

# 스마트 공장에서 의사결정 모델을 이용한 순차 마이닝 기반 제조공정

김주창<sup>1</sup>, 정호일<sup>2</sup>, 유현<sup>3</sup>, 정경용<sup>1\*</sup>

<sup>1</sup>경기대학교 컴퓨터공학부, <sup>2</sup>원광대학교 컴퓨터·소프트웨어공학과, <sup>3</sup>상지대학교 컴퓨터정보공학과

## Sequence Mining based Manufacturing Process using Decision Model in Cognitive Factory

Joo-Chang Kim<sup>1</sup>, Hoill Jung<sup>2</sup>, Hyun Yoo<sup>3</sup>, Kyungyong Chung<sup>1\*</sup>

<sup>1</sup>Division of Computer Science and Engineering, Kyonggi University

<sup>2</sup>Department of Computer·Software Engineering, Wonkwang University

<sup>3</sup>Department of Computer Information Engineering, Sangji University

요 약 본 논문에서는 스마트 공장에서 의사결정 모델을 이용한 순차 마이닝 기반 제조공정을 제안한다. 제안하는 모델은 소규모의 제조공정에서 순차 마이닝 의사결정 모델을 적용하여 제조 효율을 높이는 방법이다. 제조 단계 중 제품 제조 과정에서 나타나는 데이터를 입력 변수들로 구성하고, 시간당 제조량과 불량률을 출력 변수로 구성한다. t-검정을 통해 유의 수준이 높은 변수만을 사용하여 GSP 알고리즘과 REPTree 알고리즘을 이용한 규칙과 모델을 생성한다. 의미있는 순차 규칙과 의사결정 모델은 정확도, 민감도, 특이성, 예측도를 통해 유의미함을 확인한다. 결과적으로, 실제 제조에 적용한 결과 불량률은 0.38%가 개선되었고, 시간당 제조량은 평균 1.89/h 증가되었다. 이는 소규모 제조 공정에서 데이터 마이닝 분석을 통한 제조 효율을 높이기 위한 의미있는 결과를 나타낸다.

주제어 : 데이터 마이닝, 제조, 의사결정 모델, 의사결정 트리, 엔트로피

**Abstract** In this paper, we propose a sequence mining based manufacturing process using a decision model in cognitive factory. The proposed model is a method to increase the production efficiency by applying the sequence mining decision model in a small scale production process. The data appearing in the production process is composed of the input variables. And the output variable is composed the production rate and the defect rate per hour. We use the GSP algorithm and the REPTree algorithm to generate rules and models using the variables with high significance level through t-test. As a result, the defect rate are improved by 0.38% and the average hourly production rate was increased by 1.89. This has a meaning results for improving the production efficiency through data mining analysis in the small scale production of the cognitive factory.

**Key Words** : Data Mining, Production, Decision Model, Decision Tree, Entropy

### 1. 서론

4차 산업혁명이 시작되면서 제조 산업은 자동화 공정

에서 스마트 공정으로 변화하고 있다. 스마트 제조는 제품의 기획부터 판매까지 제조 모든 공정 과정에 인공지능, 센서, 사물 인터넷, 클라우드 컴퓨팅, 데이터 마이닝

\*This work was supported by the GRRC program of Gyeonggi province.

[2017-B03, Intelligence Information-based Security and Network Technology Research].

\*Corresponding Author : Kyungyong Chung (dragonhci@hanmail.net)

Received January 12, 2018

Revised February 14, 2018

Accepted March 20, 2018

Published March 28, 2018

같은 정보 기술을 적용하는 차세대 제조 산업이다[1]. 현재 제조과정에서 발생하는 대량의 데이터들이 전산화되고 있으며, 이러한 데이터는 제품 기획, 생산, 유통, 평가 등 공정단계에 따라 서로 다른 특성을 가진 정형/비정형 데이터로 구성되어 있다[2]. 정형 데이터는 센서, 사물인터넷, 통계, 실험 등 일정한 구조로 수집되는 데이터이다. 이는 주로 빈발패턴 마이닝, 순차패턴 마이닝, 연관패턴 마이닝 등을 통해 유용한 정보를 탐색한다. 비정형 데이터는 SNS, 웹 문서, 이미지, 음성 등 형태가 구조화되어 있지 않은 데이터이다. 비정형 데이터는 텍스트 마이닝, 소셜 마이닝, 오피니언 마이닝 등을 통해 유용한 정보를 탐색한다[3,4]. 이에 따라 각각의 상황에 적합하게 정보통신 기술을 응용하고, 데이터를 효율적으로 활용하기 위한 고도화 연구가 진행되고 있다[2-4]. 기존의 자동화 공정은 제조 공정에서 필요한 의사결정에서 작업자, 관리자, 결정권자 등 사람의 작업 숙련도나 경험에 의존한다[5]. 이에 따라 동일한 작업이라도 숙련도, 주변 상황, 심리적 상황 같이 개인의 차이에 따라 상이한 결과가 나타나는 문제가 발생한다. 특히, 수작업이 필요한 제조 공정에서 개인의 주관적인 의사결정은 같은 공정을 통한 제품이 서로 상이한 품질을 나타내는 문제가 발생할 수 있다[6]. 또한, 소규모의 제조 공정은 개인에 의한 작은 변화가 제품 품질에 미치는 영향이 크기 때문에 품질을 높이고 유지하기 위한 분석이 필요하다. 이러한 문제를 해결하기 위해 데이터 분석, 지식 탐색, 지식 시각화에 따른 의사결정 모델이 필요하다[4]. 스마트 제조에서 데이터 마이닝을 이용한 제조 데이터 분석을 통해 리스크 관리, 품질 관리, 트렌드 예측, 재고 관리가 가능하며 이를 위한 지속적인 연구와 시도가 필요하다[5,6].

본 논문에서는 스마트 공장에서의 의사결정 모델을 이용한 순차 마이닝 기반 제조공정을 제안한다. 제안하는 방법은 제조 과정에서 제품종류, 작업시간, 작업자, 원자재 등의 여러 변수들에 대한 분석을 통해 제품 제조에서 효율을 높이기 위한 순차 분석이다. 제조에서 대표적인 변수인 시간당 제조량과 불량률을 출력변수로 구성하고, 이에 영향을 주는 입력 변수들의 유의 수준을 분석한다.

## 2. 관련연구

### 2.1 스마트 팩토리

스마트 팩토리는 4차 산업 혁명의 초연결성 및 초지능성 예측 가능성을 기반으로 제품의 기획, 설계, 제조, 유통, 물류 등 제품과 관련된 모든 과정에서 정보통신 기술이 적용된 공장이다[5]. 기존의 자동화 제조공장은 각각의 공정별로 자동화가 이루어 졌기 때문에 단계별 공정 간의 소통이 어려운 단점이 있었다. 스마트 팩토리는 제조 과정에서 사물인터넷, 센서, 카메라 등을 모든 설비를 유/무선 통신을 통해 모든 공정을 연결하여 발생하는 크고 작은 데이터 전체를 수집한다. 스마트 팩토리에서 사물 인터넷을 통해 제조 현장에서 발생하는 변화를 실시간으로 모니터링하고 이를 통해 수집된 데이터에 대한 분석으로 제조 공정의 체계화와 제품 불량률의 원인 분석 등을 통해 최적의 환경을 유지한다[7]. Fig. 1은 스마트 팩토리에서 데이터 흐름을 나타낸다.

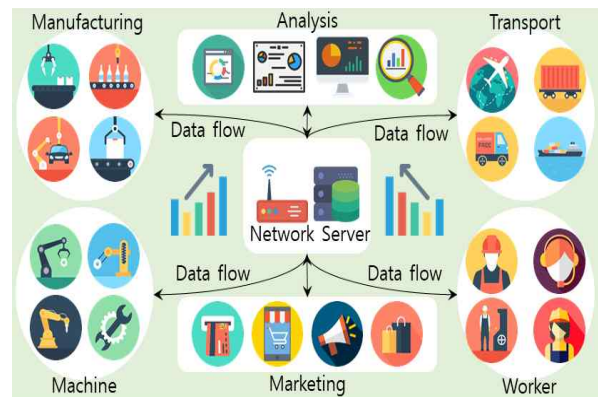


Fig. 1. Data flow in smart factory

### 2.2 마이닝 기반 제조공정

스마트 제조에서 데이터 마이닝은 빅데이터가 IT적 관점인 속도, 크기, 다양성의 기준에서 벗어나 마이닝 프로세스를 통한 공정 이상 감지, 품질 예측, 마케팅 등 가치 창출에 초점을 두고 있다[8]. 또한, 완전한 자동화 공정이 도입되지 않은 소규모 제조에서 발생하는 데이터에 대한 수집과 분석을 통해 제품불량의 원인 판단과 불량을 예방하고, 제조 효율을 높이는 제조 프로세스 개선이 필요하다. 소규모 제조에서는 분석한 결과의 즉각적인 반영이 가능하며 지속적인 피드백이 용이하다. 이에 따라 소규모 제조에서 데이터 마이닝을 통한 프로세스 개선으로 큰 변화를 기대할 수 있다[9,10].

### 3. 의사결정 모델을 이용한 순차 마이닝 기반 제조공정

#### 3.1 의료용 제품 트랜잭션에서 변수 요인 분석

본 연구는 스마트 공장에서 리프팅 시술에 들어가는 봉합사의 제조 공정에서 제조 효율과 불량률 저하를 위해 마이닝 분석을 적용한다. 리프팅용 봉합사는 리프팅 시술에 사용되는 의료용 실이며, 가공 방법에 따라 Cog, Screw, Cavem 등 여러 종류로 구분한다[9]. 이는 제품 특징 상 다양한 제품 종류, 복잡한 공정, 세밀한 작업으로 인해 수작업이 요구되며 자동화 공정을 적용하기 어려운 문제가 있다. 또한, 제품종류나 작업자 숙련도에 따라 제조시간 지연이나 제품불량이 발생할 가능성이 높다. 마이닝 분석을 통해 불량률을 최소화하는 작업 스케줄링을 통해 작업 효율을 향상시킬 수 있다[11]. 리프팅용 봉합사 제조에서 수집되는 데이터로 날짜, 자재 사용량, 제품 품목, 제조 수량, 불량수량, 작업시간 등 수집 가능한 모든 데이터를 포함한다. 이는 작업일지의 형태로 일일 단위로 작업자의 관련 데이터를 문서형태 기록한다.

Table 1은 리프팅용 봉합사 제조 데이터를 나타낸다. 리프팅용 봉합사 제조에서 출력변수는 마이닝 분석을 통해 개선하고자 하는 목표변수를 나타낸다. 입력변수는 출력변수에 영향을 미칠 것이라 예상되는 변수를 나타낸다.

제조효율과 불량률은 제조 산업에서 큰 이슈로 이를 개선하기 위해 출력변수로 구성한다. 입력변수는 제품 제조에서 필요한 소모성 자재들 같은 원재료와 제품 제조에서 필요한 정보를 포함하는 작업 데이터로 구성한다. 출력변수 2개와 입력변수 15개로 총 17개의 변수로 구성한다.

Table 1의 제품 타입에서 M=Mono type, S=Screw type, DS=Double Screw type, VS=Cavem Screw type, DVS=Double Cavem Screw type, TVS=Tripple Cavem Screw type, D=Double type, TR=Triple type을 나타낸다. 급여에서 수당은 제조 단위 AY에 따라 급여를 지급하는 방법이고, 시급은 작업 시간에 따라 급여를 지급하는 방법이다. 특이사항은 제품 제조 과정에서 발생한 예외사항에 대한 기록을 나타낸다. 작업자는 Table 1을 기준으로 자신의 작업 상황을 일일 단위로 작성한다. 관리

Table 1. Production data of suture

Data type		Variable			
Discision	Classification	Name(Variable Name)	Range	Note	
Input	Job	Worker	Worker Code(wocode)	[w01, w02, w03, ..... ]	
		Date	Job Date(date)	[yyyy.mm.dd]	
		Assembly	Product Type(type)	[M, S, DS, VS, DVS, TVS, D, TR]	
		Amount	Assignment(asm)	[100, 200, 300, 400]	Quantity : AY
			Make(make)	[1~Assignment]	Worker's output
			Defect(defect)	[AY]	Number of defective products
		Working Time	Start Time(stime)	[hh,mm]	
			End Time(etime)	[hh,mm]	
			Work Time(wtime)	[etime-stime]	min.
	Payment	Payment(pay)	[Extra Pay, Hourly Wage]	[ep, hw]	
	ETC	Notice(notice)	[Record notice between jobs]		
	Raw Materials	Suture	Suture(suture)	[pdo, pci, pla, plla]	
			Suture Used(suse)	[Usage history]	Quantity : m
		Needle	Cannula(can)	[Usage history]	Quantity : EA
			Sheath(sheath)	[Usage history]	Quantity : EA
Protective Cap(cap)			[Usage history]	Quantity : EA	
Lubricant(lub)			[Usage history]	Quantity : ml	
Glue(glue)	[Usage history]	Quantity : ml			
Output	Evaluation	Efficiency	Production for Hour(pfh)	[make/wtime]	Quantity : AY
		Defect	Defect Rate(drate)	[defect/make*100]	Quantity : %

자는 작성된 작업자의 작업일지를 전산화를 통해 수집된 데이터는 발생 순서에 따라 각각 트랜잭션으로 구성한다. 데이터 분석을 위해 수집된 데이터 전처리 과정을 통해 결측값, 중복값 등을 제거한다. 불량품 수량 항목 또한 출력변수 불량률에 포함되며, 전처리를 통해 불량품 수량을 제거한다. 의미가 중복되는 시작시간과 마감시간을 작업시간으로 병합하여 트랜잭션에 포함한다. 원재료에서 캐놀리, 침기, 보호덮개 등 완제품당 하나씩 소모되어 제조량과 동일한 수치를 나타내는 데이터를 병합한다. Table 2는 전처리 과정을 통해 구성된 제조 데이터 트랜잭션을 나타낸다.

Table 2. Production data transaction

Variable \ TID	T001	T002	T003	.....
wcode	w01	w02	w03	.....
date	2017.12.04	2017.12.04	2017.12.04	.....
type	M	DVS	DVS	.....
asm	300	200	200	.....
make	300	200	185	.....
wtime	240	260	300	.....
pay	hw	ep	hw	.....
suture	pcl	pla	pla	.....
suse	15.50	14.00	12.95	.....
lub	6	4	3.7	.....
glue	6	4	3.7	.....
pfh	75	46.18	37	.....
drate	6.33	5.50	2.16	.....

### 3.2 순차 마이닝 기반 제조공정에서 패턴 발견

리프팅용 봉합사는 제품에 따라 상이한 제조 공정을 거치게 되며, 제품 제조 시간이나 불량률이 다르게 나타난다. 제품 제조의 효율을 결정하는 출력변수 시간당 제조량(pfh)와 불량률(drate)에 영향을 주는 요인을 t 검정 [12]과 GSP알고리즘[13]을 이용하여 분석한다. t 검정에서 유의 수준은 일반적으로 사용되는 0.05로 정의한다. 추가로 사용자별 숙련도 수준을 고려하기 위해 제품 타입별 불량률을 탐색한다. Table 3은 t 검정을 통해 나타난 변수별 유의 수준을 나타낸다. Table 3에서 sim.은 상관계수를 나타내고, sig.는 유의 수준을 나타낸다. t 검정 결과 입력변수 wtime, suture, suse, lub, glue는 pfh에 대한 유의 수준이 0.05를 만족하지 않기 때문에 pfh 예측에

서 제외한다. 제조 효율 요인 분석을 통해 나타나는 주요 요인들을 GSP 알고리즘을 이용하여 규칙을 생성한다. Table 4는 GSP 알고리즘 기반의 요인에 따른 제조 규칙을 나타낸다. GSP 알고리즘은 시퀀스 데이터에서 여러 트랜잭션에 걸쳐 공통으로 나타나는 순차패턴을 탐색하는 방법이다. 알고리즘 적용을 위해 수치형 데이터는 구간을 나누어 정규화한다. 알고리즘 적용 결과 총 681개의 규칙이 생성된다. Table 4의 규칙 1에서 [type = M] & [pay = hw] --> [drate = 3~4]는 mono type 제품을 시급으로 제조할 경우 불량률이 3~4%를 나타낸다. 규칙 생성을 위한 데이터는 총 350개의 트랜잭션 중에서 70%인 245개의 트랜잭션을 제조 규칙 생성을 위한 트레이닝 데이터로 활용하고, 나머지 30%를 105개의 트랜잭션을 의사결정 모델 평가에 활용한다.

Table 3. Significance level according to variables

Variable	pfh			drate		
	sim.	t	sig.	sim.	t	sig.
type	-0.563	67.25	0.034	0.162	-2.14	0.002
asm	0.958	-159.92	0.000	-0.202	-129.95	0.000
make	0.961	-157.12	0.000	-0.190	-128.36	0.000
wtime	-0.012	-102.71	<b>0.824</b>	0.131	-176.11	0.014
pay	-0.892	79.83	0.000	0.456	18.35	0.000
suture	0.007	74.67	<b>0.893</b>	0.198	12.39	0.000
suse	-0.649	49.74	<b>0.071</b>	0.128	-72.70	0.016
lub	-0.616	79.49	<b>0.056</b>	-0.149	7.76	<b>0.064</b>
glue	0.484	79.49	<b>0.056</b>	-0.149	7.76	<b>0.064</b>

Table 4. Production sequential pattern rules

No.	Rule
1	[type = M] & [pay = hw] --> [drate = 3~4]
2	[type = DVS] & [pay = hw] --> [drate = 4~5]
3	[type = M] & [pay = hw] --> [pfh = 70~80]
4	[type = M] & [pay = ep] --> [pfh = 80~90]
5	[type = S] & [suture = PCL] --> [drate = 4~5]
6	[type = DS] & [make 300] & [pay = ep] --> drate = 7~8
7	[type = DS] & [make 300] & [pay = hw] --> drate = 5~6
8	[type = VS] & [make 300] & [pay = ep] --> drate = 6~7
.....	.....
681	[suture = PLA] & [pay = hw] % [suse = 11~12] --> [drate = 5~6]

## 4. 제조공정에서 의사결정 모델과 성능평가

### 4.1 제조공정에서 의사결정 모델

마이닝을 통해 탐색한 요인 변수, 제품 타입별 불량률, 제조 규칙을 이용하여 의사결정 모형으로 구성된 모델을 구성한다. 의사결정 모델은 데이터 마이닝의 예측 모델링 방법 중의 하나이며, 측정값과 결과값의 인과 관계를 예측하는 모델이다. 의사결정 모델은 화이트 박스 모형으로 표현이 되며, 결과 값의 형태가 간략하기 때문에 제품 제조공정에 적용하기에 용이하다. 또한 논리 및 조건이 투명하고, 중간 변수나 임시 데이터가 없으며, 과정과 결과가 명료하기 때문에 실무자에게 있어서 이해와 해석이 쉬운 장점이 있다. 대표적인 알고리즘으로 ID3가 있으며, 이외에도 C4.5, C5.0, CART, REPTree 등이 있다. ID3, C4.5, C5.0, REPTree에서는 정보의 불순도 지표를 평가하기 위해 정보 이론 엔트로피 개념의 정보 획득량을 사용하고 있으며, CART에서는 지니불순도를 사용한다[14]. 엔트로피 개념은 데이터의 무질서도를 정량화하는 것으로 표본 집단의 엔트로피가 높을수록 무질서한 형태를 가지게 된다. 표본집단에서 총  $m$ 개의 데이터가 포함되는 영역  $f$ 에 대한  $Entropy(f)$ 는 식 1과 같다.

$$Entropy(f) = - \sum_{i=1}^m f_i \log_2(f_i) \quad \text{식 1}$$

$f_i$ 는  $f$  영역에 속하는 데이터 중  $i$ 의 영역에 포함되는 비율을 의미한다. 의사결정나무의 학습을 위해서는 식 1을 기반으로 재귀적 분기과정을 진행한다. 재귀적 분기 과정은 최초 각각의 표본데이터에 대한 정보 획득량을 순차적으로 식 1에 의해 추출하는 작업이 필요하며, 이후 추출된 값의 크기에 따라 분류 작업을 진행한다. 분류 작업은 반복적으로 이루어지며, 모든 단말의 엔트로피가 0이 될 때 까지 반복한다. 분류 작업중 일부 입력된 오작동 데이터나 노이즈 문제로 인하여 오류 모델이 구축되는 경우가 발생할 수 있다. 따라서 이를 방지하기 위하여 가지치기라고 불리는 일반화 작업이 필요하다. 일반화 작업으로 에러감소 프루닝은 노드 아래부분을 자르거나 결합한 뒤의 오류율을 비교한다. 논문에서는 REPTree를 적용하여 모델을 생성한다. REPTree는 데이터 마이닝 툴 Weka[15]에서 지원하는 알고리즘으로 C4.5 알고리즘과 유사한 형태로써 처리 속도가 빠르고, 높은 정확도를

나타내는 장점이 있다. Table 5는 변수 pfh와 drate를 출력으로 하는 REPTree 알고리즘 기반의 의사결정 모델을 나타낸다. type = DVS에서 pay = hw : [pfh 38.78]은 DVS 제품을 Hourly Wage로 제조할 경우 시간당 제조량이 60 AY임을 나타낸다.

Table 5. REPTree based decision model

production for hour (pfh) AY
<pre> type = M     drate &lt; 4.5       drate &lt; 3.21         drate &lt; 2.83           drate &lt; 1.35 : [pfh 75]           drate &gt;= 1.35             drate &lt; 2.17 : [pfh 70.69]             drate &gt;= 2.17 : [pfh 75]             drate &gt;= 2.83 : [pfh 69.59]             drate &gt;= 3.21 : [pfh 75]       drate &gt;= 4.5         pay = hw : [pfh 60]         pay = ep : [pfh 75] type = DVS     pay = hw : [pfh 38.78]     pay = ep : [pfh 39.73] type = S : [pfh 66.01] type = DS : [pfh 56.79] type = VS     drate &lt; 1.59 : [pfh 41.57]     drate &gt;= 1.59 : [pfh 43.28] type = D : [pfh 58.98] type = TR : [pfh 50] type = TVS     make &lt; 195.5 : [pfh 38.05]     make &gt;= 195.5 : [pfh 39.88]                     </pre>
defect rate (drate) %
<pre> pay = hw : 2.89 [drate 3.31~3.25] pay = ep     type = M : [drate 3.67~6.54]     type = DVS : [drate 10.5~18.48]     type = S : [drate 5.05~10.34]     type = DS : [drate 6.62~9.81]     type = VS : [drate 5.97~7.26]     type = D : [drate 8.26~10.88]     type = TR       wtime &lt; 313.5 : [drate 7.5~10.22]       wtime &gt;= 313.5 : [drate 0.21~3.64]     type = TVS       pfh &lt; 39.1 : [drate 1.15~2.07]       pfh &gt;= 39.1 : [drate 4.28~13.5]                     </pre>

### 4.2 성능 평가

생성된 의사결정 모델은 정확도(Accuracy), 민감도(Sensitivity), 특이성(Specificity), 예측도(Predictive Value:PV)의 측정을 통해 성능을 평가한다. 예측도는 양성(P-PV)과 음성(N-PV)으로 구분하여 평가한다. 실험을 위해 총 350개의 트랜잭션 중 트레이닝에 사용한 245개의 트랜잭션을 제외한 105개의 트랜잭션을 이용하여

성능평가를 진행한다. 실험을 위한 모델은 GSP 알고리즘 기반의 제조 순차 패턴 규칙과 REPTree 기반의 의사결정 모델이다. Table 6은 모델에 따른 평가 결과를 나타낸다.

Table 6. Performance evaluation results(%)

Evaluation	GSP			REPTree		
	pfh	drate	Avg.	pfh	drate	Avg.
Accuracy	51.23	57.82	54.52	72.54	55.91	64.22
Sensitivity	72.22	69.47	70.84	96.71	81.37	89.04
Specificity	23.88	54.39	39.13	17.55	21.41	19.48
P-PV	69.09	67.48	68.28	72.78	61.14	66.96
N-PV	29.11	51.74	40.42	71.49	50.21	60.85
Avg.	49.10	60.18	54.64	66.21	54.00	60.11

평가 결과 GSP는 54.64%, REPTree는 평균 60.11%로 나타난다. REPTree는 출력변수 pfh에서 평균 66.21%로 더 유의한 것으로 나타나고, GSP는 출력변수 drate에서 60.18%로 유의한 것으로 나타난다. 이에 따라 두 모델을 통해 추출된 규칙을 사용할 경우, 시간당 제조량과 불량률의 개선을 기대할 수 있다. Table 7은 제안하는 GSP 알고리즘과 REPTree를 통해 추출한 제조 규칙을 의료용 제품 제조공정에 적용하여 3주간 제품별 시간당 제조량과 불량률을 측정한 결과를 나타낸다. GSP 알고리즘 기반의 제조 순차 패턴 규칙을 이용하여 불량률이 낮게 나타날 것으로 예상되는 작업 스케줄을 우선적으로 구성하고, REPTree 기반의 의사결정 모델을 이용하여 시간당 제조량이 높은 순서로 작업을 할당한다. 예를들어 제품당 단가가 높아 불량에 따른 손실이 큰 VS, DVS, TVS 제품은 제조 순차 패턴 규칙을 이용하여 불량률을 최소화 할 수 있는 방법으로 변수에 대해 의사결정을 하고, 수요가 높아 제조량이 중요한 M, S, DS 제품은 REPTree 기반의 의사결정 모델을 통해 변수에 대한 결정을 한다. 제안하는 방법을 적용한 결과 전체적인 제조량은 1.89/h이고 불량률은 0.38%가 개선되었다.

Table 7. Changes in drate and pfh

drate=% pfh=YA		M	S	DS	VS	DVS	TVS	avg.
before	drate	2.60	2.76	4.12	4.13	4.60	4.71	3.82
	pfh	73.52	66.00	56.79	42.98	39.25	39.71	53.04
after	drate	2.12	2.55	3.10	4.01	4.27	4.59	3.44
	pfh	75.66	70.82	58.11	44.68	40.33	40.01	54.93

## 5. 결론

본 연구에서는 스마트 공장에서 의사결정 모델을 이용한 순차 마이닝 기반 제조공정을 제안하였다. 이는 수작업 기반의 소규모 제조 공정에서 나타나는 작업자들의 데이터를 분석하여 시간당 제조량과 불량률을 개선하기 위한 순차 마이닝 기반 제조 방법이다. 트랜잭션 구성을 위해 발생하는 데이터의 변수를 구성하고, 변수간의 유의 수준을 분석하여 관련성이 낮은 변수들을 제거하였다. 구성된 트랜잭션은 데이터 마이닝에서 GSP 알고리즘과 REPTree 알고리즘을 통해 규칙과 모델을 생성하고, 실제 제조에 적용하였다. 적용결과 불량률은 3.82%에서 3.44%로 0.38%개선되었고, 시간당 제조량은 53.04/h에서 54.93/h로 평균 1.89/h 증가되었다. 이를 통해 소규모 제조 공정에서 순차 마이닝 기반 분석으로 제조효율을 높일 수 있음을 확인하였다.

## REFERENCES

- [1] J. M. Park. (2015). Technology and Issue on Embodiment of Smart Factory in Small-Medium Manufacturing Business. *The Journal of Korean Institute of Communications and Information Sciences*, 40(12), 2491-2502.
- [2] H. Jo. C. Kim & Y. Cho. (2014). Development of Mobile Dashboard System for Manufacturing Data Visualization. *Journal of the Korean Society for Precision Engineering*, 31(4), 311-317.
- [3] E. H. Jeong & B. K. Lee. (2015). A design of SNS emotional information analysis strategy based on opinion mining. *The Journal of Korea Institute of Information, Electronics, and Communication Technology*, 8(6), 544-550.
- [4] B. H. Back. I. Ha & B. Ahn. (2014). An extraction method of sentiment information from unstructured big data on SNS. *Journal of Korea Multimedia Society*, 17(6), 671-680.
- [5] J. Lee. (2003). E-manufacturing - fundamental, tools, and transformation. *Robotics and Computer-Integrated Manufacturing*, 19(6), 501-507.
- [6] C. Jun. B. H. Kim. & J. Y. Lee. (2017). A Big Data Analysis Platform based on the Manufacturing Specialized Library: A Case Study on Implementation of the Platform for Quality Problems. *Journal of the*

*Korean Institute of Industrial Engineers*, 43(5), 380-387.

[7] H. S. Kang, J. Y. Lee, S. Choi, H. Kim, J. H. Park, J. Y. Son & S. D. Noh (2016). Smart manufacturing: Past research, present findings, and future directions. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 3(1), 111-128.

[8] B. Choi, H. Kim, J. Kim & S. Jin. (2012). Data Analytics for CRM in the Age of Big Data. *Entrue Journal of Information Technology*, 11(1), 19-27.

[9] Y. L. James, C. S. Y. Thong & K. S. Raman. (1996). Top Management Support, External Expertise and Information Systems Implementation in Small Business. *Information Systems Research*, 7(2), pp.248-267.

[10] K. C. Ahn, C. B. Moon, B. M. Kim, Y. S Shin & H. Kim. (2012). POS Data Analysis System based on Association Rule Analysis. *Journal of the Korea Industrial Information Systems Research*, 17(5), 9-17.

[11] S. J. Lee, H. R. Choi & H. C. Lee. (2011). Platform development of adaptive production planning to improve efficiency in manufacturing system. *Journal of the Korea Industrial Information Systems Research*, 16(2), 73-83.

[12] A. Agresti & B. Finlay. (1997). *Statistical methods for the social science* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

[13] R. Srikant & R. Agrawal (1996). Mining sequential patterns: Generalizations and performance improvements. *In International Conference on Extending Database Technology*. Springer, Berlin, Heidelberg, 1-17.

[14] S. Kalmegh. (2015). Analysis of WEKA data mining algorithm REPTree, Simple CART and RandomTree for classification of Indian news. *Int. J. Innov. Sci. Eng. Technol*, 2(2), 438-446.

[15] Machine Learning Group at the University of Waikato, *Weka 3: Data Mining Software in Java*. <https://www.cs.waikato.ac.nz/ml/weka/>.

김 주 창(Kim, Joo-Chang) [정회원]



- 2014년 2월 : 상지대학교 컴퓨터 정보공학부 (공학사)
- 2016년 8월 : 상지대학교 컴퓨터 정보공학과 (공학석사)
- 2017년 9월 ~ 현재 : 경기대학교 컴퓨터과학과 박사과정

▪ 관심분야 : 데이터 마이닝, 헬스케어  
 ▪ E-Mail : kjc2232@naver.com

정 호 일(Jung, Hoill) [정회원]



- 2010년 2월 : 상지대학교 컴퓨터 정보공학부 (공학사)
- 2013년 2월 : 상지대학교 컴퓨터 정보공학과 (공학석사)
- 2017년 2월 : 상지대학교 컴퓨터 정보공학과 (공학박사)

▪ 2018년 3월 ~ 현재 : 원광대학교 컴퓨터·소프트웨어 공학과 교수  
 ▪ 관심분야 : 데이터 마이닝, 지능시스템, 헬스케어  
 ▪ E-Mail : hijung1982@gmail.com

유 현(Hyun, Yoo) [정회원]



- 1999년 2월 : 상지대학교 전산학과 (이학사)
- 2011년 8월 : 상지대학교 컴퓨터 교육학과 (공학석사)
- 2006년 ~ 2015년 : 상지대학교 시스템운영부

▪ 2016년 3월 ~ 현재 : 상지대학교 컴퓨터정보공학과 박사과정  
 ▪ 관심분야 : 데이터 마이닝, 빅데이터, 머신러닝, 헬스케어  
 ▪ E-Mail : rhpa0916@gmail.com

정 경 용(Chung, Kyungyong) [종신회원]



- 2000년 2월 : 인하대학교 전자계산공학과 (공학사)
- 2002년 2월 : 인하대학교 전자계산공학과 (공학석사)
- 2005년 8월 : 인하대학교 컴퓨터 정보공학부 (공학박사)

▪ 2006년 3월 ~ 2017년 2월 : 상지대학교 컴퓨터정보공학부 교수  
 ▪ 2017년 3월 ~ 현재 : 경기대학교 컴퓨터공학부 교수  
 ▪ 관심분야 : 데이터 마이닝, 헬스케어, 빅데이터, 지능시스템, 인공지능, HCI, 정보검색, 추천 시스템  
 ▪ E-Mail : dragonhci@hanmail.net