation:** Zhao Guosheng, Shao Zihao, Wang Jian, et al. A recognition model of survival situations for survivable systems[J]. Journal of Southeast University (English Edition), 2018, 34(3): 288 – 294. DOI: 10.3969/j.issn.1003-7985.2018.03.002.

ta into survival clusters with different service levels, and then the survival clusters are classified and recognized by means of the error-eliminating decision-making method<sup>[10]</sup>, which realizes the pre-recognition of the system's survival situation.

### 1.1 Ward method

At the beginning of clustering, each of the situation data is regarded as a class. Two classes with the smallest sum of squares of deviations are selected for merging. In the end, all the situation data is classified as one class. Assuming that  $n$  situation data is divided into  $k$  classes ( $G_1, G_2, \dots, G_k$ ), the sum of the squares of data deviations and the total data deviations are

$$S_i = \sum_{j=1}^n (y_{ij} - y_i)'(y_{ij} - y_i) \quad (1)$$

$$S = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - y_i)'(y_{ij} - y_i) \quad (2)$$

where  $y_i$  is the center of gravity;  $y_{ij}$  is the  $j$ -th data in  $G_i$ ; and  $n_i$  is the amount of data.

### 1.2 Error-eliminating decision-making method

In the multi-attribute decision making problem, it is assumed that the survival situation data is indicated as  $A = \{a_1, a_2, \dots, a_m\}$ , and attributes are  $D = \{d_1, d_2, \dots, d_n\}$ . The decision matrix is  $X = (x_{i,j})_{m \times n}$ , and  $x_{i,j}$  is the measured value. Attributes are divided into cost-type attribute and benefit-type attribute. The recognition steps are as follows:

**Step 1** The error value of the survival situation data can be calculated by  $t_{i,j} (i = 1, 2, \dots, m, j \in \mathbf{N})$ .

For cost-type attribute,

$$t_{i,j} = \begin{cases} 1 & x_{i,j} \geq z'_j \\ \frac{z_j - x_{i,j}}{z_j - z'_j} & z_j < x_{i,j} < z'_j \\ 0 & x_{i,j} \leq z_j \end{cases} \quad (3)$$

For benefit-type attribute,

$$t_{i,j} = \begin{cases} 1 & x_{i,j} \leq z'_j \\ \frac{z_j - x_{i,j}}{z_j - z'_j} & z'_j < x_{i,j} < z_j \\ 0 & x_{i,j} \geq z_j \end{cases} \quad (4)$$

According to Eqs. (3) and (4), the error value sequence of the data  $a_i = \{t_{i,1}, t_{i,2}, \dots, t_{i,n}\}$ .

**Step 2** The maximum error value can be calculated by  $t_i^*$ . When the value of  $t_i^* = 1$ , it represents that  $a_i$  is not suitable to be a sample data. The formula is

$$t_i^* = \max_{j \in \mathbf{N}} \{t_{i,j}\} \quad (5)$$

**Step 3** The error loss value of the survival situation data can be calculated as

$$k_{i,j} = a_i d_j \quad i \in M'; j = 1, 2, \dots, n \quad (6)$$

**Step 4** The error loss sequence needs to be sorted. The error loss sequence  $k_{i,1}, k_{i,2}, \dots, k_{i,n}$  can be seen as the point  $R_i$  in the  $n$ -dimensional space. The closer to the origin, the better the data.

$$R_i = \sqrt{\sum_{j=1}^n (k_{i,j})^2} \quad i \in M' \quad (7)$$

## 2 Post-Prediction

The post-prediction model based on autoregressive integrated moving average (ARIMA) is introduced<sup>[11]</sup>. The basic idea of the ARIMA model is that the data sequences formed by the predictive objects over time are regarded as a random sequence, and a certain mathematical model is used to approximate the sequence. Once the model is identified, it can predict the future value from the past and the present value of the time series.

### 2.1 Selection of model

In terms of usability, Kavousi-Fard et al.<sup>[12]</sup> demonstrated the feasibility of this model. Compared with the gray model<sup>[13]</sup> and the BP neural network model<sup>[14]</sup>, it has the advantage of simple application, easy operation, high recognition of the result and support of mature software. This model does not require any assumed conditions and can be entered into any form of survival data time series for prediction.

The construction of the model is as follows. First, survival situation data can be identified as the situation data sequence ( $Y_t$ ). Secondly, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the situation data sequence are calculated. If the model is not a stationary sequence, it is necessary to make  $d$  times difference operation into stationary sequence ( $Y'_t$ ), and carry out the test of stationarity and reversibility. Finally, we select the appropriate ARIMA model for the preliminary post-prediction of survival situations.

### 2.2 Residual correction

Since the data residuals are unavoidable, the fuzzy information granularity<sup>[15]</sup> and the support vector regression (SVR) model are introduced to deal with residual data in the ARIMA model.

Fuzzy information granularity includes window division and fuzzification. For the convenience of calculation, triangular fuzzy particles are used. The formula is

$$F(x, \min, \text{avg}, \max) = \begin{cases} 0 & x < \min \\ \frac{x - \min}{\text{avg} - \min} & \min \leq x \leq \text{avg} \\ 0 & x > \max \end{cases} \quad (8)$$

where  $x$  is the time series of the input data. For a single fuzzy particle,  $\min$ ,  $\text{avg}$  and  $\max$  describe the minimum, average and maximum values of the survival situation data, respectively. The prediction residuals of the model are

mapped to minimum, average and maximum values.

Next, the SVR machine model needs to be established. SVR is a support vector machine for regression analysis.

$\{(x_1, y_1), \dots, (x_n, y_n)\}$  is a set of known training samples. The optimal decision function is constructed in the high-dimensional feature space. The formula is

$$\begin{aligned} \min_{\omega, b, \xi, \xi^*} \quad & \frac{1}{2} \omega^2 + C \sum_{i=1}^l (\xi_i^* + \xi_i) \\ \text{s.t.} \quad & (\omega \varphi(x_i) + b) - y_i \leq \varepsilon + \xi_i \\ & y_i - (\omega \varphi(x_i) + b) \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, \quad x_i \in \mathbf{R}^n, \quad y_i \in \mathbf{R} \end{aligned} \quad (9)$$

where  $\omega$  is a weight vector;  $b$  is the deviation value;  $\xi_i$  and  $\xi_i^*$  are the relaxation variables;  $C$  is a penalty factor;  $\varepsilon$  is the error requirement of regression function. Since the radial basis function has strong nonlinear prediction ability, we choose a radial basis function to construct the predictive function.

### 2.3 Realization of post-prediction

The situation data of the survivable system during a period of time is collected to constitute the time series data. Non-stationary time series data is transformed into the stationary time series data by the difference method.  $y_{i-1}^*$  is the predicted value of the ARIMA model.

Fuzzy particles are divided into three forms: low, medium and high, which are indicated by min, avg and max, respectively. The residual series data is divided into several sub sequences and classified by window. The SVR model

is constructed from these input variables. These variables are formed by the previous window's fuzzy information particle data set. The optimal decision function is constructed by predicting the fuzzy information particles and residual mean. The residual prediction value is  $e_{\text{svr}-1}$ .

The formula of post-prediction for the survival situation is  $y_p = y_{t-1}^* + e_{\text{svr}-1}$ .

## 3 Simulation Experiment

### 3.1 Example of pre-recognition

#### 3.1.1 Index selection

Since the recognition emphasizes the ability of a survivable system to recognize its own current survivable situation, the index system is mainly established according to the method proposed in Ref. [16]. This paper mainly considers four factors in recognition: integrity, usability, emergency and perception. The four factors can be divided into eight performance indices.

#### 3.1.2 Data clustering

In the experiment, it is assumed that the survivable system provides five levels of service: A1 (highest), A2, A3, A4, A5 (lowest), and the performance of survival situation data is analyzed according to the above evaluation indices. The allow access level and deny access level will be defined by an expert scoring system with a discrete value, for 1 to 4, and the performance reduces gradually. Cost-type attributes include the data reuse rate and channel delay, and the rest are benefit-type attributes. The initial data is shown in Tab. 1.

**Tab. 1** The initial data of the selected index

Service level		A1	A2	A3	A4	A5
Integrity	Data multiplex rate	[0.30, 0.40]	[0.35, 0.50]	[0.45, 0.60]	[0.55, 0.75]	[0.70, 0.90]
	Test strength	[900, 1000]	[800, 950]	[700, 850]	[550, 750]	[500, 600]
Usability	Channel delay	[0.40, 0.55]	[0.45, 0.65]	[0.55, 0.75]	[0.65, 0.85]	[0.80, 0.90]
	Channel throughput	[800, 900]	[700, 850]	[550, 700]	[500, 650]	[500, 600]
	Channel utilization	[0.80, 0.90]	[0.70, 0.85]	[0.70, 0.80]	[0.65, 0.75]	[0.60, 0.70]
Emergency	Deny access level	[3, 4]	[2, 4]	[2, 3]	[1, 3]	[1, 2]
	Allow access level	[3, 4]	[3, 4]	[2, 3]	[1, 3]	[1, 2]
Perception	Perception rate	[0.90, 0.99]	[0.75, 0.90]	[0.70, 0.80]	[0.60, 0.75]	[0.60, 0.70]

Tab. 1 illustrates that the data multiplex rate and the channel delay may be reduced with the higher level of service, while the rest indices will be increased. At the same time, various service levels will have different survival performance indicators. The data of 250 normal users is randomly selected, among which the number of A1 level users, A2 level users, A3 level users, A4 level users and A5 level users are 15, 25, 70, 70 and 70, respectively.

By using the Ward method, the sum of the squares of deviations can be calculated by Eqs. (1) and (2). The SPSS19.0 software is used to obtain five kinds of data clustering. The clustering results generated by statistics are shown in Tab. 2.

According to the clustering method and the comparison

with Tab. 1, it is found that the clustered service levels A1', A2', A3', A4' and A5' are approximately equal to the clustered service levels A3, A1, A2, A5 and A4, respectively. The range of index attributes are basically consistent with the initial data classification.

#### 3.1.3 Data recognition

We randomly selected ten groups of test data. The numbers of A1, A2, A3, A4 and A5 are 1, 1, 2, 2, 2, respectively, and two illegal users are selected. Through the calculation of situation data, the recognition results of the service level are realized in Tab. 3.

According to Eqs. (3) to (5) and the decision alternative ratio evaluation (DARE) method, the limit value of each data is  $t_1^* = 0.18$ ,  $t_2^* = 0.95$ ,  $t_3^* = 0.97$ ,  $t_4^* = 0.72$ ,  $t_5^* = 0.88$ ,  $t_6^* = 0.72$ ,  $t_7^* = 1$ ,  $t_8^* = 0.89$ ,  $t_9^* = 0.49$ , and

**Tab. 2** The data clustering of service level

Service level		A1'	A2'	A3'	A4'	A5'
Number of users		72	15	23	72	68
Integrity	Data multiplex rate	[0.47, 0.62]	[0.30, 0.45]	[0.40, 0.50]	[0.68, 0.90]	[0.56, 0.80]
	Test strength	[697, 820]	[882, 1000]	[811, 926]	[500, 649]	[511, 670]
Usability	Channel delay	[0.56, 0.75]	[0.41, 0.54]	[0.53, 0.60]	[0.78, 0.90]	[0.68, 0.88]
	Channel throughput	[551, 700]	[806, 894]	[704, 846]	[500, 575]	[505, 616]
	Channel utilization	[0.66, 0.80]	[0.81, 0.90]	[0.70, 0.85]	[0.61, 0.71]	[0.61, 0.75]
Emergency	Deny access level	[2, 3]	[3, 4]	[2, 4]	[1, 3]	[1, 3]
	Allow access level	[2, 3]	[3, 4]	[3, 4]	[1, 3]	[1, 3]
Perception	Perception rate	[0.69, 0.82]	[0.88, 0.99]	[0.78, 0.90]	[0.61, 0.70]	[0.60, 0.75]

**Tab. 3** Decision matrix of test data

Data index	Data multiplex rate	Test strength	Channel throughput	Channel delay	Channel utilization	Perception rate
$a_1$	0.35	990	881	0.44	0.87	0.92
$a_2$	0.79	569	520	0.83	0.64	0.62
$a_3$	0.82	592	514	0.81	0.66	0.61
$a_4$	0.47	800	613	0.69	0.72	0.72
$a_5$	0.72	574	600	0.84	0.70	0.70
$a_6$	0.56	774	703	0.59	0.75	0.71
$a_7$	0.99	934	877	0.41	0.83	0.88
$a_8$	0.70	623	576	0.75	0.75	0.64
$a_9$	0.37	936	702	0.45	0.82	0.90
$a_{10}$	0.71	501	566	0.80	0.50	0.72
Lower limit	0.90	500	500	0.90	0.60	0.60
Upper limit	0.30	1 000	900	0.40	0.90	0.99

$t_{10}^* = 1$ . Through Eq. (5), the data  $a_1, a_2, a_3, a_4, a_5, a_6, a_8, a_9$  are feasible data, while the others are wrong data. According to Eq. (6) and Ref. [17], the error limit loss and loss sequence of each attribute is  $k_{a1}^* = 0.30, k_{a2}^* = 0.17, k_{a3}^* = 0.13, k_{a4}^* = 0.21, k_{a5}^* = 0.07, k_{a6}^* = 0.12,$

$$a_1 = \{0.025, 0.003, 0.007, 0.016, 0.007, 0.021\}$$

$$a_2 = \{0.245, 0.146, 0.124, 0.181, 0.061, 0.114\}$$

$$a_3 = \{0.259, 0.138, 0.126, 0.173, 0.056, 0.117\}$$

$$a_4 = \{0.084, 0.068, 0.093, 0.121, 0.042, 0.083\}$$

$$a_5 = \{0.211, 0.145, 0.098, 0.184, 0.047, 0.089\}$$

$$a_6 = \{0.131, 0.077, 0.064, 0.079, 0.035, 0.086\}$$

$$a_8 = \{0.199, 0.128, 0.105, 0.147, 0.035, 0.107\}$$

$$a_9 = \{0.036, 0.002, 0.064, 0.021, 0.019, 0.028\}$$

According to Eq. (7), the eccentric distance is  $R_1 = 0.038, R_2 = 0.382, R_3 = 0.386, R_4 = 0.209, R_5 = 0.345, R_6 = 0.205, R_8 = 0.319,$  and  $R_9 = 0.086$ . It also can be seen that the performance of eight data is sorted as  $a_1, a_9, a_6, a_4, a_8, a_5, a_2,$  and  $a_3$ . The error-eliminating decision-making method can not only recognize the error data, but also calculate the range of eccentric distances for each service level. The calculation results are shown in Tab. 4, where it can be concluded that  $a_1$  is A1 level users,  $a_9$  is A2 level users,  $a_6$  and  $a_4$  are A3 level users,  $a_8$  and  $a_5$  are A4 level users,  $a_2$  and  $a_3$  are A5 level users, and others are illegal users.

**Tab. 4** Eccentric distance

Service level	Range of eccentric distance
A1	$0 \leq R_i \leq 0.08$
A2	$0.08 < R_i \leq 0.19$
A3	$0.19 < R_i \leq 0.29$
A4	$0.29 < R_i \leq 0.38$
A5	$0.38 < R_i \leq 0.44$

The Ward method can be used to cluster the survival situation data into five clusters of the service level. The error-eliminating decision-making method can be used to classify the clusters of the survival data and recognize the service level of the user data. With the upgrading of the service level, the performance of the survival situation data will be improved. It means that the higher the level of service, the better the overall performance of the data.

In order to further illustrate the feasibility of this method, two traditional common decision-making methods (the ideal point method and the weighted average method) are used to sort these ten sets of data. The sorting results are shown in Tab. 5.

**Tab. 5** Comparison of decision-making methods

Methods	Sorted results
Error-eliminating decision-making method	$a_1 > a_9 > a_6 > a_4 > a_8 > a_5 > a_2 > a_3,$ $a_7$ and $a_{10}$ are illegal users
Ideal point method	$a_1 > a_9 > a_7 > a_6 > a_4 > a_8 > a_5 > a_3 > a_2 > a_{10}$
Weighted average method	$a_1 > a_7 > a_9 > a_6 > a_4 > a_8 > a_5 > a_3 > a_2 > a_{10}$

From the sorting results, the traditional decision-making methods can only sort the data, while they cannot verify the ability of feasible data. In this paper, some of

the erroneous data is successfully recognized by calculating the limit loss value, which not only increases the accurate processing ability and response time of the data, but also has consistent results with the traditional decision-making method.

3.2 ARIMA combinatorial model

3.2.1 ARIMA model

The “corrected” data in KDD99 is selected for analysis and prediction, which is used to predict the number of abnormal situations that may occur in survivable systems for the next period of time. We select 1 500 daily data, then choose the tag attribute in the last column and count the number of abnormal situations. The results are shown in Fig.1 (a). The collected data sequence map illustrates that the time series data is random and non-stationary, which requires first difference processing, as shown in Fig. 1(b).

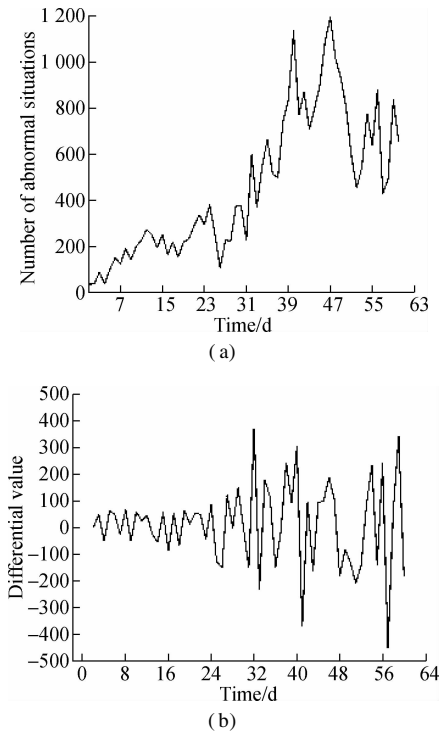


Fig.1 Processing of abnormal situation with first difference. (a) Abnormal time series; (b) First difference time series

Since the difference sequence is basically distributed around 0, upper and lower sides of the scale line, it can be judged that the differential sequence is stable. The most appropriate ARIMA model is automatically generated by the use of SPSS software. The prediction results are shown in Fig. 2.

In Fig. 2, ARIMA (1, 1, 0) is the optimal prediction model. The solid line indicates the actual number of abnormal situations and the dotted line indicates the number of abnormal situations by the ARIMA (1, 1, 0) model. In general, the predictive results of the model are fitted with the actual situation, but accuracy is insufficient and there

is a slight delay in the results.

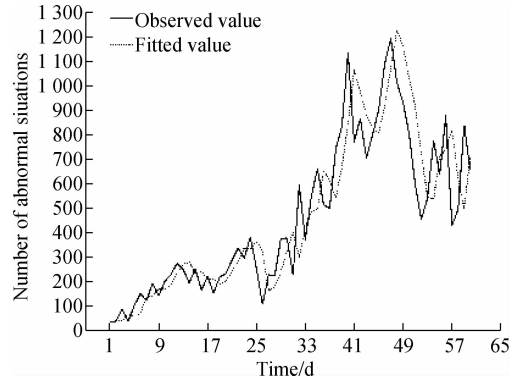


Fig. 2 ARIMA (1, 1, 0) model prediction results

3.2.2 Information granulation and SVR model

The data of 60 d is used as training sets to predict the number of abnormal situations for the next three days. Every three days’ amount of data will become a granular information window and each group of the data will be blurred into three parameters (low, medium, high). This means that the sample sets for the training sets contain 20 windows and the output samples contain 21 windows, as shown in Fig. 3, where low, medium and high describe the minimum, average, and maximum value of changes to the abnormal situation, respectively.

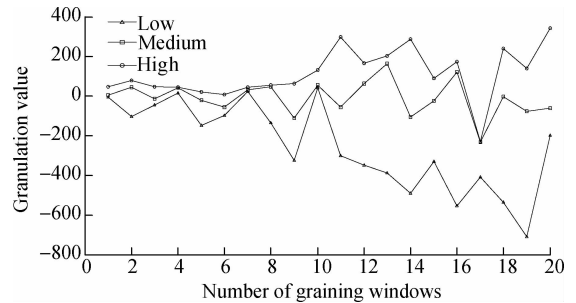
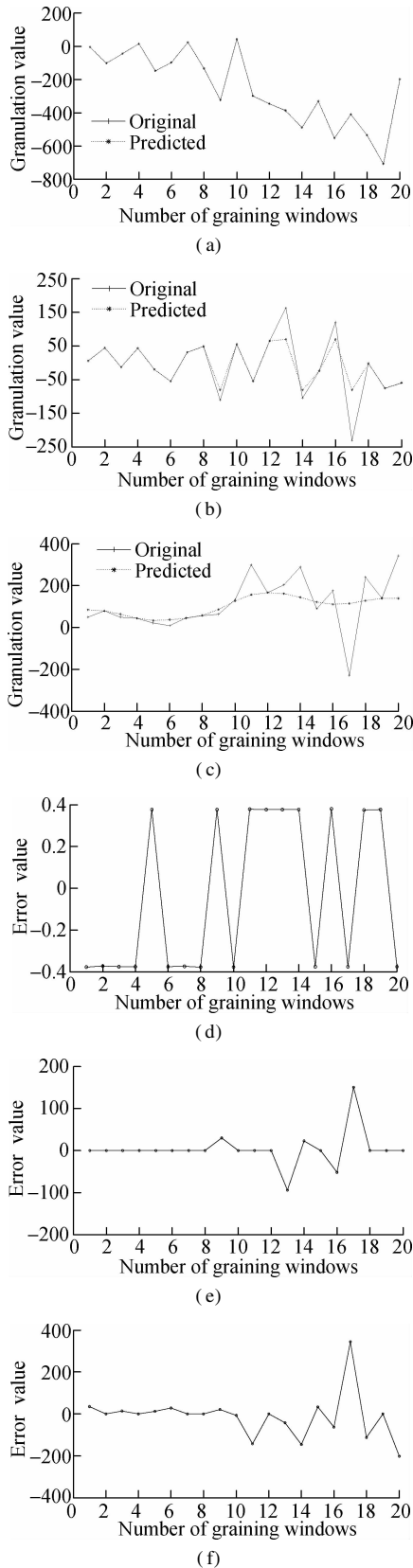


Fig. 3 Granulation results

The SVR model is constructed by using fuzzy information particle data sets. Fuzzy information particles and window residuals are predicted so that the data can be normalized. Parameters are optimized by the Gale-Shapley algorithm. The output is shown in Fig. 4.

It can be seen from Fig. 4 that the overall residual prediction accuracy of the SVR model is high, where the low parameter prediction results are more accurate. However, when the windowed residual value changes greatly, the prediction values of the medium and high parameters will be inaccurate and the predicted values will be small, such as No. 13, 16, 17 window. The SVR model uses the data from the previous window to complete the prediction of the data for the next window. This means that if the number of abnormal situations has a large fluctuation, the accuracy of the forecast will decrease.

The range of residual variation for the last three days is predicted. The low, medium, and high parameter predic-



**Fig. 4** Forecast results and errors of low, medium and high parameters. (a) Low parameter; (b) Medium parameter; (c) High parameter; (d) Low error; (e) Medium error; (f) High error

tions are  $-1.7$ ,  $-5.9$ , and  $127.2$ , respectively. Therefore, the SVR model is feasible for predicting the residual data, but there are also some shortcomings.

### 3.2.3 Combination model

The prediction value of the ARIMA model and the residual prediction value of the SVR are combined to predict the number of abnormal situations in the next three days (Time window is from the 61st to 63rd), as shown in Tab. 6.

**Tab. 6** Prediction results of combined model

Parameter	Value
Number of abnormal situations	2 609
ARIMA prediction $y_{t-1}^*$	2 344
ARIMA relative error/%	10.2
SVR residual prediction $e_{SVR-1}$	120
Combined model prediction $y_p$	2 464
Combined model relative error/%	5.6
Residual variation range	$[-1.7, -5.9, 127.2]$

The experimental results show that the number of abnormal situations by the ARIMA model is 778, 761 and 805. It is similar to the actual numbers of abnormal situations (893, 758, 958), and the trends are consistent. The SVR model is used to correct the residual data. Through the combination of the two models, the overall prediction accuracy is increased by 4.6%. The prediction accuracy of the combined model is up to 94.4%.

## 4 Conclusion

This paper proposed a recognition model of survival situations, which focuses on the pre-recognition and the post-prediction. Experimental results show that the combined model effectively improves the accuracy of pre-recognition and post-prediction. The overall prediction accuracy is increased by 4.6%. The prediction accuracy of the combined model is up to 94.4%. However, there are still some flaws in our method. First, this paper mainly studies the internal environment of survivable system, but lacks the recognition of external attacks. Secondly, when the number of abnormal situations has a large fluctuation, the accuracy of the recognition will decline. Finally, the model is not suitable for predicting long-term survivability due to the uncertainty of situation data.

## References

- [1] Westmark V R. A definition for information system survivability [C]//*Proceedings of the 37th Hawaii International Conference on System Sciences*. Big Island, HI, USA, 2004: 2086 – 2096.
- [2] Yaghlane A B, Azaiez M N. Systems under attack-survivability rather than reliability: Concept, results, and applications [J]. *European Journal of Operational Research*, 2017, **258**(3): 1156 – 1164. DOI: 10.1016/j.ejor.2016.09.041
- [3] Zhao L, Zou H, Zhang X H. Survivability model for reconfigurable service carrying network based on the stochastic Petri net [J]. *Journal on Communications*, 2016, **37**(3): 71 – 78. (in Chinese)
- [4] Raja U, Tretter M J. Defining and evaluating a measure of open source project survivability[J]. *IEEE Transactions on*

- Software Engineering*, 2012, **38**(1): 163 – 174. DOI: 10.1109/tse.2011.39.
- [5] Zhao G S, Liu H L, Wang J. Study on the autonomous recognition mechanism for survivable systems [J]. *Chinese High Technology Letters*, 2014, **24**(10): 999 – 1006. (in Chinese)
- [6] Wang J, Zhao G S. Cognitive model and quantitative analysis for survivable system based on SM-PEPA [J]. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2015, **43**(5): 99 – 103. (in Chinese)
- [7] Dharmaraja S, Vinayak R, Trivedi K S. Reliability and survivability of vehicular ad hoc networks: An analytical approach[J]. *Reliability Engineering & System Safety*, 2016, **153**: 28 – 38. DOI: 10.13039/501100007488.
- [8] Duessel P, Gehl C, Flegel U, et al. Detecting zero-day attacks using context-aware anomaly detection at the application-layer [J]. *International Journal of Information Security*, 2017, **16**(5): 475 – 490. DOI: 10.1007/s10207-016-0344-y.
- [9] Greve B, Pigeot I, Huybrechts I, et al. A comparison of heuristic and model-based clustering methods for dietary pattern analysis[J]. *Public Health Nutrition*, 2015, **19**(2): 255 – 264. DOI: 10.1017/s1368980014003243
- [10] Huang H R, Jiang S L, Cai K. Key important multiple attribute error-eliminating decision-making method [J]. *Mathematics in Practice and Theory*, 2015, **45**(11): 15 – 20. (in Chinese)
- [11] Pati J, Kumar B, Manjhi D, et al. A comparison among ARIMA, BP-NN, and MOGA-NN for software clone evolution prediction [J]. *IEEE Access*, 2017, **5**: 11841 – 11851. DOI: 10.1109/access.2017.2707539
- [12] Kavousi-Fard A, Kavousi-Fard F. A new hybrid correction method for short-term load forecasting based on ARIMA, SVR and CSA[J]. *Journal of Experimental & Theoretical Artificial Intelligence*, 2013, **25**(4): 559 – 574. DOI: 10.1080/0952813x.2013.782351
- [13] Liu S F, Zeng B, Liu J F, et al. Four basic models of GM(1, 1) and their suitable sequences[J]. *Grey Systems: Theory and Application*, 2015, **5**(2): 141 – 156. DOI: 10.1108/gs-04-2015-0016
- [14] Wu B, Han S J, Xiao J, et al. Error compensation based on BP neural network for airborne laser ranging[J]. *Optik—International Journal for Light and Electron Optics*, 2016, **127**(8): 4083 – 4088. DOI: 10.13039/501100004750
- [15] Huang W W, Zhao Y, Huangpeng Q. SOC prediction of Lithium battery based on fuzzy information granulation and support vector regression[C]// *International Conference on Electrical and Electronic Engineering*. Ankara, Turkey, 2017: 177 – 180.
- [16] Zhao G S, Wang H Q, Wang J. Study on situation evaluation for network survivability based on grey relation in analysis [J]. *Mini-Micro Systems*, 2006, **27**(10): 1861 – 1864. DOI: 10.3969/j.issn.1000-1220.2006.10.015. (in Chinese)
- [17] Huang H R. The research of multiple attribute error-eliminating decision-making method[D]. Guangzhou: School of Management, Guangdong University of Technology, 2014. (in Chinese)

## 一种可生存系统生存态势的可识别性模型

赵国生<sup>1</sup> 邵子豪<sup>1</sup> 王 健<sup>2</sup> 李英梅<sup>1</sup>

(<sup>1</sup> 哈尔滨师范大学计算机科学与信息工程学院, 哈尔滨 150025)

(<sup>2</sup> 哈尔滨理工大学计算机科学与技术学院, 哈尔滨 150080)

**摘要:** 由于现有可生存系统中缺乏对系统生存态势的事前识别和事后预测, 提出一种可生存系统生存态势的可识别性模型. 首先, 基于 Ward 方法将生存态势数据聚类为不同服务等级的生存簇, 然后利用消错决策方法对生存态势数据所属簇进行分类和识别, 实现系统生存态势的事前识别; 其次, 利用差分生存态势数据生成平稳预测序列, 构建 ARIMA 模型并通过自相关函数和偏自相关函数校验模型的平稳性、随机性和可逆性指标; 最后, 通过模糊粒子和 SVR 模型的残差修正实现生存态势的事后预测识别. 仿真实验表明, 模型的事前评估模块相较于传统决策方法不仅可对生存态势数据进行聚类和服务等级识别, 还可对非法用户进行识别; 模型通过对异常态势次数的预测与残差修正, 有效实现对可生存系统生存态势的事后预测.

**关键词:** 可生存性; 可识别性; 模糊粒子; 残差修正

**中图分类号:** TP393