

Article

Identification of Defect Generation Rules among Defects in Construction Projects Using Association Rule Mining

Jungeun Park ¹, Yongwoon Cha ² , Hamad Al Jassmi ³ , Sangwon Han ^{1,*} and Chang-taek Hyun ¹

¹ Department of Architectural Engineering, University of Seoul, Seoul 02504, Korea; freedom8989@nate.com (J.P.); cthyun@uos.ac.kr (C.-t.H.)

² Department of Construction Policy Research, Korea Institute of Civil Engineering and Building Technology, Gyeonggi-Do 10223, Korea; ywcha@kict.re.kr

³ Department of Civil and Environmental Engineering, United Arab Emirates University, Al Ain 15551, UAE; h.aljassmi@uaeu.ac.ae

* Correspondence: swhan@uos.ac.kr; Tel.: +82-2-6490-2764

Received: 7 April 2020; Accepted: 7 May 2020; Published: 9 May 2020



Abstract: This study aims to identify the defect generation rules between defects, to support effective defect prevention at construction sites. Numerous studies have been performed to identify the relations between defect causes, to prevent defects in construction projects. However, identifying the inter-causal pattern does not yet guarantee an ultimate grasp of what constitutes proper defect mitigation strategies, unless the underlying defect-to-defect generation rules are thoroughly understood too. Specifically, if a defect generated in a work process is ignored without taking necessary corrective action, then additional defects could be generated in its following works as well. Thus, to minimize defect generation, this study analyzes the defects in the sequence of a construction work. To achieve this, the authors collected 9054 defect data, and association rule mining is used to analyze the rules between the defects. Consequently, 216 rules are identified, and 152 rules are classified into 3 categories along with 4 experts (71 expected rules, 22 unexpected but explainable rules, and 59 unexpected and unexplainable rules). The generation rules between the defects identified in this study are expected to be used to regularize various defect types to determine those that require priority management.

Keywords: defect generation rule; defect prevention; construction quality management; association rule mining; multi-family housing

1. Introduction

In construction projects, quality defects are major factors degrading the project performance. Quality defects at a construction site are extremely important issues for management, because they cause economic losses, such as schedule delays, cost overruns, and disputes because of rework. [1–4]. Love et al. [3] argued that the cost for the rework required owing to defects can reach approximately 10% of the construction cost. Therefore, it is important to identify the causes of such defects in an early stage and remove them in advance to prevent quality defects from occurring.

In this context, construction defects are recognized as factors that have to be prevented in advance, instead of being treated at a subsequent stage. However, because quality management in the construction stage depends on repairing based on inspections of the work results, rather than preventing defects earlier, similar forms of defects are continuously formed. Because these defects occur in numerous combinations, owing to their complex relations, their management is difficult [3,5–8].

In addition, there are numerous defect types to be managed and various combinations of defect causes; therefore, quality managers cannot appropriately determine the rules of defect generation that can be correlated.

To address this issue, numerous studies have been performed to identify the causes of quality defects and minimize them by eliminating the causes [1–3,7,8]. Previous studies have made efforts to identify relations between the defect causes, to prevent defects in construction projects [1–3,7,8]. These studies have attempted to prevent defects by identifying their single causes using statistical analysis and analyzing the relationship between the causes of the defects to determine the combinations that cause them.

These studies contributed to a better understanding of the relationship between the defects causes for the sake of drawing optimum proactive defect mitigation strategies. Nonetheless, the full picture of what constitutes a better prioritization strategy should also realize the implications of the various types of defects in terms of their evolving defect-to-defect generative mechanisms. Actually, at a construction site, if a defect generated in a construction work process is ignored without taking corrective action, another defect could be generated in connection with it in its following construction works. Understanding these implications would provide grounds for measuring defect severities, to complement root cause analyses for drawing optimum defect mitigation strategies. Therefore, in addition to the study that attempts to identify the complex generation structure between the causes, a study that attempts to prevent defects in advance by determining the generation rules between the possible defects is also necessary. Based on this recognition, the authors aim to develop a framework that can identify and analyze the complex generation structure between the various causes of construction defects and their systematic classification.

2. Literature Review

For management, it is important to understand the location and the characteristics with which defects are generated in a construction phase. To understand defect generation, the existing studies attempted to identify and analyze defects by the process, type, cause, and time. Kim [9] performed an analysis of defect types and causes using the defect performance data of Korean construction companies targeting high-rise residential buildings. Furthermore, he suggested a defect management measure in view of its results. Kwon [10] classified the defects that can be generated in a multi-family housing into those before and after occupation, and analyzed the generation type by the process. He further suggested five measures for defect reduction: applying a fair construction schedule, training, improving recognition for supervisors, and appropriate manpower arrangement according to the construction scale. Lee [11] identified defect generation causes by the type and process, using a survey on the annual defect generation status of a multi-family housing, and proposed measures to deal with each cause and prevent defects. Song [12] determined the types and causes in finishing a work based on the defects generated in interior finishing works in a multi-family housing. Furthermore, he divided the causes by the type into defects by materials and construction workers, defect by follow-up workers, and defect by equilibrium and furniture placement, and analyzed them. Kang [13] identified priority management targets by analyzing 1.71 million defects of 105 multi-family housings supplied by large construction companies. The existing studies [9–13] analyzing the types and characteristics of defect generation classify defect generation types by the construction process, in view of the collected cases, derived preventive measures, and priority management targets by estimating the impact and importance of the causes. In addition, the causes were derived by the classification of the defect types generated, and each cause was considered a factor that generates defects independently, because the analysis between the causes was not considered.

Defects can be generated by the root causes with the most influence on defect generation, and also by a combination of numerous causes. Existing studies that attempted to prevent defects by analyzing the relation between the root cause and the causes of defects have provided knowledge based on which an effective performance can be realized during quality management in a construction site. This is

achieved by estimating the importance and influence of the causes of defects to derive the root cause and by deriving a defect generating mechanism considering the relation between the causes.

Mireille [14] classified the causes of defects based on the condition, effect, nonconformance, and distress, to analyze the root cause of the defects that can be generated, and identified a sequential mechanism suggesting that causes are generated sequentially and not independently (Condition → Effect → Nonconformance → Distress). Love [3] proposed a causal model that causes reconstruction by focusing on the factors that are difficult for humans to identify based on defect cases, to identify the defect generation structure in terms of missing errors. Cui [15] surveyed by distinguishing the defect generation types, frequencies, and responsible agents for a defect-generated process in view of the defect cases to identify the causes of defects and the factors that affect them. In addition, each influence factor was subcategorized to derive the degree of influence on the defect generation that can be caused by the responsible agents, and the causes were classified into high, middle, and low levels, according to the degree of influence. Al Jassmi et al. [5] derived a defect generation mechanism using the “fault tree analysis”, with which they analyzed the root causes of defects. Furthermore, they identified the defect generation structure that the defects are generated sequentially (organization influence → defective supervision → preconditions for defective acts → defective acts) under the assumption that they are generated not by a single cause but by the causal relations between the causes. Furthermore, each of the causal factors was classified into the mechanism, and the influence of each cause was evaluated to select the priority management targets [7].

The existing studies [3,5,7,14,15] mentioned above attempted not only to analyze the root cause of defects but also to identify the mechanism in which the causing factors occur sequentially to generate defects by analyzing the relation between the causes and determining the defect generation structure. Most of the causes were identified as man, equipment, construction methods, materials, etc.; however, if defects are generated in the previous works, they will inevitably result in defects in the following works, and it is considered that there is a limit in resolving unpredictable defects only with the derived causes. In addition, because there are numerous combinations of causes that generate the defects derived from the existing studies, managing all the causes at the construction sites where the construction costs and period are to be considered is still limited. Particularly, numerous studies have attempted to derive the causes and identify the relations between the causes; however, a study that attempts to assist effective quality control work by identifying the generation structure between the possible defects is still lacking.

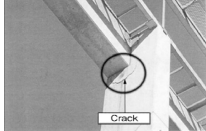
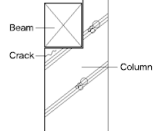
3. Data Collection

To identify the defect generation structure, 9054 cases of non-conformance reports (NCRs) and field action notices (FANs) were collected from 31 construction projects in Korea. The detailed information obtained from the NCRs and FANs include the defect cases, classification (for example, work type and facility), keyword (for example, crack and vertically faulty), and circumstantial description (see Table 1).

All the 9054 cases of NCRs and FANs collected were carefully analyzed and classified by the work type. The work type was classified into 14 types (for example, reinforced concrete work, brick work, carpenter work, furniture, and fixture work) specified in the NCRs and FANs. The most frequent defects were generated in the furniture and fixture works in 2984 cases (33.0%), followed by window work in 1803 cases (19.9%), and tile work in 1481 cases (16.4%) (Table 2).

Based on this classification, the defect types were refined. The items that were less related to the defects generated in other work types, such as scratches and damages, were excluded from the defect types, because they were unsuitable for the purpose of present study. Finally, a total of 66 defect types were reorganized considering the characteristics by the work type. With the reorganized defect types by work type, a basic data set was constructed based on the 9054 cases of NCR and FAN.

Table 1. Example of defect data (NCR).

Case	A Crack on a Column Damaged during Installation of Precast Beam		Code	W3031F421-001
Classification	Activity: W3031 Reinforced Concrete Pouring concrete	Facility: F421 Housing Apartments	Area: S19 Structural Section	Part: E222 Beam Concrete beam
Keyword Attributes	Crack, P.C beam Building Area: 5462 m2 / Structure: P.C			
Photo			Drawing	
Cause Analysis	Design Phase	[CD01] Omission [CD02] Error [CD03] Absent Countermeasure [CD04] Interference [CD05] Others CD03: Unwritten details concerning reinforcement for a weak part in the specification.		
	Construction Phase	[CC01] Insufficient Inspection [CC02] Inappropriate Material [CC03] Inappropriate Method [CC04] Poor installation [CC05] Interference [CC06] Others CC04: While the P.C column was shifted, the edge of the column was damaged, but installed. CC06: Training on transportation was not provided.		

Note : CD = Cause in terms of Design; CC = Cause in terms of Construction.

Table 2. Defect type classification by work type.

Work Type	Defect Type	No.	Sub Total	Ratio
[W1] Reinforced concrete work	[D1] Material separation	59	385	4.3%
	[D2] Crack	27		
	[D3] Fracture	16		
	[D4] Pockmark and surface blister	36		
	[D5] Poor covering depth	57		
	[D6] Poor horizontality and verticality	50		
	[D7] Cold joint	7		
	[D8] Deformation of formwork	21		
	[D9] Burst of formwork	18		
	[D11] Poor reinforcing bar placement/assembly	76		
	[D11] Swell	18		
[W2] Brick work	[D12] Crack	6	27	0.3%
	[D13] Poor slope	3		
	[D14] Separation	18		
[W3] Wood work	[D15] Deflection	42	80	0.9%
	[D16] Distortion	9		
	[D17] Gap	29		
[W4] Window work	[D18] Gap and separation	664	1803	19.9%
	[D19] Poor opening and closing	268		
	[D20] Deformation	134		
	[D21] Poor horizontality and verticality	453		
	[D22] Poor caulking	284		
[W5] Roof work	[D23] Breakage	2	16	0.2%
	[D24] Leakage	2		
	[D25] Poor drainage	12		

Table 2. Cont.

Work Type	Defect Type	No.	Sub Total	Ratio
[W6] Waterproof work	[D26] Separation	11	42	0.5%
	[D27] Contamination	2		
	[D28] Poor surface treatment	8		
	[D29] Crack	21		
[W7] Plastering work	[D30] Crack	49	103	1.1%
	[D31] Lack of plastering thickness	23		
	[D32] Separation	31		
	[D33] Separation	203		
[W8] Paper hangers work	[D34] Bad surface	97	452	5.0%
	[D35] Poor adhesion	129		
	[D36] Poor verticality	23		
	[D37] Condensation	174		
[W9] Insulation work	[D38] Gap and separation	142	399	4.4%
	[D39] Lack of thickness	83		
	[D40] Poor surface	83		
	[D41] Pockmark	46		
[W10] Painting work	[D42] Porosity	184	493	5.5%
	[D43] Separation	71		
	[D44] Crack	109		
	[D45] Step difference	61		
[W11] Tile work	[D46] Joint defect	103	1481	16.4%
	[D47] Separation	304		
	[D48] Crack	273		
	[D49] Fracture	648		
[W12] Furniture and fixture work	[D50] Poor horizontality and verticality	92	2984	33.0%
	[D51] Crack in joint	204		
	[D52] Poor size	797		
	[D53] Gap	924		
[W13] Metal work	[D54] Separation	181	746	
	[D55] Poor horizontality and verticality	103		
	[D56] Poor opening and closing	746		
	[D57] Deflection	29		
[W14] Stone and board work	[D58] Poor fixation	9	48	0.5%
	[D59] Corrosion/Rust	21		
	[D60] Poor welding state	18		
	[D61] Crack	129		
[W14] Stone and board work	[D62] Fracture	173	741	8.1%
	[D63] Contamination	69		
	[D64] Separation	275		
	[D65] Joint defect	11		
	[D66] Discoloration	84		
Total			9054	100%

Note : Wi = Work type; Di = Defect type.

4. Data Analysis

4.1. Analysis of Defect Generation Rules Using Association Rule Mining

Association Rule Mining (ARM) is defined as a methodology to discover interesting relationships hidden in the large amount of data sets [16]. ARM is one of the data mining techniques that acquire unlearned knowledge from a specific set of data and determine patterns that can explain the acquired knowledge [17,18]. In general, it is useful when determining the association of two or more groups. In other words, ARM is a useful method for not only verifying common sense and well-known rules (e.g., expected rules), but also finding hidden new rules (e.g., unexpected rules).

ARM can use the dataset to identify the rules that occur together frequently. The values derived by ARM are *Support*, *Confidence*, and *Lift*. '*Support*' is the probability value (i.e., $X \rightarrow Y$) that Y occurs when X occurs within the dataset and is obtained from Equation (1). For example, if horizontality and verticality faults are generated in the reinforced concrete work, there is a probability of generating a verticality fault in the window work simultaneously. The "*Support*" value is an index to measure how frequently two defects are generated simultaneously, and a standard to determine how important the rules are.

$$\text{Support}(x \rightarrow y) = P(x \cap y) \quad (1)$$

"*Confidence*" is a conditional probability value that "Y" is generated when "X" is generated in the ' $X \rightarrow Y$ ' relationship and is obtained from Equation (2). For example, when horizontal and vertical faults are generated in the reinforced concrete work, there is a probability that a verticality fault in the window work will occur. The "*Confidence*" value is a standard to present how much the rules can be trusted. Owing to the "*Support*" and "*Confidence*" being probability values, the closer they are to 1, the more connection they have.

$$\text{Confidence}(x \rightarrow y) = P(y | x) = \frac{P(x \cap y)}{P(x)} \quad (2)$$

The "*Lift*" value is a measure that denotes how much the conditional probability of "Y" increases when "X" is generated from the probability of "Y" relative to the total, and is obtained from Equation (3) [19]. Thus, it depicts how much "X" influences the probability of "Y" in the ' $X \rightarrow Y$ ' relationship. For example, it is a value that demonstrates how much the horizontality and verticality faults influence in reinforced concrete work, i.e., the probability of generating a verticality fault in the window work. The "*Lift*" value can be used to infer only those rules with higher correlations in the ' $X \rightarrow Y$ ' relationship. If the "*Lift*" value is higher than 1 (i.e., "*Lift*" > 1), it is considered to be in a positive correlation, and if it is identical to 1 (i.e., "*Lift*" = 1), it is regarded as an independent relationship, and if it is between 0 and 1 (i.e., $0 < \text{Lift} < 1$), it is a negative correlation with a low relationship.

$$\text{Lift}(x \rightarrow y) = \frac{P(y | x)}{P(y)} = \frac{P(x \cap y)}{P(x)P(y)} \quad (3)$$

ARM can derive the rules over three steps: (1) users define the min-support of '*Support*' to select rules beyond the min-support; (2) users define the min-confidence of '*Confidence*' to select rules beyond the min-confidence value; (3) select rules that the '*Lift*' value is greater than 1 (i.e., '*Lift*' > 1). Here, there is no evidence that a specific value is correct because the criteria for the min-support and min-confidence required for the ARM application are arbitrarily set by researchers in consideration of the study purpose and data characteristics [8,19].

Therefore, in present study, all the rules are considered important in that the min-support value is not set. Further, for the min-confidence value, 0.7 was applied, and the standard for that is in view of the value used in the existing studies [20,21]. In addition, owing to the purpose of present study, which is to identify the rules such that the defects in previous works may lead to defects in the future works, the rules in which the order of construction work process is reversed are excluded. A process of deriving the rules using ARM is displayed in Figure 1.

A total of 1237 rules were derived in the first ARM analysis stage using the basic dataset previously prepared. Subsequently the rules were filtered using the second stage (i.e., min support > 0.7), third stage (i.e., Lift > 1), fourth stage (i.e., work process). Using the filtering process, 216 significant rules were finally identified.

Table 3 shows 216 rules arranged according to the 'confidence' level measuring the strength of the relationship between defect causes. Rule 120 identified as the strongest rule shows that when a pockmark and surface blister (D4) occurs in the reinforced concrete work (W1), a discoloration (D66)

is 89% likely to occur in stone board work (W14). Rule 172, which turns out to be the second most powerful rule, suggests that D2 in W11 can generate D27 in W6 with a 86.2% probability.

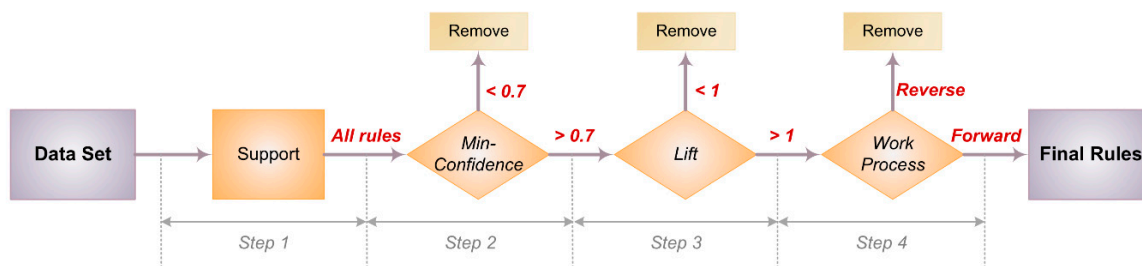


Figure 1. Defect rules filtering process.

Table 3. A strongest rules by confidence level.

Rule	Predecessor (i)		Successor (j)		Confidence
	Di	Wi	Dj	Wj	
120	D4	W1	D66	W14	0.890
172	D2	W11	D27	W6	0.862
57	D4	W1	D14	W7	0.861
30	D6	W1	D18	W4	0.857
33	D8	W1	D21	W4	0.857
17	D16	W3	D1	W1	0.856
20	D16	W3	D6	W1	0.856
10	D5	W1	D4	W1	0.854
92	D6	W1	D15	W12	0.853
132	D17	W3	D27	W6	0.853
109	D6	W1	D23	W5	0.846
207	D31	W7	D14	W12	0.845
129	D24	W5	D14	W2	0.843
168	D2	W7	D27	W6	0.843
16	D15	W3	D6	W7	0.842
67	D31	W7	D2	W7	0.842
....
178	D28	W6	D45	W11	0.710
188	D2	W7	D37	W9	0.710
74	D4	W1	D37	W9	0.703

4.2. Classification of Defect Generation Rules

The ARM technique, which is applied in view of the large amount of data, can derive numerous rules according to the number of data, owing to which useful rules were filtered using the min-support, min-confidence, and lift values. Furthermore, 216 rules were finally identified from 9054 cases of defect data, except for the rules with reversed construction work processes. However, there is a limit in analyzing only the rules with the probability (i.e., confidence) of generating rules and correlation (i.e., lift) in the ARM technique. Therefore, it is necessary to distinguish whether the identified rules are the expected rules or unexpected rules for construction practitioners to apply in real construction projects.

For this, in-depth interviews were conducted with four experts having domain knowledge and field experience to distinguish between expected rules and unexpected rules driven out of the data analysis through ARM (i.e., extracted rules) [16]. The interviewees consisted of one expert (MSc, and Professional Engineer) with more than 20 years in the field experiences, two experts (BSc) with more than 15 years in the quality management area (dealing with all of work type), and the other one (Ph.D.) with more than 10 years in the construction management research area. Results obtained from interviews with four people may not be sufficient to obtain statistical validity. However, given that

216 defect generation rules should be carefully analyzed, it is more effective to have in-depth interviews with a small number of experts than to survey many respondents.

The experts were asked to analyze 216 rules by classifying them into three categories: “Expected Rule”, “Unexpected but Explainable Rule”, and “Unexpected and Unexplainable Rule”. In this section, the process in Figure 2 is used to classify 216 rules into three types of useful rules (i.e., Expected Rule, Unexpected but Explainable Rule, and Unexpected and Unexplainable Rule).

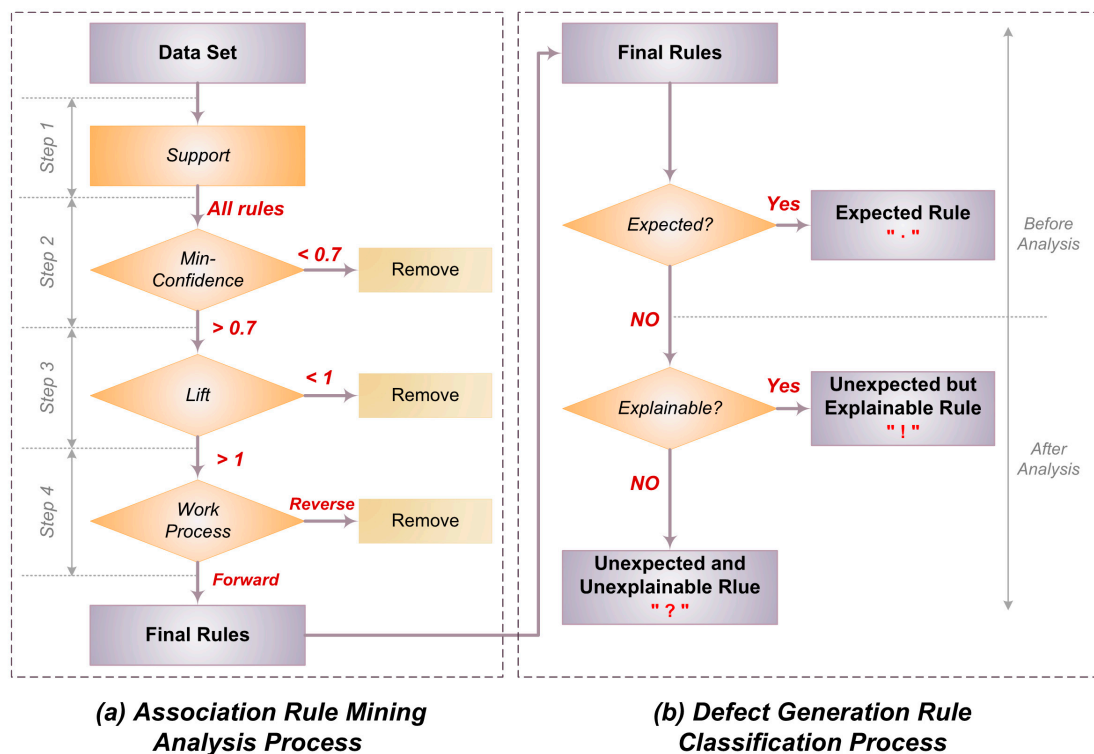


Figure 2. Integration process for defect generation rules classification.

First, the “Expected Rule” is a rule that they can easily expect without looking at the analysis results. For example, Rule 70, “a separation (D26) in the waterproofing work following a crack (i.e., D2) is generated in the reinforced concrete work in Table 3 can be referred to as “Expected Rule”.

Second, “Unexpected but Explainable Rule” is a rule that can be explained based on the analysis results among the rules unexpected before looking at the analysis results. Specifically, it can be referred to as a piece of information that was not known before but came to be known recently. For example, it is “Rule 76” in Table 3, the case that a separation (D26) is generated in the waterproofing work following horizontality and verticality faults (i.e., D6) that are generated in the reinforced concrete work”.

Finally, the “Unexpected and Unexplainable Rule” is a rule that is still difficult to explain among the rules never expected before looking at the analysis results. Similarly, to classifying rules into three, it is required to see if one is an expected rule before looking at the analysis results (i.e., Expected? in Figure 2b) and determine whether it is possible to explain the unexpected rules (i.e., Explainable? in Figure 2b).

For an accurate classification of rules, ARM and the details of the three useful rules were thoroughly explained to the experts before the classification. To classify easily into three types when interviewing with the experts, the interview paper was composed in such an approach that it can be marked by “•” (i.e., Expected Rule), “!” (i.e., Unexpected but Explainable Rule), and “?” (i.e., Unexpected and Unexplainable Rule) separately. The result of classifying the 216 rules for defect generation by the interview with the experts is as listed in Table 4.

Table 4. Example of defect generation rule classification by expert interview.

Rule	X (Di-Wi)	Y (Di-Wi)	Confidence	Experts Interview			
				A	B	C	D
1	D8-W1	D2-W1	0.714	●	●	●	●
2	D8-W1	D4-W1	0.800	!	●	!	!
...
28	D2-W1	D17-W4	0.751	?	?	?	?
29	D6-W1	D20-W4	0.834	●	●	●	●
...
54	D2-W1	D30-W7	0.721	●	●	●	●
...
64	D10-W1	D30-W7	0.729	!	?	!	!
65	D10-W1	D32-W7	0.791	!	!	!	!
...
154	D22-W4	D66-W14	0.754	?	?	?	?
155	D22-W4	D64-W14	0.765	!	?	?	!
...
206	D31-W7	D56-W12	0.751	?	?	?	?
...
216	D47-W11	D56-W12	0.754	?	?	?	?

Note: "●" Expected Rule; "!" Unexpected but Explainable Rule; "?" Unexpected and Unexplainable.

The rules marked as "●", "!", and "?" by all the four experts (see Table 4, Rules 1, 28, 29, 54, 65, 154, 206, and 216) indicate that their opinions were all in agreement. First, taking Rule 1 as an example in "Expected Rule" (Rule 1, 29, and 54), it depicts that if a form deformation (i.e., D8) occurs in the reinforced concrete work (i.e., W1), cracks will be generated in the concrete following the formwork. Specifically, it can be classified as an "Expected Rule" by anyone without the analysis results.

Further, "Rule 65" was classified as "Unexpected but Explainable Rule". To analyze it, if a reinforcing bar placement/assembly fault (i.e., D10) is generated in the reinforced concrete work (i.e., W1), cracks and pockmarks can be generated, causing separations (i.e., D32) in the plastering work (i.e., W7). Specifically, the rules presented an unexpected result in view of the experience of the experts' but can be explained logically.

Finally, Rules 28, 154, 206, and 216 were classified as "Unexpected and Unexplainable Rule". These rules are new information derived by the ARM analysis but are unexpected and unexplainable rules. However, they obtained significant values of "Confidence" and "Lift" from the ARM analysis, requiring additional analysis.

Not all rules have been agreed upon by four experts. There are also rules where there are differences of opinion among experts. For example, in Rule 2, expert B classified as "Expected" but the other three as "Unexpected but Explainable". This is because their experiences and knowledge levels are different. The rule 2 can be reclassified to a unanimous "Expected" after being explained by expert B. Another example of disagreement is rule 155. Experts A and D classified the rule as "Unexpected but Explicit", while experts B and C classified it as "Unexpected and Unexclusive". This means that experts A and D have not been fully persuaded by the explanations of experts B and C. We found it difficult to resolve these disagreements completely even after post-mortem. Furthermore, if we forced them to reach an agreement, prejudice could rather affect classification. Based on this recognition, we decided to analyze only the rules unanimously agreed upon by four experts and deal with the other remaining rules in our subsequent studies. As listed in Table 5, there are 152 rules that the opinions of the experts are all in agreement (70.4%) and the rules without agreement are 64 (29.6%).

Table 5. Expert interview results.

Classification	Criteria	Expert Opinion	No. of Rules	Rate (%)
Agreed experts' opinions	Expected Rule	•	71	32.9
	Unexpected Rule but Explainable Rule	!	22	10.2
	Unexpected and Unexplainable Rule	?	59	27.3
	Sub-total		152	70.4
Disagreed experts' opinions			64	29.6
Total			216	100

Note: "•" Expected Rule; "!" Unexpected but Explainable Rule; "?" Unexpected and Unexplainable.

5. Result and Discussion

The purpose of present study was to identify rules that the defects generated in previous works lead to defects in the future works to prevent the quality defects generated. Therefore, 152 rules were finally identified by conducting the ARM analysis on 9054 cases of defect data and expert interviews. The 152 rules were arranged in a matrix form so that they could be identified at an overview based on the type and defect (Figure 3).

As displayed in Figure 3, the "X" value of ARM is marked on the vertical axis and "Y" value on the horizontal axis; the rules of the work type (i.e., W_i) and defect type (i.e., D_i) are indicated as "•", "!", "?". In terms of the analysis of the rules, for example, if W_1-D_2 (i.e., X) occur in the Figure 3 " $X \rightarrow Y$ " form of ARM, it can be interpreted as "Expected Rule (i.e., "•") in which W_1-D_3 (i.e., Y) are generated. From Figure 3, the defect generation rules can be interpreted largely in three ranges arranged by the matrix (see Figure 3, ①, ②, and ③).

First, in ① (i.e., red section), most of the rules are "Expected Rules", which can be considered that the experts in quality management understand the relations between the defects that can be generated in the same work process or adjacent work process. However, in ②, it can be considered that owing to the rules in the work processes being remote from each other, over time they are identified as "Unexpected but Explainable Rule" and "Unexpected and Unexplainable Rule", and that the experts in quality management have not identified them yet. Therefore, it is expected to provide a new piece of information, i.e., the generation rules between the defects that can be generated in the distant (i.e., not close in sequence) work processes other than adjacent processes suggest a new direction for the quality management.

Next, in ② (i.e., gray section), it can be calculated how much the defects generated in the previous work process influence the defects in the future works process. For example, the reinforced concrete work is the previous process, and if a defect is generated in the process, numerous defects will be generated in the future work process. Specifically, the reinforced concrete work has a tremendous impact on the defects that can be generated in the future work's process. Therefore, it is expected to assist in determining and managing the previous work processes in advance and have a tremendous influence on the defect generation in the future process by using this matrix.

Lastly, ③ (i.e., blue section) presents the defects to be managed in the previous works to prevent defects generated in a specific work process. It also depicts that the defects that induce W_8-D_{33} in ③ are W_1-D_3 , W_7-D_{30} , D_{32} , and $W_8-D_{34, 35}$ (white box in ③ section). Therefore, it is expected to minimize the defects in the specific work process by identifying the type of the defects generated in the previous process.

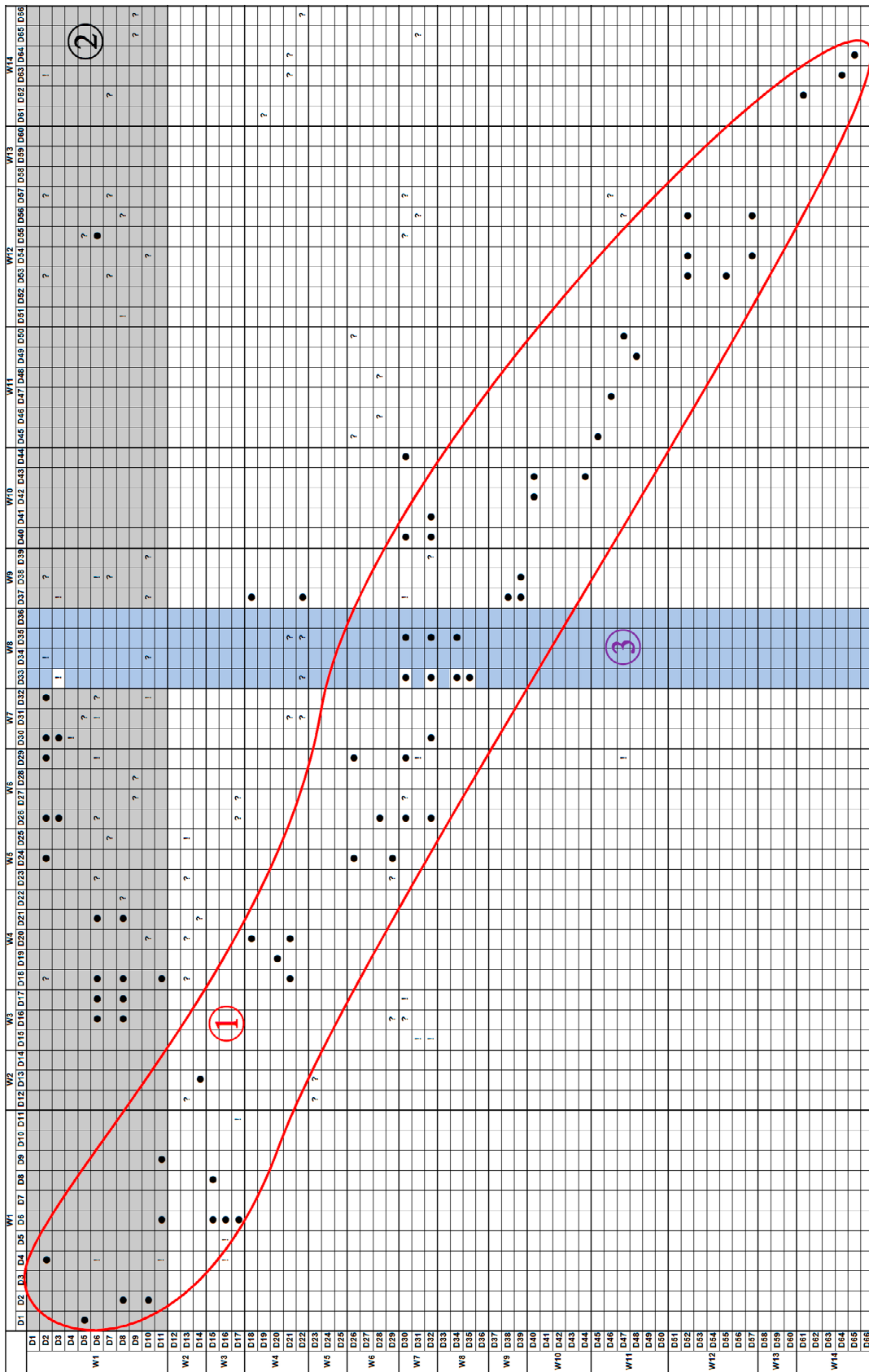


Figure 3. Defect generation rule matrix.

6. Conclusions

Defects in a construction site can cause economic losses, such as schedule delay and cost overrun owing to rework. Construction defects are recognized as factors that have to be prevented in advance, rather than be treated in the future. The aim of the study is to identify the rules for the defect generation that can be generated in the following construction works following being generated in the previous works. To achieve this, the authors used an association rule mining (ARM) to easily identify the relevant defect rules amongst a large amount of data.

A total of 1237 rules were derived through the ARM analysis of 9043 defects, followed by 216 rules selected through filtering process. The selected rules were classified into 71 “Expected” rules, 22 “Unexpected but Explainable” rules, 59 “Unexpected and Unexplainable” rules and 64 disagreements through in-depth interviews with four experts. Finally, 152 rules were presented in a matrix for preventive quality management in the construction phase. Using this matrix, when a defect occurs, it is easy to identify which defects are likely to generate in the following work process. In view of these determined facts, present study has identified the generation rules between defects considering the characteristics of the construction work that defects can be overlapped.

Furthermore, the study touched on a problem that quality management efforts are usually drawn based on quality managers’ experience, where in fact there are inter-relations between the various types of defects which are not necessarily comprehended by mere expert judgment. The rules between defects identified in present study can be utilized to prioritize defect prevention efforts to tackle those un-noticeably severe defects that has potential of propagating to create a series of other defects. In particular, it is relevant that the quality manager has identified the rules that were not known by his experience to suggest a new direction in the researches and practices of quality management.

Meanwhile, the study revealed some limitations. First, a deeper consideration is needed for the classification of the rules that experts did not agree on. Delphi techniques, Focus group interviews, etc., are expected to be effective in reaching their agreements. This agreement is expected to further present new management measures in quality control. Further in-depth interviews with more experts with extensive experience and knowledge are expected to elicit statistically validity on the analysis results. Finally, more research will also need to be conducted on the relationship between the defect generation rules and the cost of recovering their defects. This will greatly contribute to the development of a more cost-effective quality control strategy.

Author Contributions: Conceptualization, J.P. and S.H.; Methodology, J.P., H.A.J. and Y.C.; Validation, J.P., H.A.J. and S.H.; Formal Analysis, H.A.J. and S.H.; Investigation, C.H.; Data Curation, J.P. and Y.C.; Writing—Original Draft Preparation, J.P.; Writing—Review and Editing, H.A.J. and S.H. and C.-t.H.; Visualization, J.P. and Y.C.; Supervision, S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Josephson, P.E.; Hammarlund, Y. Causes and Costs of Defects in Construction a Study of Seven Building Projects. *Autom. Constr.* **1999**, *8*, 681–687. [[CrossRef](#)]
2. Mills, A.; Love, P.E.; Williams, P. Defect Costs in Residential Construction. *J. Constr. Eng. Manag.* **2009**, *135*, 12–16. [[CrossRef](#)]
3. Love, P.E.D.; Edwards, D.J.; Irani, Z.; Walker, D.H.T. Project Pathogens: The Anatomy of Omission Errors in Construction and Eesource Engineering Project. *IEEE Trans. Eng. Manag.* **2009**, *56*, 425–435. [[CrossRef](#)]
4. Hwang, B.-G.; Thomas, S.R.; Haas, C.T.; Caldas, C.H. Measuring the Impact of Rework on Construction Cost Performance. *J. Constr. Eng. Manag.* **2009**, *135*, 187–198. [[CrossRef](#)]
5. Al Jassmi, H.; Han, S. Analysis of Causes of Construction Defects Using Fault Trees and Risk Importance Measures. *J. Constr. Eng. Manag.* **2013**, *139*, 870–880. [[CrossRef](#)]
6. Al Jassmi, H.; Han, S.; Davis, S. Project Pathogens Network: New Approach to Analyzing Construction-Defects-Generation Mechanisms. *J. Constr. Eng. Manag.* **2014**, *140*, 04013028. [[CrossRef](#)]

7. Al Jassmi, H.; Han, S.; Davis, S. Analysis of the Complex Mechanisms of Defect Generation in Construction Projects. *J. Constr. Eng. Manag.* **2015**, *142*, 04015063. [[CrossRef](#)]
8. Lee, S.; Cha, Y.; Han, S.; Hyun, C. Application of Association Rule Mining and Social Network Analysis for Understanding Causality of Construction Defects. *Sustainability* **2019**, *11*, 618. [[CrossRef](#)]
9. Kim, H.J. The Analysis of Defects Type and Cause in High-Rise Residential Buildings. Master's Thesis, Dong-A University, Busan, Korea, 2015.
10. Kwon, J.H. Defect Diagnosis and Research on Reduction Countermeasure of Apartment House. Master's Thesis, Jeonbuk National University, Jeonju, Korea, 2010.
11. Lee, H.J. A Study on Defects Occurrence and Preventive Measures in Apartment Building Construction. Master's Thesis, Hanyang University, Seoul, Korea, 2011.
12. Song, G. An Analysis and Preventive Measures on the Defect Causes of Interior Finishing Works to New Apartment Building. Master's Thesis, Chung-Ang University, Seoul, Korea, 2013.
13. Kang, T. A Study on the Changes and Characteristics of Types of Defects According to the Changing Designs of Apartment Houses by Period. Master's Thesis, Yonsei University, Seoul, Korea, 2015.
14. Battikha, M. Reasoning Mechanism for Construction Nonconformance Root-Cause Analysis. *J. Constr. Eng. Manag.* **2008**, *134*, 280–288. [[CrossRef](#)]
15. Cui, J. Analysis of Construction Quality Accident Causes of Public Buildings based on Failure Study Theories. In Proceedings of the 2011 International Conference on Computer and Management (CAMAN), Wuhan, China, 19–21 May 2011; pp. 1–4.
16. Tan, P.-N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*; Pearson: New York, NY, USA, 2005.
17. Kamsu-Foguem, B.; Rigal, F.; Mauget, F. Mining association rules for the quality improvement of the production process. *Expert Syst. Appl.* **2013**, *40*, 1034–1045. [[CrossRef](#)]
18. Witten, I.H.; Frank, E.; Hall, M. *Data Mining: Practical Machine Learning Tools and Techniques (Google eBook)*, 3rd ed.; Morgan Kaufmann: Burlington, MA, USA, 2013; ISBN 0080890369.
19. Brijs, T.; Vanhoof, K.; Wets, G. Defining Interestingness for Association Rules. *Int. J. Inf. Theory Appl.* **2003**, *10*, 370–376.
20. Lee, E. *The Method of Wireless Mobile Networks Analysis Using the Association Rule Mining*; Soongsil University: Seoul, Korea, 2007.
21. Kim, S. *An Analysis of Relation between the Schoolwork Achievement and Computer Usage Pattern of the Students Using Association Rules and Correlations*; Chonam National University: Kwangju, Korea, 2008.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).