Journal of Software Engineering

ISSN 1819-4311

Academic Journals Inc.

www.academicjournals.com

OPEN ACCESS Journal of Software Engineering

ISSN 1819-4311
DOI: 10.3923/jse.2016.424.430

Research Article
Application of Partial Least-squares Regression to Material Consumption Prediction

[1]Si Li, [2]Shenyang Liu, [1]Xinzhong Li, [1]Zhen Li and [1]Yuan Wang

[1]Department of Equipment Command and Management of Mechanical Engineering College, Xinhua District, 050003 Shijiazhuang, China
[2]Department of Aerial Four Stations of Air Force Logistics College, Gulou District, 221000 Xuzhou, China

Abstract
Background: Nearly all the segments about material include acquisition, storage, supplying and management have close connections with the maintenance material consumption information. The material consumption rule has a great significance on all the segments about material include acquisition, storage, supplying, management and improve the scientificalness of material support. **Materials and Methods:** Through making an analysis of the character of material consumption and some factors that affects material consumption, this study applies partial least-squares regression to solve the problem of material consumption prediction when the sample is small. **Results:** The example indicates that partial least-squares regression is much more accurate than multiple linear regressions. **Conclusion:** The models provide a theoretical basis for calculating reserves of material scientifically and have the vital important guiding significance.

Key words: Maintenance material, material consumption, partial least-squares regression, multiple linear regressions, small sample

Received: December 18, 2015 Accepted: May 09, 2016 Published: September 15, 2016

Citation: Si Li, Shenyang Liu, Xinzhong Li, Zhen Li and Yuan Wang, 2016. Application of partial least-squares regression to material consumption prediction. J. Software Eng., 10: 424-430.

Corresponding Author: Si Li, Department of Equipment Command and Management of Mechanical Engineering College, Xinhua District, 050003 Shijiazhuang, China

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

In recent years, with the wider application of high and new technology, the equipment types of units are getting more various and the equipment structure is getting more and more complex, which leads to the difficulty in mastering the consumption rule of equipment maintenance materials, the prediction work of types, quantities and cost of consumed maintenance material is becoming more heavy. In order to achieve high efficiency of equipment maintenance, the units must realize the accurate stock of equipment maintenance material and establish back-ups of a certain quantity of pre-stocked material. If the stock is too small, the equipment maintenance work can hardly be satisfactorily completed; while if the stock is too large, it will cause overstock, the shortened lifespan of materials being stocked for too long and the invalidity of material even during reserve period. The protection activities including financing, storage and supply of equipment maintenance materials are closely related to the materials consumption and base on the consumption rule of maintenance materials. The key point in this study is to master the material consumption rule, to predict consumption quantity in a short period and to determine a proper stock of equipment maintenance material. Therefore, it is of vital importance to analyze the consumption rule of equipment maintenance materials[1].

Researchers from home and abroad have done large amounts of study on the consumption rule of equipment maintenance materials and have achieved fruitful results. Tao *et al.*[2] did comprehensive analysis on the impacting factors for material secure probability which was divided into subsystem and replaceable unit level of out-field and established prediction model for material consumption when the product failure rate is in line with index distribution. Through a case of calculating the material consumption of certain type of plane, the practicability and applicability of such method were verified. Wang and Kang[3] and Dekker *et al.*[4] proposed exponential type lifespan distribution, which is normally applicable to electronics, complex systems and products being sophisticatedly tested and periodically maintained, while in the calculation of material quantities of products of which the lifespan is in index distribution, the adopted model is poisson distribution and the calculation model of material in poisson distribution was established. Cao *et al.*[5] analyzed the consumption rule of equipment maintenance material by adopting the time series-based exponential smoothing based on the characteristics of equipment maintenance and the collected consumption data of certain type of material. Chen *et al.*[6] divided the consumption of material for preventive maintenance into two kinds: Replacement consumption of random failure parts and replacement consumption of hidden parts for preventive maintenance in view of various factors that impact the material ordering strategy of multi-unit system. Under the condition of periodic preventive maintenance, the material consumption model of units with exponential lifespan was established.

The prediction of maintenance materials consumption was conducted during the whole processes including the financing, storage and supply of equipment maintenance materials when conducting equipment maintenance. Currently, there are numerous methods for predicting equipment maintenance material consumption, when considering the single impacting factor of material consumption, the representative prediction methods are unary linear regression method, grey prediction method and curve fitting method; the multi factors are taken into consideration, the relatively conventional methods include multiple linear regression method and multiple nonlinear regression method. However, the multiple linear regression prediction entails a high demand of sample data, which can achieve ideal prediction result only when the sample quantity is large and moreover, multiple correlations of variables often occurs when using this method. In the practical process of equipment maintenance, the reliable prediction data of material consumption is often scarce, in this occasion, big error will be caused if using conventional multiple regression prediction method. As a result, this study proposed the partial least squares regression method to solve the problem in consumption prediction of maintenance materials.

MATERIALS AND METHODS

The partial least squares regression method can solve the multicollinearity problem that the multiple linear regression model cannot solve, which is a supplementation for multiple linear regression model. The partial least squares regression method is an integrated method which organically combines the multiple linear regression analysis, principal component analysis and canonical correlation analysis among different variables[7-9], which simultaneously realizes the regression modeling, data structure simplification and correlation analysis between two groups of variables. Compared with conventional multiple linear regression models, the method possesses following advantages: (1) Regression analysis is available even when multiple correlation between independent variables is existed, (2) The regression model can be constructed even in condition of variable numbers is more than sample numbers, (3) The regression model includes all independent variables, (4) The regression model is more sensitive in identifying the system information and noise and (5) The regression coefficient of each independent variable in the model is more easily to be explained.

Such characteristics of Partial Least Square (PLS) were analyzed in this study, which serves as a supplementation to conventional multiple linear regression models.

Suppose the material consumption is c, p regression independent variables are $x_1, x_2, ..., x_p$ the number of samples is n, so the data sheet is $y = [y]_{n\times 1}$ and $X = [x_1, x_2, ..., x_p]_{n\times p}$.

The specific steps for construction the partial least squares regression model are as follow:

- Remove data distortion in the sample. After conducting normalization of the X and y, normalized independent variable matrix E_0 and dependent variable F_0 can be obtained
- Determine the number of main component using leave-one out cross-validation and determine the regression equation to calculate Rd(X) and Rd(y) for precision analysis

Suppose, $t_1 = Xw_1$, $w_1 = (w_{11}, ..., w_{lp})^T \in R^p$ and get following optimization problems[10,11]:

$$\max_{|w_1|=1} \sum_{j=1}^{p} r^2(y, Xw_1)\,\mathrm{Var}(Xw_1) \tag{1}$$

After calculation, the optimal solution to this issue is:

$$w_1 = \frac{1}{\sqrt{\sum_{j=1}^{p} r^2(x_j, y)}}\left(r(x_1,y), r(x_2,y), \cdots, r(x_p,y)\right)^T \tag{2}$$

$$t_1 = Xw_1 = \frac{1}{\sqrt{\sum_{j=1}^{p} r^2(x_j, y)}}\left[r(x_1,y)x_1 + r(x_2,y)x_2 + \cdots + r(x_p,y)x_p\right] \tag{3}$$

After extracting the first principal component t_1 and conduct regression analysis on $X = t_1p_1+X_1$, $y = r_1t_1+y_1$. Where, P_1 is the regression coefficient vector, x_1 is residual matrix, r_1 regression coefficient and y_1 residual vector.

In addition:

$$r_1 = \frac{y^T t_1}{\|t_1\|^2},\ P_1 = \frac{X^T t_1}{\|t_1\|^2} \tag{4}$$

Calculate Rd(X) and Rd(y), if the regression equation can achieve satisfactory precision, then stop the equation; if no satisfactory precision is achieved, then extract the second principal component t_2, respectively for y and X.

- Based on the inverse process of standardization, revert standardized variable into original variable and get the final model

Since t is the linear combination of E_0, if extract m components $t = (t_1, t_2, \cdots, t_m)$ from X, so the regression model of F_0 based on component E_0 is:

$$F_0 = r_1E_0w_1^* + r_2E_0w_2^* + \cdots + r_mE_0w_m^* \tag{5}$$

In the equation:

$$w_h^* = \prod_{j=1}^{h-1}\left(I - w_jP_j'\right)w_h$$

where, I is the unit matrix.

Reverted into original variable, so:

$$y = \alpha_1x_1 + \alpha_2x_2 + \cdots + \alpha_px_p \tag{6}$$

where, $\alpha_j = \sum_{h=1}^{m} r_h w_{hj}^*$ is the regression coefficient, w_{hj}^* is the jth variable of w_h^*.

Conducting regression till reaching the satisfactory precision.

- Auxiliary analysis technique the number of principal components can be determined by cross-validation. Suppose the cross-validation of kth principal component t_k is Q_k^2, then in general condition, when $Q_k^2 < 0.0975$, introducing new component t_{k+1} will make no difference in improving the prediction ability of the model, when the calculation is finished, the number of components is k

Through calculating the Rd(X) and Rd(y), it can test explanatory ability of t, the larger values of Rd(X) and Rd(y) are the stronger explanatory ability of t is. Wherein, Rd(X) represents the explanatory ability of t towards X, Rd(y) represents the explanatory ability of t towards y.

The explanatory ability of independent variables towards dependent variables is measured by the variable importance in projection Variable Importance in Projection (VIP_j), the bigger value of VIP_j represents the stronger influence of the responding x_j upon the equipment maintenance material consumption.

Application analysis: Given that there a 5 certain type of equipments in some unit, the consumption situation of the maintenance materials is related to the attended time, travelled distance and traveling time. Table 1 lists the maintenance consumption and operating parameters of the equipment. Suppose the number of the equipment remains unchanged, try to establish the prediction model for the maintenance material consumption:

- **Remove the abnormal points in the sample:** The x_1, x_2, x_3 are the variables respectively representing the attended time, travelled distance and travel time, y is the maintenance material consumption. Two components t_1 and t_2 are extracted by partial least squares regression and calculated the variances of t_1 and t_2, it proves that there is no abnormal points in sample. In order to remove the abnormal points (distorted data) of collected samples that will affects the prediction precision of the model, firstly remove this abnormal points by establishing[12] the oval diagram T^2. Through partial least squares regression, two component t_1 and t_2 can be extracted and the variance of t_1 and t_2 could be calculated. Suppose taking confidence coefficient 95%, drow the oval diagram on the plate of t_1-t_2, it can be found that all samples are within the oval (Fig. 1), therefore there is no abnormal points in the sample
- **Standardized processing of original samples:** In the model, variables represents the operation parameter such as attended time, travelled distance and travel time in order to remove the negative influence to the variables due to the application of different measurement units, it is needed to standardize the original data. The standardized independent variable matrix:

$$E_0 = (E_{01}, E_{02}, \cdots, E_{0m})_{n\times m}$$

The dependent variable:

$$F_0 = (F_{0i})_{n\times 1}$$

Through standardizing the data in Table 1, get the correlation coefficient matrix of variables. It is shown in Table 2.

Table 1: Maintenance material consumption and operating parameters of a certain equipment

Attended time (year)	Travelled distance (km)	Travel time (h)	Material consumption
12	46500	4790	6
12	52000	5090	7
5	24690	2450	1
10	42050	4150	5
10	37800	3860	4

Table 2: Correlation coefficient matrix of original variables

	x_1	x_2	x_3	y
x_1	1.000	0.998	0.996	0.944
x_2	0.998	1.000	0.964	0.946
x_3	0.996	0.964	1.000	0.959
y	0.944	0.946	0.959	1.0000

Table 3: Value table of x_1

	x_1	x_2	x_3
w_1	0.584	0.576	0.575

Table 2 shows that, there is strong correlation among the three variables; constructing models using multiple regression methods will necessarily give birth to the problem of multicollinearity, which causes difficulty in getting ideal prediction results. So, it is needed to use partial least squares regression instead:

- **Extraction of principal components:** Firstly determine the number of components by leave-one-out cross-validation, so as to determine the regression regression equation. The calculation results shows that during the modeling process of partial least squares regression, the first and the second components are 0.81 and -0.19, respectively, so only one component is extracted. According to Eq. 2, the value of w_1 is obtained. It is shown in Table 3

Since there is only I principal component extracted from the model, so $w^*_1 = w_1$. According to Eq. 3, get:

$$t_1 = (0.4518, 1.4961, -3.3462, 0.8403, -0.0523)$$

Then get:

$$P_1 = (0.582, 0.591, 0.578), r_1 = 0.5494$$

Establish the regression model of F_0 based on E_0 and calculate the value of Rd(X) and Rd(y), the results are:

$$\begin{cases} F_0 = 0.332E_{01}+0.321E_{02}+0.325E_{03}+1.886 \\ Rd(X) = 0.983;\ Rd(y) = 0.922 \end{cases}$$

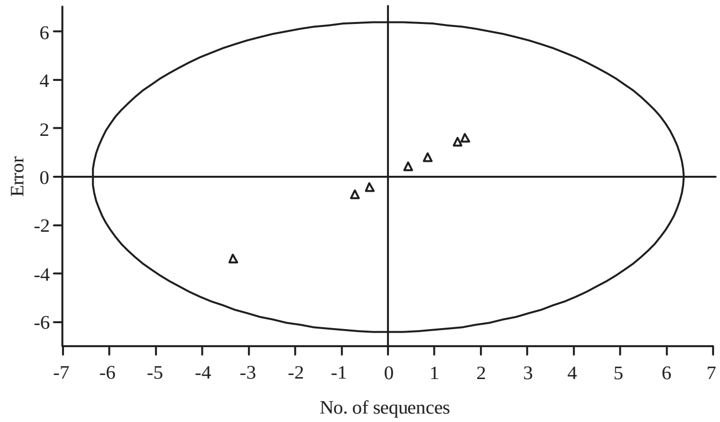

Fig. 1: Oval diagram

- **Model checking:** Calculated Rd(X) and Rd(y) and the conducted precision analysis, the calculate results of Rd(X) and Rd(y) show that the extracted components can reflect 98.3% of variation information of independent variables as well as 92.2% of variation information of dependent variables, which indicates that it is reasonable to establish prediction model for equipment maintenance material consumption using partial least squares regression. Then judge the reasonability of the model and calculate the VIP values of all independent variables

The VIP diagram of explanatory variables is shown in Fig. 2.

According to above column diagram, there is no obvious difference in the importance indexes in projection of the three explanatory variables, which indicates that the three factors are approximately equivalent in affecting maintenance material consumption, so any of explanatory variables should not be randomly deleted.

- **Restore to original equation:** According to the inverse standardization process, revert standardized variables into original variables and obtain the partial least squares regression model between y and x_1, x_2, x_3, which is shown below:

$$y = 0.271x_1 + 0.00007331x_2 + 0.00073x_3 - 4.0056$$

If the specific values of attended time, travelling distance and travel time of equipments in later period are known, the consumption quantity of maintenance material can be predicted in later operation period.

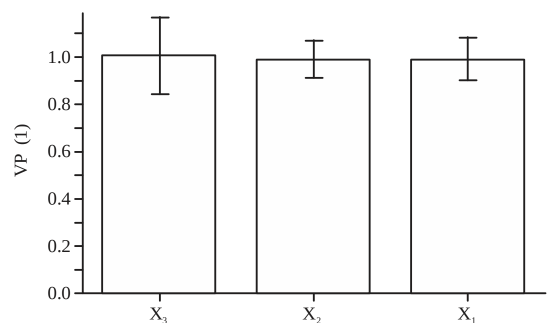

Fig. 2: Column diagram of VIP

RESULTS

The numerous methods for predicting equipment maintenance material consumption, such as unary linear regression method, multiple linear regression method, grey prediction method and curve fitting method need large sample data, which can achieve ideal prediction result. In the practical process of equipment maintenance, the reliable prediction data of material consumption is often scarce, in this occasion, big error will be caused if using conventional multiple regression prediction method. After analyzing the latest previous studies, it is found that many models have been used for forecasting material consumption, including time series models, grey models, neural networks models, support vector machine and combined models. In many times, it is difficult for us to acquire satisfactory effectiveness using these models to forecast material consumption. Compared with these models used for forecasting material consumption, the partial least-squares regression is much more accurate. The partial least-squares regression model includes all independent variables, the partial least-squares regression model can be constructed even in condition of variable numbers is more than sample numbers and partial least-squares regression analysis is available even when multiple correlation between independent variables is existed.

This study presents a partial least-squares regression method to master the material consumption law. Then we could predict the material consumption in a period of time and determine a reasonable number of stored material.

DISCUSSION

Previous research results indicates that partial least squares regression analysis is available even when multiple correlation between independent variables is existed. The partial least squares regression model is more sensitive in identifying the system information and noise. Through this study, it is found that the partial least-squares regression model can be constructed even in condition of variable numbers is more than sample numbers and the model includes all independent variables. What's more, the predictive precision of the partial least-squares regression model applied to predict the material consumption is higher than conventional multiple regression methods and the regression coefficient of each independent variable in the model is more easily to be explained.

CONCLUSION

In this study, the characteristics of maintenance material consumption of a certain type of equipment are analyzed and various factors impacting the material consumption are considered. Regarding the problem of insufficient samples for collecting consumption data, partial least squares regression method is used to build models and predict the maintenance material consumption in later period of equipment operation. Compared with conventional multiple regression methods, the partial least squares regression method possesses obvious superiority, which should be reasonably used in prediction of material consumption, improving the guarantee efficiency of maintenance material in some degree. Nearly all the segments about material include acquisition, storage, supplying and management have close connections with the material consumption information.

Through making a analysis of the related problems, the material consumption models can be derived, of which the practicability is verified by an example. The method proposed in this study can help equipment management personnel to grasp the material consumption rule and to accurately predict material consumption amount, which provide theoretical basis for proper stock amount of material.

The application of the material consumption models based on the partial least-squares regression method could be extended and the material consumption models could also be improved aiming at solving different problems.

ACKNOWLEDGMENT

This study is supported by the Education Science fund of the Education Department of Shijiazhuang, China (No. 2013JGA127).

REFERENCES

1. Zhao, J.Z., T.X. Xu, Y. Liu and Y.T. Yin, 2012. Consumption forecasting of missile spare parts based on rough set, entropy weight and improved SVM. Acta Armamentarii, 33: 1258-1265.
2. Tao, X., L. Guo, B. Xiao and R. Liu, 2012. The prediction model of spare sarts demand based on probability distribution of spare parts guarantee. J. China Ordnance, 33: 975-979.
3. Wang, N. and R. Kang, 2008. Research of the generation, transmission and analytical algorithm of spare parts demand. Chin. J. Aeronaut., 29: 1163-1167.
4. Dekker, R., M.J. Kleijn and P.J. de Rooij, 1998. A spare parts stocking policy based on equipment criticality. Int. J. Prod. Econ., 56-57: 69-77.
5. Cao, J., H. Du, X. Chen and Q. Wang, 2013. Forecasting research for maintenance support materials of armored equipments based on simulation optimization of eoefficient of exponential smoothing method. J. Syst. Simul., 25: 1961-1965.
6. Chen, X.H., T.W. Sheng and S.P. Yi, 2009. Ordering policy of spare parts in multi-unit system for equivalent-cycle preventive maintenance. J. South China Univ. Technol. (Nat. Sci. Edn.), 37: 95-99.
7. Wang, H., 1999. Partial Least Squares Regression and the Application. National Defence Industry Press, Beijing, China.
8. Wang, H., Z. Wu and J. Meng, 2006. Linear and Non-linear Partial Least Squares Regression Method. National Defence Industry Press. Beijing,.
9. Dong, M., 2009. The relationship between input and output of agriculture in China: Analysis based on partial least squares regression model. Technol. Econ., 28: 37-41.

10. Zhou, L., Z. Fu and Z. Ge, 2005. Analysis of partial least squares regression and its application in unit parameter prediction. J. Power Eng., 25: 496-499.
11. Zhang, M.H., Q.S. Xu and D.L. Massart, 2004. Averaged and weighted average partial least squares. Analytica Chimica Acta, 504: 279-289.
12. Ramadan, Z., P.K. Hopke, M.J. Johnson and K.M. Scow, 2005. Application of PLS and back-propagation neural networks for the estimation of soil properties. Chemom. Intell. Lab. Syst., 75: 23-30.