IOP Conf. Series: Journal of Physics: Conf. Series **892** (2017) 012008 doi:10.1088/1742-6596/892/1/012008

Association Rule-based Predictive Model for Machine Failure in Industrial Internet of Things

Jung-Hyok Kwon[1], Sol-Bee Lee[1], Jaehoon Park[2] and Eui-Jik Kim[1*]

[1] Department of Convergence Software, Hallym University, 1 Hallymdaehak-gil, Chuncheon-si, Gangwon-do 24252, South Korea
[2] Department of Electronic Engineering, Hallym University, 1 Hallymdaehak-gil, Chuncheon-si, Gangwon-do 24252, South Korea

ejkim32@hallym.ac.kr

Abstract. This paper proposes an association rule-based predictive model for machine failure in industrial Internet of things (IIoT), which can accurately predict the machine failure in real manufacturing environment by investigating the relationship between the cause and type of machine failure. To develop the predictive model, we consider three major steps: 1) binarization, 2) rule creation, 3) visualization. The binarization step translates item values in a dataset into one or zero, then the rule creation step creates association rules as IF-THEN structures using the Lattice model and Apriori algorithm. Finally, the created rules are visualized in various ways for users' understanding. An experimental implementation was conducted using R Studio version 3.3.2. The results show that the proposed predictive model realistically predicts machine failure based on association rules.

1. Introduction

Industrial Internet of things (IIoT) is a key enabler for realizing the fourth wave of the industrial revolution, since it can greatly improve manufacturing enterprises' operational efficiencies, such as performance optimization, safety improvement, time and cost savings, etc. [1–3]. IIoT generally employs many machines, and generates a huge amount of data by detecting and measuring changes in their surrounding environments. This data can be processed and analyzed to provide useful information for end users to manage the machines efficiently. In recent years, many manufacturing enterprises and research institutes have investigated big data analytics to extract meaningful values from the collected data [4].

The prediction of machine failure is one of the most challenging issues faced by manufacturing enterprises, since machine failures may seriously damage the manufacturing process and increase maintenance costs. Big data analytics can provide valuable insight into failure causes, and thus it is widely used to predict machine failure. The types and causes of machine failure vary widely. Therefore, for accurate prediction, it is very important to create rules based on relationships between the cause and type of machine failure. However, most existing predictive models for machine failure only focus on a single type of machine failure [5]. Therefore, these models are difficult to be applied to real enterprise manufacturing processes.

In this paper, we propose an association rule-based predictive model for machine failure in IIoT, which provides more realistic prediction results by considering various machine failure types. The predictive model was developed based on association rules considering relationships between failure

causes and types. We considered three major steps: 1) binarization, 2) rule creation, 3) visualization. The first step creates a binary representation of item values. The second step creates the association rules, represented as an IF-THEN structure. The third step visualizes the rules in various ways for users' understanding.
To verify the feasibility of the proposed predictive model, we conducted an experimental implementation using R Studio version 3.3.2 [6]. As the dataset, we use the log data of 1,000 machines. The data for each machine includes four items for the cause of machine failure (i.e., CPU usage ratio, memory usage ratio, machine temperature, and low battery) and three items for the types of machine failure (i.e., long latency, shutdown, and freezing). The results show that the proposed predictive model provides the realistic prediction results based on the association rules.
The rest of this paper is organized as follows. In Section 2, the predictive model for machine failure is described in detail. The description for dataset and the implementation results are presented in Section 3. We conclude this paper in Section 4.

2. Predictive model for machine failure

To develop the machine failure prediction model, we used association analysis to create association rules identifying how the items are related [7]. The predictive model for machine failure is represented in the form of an IF-THEN structure based on the created association rules. The model development incorporated three major steps, as follows.

2.1. Binarization

The value of categorical items and continuous items were transformed into binary representation. Each item value was represented as 1 when it was present, and 0 otherwise. Thus, the continuous values were converted to categorical items by dividing into multiple categories, i.e., discretization. The number of categories and the range of each category can be predetermined. Figure 1 shows an example of binarization, where the number of categories for CPU usage ratio was set to 3: [0, 50], (50, 80], and (80, 100].

Machine ID	CPU usage ratio	Residual energy
1	22%	No
2	58%	Yes
…	…	…
N	82%	Yes

Binarization

Machine ID	CPU usage ratio = [0, 50]	CPU usage ratio = (50, 80]	CPU usage ratio = (80, 100]	Residual energy = Yes	Residual energy = No
1	1	0	0	0	1
2	0	1	0	1	0
…	…	…	…	…	…
N	0	0	1	1	0

Discretization

Figure 1. Binarization.

2.2. Rule creation

Association rules were created, represented as $X \rightarrow Y$, where X is the cause subset for machine failure, and Y is the subset for machine type. The itemsets of association rules become the predictive model for machine failure. To create the association rules, frequent itemsets should be extracted from the binarized dataset. We used the Lattice model, as shown in Figure 2, to list all possible itemsets (i.e., candidates), and the itemset of each machine in the binarized dataset was compared with every

IOP Conf. Series: Journal of Physics: Conf. Series **892** (2017) 012008 doi:10.1088/1742-6596/892/1/012008

candidate to identify frequent itemsets. However, this procedure may lead to significant computational burden on the system, since the number of comparisons increases exponentially as the number of items increase. Therefore, we applied the Apriori algorithm to reduce the number of candidates compared. The Apriori algorithm follows two principles: i) If an itemset occurs frequently, then all of its subsets must also be frequent, and ii) If an itemset occurs infrequently, then all of its subsets must also be infrequent. This reduces the number of comparisons by pruning subsets of infrequent itemsets, as shown in Figure 3. Frequent items can be identified depending on the value of support, confidence, and lift;

$$s(X \to Y) = \frac{n(X \cup Y)}{N}, \tag{1}$$

$$c(X \to Y) = \frac{n(X \cup Y)}{n(X)}, \tag{2}$$

and

$$Lift(X,Y) = \frac{c(X \to Y)}{s(Y)}, \tag{3}$$

respectively; where N is the number of machines, $n(X \cup Y)$ is the number of itemsets including X and Y, $n(X)$ is the number of itemsets including X, and $s(Y)$ is the support for Y.

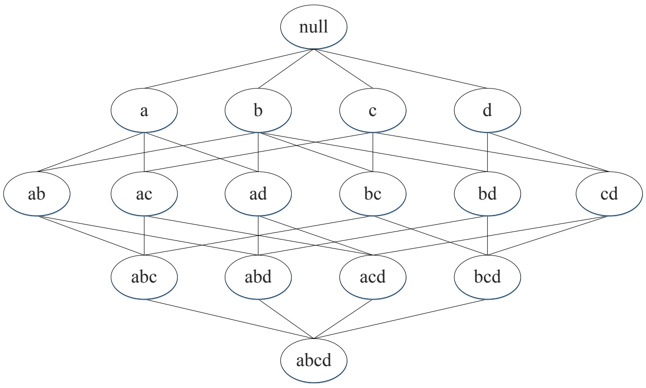

Figure 2. Lattice model.

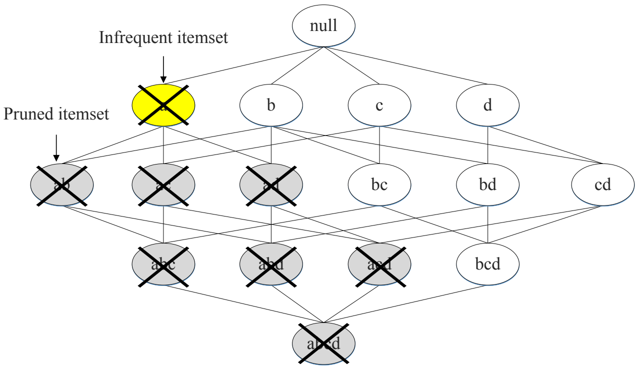

Figure 3. Itemset pruning.

2.3. Visualization

Finally, the created association rules were visualized in various ways, e.g. graphs, tables, matrixes, etc., to help the end user understand the rule patterns. Visualization included sorting and filtering to simplify finding specific rules.

3. Implementation

To evaluate the feasibility of the proposed machine failure prediction model, we conducted an experimental implementation using R Studio version 3.3.2. In the following subsections, we describe the dataset and discuss the results of the implementation, in detail.

3.1. Dataset

We used log data collected from 1,000 machines, consisting of the failure cause and failure type items for each machine. The cause items were CPU usage ratio, memory usage ratio, machine temperature, and residual battery; and the type items were long latency, shutdown, and freezing.
CPU and memory usage ratios are continuous items, and are commonly divided into three categories: low, middle, and high level. However, the category ranges are different: CPU usage ratio categories were [0, 30], (30, 50], and (50, 100]; whereas memory usage ratio categories were [0, 50], (50, 80], (80, 100], respectively. Machine temperature was divided into two categories: low and high, being < 60°C and ≥ 60°C, respectively. Other categorical items were represented as 1 or 0, corresponding to Yes or No, respectively.

3.2. Implementation results

Table 1 summarizes the implementation when support, confidence, and lift for frequent items were greater than 0.07, 0.5, and 2, respectively. In the table, the left-hand-side (LHS) lists the cause itemsets for machine failure, and the right-hand-side (RHS) lists the itemsets for machine failure type. Eleven association rules were identified. Only low battery affects machine shutdown, and machine freezing occurred when the machine temperature and CPU usage ratio become high. Long latency was affected by multiple items, including CPU usage ratio, memory usage ratio, and machine temperature, and others, so it is difficult to identify the cause of a specific long latency. Table 2 shows the association rules when support, confidence, and lift for frequent items were greater than 0.09, 0.5, and 2, respectively. The number of association rules is less than the previous case (Table 1). There are only 3 association rules for long latency, and we can identify that CPU and memory usage ratio have the largest impacts on long latency.

Table 1. Association rules (Support: 0.07, Confidence: 0.5, Lift: 2).

LHS	RHS	Support	Confidence	Lift
{Low.battery=O}	{Shut.down=Y}	0.099	1	10.101
{CPU_cd=CPU_high, Temperature_cd=Temperature_high}	{Freezing=Y}	0.099	0.831	8.403
{CPU_cd=CPU_mid, Memory_cd=Memory_mid}	{Long.latency=Y}	0.088	0.838	3.862
{CPU_cd=CPU_mid, Temperature_cd=Temperature_low}	{Long.latency=Y}	0.087	0.527	2.429
{Low.battery=X, CPU_cd=CPU_mid}	{Long.latency=Y}	0.139	0.522	2.408
{Memory_cd=Memory_mid, Temperature_cd=Temperature_low}	{Long.latency=Y}	0.109	0.521	2.403
{Low.battery=X, CPU_cd=CPU_high, Temperature_cd=Temperature_high}	{Freezing=Y}	0.099	0.942	9.523
{Low.battery=X, CPU_cd=CPU_high, Temperature_cd=Temperature_low}	{Long.latency=Y}	0.075	0.528	2.433
{Low.battery=X, CPU_cd=CPU_mid, Memory_cd=Memory_mid}	{Long.latency=Y}	0.088	0.907	4.180

IOP Conf. Series: Journal of Physics: Conf. Series **892** (2017) 012008 doi:10.1088/1742-6596/892/1/012008

{Low.battery=X, CPU_cd=CPU_mid, Temperature_cd=Temperature_low}	{Long.latency=Y}	0.087	0.583	2.690
{Low.battery=X, Memory_cd=Memory_mid, Temperature_cd=Temperature_low}	{Long.latency=Y}	0.109	0.567	2.616

Table 2. Association rules (Support: 0.09, Confidence: 0.5, Lift: 2).

LHS	RHS	Support	Confidence	Lift
{Low.battery=O}	{Shut.down=Y}	0.099	1	10.101
{CPU_cd=CPU_high, Temperature_cd=Temperature_high}	{Freezing=Y}	0.099	0.831	8.403
{Low.battery=X, CPU_cd=CPU_mid}	{Long.latency=Y}	0.139	0.522	2.408
{Memory_cd=Memory_mid, Temperature_cd=Temperature_low}	{Long.latency=Y}	0.109	0.521	2.403
{Low.battery=X, CPU_cd=CPU_high, Temperature_cd=Temperature_high}	{Freezing=Y}	0.099	0.942	9.523
{Low.battery=X, Memory_cd=Memory_mid, Temperature_cd=Temperature_low}	{Long.latency=Y}	0.109	0.567	2.616

Figures 4–7 visualize the created association rules. Figures 4 and 5 show the graph for 11 and 6 rules (corresponding to Tables 1 and 2), respectively, where green circles are items, and pink circles represent association rules where the lift is indicated by the circle size. Rule relationships are represented by arrows and circle positions. Figures 8 and 9 show the grouped matrixes for the same association rules, respectively, where the item support is represented by the circle size, and lift by circle color.

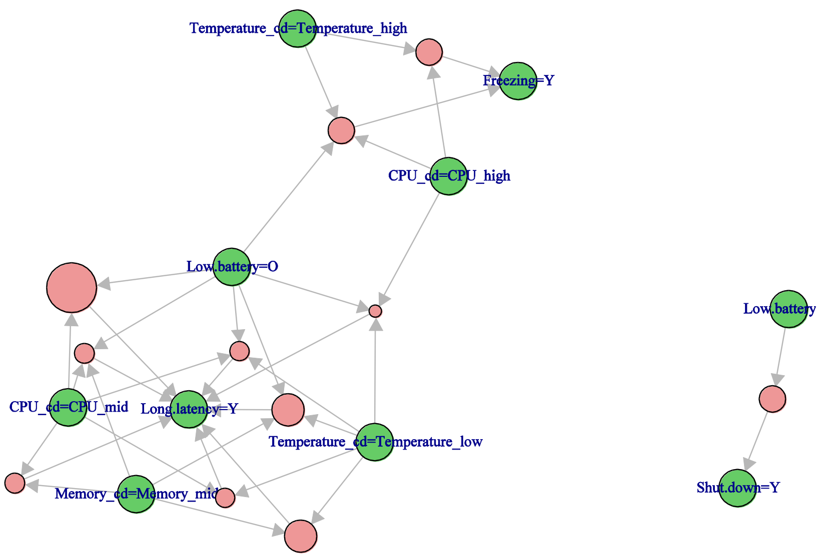

Figure 4. Graph for 11 rules.

IOP Conf. Series: Journal of Physics: Conf. Series **892** (2017) 012008 doi:10.1088/1742-6596/892/1/012008

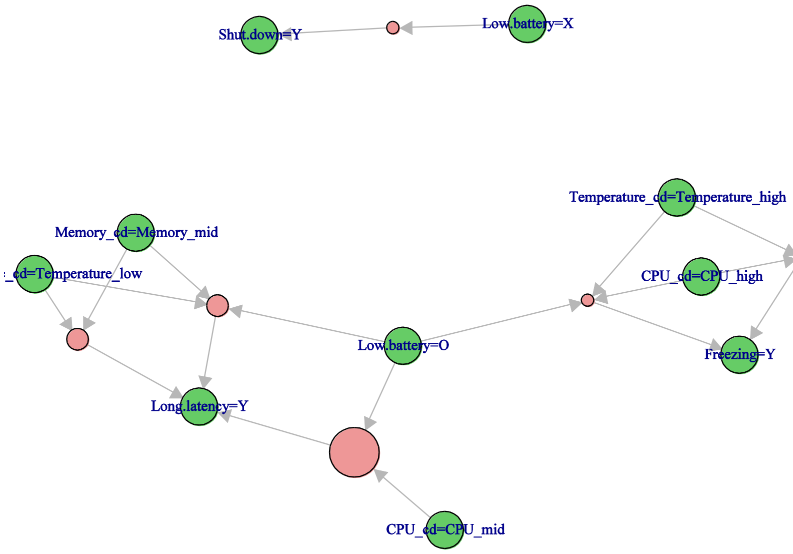

Figure 5. Graph for 6 rules.

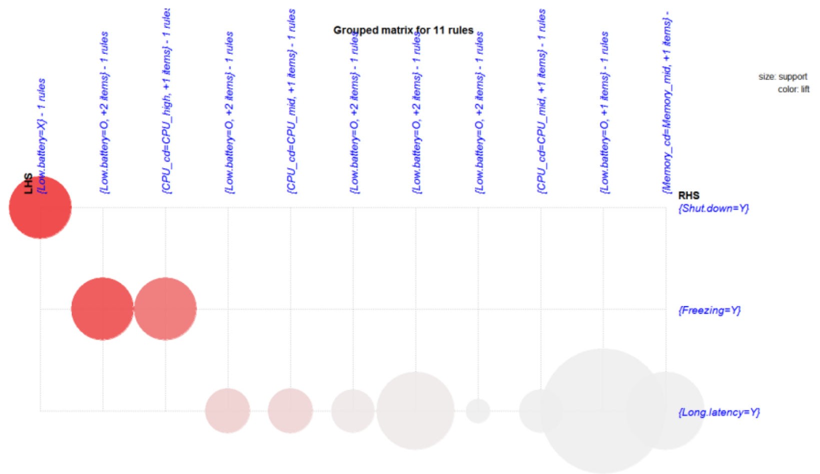

Figure 6. Grouped matrix for 11 rules.

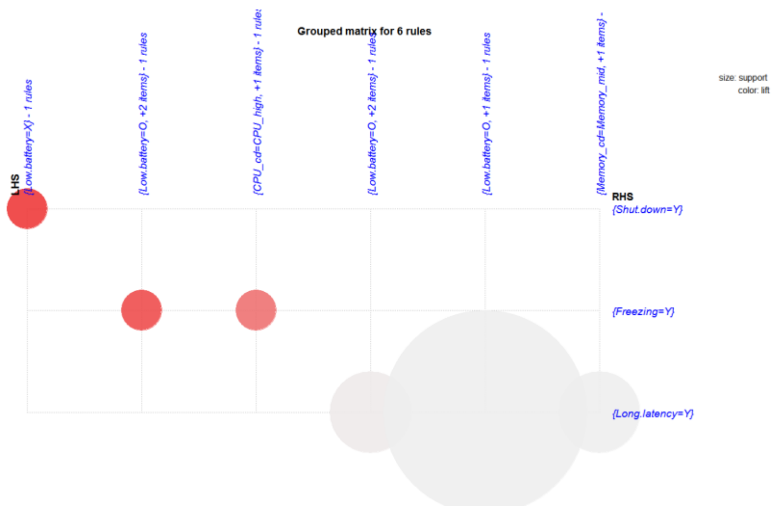

Figure 7. Grouped matrix for 6 rules.

4. Conclusions

We proposed an association rule-based predictive model for machine failure in industrial Internet of things (IIoT), which provides realistic association rules. Three major steps were considered in developing the model: 1) binarization, 2) rule creation, 3) visualization. Binarization transforms item values to a binary representation, and association rules were created using the Lattice model and Apriori algorithm. Visualization displays the identified rules in various ways. An experimental predictive model was implemented using R Studio version 3.3.2, and the identified association rules visualized using tables, graphs, and grouped matrixes. The implementation verified that the proposed predictive model for machine failure can provide realistic prediction results.

Acknowledgements

This work was supported by the Hallym Leading Research Group Support Program of 2017 (HRF-LGR-2017-0003).

References

[1] Hossain M S and Muhammad G 2016 *Comput. Netw.* **101** 192–202
[2] Posada J, Toro C, Barandiaran I, Oyarzun D, Stricker D, de Amicis R and Vallarino I 2015 *IEEE Comput. Graph. Appl. Mag.* **35** 26–40
[3] Kwon J -H, Lee H -H, Lim Y and Kim E -J 2016 *Appl. Sci.-basel* **6** 427:1–427:17
[4] Hashem I A T, Yaqoob I, Anuar N B, Mokhtar S, Gani A and Khan S U 2014 *Inform. Syst.* **47** 98–115
[5] Sahoo S K, Rodriguez P and Savinovic D 2015 *Proc. IEEE Int. Conf. on Electric Machines & Drives Conference (IEMDC)* pp 1398–1404
[6] https://www.rstudio.com/
[7] Zaki M J, Meira Jr W and Meira W, 2014 Itemset mining *Data mining and analysis: fundamental concepts and algorithms* (Cambridge: Cambridge University Press) chapter 8 pp 240–263