A comparison between some classical and artificial intelligence methods for estimating missing values in univariate time series

Firas A. Mohammed ALmohana **Wasn Saad Mahdi**

University of Baghdad – College of Administration and Economics – Department of Statistics

يهدف هذا البحث الى مقارنة بعض الطرائق الكلاسيكية مع الطرائق الذكاء الأصطناعي في تقديرالقيم المفقودة في بيانات السلاسل الزمنية أحادية المتغير لتحديد أدق الطرائق التي تعالج فقدان القيم في مجموعة بيانات السلاسل الزمنية أحادية المتغير حيث تم أختيار حجوم عينات مختلفة (٣٠٠,١٠٠,٦٠) مع حذف مايعادل أربع نسب للفقدان (٥٪ ، ١٠٪ ،١٥٪ ،٢٠٪) بطريقة تحقق شروط الفقدان عشوائياً MAR ويتم ذلك من خلال أستعمال أسلوب المحاكاة ومن أجل الحصول على دقة أداء الطرائق تم أستعمال معيار الدقة متوسط مجموع مربعات الخطأ (MSE) . وأشارت النتائج االى أن أكثر الطرائق دقة في تقدير القيم المفقودة هي طريقة (RBF) لأنها تنتج أقل قيمة من متوسط مجموع مربعات الخطأ مقارنة بالطرائق الأخرى.

الكلمات المفتاحية : السلاسل الزمنية أحادية المتغير ، القيم المفقودة ، الأستكمال الخطي الداخلي LI، أستكمال الجار الأقربNNI ، ترحيل اخر ملاحظة للأمام LOCF ، التمهيد الأسي (المعامل التكيفي) ARRES، k-means، دالة الأساس الشعاعي RBF ، الشبكة العصبية المتكررة ثنائية الأتجاه Bi-RNN ، آلة متجه دعم الأنحدارSVR .

Abstract: -

This research aims to compare the some classical methods with the artificial intelligence methods in estimating the missing values in the univariate time series data to determine the most accurate methods that treat the missing values in the univariate time series data set, where different sample sizes (300,100,60) were chosen with the deletion of four Loss percentages (5%, 10%, 15%, 20%) in a way that meets the MAR random missing conditions, and this is done by using the simulation method, and in order to obtain the accuracy of the performance of the methods, the accuracy standard of the mean squared error (MSE) was used. The results indicated that the most accurate method in estimating the missing values is the (RBF) method because it produces the lowest value of mean squared error compared to other methods.

Keywords: univariate time series, Linear Interpolation(LI), nearest – Neighbor Interpolation(NNI), Last observation carried forward(LOCF), K-Means, Radial Basis Function (RBF), Bi-directional Recurrent Neural Networks (Bi-RNN), Support vector machine – Regression(SVR)

1- Introduction

The missing values and what is meant (the missing of a value or a set of values of information) in the univariate time series data set has become a realistic problem facing researchers in many scientific fields such as (medicine - engineering - science ...) because it directly affects the process of building mathematical models. And statistical, which affects the accuracy of the final results through which the correct decision is reached in the future. The reason for the missing values in the data set is due to a malfunction in the measuring device or errors in the sensor. In recent years there have been many studies and remarkable progress with the interest of researchers in developing and improving Methods to treat missing values before entering the statistical analysis process in order for researchers to obtain complete data to make accurate and correct decisions. In the year(Mahir and Al-Khazaleh, 2008)the researchers proposed a filtering process to estimate the missing values in the time series, where Box-Jenkins models were used to predict the monthly rainfall rate by using two sets of data, a group without missing data and missing data randomly according to ARIMA models (1,0,0) and (1,1,0) and to verify the result, the Naive test was used, which is Thiel's statistic. It turned out to be good models. In (Moustris *et al.*, 2012) researchers used artificial neural network (ANN) models to estimate the missing values of the average daily concentration (PM10) and it was found that the models are excellent predictive methods for estimating the missing values in the time series data of atmospheric pollutants. In (Iwueze *et al.*, 2018) researchers proposed new methods (RMI) Row Mean Imputation, (CMI) Columns Mean Imputation, and (DWMV) Decomposing Without the Missing Value to estimate missing values in time series data based on rows, columns, and total averages of the series data. The times are arranged in the Buys-Ballot table with m rows and s columns and are compared with the currently used methods (MI) Mean Imputations, (SM) Series Mean, (LI) Linear Interpolation and (RI) Regression Imputation, and the methods are evaluated through the accuracy measures (MAE, MAPE and RMSE) and the DWMV method achieved the best estimates in terms of accuracy measures to estimate the missing values compared to the current and new methods. In the same year (Mahboob *et al.*, 2018) the researchers used three machine learning algorithms (K-Means, K-Nearest Neighbors, and K-Medoids Clustering) to estimate the missing values. The performance evaluation of the three algorithms was carried out by applying the Decision Tree and Random Forest algorithms. Forest algorithms and found that the K-Nearest Neighbors algorithm has a high accuracy in estimating missing values compared to other algorithms.(Yen *et al.*, 2020) researchers used linear regression, support vector regression, artificial neural networks, and long short-term memory to estimate missing values in the time series. Deep learning models achieved better performance compared to with traditional models. In (Li *et al.*, 2020) the researchers proposed a Multimodal Deep Learning Model to estimate missing values in heterogeneous traffic speed data. The model (MMDL) was compared with other models and through accuracy measures (MAE, MAPE, and RMSE). It was found that the proposed model has the best performance in estimating missing values. In (Yang *et al.*, 2022)researchers proposed a generative adversarial network (ST-FVGAN) to estimate missing values in time series of taxi traffic data, where This network depends on the spatial and temporal correlation and external factors, where the model was built from a generator network consisting of a convolutional layer, a residual block, and a pixelhuffle block to distribute the data well in order to improve the accuracy of estimating the missing values, and then a network (Discriminator) was used to control the input data. The accuracy of the network was assessed by the root mean squared error (RMSE) and they found that this network (ST-FVGAN) has high accuracy in estimating the missing values .

Materials and Methods

Classical Methods 1-

1.1. Linear Interpolation

This method is one of the simplest forms of interpolation (estimated a missing value or several missing values) that has an arithmetic nature that depends on data before and after the missing value ,Which is a single straight line connecting two points $p_0 = (t_0, y_0)$, $p_1 = (t_1, y_1)$ or more , Where interpolation is good in handling missing values in the data set where the straight line distance between them is short This missing values can be calculated through the equation below (Chapra and Canale, 2011):

$$\hat{Y} = \mathrm{y}_0 + \frac{\mathrm{y}_1 - \mathrm{y}_0}{\mathrm{t}_1 - \mathrm{t}_0}\,(\mathrm{t} - \mathrm{t}_0) \qquad (1)$$

Where: $\hat{Y}$ is the missing value estimation function and y_0 is the previous value that is before the missing value , and y_1 is the suffix value that is after the missing value , and $t_i = 2, \dots\dots, n-1$ be between $(t_0 < t_i < t_1)$ the time variable corresponding to the missing value, (t_0) the previous value of the time variable and (t_1) suffix value of the time variable .

1.2. Nearest – Neighbor Interpolation (NNI)

A method is a form of interpolation used to estimate missing values in univariate time series, based on the endpoints of missing values as estimates of missing values, calculated according to the equation below:

$$\hat{Y} = \begin{cases} Y_0 & if \quad t \le t_0 + \frac{(t_1 - t_0)}{2} \\ \\ Y_1 & if \quad t > t_0 + \frac{(t_1 - t_0)}{2} \end{cases} \qquad (2)$$

Where: $\hat{Y}$ the value of the nearest neighbor interpolation that is represented by the missing value, t the time point in the equation of the nearest neighbor interpolation, (Y_0, Y_1) represent the values before and after the missing value, and (t_0, t_1) the time points before and after the time point t.

1.3. Last observation carried forward(LOCF)

This method is one of the simple methods that are used with univariate time series to estimate the missing values , where each missing value is replaced by the last value seen in previous information in the same variable(Flores, Tito and Silva, 2019) the previous view is moved forward (Moritz *et al.*, 2015)

$$\hat{Y} = y_{i-1} \qquad \dots (3)$$

2. Modern methods of artificial intelligence

2.1. K-Means Clustering Algorithm

(KMC) is a very famous grouping method that can be used in time series, it is considered one of the simplest and most popular unsupervised learning algorithms, its goal is to collect observations that are similar to each other in the data set to a number of clusters K, and each cluster is formed by calculating the distance between the observation and the nearest central point, and the central point represents the basis for collecting data, the algorithm work steps as follows:

1- We have the data set $Y = \{y_1, y_2, \dots \dots, y_n\}$.

2- Choose the number of clusters K required $2 \leq K \leq n$.

3- And then identify some random points called (centroids) within groups K, we can choose these points from the data itself or from external data(Patil, Joshi and Toshniwal, 2010).

$$C = \{c_1, c_2, \dots \dots, c_k\} \qquad (4)$$

4- We calculate the distance between each view with the nearest central point from the selected central points, to measure the distance, the squared Euclidean distance is used according to the equation below:

$$E = \sum_{i=1}^{N} (y_i - C_K)^2 \qquad (5)$$

Where y_i observation in the data set i=1,…,n and C_K central point k=1,…,k , N all observation.

The method of choosing the best number of clusters K is not an easy method, and there are many ways to find it, the most famous of which is the (Elbow Method)(Patil, Joshi and Toshniwal, 2010), which is a method that determines whether the cluster was created according to a correct value calculation from Two main values are required, according to the calculation of the sum of the square of the error SSE between the cluster correlation for different K values, and the cluster correlation strength that has been collected from the data that is similar to each other can be calculated through (Within cluster sum of squares) and abbreviated by the symbol (WCSS) by calculating The distance between the observations and their center in each cluster, and then summing and squaring these distances in each cluster, and this is done according to the equation below(Cui, 2020) :

5- $WCSS = SSE_1 + SSE_2 + \cdots + SSE_k \qquad (6)$

Where :

$$SSE_1 = \sum_{y_i \ in \ cluster \ 1} distance(y_i, c_1)^2 \quad + \cdots$$

$$+ SSE_k = \sum_{y_i \ in \ cluster \ k} distance(y_i, c_k)^2$$

6- Calculate the mean to create the new center point according to the equation below

$$C_K = \frac{\sum_{i=1}^{n} z_{ik} y_i}{\sum_{i=1}^{n} z_{ik}} \qquad (7)$$

Where : z_{ik} the degree of observation i in the set k, and it takes the values$z_{ik} = \{0,1\}$ i.e. $z_{ik} = 0$ that it means the observation i is not in the set k($y_i \notin Y_k$) and $z_{ik} = 1$ that it means the observation i is in the set k($y_i \in Y_k$) , y_i observations in the dataset.

7- Calculate the basic function of k-means according to the equation below:

$$\min \sum_{k=1}^{c} \sum_{i=1}^{n} z_{ik} \ \ \|y_i - c_k\|^2 \qquad (8)$$

Where : $\sum_{i=1}^{n} \|y_i - c_k\|^2$ The Euclidean distance between each view with the nearest central point , $\sum_{k=1}^{c} z_{ik}$ Total observations in the dataset .

8- We repeat steps 3 to 5 until the ideal division of clusters is reached.
9- Values are estimated by taking the average of clusters of the same category with the number of non-missing values.

2.2. Radial Basis Function Neural Networks (RBF)

RBF is one of the most important methods of supervised machine learning it is a type of artificial neural network (ANN), It contains linear processing units that convert a non-linear, non-separable problem into a linear separable problem, used to estimate missing values in time series ,the network having of three layers: input layer ,hidden layer and output layer. in hidden layer Contains activation function Gaussian ,Its equation can be computed using Eq. (9).(Jinkun, 2013)

$$G_j = exp\left(-\frac{\|y_i-\mathrm{C_p}\|^2}{2\sigma_j^2}\right) \qquad \sigma > 0 \qquad (9)$$

Where: σ_j^2 represent the width value of Gaussian function for neural net j, y_i represent input values, $\mathrm{C_p}$ represent the center vector of the Radial Basis Function whose value is determined using the k- means clustering algorithm that will be mentioned in the following paragraphs.

But the final output of the network RBF can be computed using Eq. (10).

$$\hat{Y}_m = \sum_{i=1}^{m} w_{ji}\left[exp\left(-\frac{\|y_i-\mathrm{C_p}\|^2}{2\sigma_j^2}\right)\right]^T \qquad (10)$$

Where: w_{ji} represent the weight in the output sum , m represent the number of neurons in the hidden layer that each contain a Gaussian activation function , The other symbols have already been explained.

The work of the network RBF to apply the radial basis function in the hidden layer of a non-linear nature is by calculating the Euclidean distance between the center of the activation function and the input vector. neural in the output layer to obtain the output values of the radial basis function network, which are linear in nature.

2.3. Bi-directional Recurrent Neural Networks (Bi-RNN)

Bi-RNN It is a distinctive and advanced type of recurrent neural network. RNN It is a type of neural network that is used in deep learning as it uses its hidden units to analyze the data set(Ashour, 2022). The previous and subsequent time steps can be exploited to estimate the missing values as it consists of a combination of two separate hidden layers of the network.(Cao *et al.*, 2018)

- The output of the forward hidden layer $\overrightarrow{h_t}$

$$h_t{}^f = \sigma_{sig}\left(y_t * W_{yh}^f + h_{t-1}{}^f * W_{hh}^f + b_h^f\right) \qquad (11)$$

Where y_t input observations, σ_{sig} activation function, W_{yh}^f and W_{hh}^f weight parameter, b_h^f bias, $h_{t-1}{}^f$ The hidden state in the previous time step.

- The output of the background hidden layer $\overleftarrow{h_t}$

$$h_t{}^b = \sigma_{sig}\left(y_t * W_{yh}^b + h_{t-1}{}^b * W_{hh}^b + b_h^b\right) \qquad (12)$$

Where y_t input observations, σ_{sig} activation function, W_{yh}^{b} and W_{hh}^{b} weight parameter, b_{h}^{b} bias, ${h_{t-1}}^{b}$ The hidden state in the previous time step.

- The output bi-directional recurrent neural network

$$\hat{y}_t = \sigma_{sig}\left(w_{\hat{y}h}h_t + b_{\hat{y}}\right) \qquad (13)$$

2.4. Support Vector Regression (SVR)

SVR is a type of SVM used to deal with regression models. It is one of the supervised machine learning algorithms. It is used to estimate missing values from a data set of observations, each of which is represented by a point with multiple input values and one output value(Gazzola and Jeong, 2021). The best and optimum is called hyperplane (the hyperplane) and is symbolized by the symbol y_0. It separates the data set into two independent sets and equally distributed around the values to be estimated, provided that the observations of the data set do not exceed the two lines ($y_1,\ y_2$) and any observations outside the two lines are neglected and thus any errors will be neglected - ε The goal of this algorithm is to make the distance (margin) between the two lines ($y_1,\ y_2$) as large as possible.

- The linear form of a hyperplane is given by the equation below(Gazzola and Jeong, 2021):

$$\hat{Y} = W^{T}y + b \qquad .... (14)$$

where $\hat{Y}$ the output (real number) and y vector the input and W the weight vector that consists of real numbers whose dimensions are the same as the dimensions of the vector y and b is the bias (real number)

3. Simulation

In this research, the MATLAB program was used to generate the simulation model, which is the moving average model, once when the value β =0.5 and once when the value β =0.9, in order to compare the methods used to estimate the missing values in the time series, where the experiment was repeated (500 once) with the use of three different sizes of sample (60,100,300) and this is done through :

1- Generate variable X into numbers following the standard normal distribution .

2- Generating data through random error that has a standard normal distribution with (mean = zero, variance = 1)
e_t~N (0,1).

3- Using the simulation model of Box-Jenkins models, which is called the first-order moving averages model (Gorgess, 2017)

MA(1) $\hat{Y}_t = e_t - \beta e_{t-1}$ β_1=0.5 و β_1=0.9

4- The type of data missing is Missing at Random (MAR), Taking four different missing ratios (5%, 10%, 15%, and 20%).

5- A comparison has been made between the methods of estimating the missing values by MSE means of an explanation of the best method of estimation(سليم, ٢٠١٨ and محمد) $MSE = \frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)}{n}$

3.1. Discussion of Results

Table (1) summarizes the results (MSE) for a model MA(1) at β_1=0.5 the methods used to estimate the missing values when the sample size is (60,100,300) with missing (5%,10%,15%,20%) and with repetition (r=500)

			MSE						
			Method						
	Missing	**sample size**	**LI**	**NNI**	**LOCF**	**K-means**	**RBF**	**SVR**	**Bi-RNN**
β =**0.5**	**5%**	**60**	0.0079	1.8000	0.0410	0.057302	0.00012275	0.0217	0.1228
		100	0.0046	0.5000	0.0513	0.058678	0.00004039	0.0162	0.1425
		300	0.0017	0.4500	0.0512	0.022484	0.00001801	0.0056	0.04123
	10%	**60**	0.0084	3.6000	0.1246	0.0441488	0.00014904	0.0224	0.1523
		100	0.0050	0.1000	0.0936	0.0717561	0.00001298	0.0141	0.2492
		300	0.0017	0.9000	0.1002	0.024293	0.00001104	0.0051	0.06706
	15%	**60**	0.0082	5.4000	0.1695	0.0546437	0.00007313	0.0278	0.2140
		100	0.0052	0.1500	0.1303	0.0793784	0.00001163	0.0162	0.2954
		300	0.0016	0.135	0.1485	0.022927	0.00000322	0.0048	0.08972
	20%	**60**	0.0075	7.2000	0.1695	0.0389098	0.0001446	0.0204	0.2900
		100	0.0051	0.2000	0.2104	0.0789721	0.00001426	0.0146	0.3817
		300	0.0015	0.180	0.1879	0.024280	0.00000313	0.0055	0.01004

Table (2) summarizes the results (MSE) for a model MA(1) at β_1=0.9 the methods used to estimate the missing values when the sample size is (60,100,300) with missing (5%,10%,15%,20%) and with repetition (r=500)

			MSE						
			Method						
	Missing	**sample size**	**LI**	**NNI**	**LOCF**	**K-means**	**RBF**	**SVR**	**Bi-RNN**
β =**0.9**	**5%**	**60**	0.0080	1.8000	0.0377	0.399298	0.00125047	0.0250	0.0645
		100	0.0048	0.5000	0.0511	0.183813	0.000016653	0.0162	0.01193

		300	0.00170	0.4500	0.0458	0.139308	0.000005454	0.00560	0.010684
	10%	**60**	0.0087	3.6000	0.1437	0.16074	0.00197382	0.0206	0.1083
		100	0.0052	0.1000	0.0981	0.153439	0.000084091	0.01410	0.01647
		300	0.00171	0.9000	0.01191	0.067843	0.000042932	0.00561	0.01442
	15%	**60**	0.0083	5.4000	0.1790	0.318029	0.00072652	0.0278	0.1146
		100	0.0050	0.1500	0.1494	0.171923	0.000015573	0.0162	0.02266
		300	0.0016	0.1350	0.1316	0.10999	0.000008644	0.00440	0.01665
	20%	**60**	0.0079	7.2000	0.2305	0.077451	0.00017506	0.0196	0.2443
		100	0.0049	0.2000	0.1924	0.041226	0.000018858	0.01320	0.02119
		300	0.0016	0.1800	0.1450	0.019645	0.000013662	0.00550	0.01039

1- We note from Table 1, when β =0.5 that the best method for estimating the missing values is the radial basis function RBF method, by calculating the value of the average squares of error for all methods, so the method was characterized by high accuracy because it has the lowest value in at the missing ratios (5%,10%,15%,20%) and for all sample sizes (60,100,300).

2- We note from Table (1), when β =0.5 the method NNI was the worst at sample size 60 and for all missing ratios.

3- We note from Table (2), when β =0.9, that the best methods for estimating the missing values are RBF at the missing ratios (5%,10%,15%,20%) and for all sample sizes (60,100,300) because they have the lowest value of MSE . However, the method NNI was the worst at sample size 60 and for all missing ratios.

4- We note from Table (1), (2) that the methods are convergent in values at the missing ratios (5%, 10%, 15%, 20%) and for all sample sizes (60,100,300) because they have the lowest value for MSE, but method RBF outweighs all other methods. Except for the method NNI that was the worst among them at the sample size 60 and for all Missing ratios.

4. Conclusion

1- Simulation results showed that the best method for estimating the missing values is a method RBF where this method is characterized by high accuracy because it has the lowest rate MSE compared to other methods.

2- The sample size had a significant and clear effect on the accuracy of the results RBF because the value gradually decreases with the increase in the sample size.

3- All the methods used in the research can be used in the process of estimating the missing values in the different sizes of the samples and for all the percentages of missing. Only a method NNI that cannot be used with the small sizes because it gave the worst rate of MSE for all the percentages of missing.

References

1. Ashour, M. A. H. (2022) ‘Optimized Artificial Neural network models to time series’, *Baghdad Science Journal*, p. 899.
2. Cao, W. *et al.* (2018) ‘Brits: Bidirectional recurrent imputation for time series’, *Advances in neural information processing systems*, 31.
3. Chapra, S. C. and Canale, R. P. (2011) *Numerical methods for engineers*. Mcgraw-hill New York.
4. Cui, M. (2020) ‘Introduction to the k-means clustering algorithm based on the elbow method’,

Accounting, Auditing and Finance, 1(1), pp. 5–8.

5. Flores, A., Tito, H. and Silva, C. (2019) ‘Local average of nearest neighbors: Univariate time series imputation’, *International Journal of Advanced Computer Science and Applications*, 10(8), pp. 45–50.
6. Gazzola, G. and Jeong, M. K. (2021) ‘Support vector regression for polyhedral and missing data’, *Annals of Operations Research*, 303(1–2), pp. 483–506. doi: 10.1007/s10479-020-03799-y.
7. Gorgess, H. M. (2017) ‘Time Series Forecasting by Using Box-Jenkins Models’, *Ibn AL-Haitham Journal For Pure and Applied Science*, 26(1), pp. 337–345.
8. Iwueze, I. S. *et al.* (2018) ‘Comparison of Methods of Estimating Missing Values in Time Series’, *Open Journal of Statistics*, 08(02), pp. 390–399. doi: 10.4236/ojs.2018.82025.
9. Jinkun, L. (2013) ‘Radial basis function neural network control for mechanical systems’, *Tsinghua University Press, Beijing, China*, 10, pp. 973–978.
10. Li, L. *et al.* (2020) ‘Estimation of missing values in heterogeneous traffic data: Application of multimodal deep learning model’, *Knowledge-Based Systems*, 194, p. 105592. doi: 10.1016/j.knosys.2020.105592.
11. Mahboob, T. *et al.* (2018) ‘Handling missing values in chronic kidney disease datasets using KNN, K-means and K-medoids algorithms’, in *2018 12th International Conference on Open Source Systems and Technologies (ICOSST)*. IEEE, pp. 76–81.
12. Mahir, R. A. and Al-Khazaleh, A. M. H. (2008) ‘Estimation of missing data by using the filtering process in a time series modeling’, *arXiv preprint arXiv:0811.0659*, (April), pp. 1–12.
13. Moritz, S. *et al.* (2015) ‘Comparison of different Methods for Univariate Time Series Imputation in R’, *arXiv preprint arXiv:1510.03924*. Available at: http://arxiv.org/abs/1510.03924.
14. Moustris, K. P. *et al.* (2012) ‘Missing value estimation for PM10 concentration time series using artificial neural networks’, in *Full Proceedings, in CD-ROM, of the 3rd International Symposium on Green Chemistry for Environment and Health*.
15. Nazim, A. and Afthanorhan, A. (2014) ‘A comparison between single exponential smoothing (SES), double exponential smoothing (DES), holt’s (brown) and adaptive response rate exponential smoothing (ARRES) techniques in forecasting Malaysia population’, *Global Journal of Mathematical Analysis*, 2(4), pp. 276–280.
16. Patil, B. M., Joshi, R. C. and Toshniwal, D. (2010) ‘Missing value imputation based on k-mean clustering with weighted distance’, *Communications in Computer and Information Science*, 94 CCIS(PART 1), pp. 600–609. doi: 10.1007/978-3-642-14834-7_56.
17. Yang, B. *et al.* (2022) ‘ST-FVGAN: filling series traffic missing values with generative adversarial network’, *Transportation Letters*, 14(4), pp. 407–415.
18. Yen, N. Y. *et al.* (2020) ‘Analysis of interpolation algorithms for the missing values in IoT time series: a case of air quality in Taiwan’, *Journal of Supercomputing*, 76(8), pp. 6475–6500. doi: 10.1007/s11227-019-02991-7.
19. محمد, فراس ا. and سليم, نور. (2018) ‘Comparison Some Estimation Methods Of GM (1, 1) Model With Missing Data and Practical Application’, *journal of Economics And Administrative Sciences*, 24(103).