Sensors and Materials, Vol. 31, No. 5 (2019) 1751–1757
MYU Tokyo

S & M 1895

Machine Failure Analysis Using Nearest Centroid Classification for Industrial Internet of Things

Jung-Hyok Kwon[1] and Eui-Jik Kim[2*]

[1]Smart Computing Laboratory, Hallym University,
1 Hallymdaehak-gil, Chuncheon, Gangwon-do 24252, South Korea
[2]School of Software, Hallym University,
1 Hallymdaehak-gil, Chuncheon, Gangwon-do 24252, South Korea

(Received April 30, 2018; accepted March 13, 2019)

Keywords: big data analysis, industrial Internet of things, machine failure reasoning, nearest centroid classification, predictive model

This paper presents a predictive model for machine failure analysis, aiming to accurately analyze various causes of machine failure. The predictive model was developed in the following three steps: 1) dataset classification, 2) attribute selection, and 3) centroid calculation. In the first step, the dataset is classified into multiple subdatasets according to the cause of machine failure. Each subdataset is denoted by a cluster. In the second step, the mean of each attribute measured at the same time is calculated and compared with that of the normal case. Then, the attribute that changes most after the machine failure is selected. In the last step, the mean and variance of the selected attribute are calculated to create the elements of each cluster, and then the centroid of each cluster that maximizes the cohesion of the cluster is calculated. The causes of machine failure are determined by comparing the distance between the data of the new machine failure with the centroid of each cluster. To verify the feasibility of the predictive model, we conducted an experimental implementation. The results show that the implemented predictive model is feasible for analyzing the causes of machine failure.

1. Introduction

Recently, the industrial Internet of things (IIoT) has received a great deal of attention from both industry and academia, since it makes the manufacturing process more reliable, efficient, and safe.[1–3] In IIoT, a number of sensors are typically employed to monitor their surrounding environments.[4] Particularly, in the manufacturing environment, multiple sensors are attached to machines to detect whether the machines are operating successfully. The sensors periodically generate a large amount of data and forward it to a central monitoring server via a wireless link. The central monitoring server extracts the meaningful information from the collected data using big data analysis technologies. The use of big data analysis technologies enables a user to make better decisions by predicting the occurrence of problems such as malfunction, long delay, and failure. Therefore, in recent years, many manufacturing enterprises have been attempting to apply such technologies to their manufacturing process.

*Corresponding author: e-mail: ejkim32@hallym.ac.kr
https://doi.org/10.18494/SAM.2019.2263

ISSN 0914-4935

Machine failure analysis is one of the most challenging issues faced by manufacturing enterprises, since machine failure may lead to disastrous consequences such as an increase in maintenance cost and defective products.[5] In the manufacturing environment, machine failure can occur owing to different causes. For example, machine failure can occur owing to collision with physical objects in the vicinity (e.g., workers, machines, and other manufacturing equipment). To cope with machine failures efficiently, it is necessary to accurately analyze various causes of machine failure using the data collected from sensors. For an accurate analysis, the relationship between the cause and type of machine failure should be determined precisely. To achieve this, most existing studies proposed predictive models using various big data analysis technologies such as association analysis, regression analysis, and neural network.[6–8] However, such predictive models suffer from high complexity and high computational burden for data analysis. Therefore, their application in real manufacturing environments is difficult.

In this paper, we propose a predictive model for machine failure analysis, aiming to analyze various causes of machine failure. The predictive model is developed using the dataset consisting of the data collected from the various sensors attached to the machine. Each sensor collects data multiple times when the event (i.e., machine failure) occurs. To develop the predictive model, we conduct the following three steps: 1) dataset classification, 2) attribute selection, and 3) centroid calculation. The first step is to classify the dataset into multiple subdatasets according to the cause of machine failure. Each subdataset is represented by a cluster. In the second step, the attribute that changes most after the machine failure is selected. For this, we calculate the mean of each attribute measured at the same time and compare it with that of the normal case. Finally, the last step is to calculate the centroid of each cluster that maximizes the cohesion of the cluster. In this step, the mean and variance of the selected attribute are calculated to create the elements of each cluster, expressed as two-dimensional points. To verify the feasibility of the predictive model, we conducted an experimental implementation using R Studio version 1.0.153. In the implementation, we used the dataset containing the force and torque measurements collected after detecting the failure of robot execution, provided by the University of California Irvine (UCI) machine learning repository.[9] The results show that the implemented predictive model is feasible for analyzing the causes of machine failure.

The rest of this paper is organized as follows. In Sect. 2, the predictive model for machine failure analysis is described in detail. The implementation results are presented in Sect. 3. Finally, we conclude this paper in Sect. 4.

2. Predictive Model for Machine Failure Analysis

To develop the predictive model for machine failure analysis, we conduct the 1) dataset classification, 2) attribute selection, and 3) centroid calculation steps sequentially. All the steps are separately conducted for each cause of machine failure. For the development, we use a dataset containing the data collected from the various sensors attached to the machine, which consists of various attributes determined depending on the type of sensor. Each attribute is collected a certain number of times after the event (i.e., machine failure).

In the dataset classification step, the dataset is classified into multiple subdatasets according to the cause of machine failure. Each of the classified subdatasets is used as a cluster. The sensors periodically collect the data a specific number of times when the event occurs. Therefore, each subdataset differentiates the data according to the event. For example, if the sensor generates fifteen data whenever an event occurs, each subdataset differentiates the fifteen data into a set.

In the attribute selection step, one of the attributes that change most after the machine failure is selected. To this end, we calculate the mean of each attribute measured at the same time and compare it with that of the normal case. Then, we select the attribute that has the largest difference from the normal case.

Upon selecting the attribute, the centroid calculation step starts. In this step, the mean and variance of the selected attribute are calculated to create the elements of each cluster. The elements of each cluster are generated for each event. The mean and variance of the selected attribute for the i-th cluster can be given as

$$E(X_{i,j}) = \sum_{k=1}^{n} \frac{x_{i,j,k}}{n}, \tag{1}$$

$$\mathrm{Var}(X_{i,j}) = E(X_{i,j}^2) - E(X_{i,j})^2, \tag{2}$$

respectively, $x_{i,j,k}$ is the attribute of the k-th measurement for the j-th event in the i-th cluster, and n is the number of measurements. To create the elements of each cluster, we use the mean and variance calculated previously. Each element of each cluster is expressed as a two-dimensional point as

$$\boldsymbol{e}_{i,j} = \left(\mathrm{E}(X_{i,j}), \mathrm{Var}(X_{i,j})\right), \tag{3}$$

where $\boldsymbol{e}_{i,j}$ is the element of the j-th event for the i-th cluster. Afterward, the centroid of each cluster, which maximizes the cohesion of the cluster is calculated as

$$\boldsymbol{c}_i = \arg\min_{\mathbf{r}} \sum_{j=1}^{m} d\left(\boldsymbol{e}_{i,j}, \boldsymbol{r}\right), \tag{4}$$

where $\boldsymbol{c}_i$ is the centroid of the i-th cluster, $\boldsymbol{r}$ is the two-dimensional random point, and $d(\bullet)$ is the distance function between two points. This step is performed separately for each subdataset to create the elements and centroid of each cluster. When a new event occurs, the distance between the centroid of each cluster and the point of the event (i.e., the mean and variance of the event) is calculated using the Euclidean distance.[10] Then, the obtained distances for all clusters are compared, and the cluster with the shortest distance is selected.

3. Implementation Results

The experimental implementation was conducted using R Studio version 1.0.153 to verify the feasibility of the model. We used a dataset including the force and torque measurements collected after detecting the failure of robot execution, provided by the UCI machine learning repository. In the dataset, two causes are considered for machine failure. The dataset also contains normal-case data. Thus, we generated the subdataset for each of the normal, collision, and obstruction cases. Note that the subdataset of the normal case is compared with those of the other cases. Each subdataset commonly contains the following six attributes: forces for the x-axis (F_x), y-axis (F_y), and z-axis (F_z), and torques for the x-axis (T_x), y-axis (T_y), and z-axis (T_z). Each attribute is measured every 315 ms fifteen times after machine failure.

Tables 1–3 show the mean of each attribute measured at the same time in the normal, collision, and obstruction cases, respectively. In the tables, the attribute F_z in the normal case changes greater than the other attributes after failure of robot execution caused by collision and obstruction. Therefore, F_z is selected as the attribute for data analysis. Figures 1–3 show the scatter plots for elements of each cluster. In each figure, the x- and y-axes of the plot indicate the mean and variance of the attribute F_z, respectively. Each circle indicates an element of a cluster. In the normal case, all the points are concentrated at a certain point. This is because the F_z used in the normal case is almost the same as 15.5. On the other hand, F_z varies dynamically when machine failure occurs. In particular, F_z is more dispersed in the obstruction case than in the collision case since its variation is larger in the collision case. Each figure shows the mean and variance in the normal, collision, or obstruction case. From the figures, we can extract the centroid of a cluster. Specifically, the centroids of the clusters are (15.6, 39.5), (–29.2, 155944.1), and (–804.2, 1086131.5), respectively. By using these centroids, the causes of machine failure can be analyzed. For example, if the mean and variance of a new event are –31.7 and 260.0, the machine failure is caused by collision.

Table 1
Mean of each attribute in normal case.

No	F_x	F_y	F_z	T_x	T_y	T_z
1	–1.91	1.08	15.33	3.70	–4.79	0.37
2	–2.12	1.00	15.33	3.66	–5.20	0.58
3	–1.87	1.04	15.95	3.58	–4.87	0.37
4	–2.00	0.95	15.79	4.08	–4.91	0.37
5	–1.91	0.95	15.79	3.50	–4.75	0.41
6	–1.91	0.87	16.08	3.83	–4.62	0.20
7	–2.00	1.04	16.04	3.58	–4.87	0.25
8	–1.83	1.04	16.75	3.62	–4.58	0.33
9	–1.87	1.04	14.79	2.75	–4.70	0.29
10	–1.79	0.70	14.29	3.16	–4.41	0.04
11	–1.58	0.58	16.08	3.41	–3.91	–0.20
12	–2.04	0.79	13.75	3.41	–4.58	0.04
13	–2.04	0.83	16.62	3.00	–4.95	–0.20
14	–1.62	0.58	14.50	3.79	–4.45	0.04
15	–1.37	1.00	15.25	2.41	–4.45	–0.20

Table 2
Mean of each attribute in collision case.

No	F_x	F_y	F_z	T_x	T_y	T_z
1	2.86	–0.15	–14.79	–9.09	–3.95	–1.45
2	1.77	0.29	–20.11	–8.65	–3.77	–0.31
3	0.27	–4.34	–40.47	–9.77	1.13	–0.75
4	0.63	–2.56	–11.13	–3.77	–6.11	–0.40
5	–0.90	0.36	–13.77	–10.09	–5.93	–2.15
6	1.86	–1.95	–18.54	–3.84	–1.31	–1.77
7	2.70	–2.75	–28.31	–3.38	–2.72	–0.59
8	4.54	–3.09	–51.88	4.70	11.59	–4.90
9	–0.81	–0.86	–9.77	0.61	7.06	–1.09
10	–0.29	–1.95	–31.65	1.50	9.43	–1.81
11	0.02	–1.79	–51.22	0.43	13.75	–1.52
12	2.50	–5.38	–44.50	27.93	27.47	–7.47
13	0.22	0.13	–28.11	4.13	1.09	–3.25
14	–1.54	–1.40	–31.86	3.75	–0.18	–2.52
15	–1.52	–1.13	–48.95	3.70	4.97	–4.09

Table 3
Mean of each attribute in obstruction case.

No	F_x	F_y	F_z	T_x	T_y	T_z
1	40.30	–25.65	–510.80	70.25	–17.15	5.10
2	42.45	–24.45	–568.95	77.60	–25.10	5.90
3	38.45	–14.80	–566.65	52.95	29.45	5.25
4	34.40	–19.40	–740.05	55.55	39.05	1.95
5	15.65	–9.75	–881.40	12.40	–11.30	–0.05
6	5.80	–24.40	–687.45	64.35	45.70	3.10
7	–22.40	–41.70	–705.35	26.55	–16.30	–11.80
8	–0.45	–31.25	–594.35	–23.90	–9.00	–7.55
9	1.55	–34.00	–714.50	–2.20	–29.85	–5.90
10	21.35	–33.05	–522.40	–18.00	–5.95	–7.30
11	9.25	–90.60	–1147.95	2.75	–38.70	–23.00
12	–28.05	–99.70	–1377.40	–3.10	–47.45	–14.10
13	–36.35	–97.30	–1450.90	12.05	–54.00	–11.75
14	11.00	–70.10	–1373.50	5.65	–29.10	–10.35
15	–6.55	–49.10	–975.15	–19.85	–59.10	–1.00

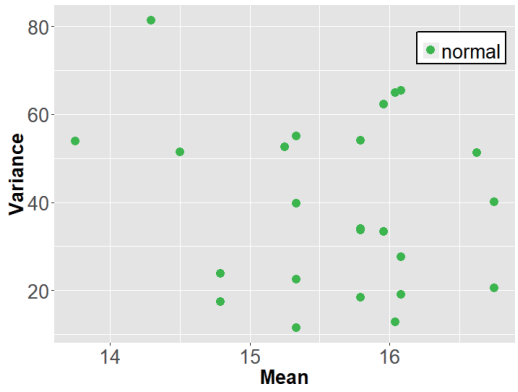

Fig. 1. (Color online) Scatter plot for elements of normal case.

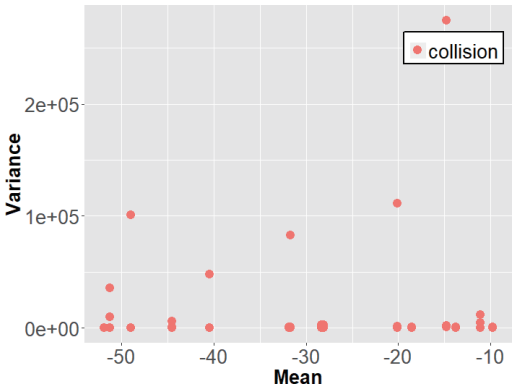

Fig. 2. (Color online) Scatter plot for elements of collision case.

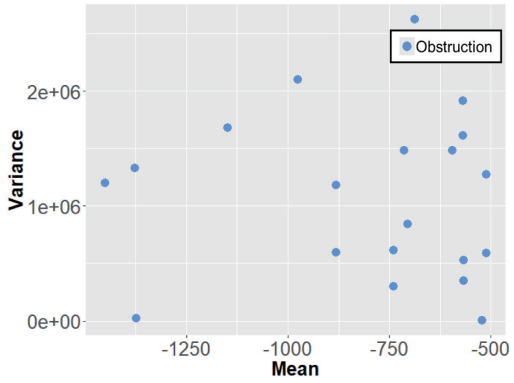

Fig. 3. (Color online) Scatter plot for elements of obstruction case.

4. Conclusions

In this paper, we propose a predictive model for machine failure analysis, aiming to analyze various causes of machine failure. To develop the predictive model, we conducted the following three steps: 1) dataset classification, 2) attribute selection, and 3) centroid calculation. The dataset classification step classifies the dataset into multiple subdatasets according to the cause of machine failure, and the attribute selection step selects one of the attributes that change most after the machine failure. The centroid calculation step calculates the centroid of each cluster that maximizes cluster cohesion. To verify the feasibility of the predictive model, an experimental implementation was conducted using R Studio 1.0.153. We used the dataset containing the force and torque measurements collected after detecting the failure of robot execution, provided by the UCI machine learning repository. In the experiment, the clusters in the normal, collision, and obstruction cases were created, with the centroids (15.6, 39.5), (–29.2, 155944.1), and (–804.2, 1086131.5), respectively. The results show that the implemented predictive model is feasible for analyzing various causes of machine failure.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A1B03031055).

References

1 J. Wan, S. Tang, Z. Shu, D. Li, S. Wang, M. Imran, and A. V. Vasilakos: IEEE Sens. J. **16** (2016) 7373. https://doi.org/10.1109/JSEN.2016.2565621
2 J. -H. Kwon, H. -H. Lee, Y. Lim, and E. -J. Kim: Appl. Sci.-Basel **6** (2016) 427:1. https://doi.org/10.3390/app6120427
3 H. -H. Lee, J.-H. Kwon, and E. -J. Kim: J. Supercomput. **74** (2018) 4385. https://doi.org/10.1007/s11227-016-1915-4
4 T. M. Chiwewe, C. F. Mbuya, and G. P. Hancke: IEEE Trans. Ind. Inf. **11** (2015) 1446. https://doi.org/10.1109/TII.2015.2491267

5 A. Rodriguez, D. Bourne, M. Mason, G. F. Rossano, and J. Wang: Proc. 2010 IEEE Conf. Automation Science and Engineering (IEEE, 2010) 210. https://doi.org/10.1109/COASE.2010.5584452
6 Z. Zheng, W. Xu, Z. Zhou, D. T. Pham, Y. Qu, and J. Zhou: Procedia Manuf. **10** (2017) 15. https://doi.org/10.1016/j.promfg.2017.07.005
7 A. Diryag, M. Mitić, and Z. Miljković: J. Mech. Eng. Sci. **228** (2014) 1444. https://doi.org/10.1177/0954406213507704
8 F. Lopez, M. Saez, Y. Shao, E. C. Balta, J. Moyne, Z. M. Mao, K. Barton, and D. Tilbury: IEEE Rob. Autom. Lett. **2** (2017) 1885. https://doi.org/10.1109/LRA.2017.2714135
9 University of California Irvine (UCI) Machine Learning Repository: http://archive.ics.uci.edu/ml/datasets/Robot%20Execution%20Failures (accessed October 2018)
10 S. Scardapane, R. Altilio, M. Panella, and A. Uncini: Proc. 2016 IEEE Int. Joint Conf. Neural Networks (IEEE, 2016) 3093. https://doi.org/10.1109/IJCNN.2016.7727593

About the Authors

Jung-Hyok Kwon received his B.S. degree in Electronic Engineering from Soongsil University, Seoul, Korea in 2010, his M.S. degree in Electronics and Computer Engineering from Korea University, Seoul, Korea in 2012, and his Ph.D. degree in Convergence Software from Hallym University, Chuncheon, Korea in 2019. From April 2013 to June 2015, he was a research engineer at software R&D Lab., LIG Nex1 Co., Ltd. Since March 2019, he has been a research professor at the Smart Computing Laboratory, Hallym University. He was selected as a recipient of the Global Ph.D. Fellowship Program sponsored by the National Research Foundation of Korea in 2016. His research interest includes design and performance analysis of medium access control protocols for next-generation communication systems, vehicular communications (V2X), machine-to-machine communications, machine learning, and Internet of things. (jhkwon@hallym.ac.kr)

Eui-Jik Kim received his B.S., M.S., and Ph.D. degrees in Electronics and Computer Engineering from Korea University, Seoul, Korea in 2004, 2006, and 2013, respectively. From September 2005 to December 2005, he was with the Intel Korea R&D Center at Intel Corporation. From February 2006 to July 2009, he was with the DMC R&D Center at Samsung Electronics Co., Ltd., Suwon, Korea. From August 2009 to August 2013, he was with the Advanced Institute of Technology at KT Corporation, Seoul, Korea. Since September 2013, he has been a professor at the School of Software, Hallym University. His current research interests include wireless/mobile networks with an emphasis on QoS guarantee and adaptation, wireless sensor networks, wireless LAN/PAN/BAN, vehicular communications (V2X), Internet of things, machine-to-machine communications, and ICT convergence services. (ejkim32@hallym.ac.kr)