

# Root-Cause and Defect Analysis based on a Fuzzy Data Mining Algorithm

Seyed Ali Asghar Mostafavi Sabet  
Department of Industrial Engineering  
Science and Research Branch, Islamic  
Azad University  
Tehran, Iran

Alireza Moniri  
Department of Industrial Engineering  
Malayer Branch, Islamic Azad  
University  
Malayer, Iran

Farshad Mohebbi  
Department of Executive Management  
Science and Research Branch, Islamic  
Azad University  
Tehran, Iran

**Abstract**—Manufacturing organizations have to improve the quality of their products regularly to survive in today's competitive production environment. This paper presents a method for identification of unknown patterns between the manufacturing process parameters and the defects of the output products and also of the relationships between the defects. Discovery of these patterns helps practitioners to achieve two main goals: first, identification of the process parameters that can be used for controlling and reducing the defects of the output products and second, identification of the defects that very probably have common roots. In this paper, a fuzzy data mining algorithm is used for discovery of the fuzzy association rules for weighted quantitative data. The application of the association rule algorithm developed in this paper is illustrated based on a net making process at a netting plant. After implementation of the proposed method, a significant reduction was observed in the number of defects in the produced nets.

**Keywords**—Data mining; association rules; defect analysis; fuzzy sets; root cause analysis; quality

## I. INTRODUCTION

Manufacturing organizations have to improve the quality of their products regularly in order to survive in today's competitive production environment. The high quality of a product is an important factor for increasing customer satisfaction and market share; therefore, manufacturing organizations should have an extensive understanding of quality to compete in the international markets. From the ISO-9000 point of view, quality is "the totality of characteristics of an entity that bear on its ability to satisfy stated and implied need". Quality improvement means the promotion of standards and the reduction of product defects. A defect is a gap between the expected results and observed results [1]. Consequently, identifying product defects, determining their causes, and implementing corrective actions to reduce defects are essential and inevitable matters for manufacturing organizations.

It is generally difficult to identify the causes of a particular defect because the defect is not the outcome of a single cause, but occurs when a few associated causes combine [2]. There is a close relationship between occurrence of defects in the products and the manufacturing process parameters; i.e. the malfunction of these parameters can cause defects to occur in the products. The manufacturing process parameters can be categorized based on the following: man, machine, material, method, and environment. Controlling these parameters and

finding their relationships with the product defects will help Quality Improvement Teams (QIT) reduce and eliminate the defects.

This paper presents a methodology for identification of unknown patterns between the manufacturing process parameters and defects of the output products. Moreover, it identifies the relationships between the defects. Discovery of these patterns helps practitioners achieve three main goals:

- 1) Identification of the process parameters that can be used to control and reduce output product defects.
- 2) Identification of the defects that most probably have common roots.
- 3) Root cause analysis.

Since manufacturers usually face large data warehouses of manufacturing processes, data mining techniques can be used to exploit useful knowledge from these datasets. Data mining is a discipline that aims at extracting novel, relevant, valuable and significant knowledge from large databases. Data mining includes several tools such as decision trees, association rule mining (ARM), neural networks, fuzzy sets, statistical approaches, etc.

In this paper, a data mining algorithm is used to find fuzzy association rules on weighted quantitative data. The values of defects and parameters are expressed in fuzzy values, and the weights of defects and parameters are allocated according to their importance. The proposed technique will obtain interesting, understandable patterns discovered among the process parameters and output defects due to use of the concept of fuzzy sets and weights.

The rest of this paper is organized as follows: In the next section, the related works on fuzzy ARM and root cause analysis are outlined. Section 3 introduces the mathematical approach; Section 4 presents an application of the methodology in a netting plant, and provides a discussion of how to analyze defects in a net fabrication process using the results obtained from the algorithm; and finally, concluding remarks will be discussed in Section 5.

## II. RELATED WORKS

### A. Fuzzy Association Rule Mining

Association rule mining is a popular data mining technique due to its numerous applications in diverse areas. An

association rule is an expression of  $X \rightarrow Y$ , where  $X$  is a set of items, and  $Y$  is a single item [3]. For mining an association rule, two numeric values should be calculated: support and confidence. The support of an association rule is the proportion of transactions that contain both the antecedent and the consequent. The confidence of an association rule is the proportion of transactions containing the antecedent that also contains the consequent.

Agrawal et al. introduced several algorithms for extracting association rules from large databases [3], [4]. Moreover, different methods of association rule mining and their applications have been proposed by other researchers. In many algorithms for association rule mining, researchers have considered the relationships between transactions consisting of categorical attributes (categorized items) using binary values. However, transaction data in real-world applications usually consist of fuzzy and quantitative values.

In previous years, some work has been done on the use of fuzzy sets in discovering association rules. Miller and Yang applied Birch clustering to identify intervals and proposed a distance-based association rules mining process, which improves the semantics of the intervals [5]. To solve the qualitative knowledge discovery problem, Au and Chan applied fuzzy linguistic terms to relational databases with numerical and categorical attributes. Later, they proposed the F-APACS method to discover fuzzy association rules [6], [7]. Consequently Hong et al. proposed an algorithm for mining fuzzy rules from quantitative data [8]. They transformed each quantitative item into a fuzzy value using membership functions to find fuzzy rules. Fuzzy association rules are easily understandable to people because of the linguistic variables associated with fuzzy sets.

In association rule mining algorithms, minimum support value (minsup) and minimum confidence value (minconf) are used to measure the frequency and strength of the rules. In a database, some valuable items may not occur frequently; therefore, they may not be included in the final association rules. To solve this problem, some researchers have suggested reduction of the minsup and minconf values to include the rules containing valuable items. But these rules sometimes fail to comply with user objectives, because many irrelevant rules may be generated. For avoidance of this issue, some approaches have been introduced. Muyebe et al. tried to use the concept of weight in their new algorithm, and introduced a fuzzy weighted association rule mining algorithm with weighted support and confidence measures [9], [10]. Gyenesei also used weighted quantitative association rule mining based on a fuzzy approach (FWAR) [11]. However, his proposed algorithm was not suitable due to the data overflow problem. Thus, Olsen et al. proposed a method capable of solving this problem [12]. His is one of the most perfect, easy using algorithms proposed to identify association rules on fuzzy weighted data.

During recent decade a few researchers have tried to introduce more sophisticated approaches. Lin et al. introduced Compressed Fuzzy Frequent Pattern Tree (CFFPT) algorithm which integrates the fuzzy-set concepts and the FP tree-like approach to efficiently find the fuzzy frequent itemsets from

the quantitative transactions [13]. Also Moustafa et al. developed a novel technique named FFP\_USTREAM. This technique integrates fuzzy concepts with ubiquitous data streams, employing sliding window approach, to mine fuzzy association rules [14].

### B. Root Cause Analysis

Root cause analysis (RCA) is a process of analysis to define the problem, understand the causal mechanism underlying transition from desirable to undesirable condition, and to identify the root cause of problem in order to keep the problem from recurring [15]. There are a variety of methods as RCA tools: Cause-Effect Diagram, Fault Tree Analysis, Current Reality Tree, 5-Whys, Apollo Root Cause Analysis, Interrelationship Diagram, Barrier Analysis, System Process Improvement Model, Causal Factor Analysis, Event-Causal Analysis, Bayesian Inference, Failure Mode and Effects Analysis, Cause-Effect Matrix, etc.

In current century, due to development in intelligence science, some researchers have used data mining methods to analyze defects in manufacturing processes. Donauer et al. utilized a pattern recognition method to find the root causes of failures considering economic aspects [16]. Al-Salim recommended a data-mining-based methodology to assign quality improvement teams to investigate and eliminate the defects in manufacturing enterprises [17]. In the first stage, related defects are grouped based on an association-rule technique, and then, in the second stage, the groups of defects are allocated to the quality improvement teams based on a mathematical programming model that minimizes expected quality costs pertaining to the quality improvement process. A major deficiency of this algorithm is that it only uses binary datasets of defect occurrences, but does not take into account their frequency in each record.

During recent years some RCA methods have been developed based on using ARM techniques. Chen et al. introduced a method using association rule mining techniques for identification of root-cause machine sets that, most likely, are sources of defective products [18]. Sadoyan used a kind of association rule based on the rough set theory for manufacturing process control [19]. This algorithm extracts knowledge from large data sets obtained from manufacturing processes, and represents the knowledge using "if/then" decision rules. Then, the results obtained from the data mining algorithm are used for controlling the output of the manufacturing process. Lee et al. used the standard ARM algorithm to quantify the causality between defect causes, and social network analysis to find indirect causality among them [20]. Most of these researches are based on using standard ARM algorithm as a RCA tool. Since there are more information in expressing the occurrence of defects based on fuzzy values rather than binary ones, thus in this paper we introduce a novel RCA methodology based on using fuzzy weighted association rule mining algorithm.

## III. METHODOLOGY

The procedure for achieving the goals mentioned in the introduction consists of a two-stage framework. The first stage determines process breakdown, and the second stage identifies

hidden rules from manufacturing process databases using FWAR algorithm. Then, the obtained rules are analyzed to improve the process.

#### A. Process Breakdown Structure

In a process, the outputs are a function of the inputs. In a manufacturing process, as well, the product defects (outputs) are related to the process parameters (inputs). So as the first step of defect analyzing process, the input parameters of the manufacturing process should be recognized. These parameters that affect product defects can be categorized into the following main groups: man, machine, material, method, and environment. The recognized parameters and defects can be displayed through a structure (process breakdown structure) that can help practitioners to gain better perception of the process.

#### B. Relationships Recognition

In this section, we attempt to find hidden relationships between the specified process parameters and defects using Olson's modified FWAR algorithm.

##### Notation:

n: the total number of data observation records;

m: the total number of parameters;

z: the total number of sub-parameters;

$P_{jk}$ : the  $k^{\text{th}}$  quantitative sub-parameter from the  $j^{\text{th}}$  parameter, where  $j=1$  to  $m$ ,  $k=1$  to  $z$ , and  $j=1$  to  $m-1$  are parameters and  $j=m$  is a defect;

$|P_{jk}|$ : the number of fuzzy regions of  $P_{jk}$ ;

$R_{jkt}$ : the  $t^{\text{th}}$  fuzzy region of  $P_{jk}$ ,  $1 \leq t \leq |P_{jk}|$ , called item;

$w_{jkt}$ : the weight of  $R_{jkt}$ ,  $0 \leq w_{jkt} \leq 1$ ;

$E^{(i)}$ : the  $i^{\text{th}}$  record,  $1 \leq i \leq n$ ;

$V_{jk}^{(i)}$ : the quantitative value of  $P_{jk}$  for  $E^{(i)}$ ;

$f_{jkt}^{(i)}$ : the membership value of  $V_{jk}^{(i)}$  in  $R_{jkt}$ ,  $0 \leq f_{jkt}^{(i)} \leq 1$ ;

Sup ( $R_{jkt}$ ): the calculated support value of  $R_{jkt}$ ;

Sup: the calculated support value of each candidate itemset;

Conf: the calculated confidence value of each large itemset;

minsup: the predefined minimum support value;

minconf: the predefined minimum confidence value;

$C_r$ : the set of candidate itemsets with  $r$  items;

$L_r$ : the set of large itemsets with  $r$  items.

#### Algorithm:

**Input:** n, m, z,  $w_{jkt}$ , the membership function of each item, minsup and minconf;

**Output:** fuzzy association rules.

**Step 1:** Transform the quantitative value  $V_{jk}^{(i)}$  of each record  $E^{(i)}$ ,  $i=1$  to  $n$ , for each  $P_{jk}$ ,  $j=1$  to  $m$ ,  $k=1$  to  $z$ , into fuzzy membership values  $f_{jkt}^{(i)}$  ( $1 \leq t \leq |P_{jk}|$ ) using the given membership function of  $R_{jkt}$ .

**Step 2:** Calculate Sup ( $R_{jkt}$ ):

$$\text{Sup} (R_{jkt}) = \frac{\sum_{i=1}^n w_{jkt} f_{jkt}^{(i)}}{n} \quad (1)$$

for  $j=1$  to  $m$ ,  $k=1$  to  $z$ ,  $1 \leq t \leq |P_{jk}|$ , the support value of fuzzy region  $R_{jkt}$ , to form  $C_1$ , the set of candidate 1-itemsets.

**Step 3:** If  $\text{Sup} (R_{jkt}) \geq \text{minsup}$ , then store  $R_{jkt}$  in  $L_1$ , the set of large 1-itemsets.

**Step 4:** If  $L_1$  is not null, then do the next step; otherwise, exit the algorithm.

**Step 5:** The algorithm first joins together large itemsets in  $L_r$  under the condition that  $r-1$  items in the two itemsets are the same, and the other one is different; then, the algorithm retains in  $C_{r+1}$  the itemsets for which all the sub-itemsets of  $r$  items exist in  $L_r$  and which do not have any two items  $R_{jkt}$  and  $R_{jkt}$  ( $p \neq q$ ) of the same  $P_{jk}$ ; the itemsets are called candidate  $r$ -itemsets.

**Step 6:** Do the following sub-steps for each newly formed  $(r+1)$ -itemset  $S$  with items  $(S_1, S_2, \dots, S_x, \dots, S_{r+1})$  in  $C_{r+1}$ ,  $1 \leq x \leq r+1$ .

a) Calculate the fuzzy value of each record  $E^{(i)}$  of  $S$  as

$$f_s^{(i)} = \bigwedge_{x=1}^{r+1} w_{s_x} f_{s_x}^{(i)} \quad (2)$$

where  $f_{s_x}^{(i)}$  is the membership value of  $E^{(i)}$  in fuzzy region  $S_x$ ,  $w_{s_x}$  is the weight of item  $S_x$ . If the minimum operator is used for the intersection, then

$$f_s^{(i)} = \min_{x=1}^{r+1} w_{s_x} f_{s_x}^{(i)} \quad (3)$$

b) Calculate the support value Sup ( $S$ ) of  $S$  in the record as

$$\text{Sup} (S) = \frac{\sum_{i=1}^n f_s^{(i)}}{n} = \frac{\sum_{i=1}^n \min_{x=1}^{r+1} w_{s_x} f_{s_x}^{(i)}}{n} \quad (4)$$

c) If  $\text{Sup} (S) \geq \text{minsup}$ , then store  $S$  in  $L_{r+1}$ .

**Step 7:** If  $L_{r+1}$  is null, then do the next step; otherwise, set  $r=r+1$  and repeat steps 5 to 6.

**Step 8:** Collect the large itemsets together.

**Step 9:** Construct association rules for each large  $q$ -itemset  $S$  with items  $S_1, S_2, \dots, S_q$ ,  $q \geq 2$ , using the following sub-steps:

a) Form each possible association rule as follows:

$$S_1 \wedge S_2 \wedge \dots \wedge S_a \wedge S_y \wedge \dots \wedge S_q \rightarrow S_t \quad (5)$$

$$(t=1 \text{ to } q, a=t-1, y=t+1)$$

b) Calculate the confidence value of each association rule, using

$$\text{Conf}(S1 \wedge S2 \wedge \dots \wedge Sa \wedge Sy \wedge \dots \wedge Sq \rightarrow St) = \frac{\sum_{i=1}^n \min_{t=1}^q w_{s_t} f_{s_t}^{(i)}}{\sum_{i=1}^n \min(\min_{t=1}^a w_{s_t} f_{s_t}^{(i)}, \min_{t=y}^q w_{s_t} f_{s_t}^{(i)})} \quad (6)$$

**Step 10:** Output the relative and interesting association rules with

$$\text{Conf}(S1 \wedge S2 \wedge \dots \wedge Sa \wedge Sy \wedge \dots \wedge Sq \rightarrow St) \geq \text{minconf} \quad (7)$$

From Step 10, three kinds of rules can be obtained:

- 1) Process Parameter(s) → Defect
  - For controlling and reducing output product defects.
  - For root cause analysis.
- 2) Defect(s) → Defect
  - For identification of the defects that most likely have common roots.
- 3) Process Parameter(s) → Process Parameter
  - For identification of the relations between parameters to help control and reduce defects.

#### IV. AN INDUSTRIAL APPLICATION: FISH-NET MANUFACTURING PROCESS

This section presents an application of the introduced algorithm in a fish-net manufacturing plant. As it is shown in Fig. 1, the fish-net manufacturing process has five major subsections as follows:

- 1) net making,
- 2) inspection and repair,
- 3) dyeing and dehydrating,
- 4) net stretching,
- 5) packing.

The most important step in this process is net making, which is performed by special machines. If the produced nets at this stage have many defects, the cost and the time of inspection and repair in the next step will increase. Also, some defects lead to defective nets that cannot be repaired. The net making process parameters such as the performance of the machines, workers' skills, and quality of the strings could impact the net defects. Fig. 2 presents some meshes without any deficiencies. This paper is focused on using the algorithm in the net making process to identify unknown rules between the net making process parameters and the defects of the output nets and also to identify the relationships among the defects. These rules can help practitioners to find 1) the causes of the defects that have occurred; 2) interrelated defects with common roots; and 3) process parameters that can be used for controlling and reducing the output net defects.

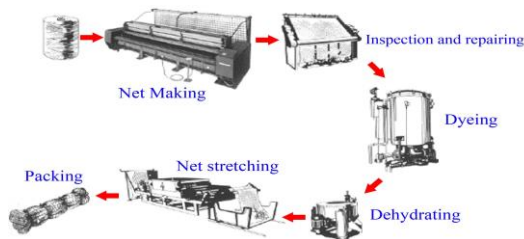


Fig. 1. Fish-net manufacturing process.

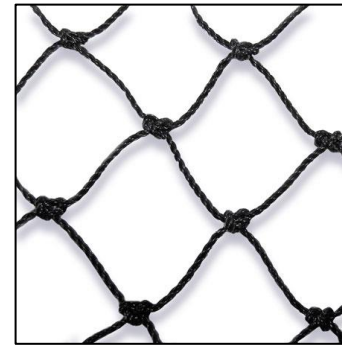


Fig. 2. Perfect fish-net meshes.

#### A. Breakdown Structure of the Net Making Process

First, the breakdown structure of the net making process is to be defined for making a standard scheme for stating the process parameters and defects. After consulting some experts, the manager of the net making section provided the breakdown structure, and specified the variables that must be considered for recognition of the relationships.

#### B. Identifying Hidden Relations

After developing the process breakdown structure, we applied the introduced algorithm to find the relationships between the net making parameters and defects. The information on the defects and process parameters is shown in Tables 1 and 2, respectively. In this section, we have used only 10 records of the net making process to show the performance of the algorithm in an industrial application (as shown in Table 3). The software program developed in MATLAB is used for execution of the rule generation algorithms introduced in this paper.

**Step 1:** The quantitative values in Table 3 are transformed into fuzzy values using the membership functions given in Fig. 3.

**Step 2:** The Sup ( $R_{jkt}$ ) values are calculated.

$$C_1 = \{(R_{111}), (R_{112}), (R_{113}), (R_{121}), (R_{122}), (R_{123}), \dots, (R_{451}), (R_{452}), (R_{453})\} \quad (8)$$

**Step 3:** The process specialists recommended  $p = 0.2$ . If  $\text{Sup}(R_{jkt}) \geq \text{minsup}$ , then  $R_{jkt}$  is stored in the set of large 1-itemsets ( $L_1$ ).

$$L_1 = \{(R_{121}), (R_{133}), (R_{223}), (R_{233}), (R_{321}), (R_{412}), (R_{422}), (R_{423}), (R_{432}), (R_{442}), (R_{443}), (R_{452}), (R_{453})\} \quad (9)$$

**Step 4:**  $L_1$  set is not null, so we go to the next step.

**Step 5:** According to the itemsets in  $L_1$ , candidate  $C_2$  is generated.

$$C_2 = \{(R_{121}, R_{133}), (R_{121}, R_{223}), (R_{121}, R_{233}), (R_{121}, R_{321}), (R_{121}, R_{412}), (R_{121}, R_{422}), (R_{443}, R_{452}), (R_{443}, R_{453})\} \quad (10)$$

Note that itemsets such as  $(R_{422}, R_{423})$ , having categorical classes of the same process parameters or defects, would not be retained in  $C_2$ .

TABLE I. NET MAKING DEFECT INFORMATION

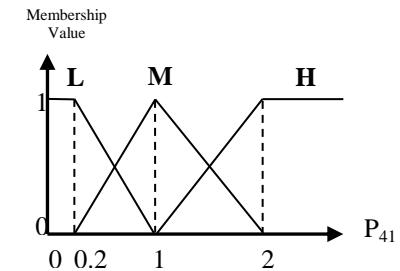
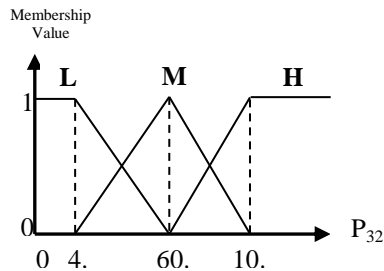
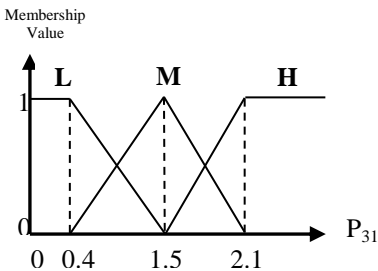
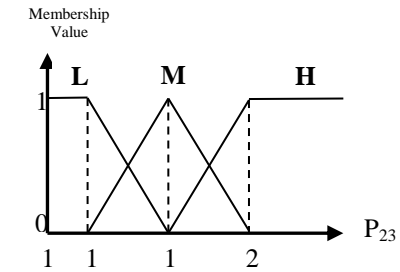
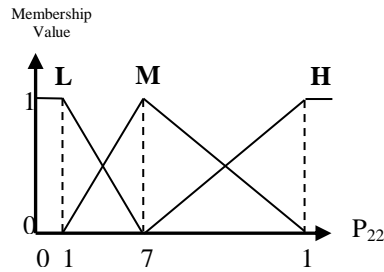
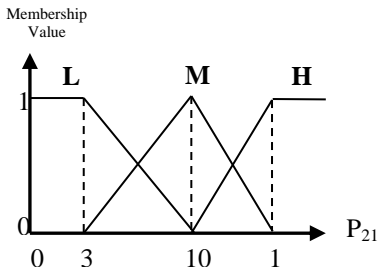
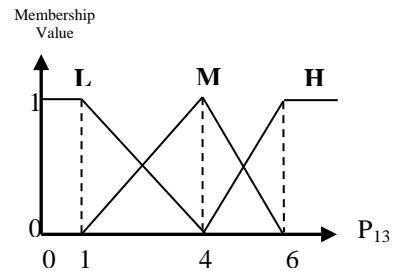
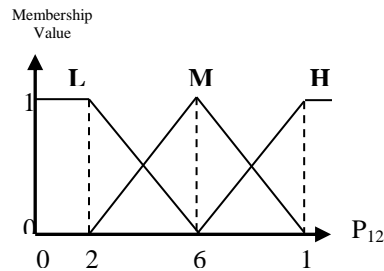
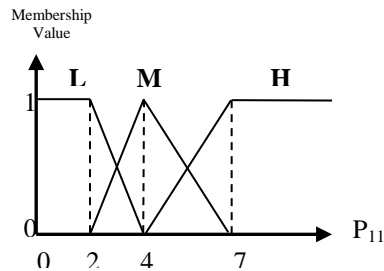
Defect	Indicator	Unit
Net tear	P <sub>41</sub>	Number in 50000 meshes
Knotless	P <sub>42</sub>	Number in 50000 meshes
Deviation from expected weight	P <sub>43</sub>	KG
Deviation from expected mesh size	P <sub>44</sub>	mm
Unexpected mesh shape	P <sub>45</sub>	Number in 50000 meshes

TABLE II. NET MAKING PARAMETER INFORMATION

Process Variable	Indicator	Unit
Time elapsed since beginning of shift	P <sub>11</sub>	Hour
Record of service	P <sub>12</sub>	Month
Number of mistakes	P <sub>13</sub>	Number in the last 3 shifts
Time for preparing machines	P <sub>21</sub>	Minute
Machine age	P <sub>22</sub>	Year
Net making speed	P <sub>23</sub>	Net row in one minute
String thickness	P <sub>31</sub>	mm
String resistance	P <sub>32</sub>	KG/Meter

TABLE III. DATA OF NET MAKING PROCESS PARAMETERS

Sample	Man (P <sub>1</sub> )			Machine (P <sub>2</sub> )			Material (P <sub>3</sub> )		Defect (P <sub>4</sub> )				
	P <sub>11</sub>	P <sub>12</sub>	P <sub>13</sub>	P <sub>21</sub>	P <sub>22</sub>	P <sub>23</sub>	P <sub>31</sub>	P <sub>32</sub>	P <sub>41</sub>	P <sub>42</sub>	P <sub>43</sub>	P <sub>44</sub>	P <sub>45</sub>
1	1.5	1.5	3	8	2	19	1.55	52	0.6	5	0.7	7	0.25
2	3.5	11	6	3	15	21	0.6	9	1.8	0.75	1.4	16	0.2
3	6.5	5.5	6	4	15	21	1.95	81	0.2	2	0.7	14	1.8
4	5	1.5	4	15	15	21	0.93	20	1.2	3.5	0.15	3	1.4
5	3.5	2	5	17	15	20.5	0.7	13	1	6	0.6	9	1.7
6	7.5	7	7	9	17	22	0.75	13.3	0.7	3	0.75	5	0.75
7	3	14	5	4	16	20	0.62	8.9	0.9	3.5	0.75	5	2.5
8	4.5	2	5	14	6	17	2.28	105	0.85	3.8	0.65	4.5	3.5
9	7	2.5	6	11	17	20	0.65	9.3	1.9	2	0.4	7.5	2.75
10	7.5	5	7	3.5	16	20	2.28	105	0.75	1	1.25	13	1.5



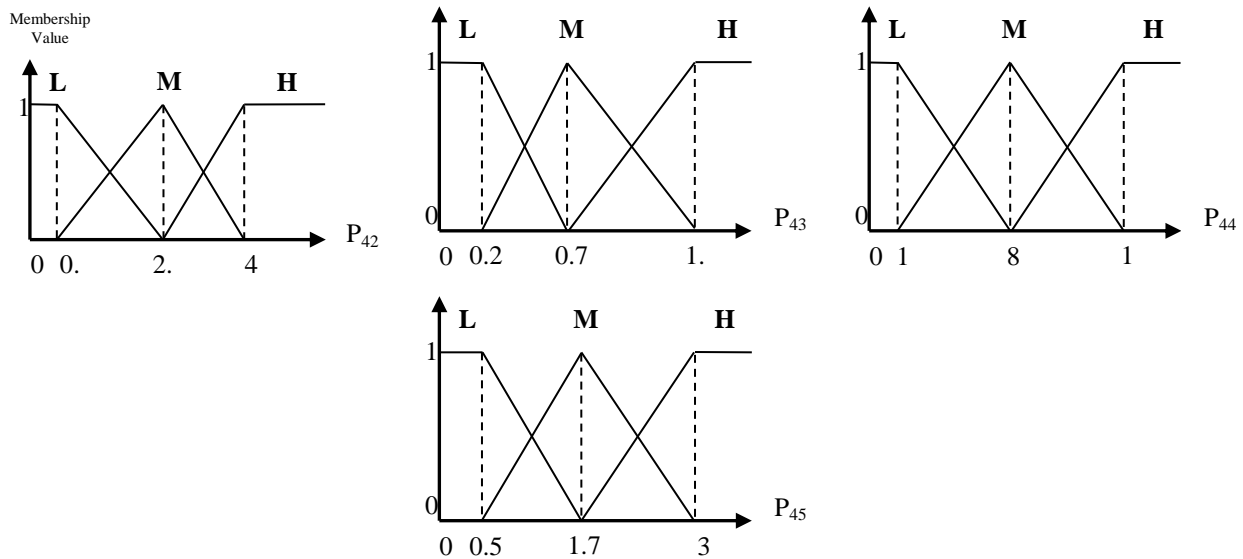


Fig. 3. Fuzzy functions of sub-parameters and defects.

**Step 6:** The following sub-steps are performed for each newly formed candidate 2-itemset.

a) The membership value of each 2-itemset is calculated. For example, consider the  $(R_{121}, R_{133})$  set. The membership function values for sample 1 are calculated as 1 for  $R_{121}$  and zero for  $R_{133}$ :

$$\min(0.4 \times 1, 0.4 \times 0) = 0 \quad (11)$$

b) The support value is calculated for each candidate 2-itemset in  $C_2$ .

$$\text{Sup}((R_{121}, R_{133})) = 0.12 \quad (12)$$

The 2-itemsets in  $C_2$  the support values of which are equal to or greater than minsup are shown in Table 4.

c) The candidate 2-itemsets the support values of which are equal to or greater than minsup are stored in  $L_2$ .

$$L_2 = \{(R_{121}, R_{442}), (R_{133}, R_{223}), (R_{133}, R_{233}), (R_{223}, R_{233}), (R_{223}, R_{321}), (R_{223}, R_{412}), (R_{223}, R_{442}), (R_{223}, R_{452}), (R_{412}, R_{423}), (R_{412}, R_{432}), (R_{412}, R_{442}), (R_{423}, R_{442}), (R_{432}, R_{442})\} \quad (13)$$

**Step 7:** Since the  $L_2$  set is not null, Steps 5 and 6 are repeated to find  $L_3$ .  $C_3$  is generated from  $L_2$ .

$$C_3 = \{(R_{121}, R_{442}, R_{133}), (R_{121}, R_{442}, R_{233}), (R_{121}, R_{442}, R_{233}), (R_{121}, R_{442}, R_{321}), (R_{423}, R_{442}, R_{432})\} \quad (14)$$

The 3-itemsets in  $C_3$  whose Support values are equal or greater than minsup are shown in Table 5.

TABLE IV. SUPPORT VALUES GREATER THAN MINSUP

Itemsets	Sup	Itemsets	Sup
$(R_{121}, R_{442})$	0.21	$(R_{223}, R_{452})$	0.20
$(R_{133}, R_{223})$	0.23	$(R_{412}, R_{423})$	0.26
$(R_{133}, R_{233})$	0.20	$(R_{412}, R_{432})$	0.22
$(R_{223}, R_{233})$	0.24	$(R_{412}, R_{442})$	0.27
$(R_{223}, R_{321})$	0.21	$(R_{423}, R_{442})$	0.27
$(R_{223}, R_{412})$	0.21	$(R_{432}, R_{442})$	0.25
$(R_{223}, R_{442})$	0.21		

TABLE V. SUPPORT VALUES GREATER THAN MINSUP

Itemsets	Sup
$(R_{133}, R_{223}, R_{233})$	0.20
$(R_{412}, R_{423}, R_{442})$	0.22
$(R_{412}, R_{432}, R_{442})$	0.21

Thus,

$$L_3 = \{(R_{133}, R_{223}, R_{233}), (R_{412}, R_{423}, R_{442}), (R_{412}, R_{432}, R_{442})\} \quad (15)$$

The 4-itemsets in  $C_4$  and their support values are shown in Table 6. All the support values are less than minsup, so  $L_4$  is null. Then, step 8 begins.

$$C_4 = \{(R_{133}, R_{223}, R_{233}, R_{412}), (R_{133}, R_{223}, R_{233}, R_{423}), (R_{133}, R_{223}, R_{233}, R_{442}), (R_{133}, R_{223}, R_{233}, R_{432}), (R_{412}, R_{423}, R_{442}, R_{432})\} \quad (16)$$

**Step 8:**  $L_1$ ,  $L_2$  and  $L_3$  are collected.

**Step 9:** All the impossible association rules for the itemsets of  $L_2$  and  $L_3$  and their confidence values are shown in Table 7.

**Step 10:**  $\text{minconf} = 0.85$  was recommended by the process specialists. The association rules the confidence values of which are equal to or greater than  $\text{minconf}$  are the outputs of the algorithm (see Table 8).

TABLE VI. SUPPORT VALUES

Itemsets	Sup
( $R_{133}, R_{223}, R_{233}, R_{412}$ )	0.12
( $R_{133}, R_{223}, R_{233}, R_{423}$ )	0.06
( $R_{133}, R_{223}, R_{233}, R_{442}$ )	0.13
( $R_{133}, R_{223}, R_{233}, R_{432}$ )	0.15
( $R_{412}, R_{423}, R_{442}, R_{432}$ )	0.17

TABLE VII. RULES OBTAINED FROM STEP 9 AND THEIR CONFIDENCE VALUES

Rule	Conf	Rule	Conf	Rule	Conf
$R_{121} \rightarrow R_{442}$	0.88	$R_{223} \rightarrow R_{442}$	0.70	$R_{432} \rightarrow R_{442}$	0.83
$R_{442} \rightarrow R_{121}$	0.53	$R_{442} \rightarrow R_{223}$	0.53	$R_{442} \rightarrow R_{432}$	0.63
$R_{133} \rightarrow R_{223}$	0.88	$R_{223} \rightarrow R_{452}$	0.67	$R_{133}, R_{223} \rightarrow R_{233}$	0.87
$R_{223} \rightarrow R_{133}$	0.77	$R_{452} \rightarrow R_{223}$	0.67	$R_{133}, R_{233} \rightarrow R_{223}$	0.95
$R_{133} \rightarrow R_{233}$	0.77	$R_{412} \rightarrow R_{423}$	0.76	$R_{233}, R_{233} \rightarrow R_{133}$	0.83
$R_{233} \rightarrow R_{133}$	0.80	$R_{423} \rightarrow R_{412}$	0.81	$R_{412}, R_{423} \rightarrow R_{442}$	0.85
$R_{223} \rightarrow R_{233}$	0.80	$R_{412} \rightarrow R_{432}$	0.65	$R_{412}, R_{442} \rightarrow R_{423}$	0.81
$R_{233} \rightarrow R_{223}$	0.96	$R_{432} \rightarrow R_{412}$	0.73	$R_{423}, R_{442} \rightarrow R_{412}$	0.81
$R_{223} \rightarrow R_{321}$	0.70	$R_{412} \rightarrow R_{442}$	0.79	$R_{412}, R_{432} \rightarrow R_{442}$	0.95
$R_{321} \rightarrow R_{223}$	1	$R_{442} \rightarrow R_{412}$	0.68	$R_{412}, R_{442} \rightarrow R_{432}$	0.78
$R_{223} \rightarrow R_{412}$	0.70	$R_{423} \rightarrow R_{442}$	0.84	$R_{432}, R_{442} \rightarrow R_{412}$	0.84
$R_{412} \rightarrow R_{223}$	0.62	$R_{442} \rightarrow R_{423}$	0.68		

TABLE VIII. FINAL RULES

	Selected rules
1	$R_{121} \rightarrow R_{442}$
2	$R_{133} \rightarrow R_{223}$
3	$R_{233} \rightarrow R_{223}$
4	$R_{321} \rightarrow R_{223}$
5	$R_{133}, R_{223} \rightarrow R_{233}$
6	$R_{133}, R_{233} \rightarrow R_{223}$
7	$R_{412}, R_{423} \rightarrow R_{442}$
8	$R_{412}, R_{432} \rightarrow R_{442}$

### C. Discussion

Association rules discover patterns in a database. Analysis and evaluation of whether or not rules are meaningful is based on the analyzer’s viewpoint. In Table 8, Rule 1 shows a relation between the net making process parameters and the defects. These kinds of rules can help net manufacturers achieve two main goals:

- Identification of the net making process parameters which can be used for controlling and reducing the output net defects.
- Root cause identification when a defect occurs.

Consider Rule  $R_{121} \rightarrow R_{442}$ . It means that if the record of service of an operator ( $P_{12}$ ) is low, the deviation from expected

mesh size defect will occur at a medium level. Rules 2 to 6 show the relations between the net making process parameters; these relations can be used to regulate the process parameters to control the net defects. Rules 7 and 8 show the relations between the net defects. These kinds of rules can be used not only in specification of the net defects that very probably have common roots but also in identification of the net defects that can impact other defects. For example, Rule  $R_{412}, R_{423} \rightarrow R_{442}$  shows that if “net tear” is medium and “knotless” is high, then the deviation from expected mesh size defect will be medium.

In this paper, we used a process dataset consisting of 10 records only to introduce the application of fuzzy weighted association rules in a net making process. Evidently, for attaining useful, valid rules from data mining algorithms, large-sized databases must be used. Although we applied the algorithm during a performance improvement project at a fish-net manufacturing plant with a dataset consisting of 850 records, the results helped the management to have a better perception of the process to control and reduce net defects and minimize the costs. After implementing the method and conducting an improvement meeting, we observed a significant reduction in the rate of defects in the produced nets.

### V. CONCLUSION AND RECOMMENDATION FOR FUTURE RESEARCH

This research clearly points out the potential of association rules as a tool for industrial application especially in manufacturing processes. In this study, an approach was presented for discovering useful patterns between process parameters and product defects using a fuzzy weighted association rule algorithm. Compared to other association rule algorithms, these obtain more understandable patterns and more interesting discovered rules using the concepts of fuzzy sets and weights. The rules obtained from the manufacturing process database can be used for controlling defects and analyzing root causes. An application of the proposed method during a net making process at a netting plant was demonstrated. A detailed discussion on how to control the manufacturing process defects using the results obtained from the algorithm was also presented. After implementing the method during a performance improvement project, we observed a significant reduction in the rate of defects in the produced nets.

This work can be applied in various areas. One of our future focuses will be on expansion of the use of association rule algorithms as a main part of quality improvement methodologies, such as six sigma. The six sigma methodology helps improve the process through finding the relations between inputs and outputs and controlling outputs using the identified relations. Therefore, association rule algorithms can be used as fast, simple tools for finding hidden relations between process variables and expedited six sigma phases.

### REFERENCES

[1] N. Dhafr, M. Ahmad, B. Burgess, and S. Canagassabady, “Improvement of quality performance in manufacturing organizations by minimization of production defects”, Robotics and Computer-Integrated Manufacturing, Vol. 22 No. 5, pp. 536-542, 2006.

- [2] Y. Cheng, W. Yu, and Q. Li, "GA-based multi-level association rule mining approach for defect analysis in the construction industry", *Automation in Construction*, Vol. 51, pp. 78-91, 2015.
- [3] R. Agrawal, and R. Srikant, "Fast algorithms for mining association rules", In *Proc. 20th int. conf. very large data bases, VLDB*, Vol. 1215, pp. 487-499, 1994.
- [4] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases", In *Acm sigmod record*, Vol. 22 No. 2, pp. 207-216, 1993.
- [5] R. J. Miller, and Y. Yang, "Association Rules over Interval Data", In *Proc. ACM SIGMOD Internat. Conf. Management of Data*, 1997, pp. 452-461.
- [6] K. C. C. Chan, and W. H. Au, "An Effective Algorithm for Mining Interesting Quantitative Association Rules", In *Proc. of the 12th ACM Symp. on Applied Computing*, San Jose, CA, Feb. 1997, pp. 88-90.
- [7] K. C. C. Chan, and W. H. Au, "Mining Fuzzy Association Rules" In *Proc. of the 6<sup>th</sup> ACM Int'l Conf. on Information and Knowledge Management*, Las Vegas, Nevada, Nov. 1997, pp. 209-215.
- [8] T. P. Hong, C. S. Kuo, and S. C. Chi, "Trade-off between computation time and number of rules for fuzzy mining from quantitative data", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 9 No.5, pp. 587-604, 2001.
- [9] M. Mueyba, M. S. Khan, and F. Coenen, "Fuzzy weighted association rule mining with weighted support and confidence framework", In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer Berlin Heidelberg, pp. 49-61, May 2008.
- [10] M. Mueyba, M. S. Khan, and F. Coenen, "Effective mining of weighted fuzzy association rules", *Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection: Technologies for Infrequent and Critical Event Detection*, 3, 47, 2009.
- [11] A. Gyenesei, "Mining weighted association rules for fuzzy quantitative items", in *Proceedings of PKDD, Conference*, Lyon, France, 2000, pp.187-219, 2000.
- [12] D. L. Olson, and Y. Li, "Mining fuzzy weighted association rules", In *the 40th Hawaii International Conference on System Sciences*, 2007, pp. 53-53, 2007.
- [13] C. W. Lin, T. P. Hong, and W. H. Lu, "An efficient tree-based fuzzy data mining approach", *Int. J. Fuzzy Syst*, Vol. 12 No. 2, pp. 150-157, 2010.
- [14] A. Moustafa, B. Abuelnasr, and M. S. Abougabal, "Efficient mining fuzzy association rules from ubiquitous data streams", *Alexandria Engineering Journal*, Vol. 54, pp. 163-174, 2015.
- [15] H. Yuniarto, "The Shortcomings of Existing Root Cause Analysis Tools", In *Proc. World Congr. Eng.* 3, 2012.
- [16] M. Donauer, P. Peças, and A. Azevedo, "Identifying nonconformity root causes using applied knowledge discovery", *Robotics and Computer-Integrated Manufacturing*, Vol. 36, pp. 84-92, 2015.
- [17] B. Al-Salim, "Optimizing the Formation of the Quality Improvement Teams through a Data Mining-Based Methodology", *Quality Engineering*, Vol.18 No.3, pp. 379-389, 2006.
- [18] W. C. Chen, S. S. Tseng, and C. Y. Wang, "A novel manufacturing defect detection method using association rule mining techniques", *Expert systems with applications*, Vol. 29 No.4, pp. 807-815, 2005.
- [19] H. Sadoyan, A. Zakarian, and P. Mohanty, "Data mining algorithm for manufacturing process control", *The International Journal of Advanced Manufacturing Technology*, Vol. 28, pp. 342-350, 2006.
- [20] S. Lee, S. Han, and C. Hyun, "Analysis of Causality between Defect Causes Using Association Rule Mining", *International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering*, Vol. 10 No. 5, 2016.