Analysis of CMM Dynamic Measurement Error Based on Decision Regression Tree

Xiaoping Wu, Mei Zhang*, Guihua Li, Haifeng Zhao, Haiting Wu
Anhui University, Hefei, Anhui, 230601, China
94149942@qq.com

Abstract. The best way to improve the accuracy of coordinate measuring machine (CMM) is to establish a correlation dynamic error correction model. However, due to the complex source of dynamic measurement error, the traditional method cannot work well. In this paper, decision regression tree is used to analyze the complex relationship between dynamic measurement error and its influencing factors, and the measurement error is predicted. The results show that the algorithm is simple, fast and superior to the least square estimation.

1. Introduction
Nowadays, the main way to improve the accuracy of CMM is using error modeling to correct the error. Researchers have proposed a variety of modeling methods such as least square estimation, polynomial regression, neural network, wavelet theory and so on [1-6]. But in practice, these methods will encounter a series of problems. For example, although linear model is easier to model, the error model of CMM has been proved to be a nonlinear model. The nonlinear model can better fit the data of CMM, but its parameters are difficult to be determined. Although the neural network overcomes the above shortcomings, its training is unstable and it is easy to fall into the local minimum. This paper proposes a decision regression tree model which combines the advantages of linear model and nonlinear model without their disadvantages. It uses the idea of differentiation to divide the data set into many subsets, and performs linear fitting on each subset to achieve global nonlinear fitting [7-12]. For CMM, the complex relationship between the measurement error and its influencing factors is difficult to be determined globally, but the linear model can achieve good results on a small subset. It is well known that the error sources of CMM all contribute to the final result in the form of 3D coordinates (x,y,z). However, in the process of dynamic measurement, DCC (Direct Computer Control) parameters including positioning speed, proximity distance and contact speed are easy to control, detect and also have important influence on errors. Considering the interaction between 3D coordinates and DCC parameters, the traditional linear and nonlinear models cannot achieve their goals in the global modeling process. This paper uses decision regression tree to model the data and predict the error. The results show our superiority comparing with the least square estimation.

2. Decision regression tree
Decision regression tree is a kind of tree data structure, which is composed of nodes and directed edges. The node divides the branch node and the leaf node. The branch node represents a feature of the data, and the leaf node represents a kind of partitioning or regression result. In other words, the branch node represents a rule by which the data set divides the data into subsets, each of which has a leaf
node that corresponds to a particular output. The process of building a regression tree is to determine the partitioning rules (branch nodes) and the corresponding outputs (leaf nodes).

Suppose that $X$ and $Y$ are input and output variables. The given training data set is $D = \{(x_i, y_i), (x_2, y_2), \ldots, (x_N, y_N)\}$, ($i = 1, 2, \ldots, N$). $x_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(n)})$ is the input instance, $n$ is the number of features and $N$ is the sample size.

Heuristic method is adopted for the partition of feature space, all values of all features in the current set are examined one by one in each partition, and the optimal one is selected as the segmentation point according to the square error minimization criterion. For example, the characteristic variable $x^{(j)}$ and its value $s$ in the training set are used as the sharding variable and the sharding point and define two regions $R_1(j, s) = \{x \mid x^{(j)} \leq s\}$ and $R_2(j, s) = \{x \mid x^{(j)} > s\}$. To find the best $j$ and $s$, solving the following equation

$$
\min_{j,s} \left[ \min_{c_1} \sum_{x \in R_1(j, s)} (y - c_1)^2 + \min_{c_2} \sum_{x \in R_2(j, s)} (y - c_2)^2 \right]
$$

That is to find the $j$ and $s$ that minimize the squared error and the two regions to be divided.

Where, $c_1, c_2$ are the fixed output value in the latter two regions, and the $\min$ means that the optimal $c_1, c_2$ are used, that is the $c_1, c_2$ minimize the squared error in each region. It is easy to know that the two optimal output values are the mean values of $Y$ in the corresponding region, so the above formula can be written as

$$
\min_{j,s} \left[ \sum_{x \in R_1(j, s)} (y - \hat{c}_1)^2 + \sum_{x \in R_2(j, s)} (y - \hat{c}_2)^2 \right]
$$

Wherein, $\hat{c}_1 = \frac{1}{N_1} \sum_{x \in R_1(j, s)} y$, $\hat{c}_2 = \frac{1}{N_2} \sum_{x \in R_2(j, s)} y$.

After finding the optimal segmentation point $(j, s)$, the input space is divided into two regions in turn, and then the above partitioning process is repeated for each region until the stop condition is satisfied, thus a regression tree is generated.

3. Data collection and quantification

3.1. Data collection

In order to establish more accurate and practical error model. In this paper, an experimental program for dynamic measurement error acquisition of measuring machine is designed. In this experiment, the dynamic measurement process of the measured part is realized by touching the probe. The validation of different 3D coordinates and DCC parameters shows the effectiveness of our method. The proposed modeling method was validated by using the moving bridge CMM MC850 (equipped with the probe RenishawTP20, nib length 20mm, nib ball diameter 4mm). The error sampling experiment is arranged as follows: 6 independent spatial coordinates: x(select 0mm, 150mm, 300mm, 450mm, 600mm, 750mm), y(select 150mm, 300mm, 450mm, 550mm), z(select -581mm, -473mm, -324mm), DCC parameters: positioning speed $v_1$(select 20mm/s, 60mm/s, 100mm/s), proximity distance $a$(select 1mm, 2mm, 5mm, 8mm), contact speed $v_2$(select 2mm/s, 4mm/s, 6mm/s, 8mm/s). The dependency is the composite space dynamic error e/m. There are a total of 3,456 combinations of variables. All errors are considered for the main engine, guide rail, environment and touch probe. Therefore, the data obtained can be used for decision tree regression modeling.
3.2. Data quantification

On the basis of the whole experimental samples, we chose 5% as the prediction sample and the rest as the modeling sample. In order to prevent the overfitting of the model and the validity of the test samples, our prediction samples were randomly selected based on the original experimental data. We think the greater the depth of the tree, the better, because depth means the number of subsets divided. But in practice, the number of subsets determine the amount of computation, so it is not necessary to have too much depth. This paper uses many experiments to find the appropriate depth. In order to eliminate the influence of different scales of independent variables, the data were normalized. For the results obtained from the test set, the predicted values corresponding to the original data were obtained through the inverse normalization. The normalization formula is

\[ x_i = \frac{x_i - \bar{x}}{\text{max}(x) - \text{min}(x)} \]  \hspace{1cm} (3)

Wherein, \( x_i \) is a multidimensional matrix, \( \bar{x} \) is the average of the matrix \( x \). \( \text{Max}(x) \) and \( \text{Min}(x) \) are the maximum and minimum values of the matrix \( x \).

The decision regression tree algorithm is implemented by the Python language using numpy and scipy libraries.

4. Experiment

First, we set up different depth training decision regression tree models to find the appropriate depth. Figure 1 shows a comparison of models with depths of 5 and 10 to show the help of large depths to the fitting effect. And the points represent the real data. By slowly increasing the depth of the tree, it can be seen from Figure 2 that when the tree depth is set to 15 and 16, the difference is not big. So let's set the tree depth to 15.

![Figure 1. The fitting results when the depths are 5 and 10](image-url)
In order to further verify the advancement of our method, we implement the least squares estimation model. Figure 3 shows the comparison between the least squares estimate and the decision regression tree estimate. In addition, we take MSE as the measurement index and calculate the MSE values of the least squares estimation and decision regression tree respectively as 12.95 and 1.03. The results show that our method is effective and advanced.

Figure 3. Comparison of decision regression tree and least squares estimation

5. Conclusion
Because of the complexity of the dynamic error source of CMM and the interaction of its influencing factors, the traditional modeling method cannot work well. In this paper, a decision regression tree algorithm is proposed. In terms of error prediction, it can be obtained from the experiments that the decision regression tree algorithm has excellent fitting ability and is proved to be significantly better than the least squares estimation model.
Acknowledgments
This research was supported by Guangdong Province Science and Technology Plan Projects (2017B010110011) and Anhui Natural Science Foundation: 1908085ME172.

References
[1] Moleiro F, Soares C M M, Soares C A M, et al. Layerwise mixed least-squares finite element models for static and free vibration analysis of multilayered composite plates[J]. Composite Structures, 2010, 92(9):2328-2338.
[2] Antonio PiratelliFilho, Nabil Anwer, Charyar Mehdi Souzani, et al. Error evaluation in reverse engineering of aspherical lenses[C]// 17 Emé Congres International De Metrologie, 2015.
[3] Zhang Y F, Nee A Y C, Fuh J Y H, et al. A neural network approach to determining optimal inspection sampling size for CMM[J]. Computer Integrated Manufacturing Systems, 1996, 9(3):161-169.
[4] Novikov I Y, Stechkin S B. Basic wavelet theory[J]. Russian Mathematical Surveys, 1998, 53(6):1159-1231.
[5] Yang H, Liu Y, Fei Y, et al. Hybrid modeling method for CMM dynamic error[J]. Yi Qi Yi Biao Xue Bao/Chinese Journal of entific Instrument, 2010, 31(8):1861-1866.
[6] Yang H T, Lin S W, Fei Y T, et al. Dynamic error modeling of CMM based on Bayesian statistical principle[J]. Proceedings of SPIE - The International Society for Optical Engineering, 2008, 7130:71300H-71300H-7.
[7] Tso G K F, Yau K K W. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks[J]. Energy, 2007, 32(9):p.1761-1768.
[8] Chen M Y. Predicting corporate financial distress based on integration of decision tree classification and logistic regression[J]. Expert Systems with Applications, 2011, 38(9):11261-11272.
[9] Nie G, Rowe W, Zhang L, et al. Credit card churn forecasting by logistic regression and decision tree[J]. expert systems with applications, 2011, 38(12):15273-15285.
[10] Sitanggang I S, Yaakob R, Mustapha N, et al. Predictive Models for Hotspots Occurrence using Decision Tree Algorithms and Logistic Regression[J]. Journal of Applied Sciences, 2013, 13(2):252-261.
[11] Kiselev M V, Ananyan S M, Arseniev S B. Regression-based classification methods and their comparison with decision tree algorithms[M]// Principles of Data Mining and Knowledge Discovery. Springer Berlin Heidelberg, 2006.
[12] Kim Y S. Comparison of the decision tree, artificial neural network, and linear regression methods based on the number and types of independent variables and sample size[J]. Expert systems with applications, 2008, 34(2):p.1227-1234.
[13] Franz E A, Chiaroni-Clarke R, Woodrow S, et al. Congenital mirror movements: Phenotypes associated with DCC and RAD51 mutations[J]. Journal of the Neurological Sciences, 2015, 351(1-2):140-145.
[14] Allegra C, Girlanda P, Morgante F. Treating congenital mirror movements with botulinum toxin[J]. Movement Disorders Clinical Practice, 2017.
[15] Liu Q, Zhang C C, Wang H P B. On the effects of CMM measurement error on form tolerance estimation[J]. Measurement, 2001, 30(1):33-47.