

---

# 고품질의 사출 성형을 위한 이상감지 방법

---

한국산업기술대학교 경영학부

2021 - 12 - 14

# 목차

---

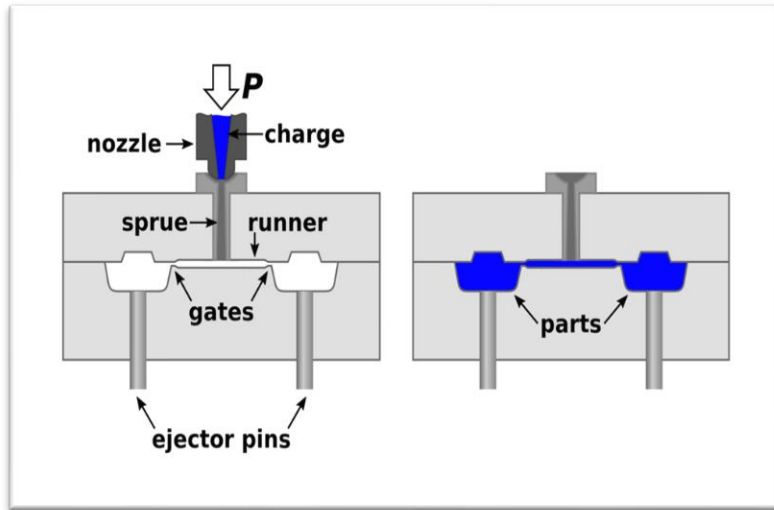
- 분석 방향
- EDA
- PCA 기반 이상감지 및 해석
- $T^2$  통계량 기반 이상감지 및 해석
- BURR 불량 이력 분석
- 요약

---

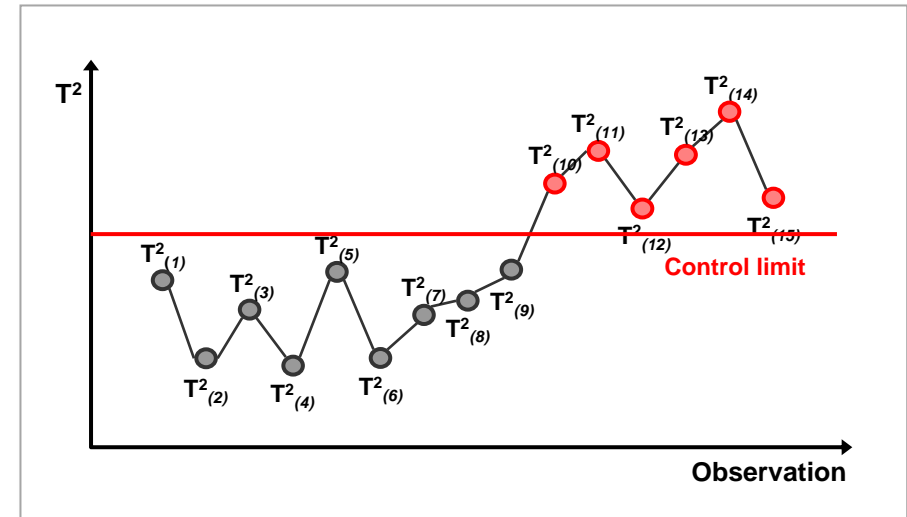
# 분석 방향

---

# 분석 방향



〈사출 공정 개념도〉

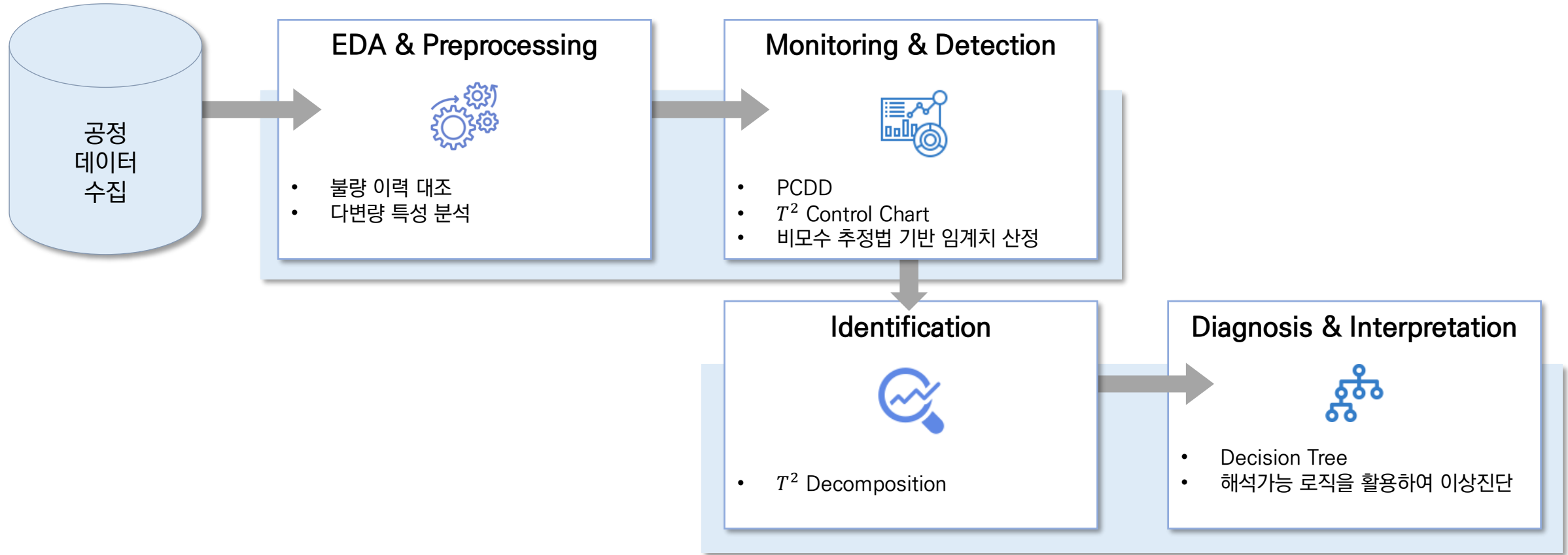


〈  $T^2$  통계량 기반 이상감지 〉

- 고품질의 사출 성형을 위해 공정 변수간 상관관계가 표준화되어야 원활한 품질관리가 이루어질 것으로 판단됨
- 불량/고장 등에 대한 Y값 데이터가 충분하지 않은 점과 변수간 상관관계가 존재한다는 점을 고려하여, 비지도 학습인 PCA와  $T^2$  Decomposition을 활용한 사출 공정의 이상감지 및 해석이 가능한 모니터링 시스템을 제안
- 향후 다양한 머신러닝 알고리즘을 활용하여 최적 제어가 가능한 방법론을 제시할 예정

# 분석 방향

## ■ 이상 원인 항목 분석 프로세스



---

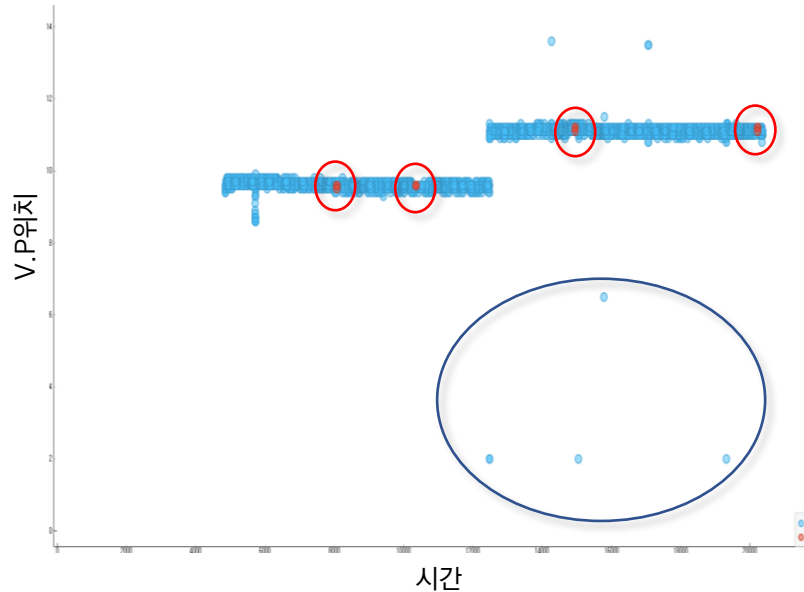
# EDA

---

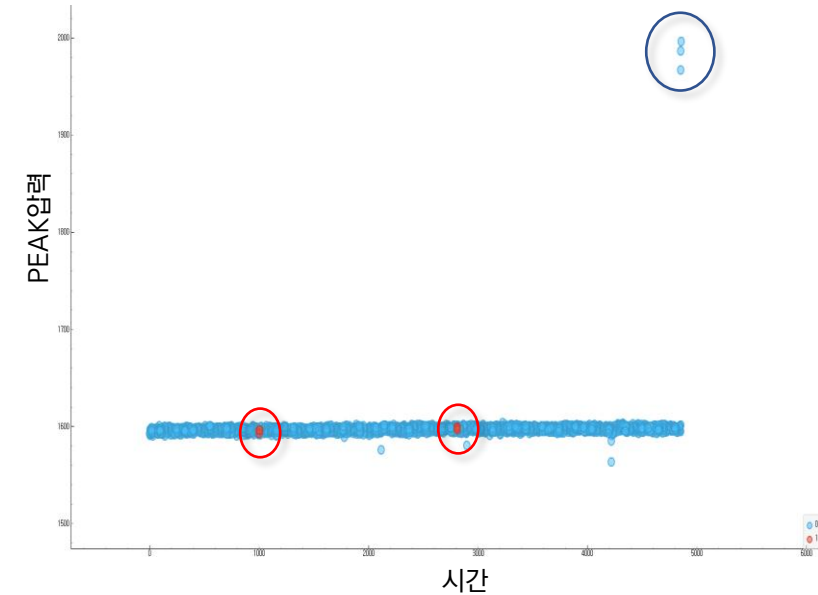
# EDA

## ▪ (IM-015\_9호기\_불량이력\_20180204) 파일을 통한 불량 대조 분석

〈2018020002-11 : V.P위치〉



〈2018020042-9 : PEAK압력〉

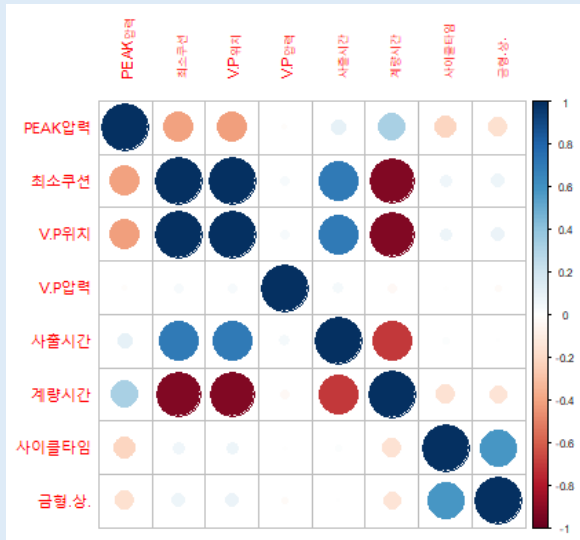


- 제공받은 소량의 불량 이력 데이터에 대해, 작업지시번호별 이벤트 대조 분석 실시
- 이상치로 보이는 값들과 실제 불량 이력 값과 크게 연관이 없다고 판단
- 불량 이력 비율이 매우 낮기 때문에, 각 이벤트에 대한 의미 부여가 힘들다고 사료됨 → 비지도학습 활용

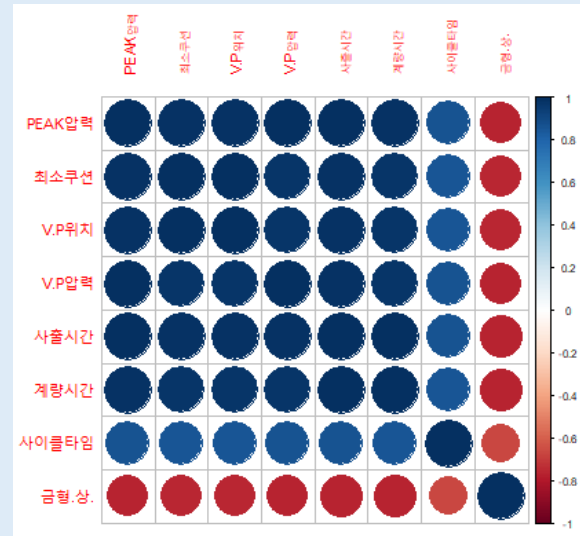
## ■ 작업지시번호별 상이한 상관관계

〈도번 1031721300〉

〈작업지시번호 2018050026-14〉

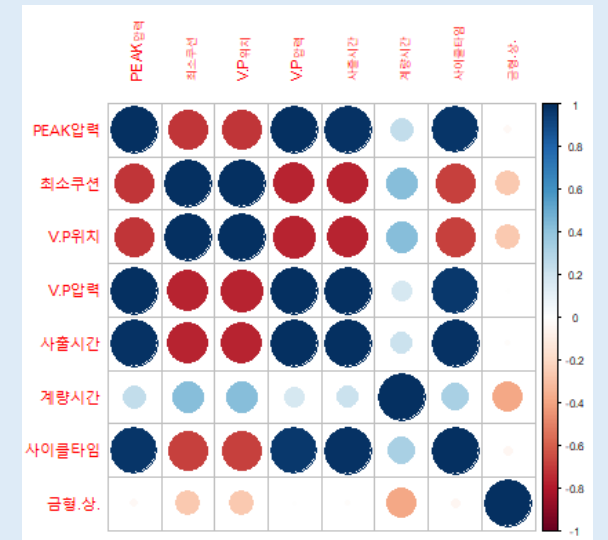


〈작업지시번호 2018070012-10〉



〈도번 299158074〉

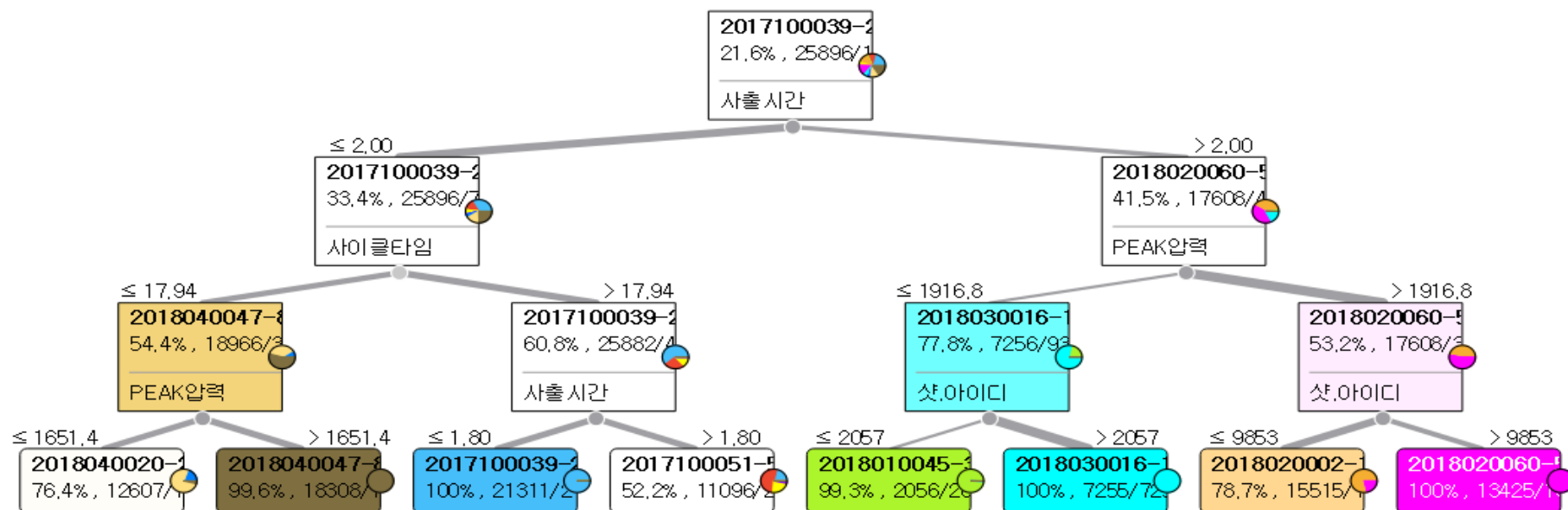
〈작업지시번호 2018070025-7〉



- 사출 공정 변수들은 서로 상관성을 지니고 있음 → 변수간 교호관계가 이상원인을 일으킬 수 있음
- 하나의 도번 안에 여러 작업지시번호가 있지만 작업지시번호별 변수간 상관관계가 상이한 패턴을 가짐
- 사출 공정에서 작업지시번호별 변수간 교호작용을 고려한 다변량 분석을 수행함



## 의사결정나무를 활용한 작업지시번호를 구분 짓는 Rule들 확인



- 의사결정나무를 통해 작업지시번호별로 구분 지을 때, 중요한 변수 및 Rule을 직관적으로 해석할 수 있음
- 트리 모델의 깊이를 3으로 알게 주어, 다변량 관계를 해석하고자 함
- 각 작업지시번호는 사출시간에 따라 크게 나누어지고, 사이클타임과 PEAK압력 순으로 나누어짐

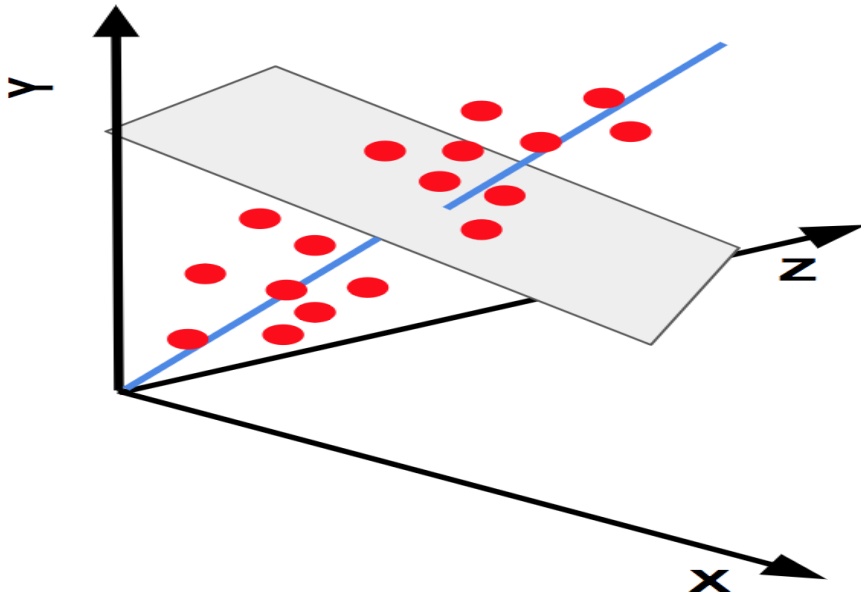
---

# PCA 기반 이상감지 및 해석

---

# PCA 기반 이상감지 및 해석

## PCA란?

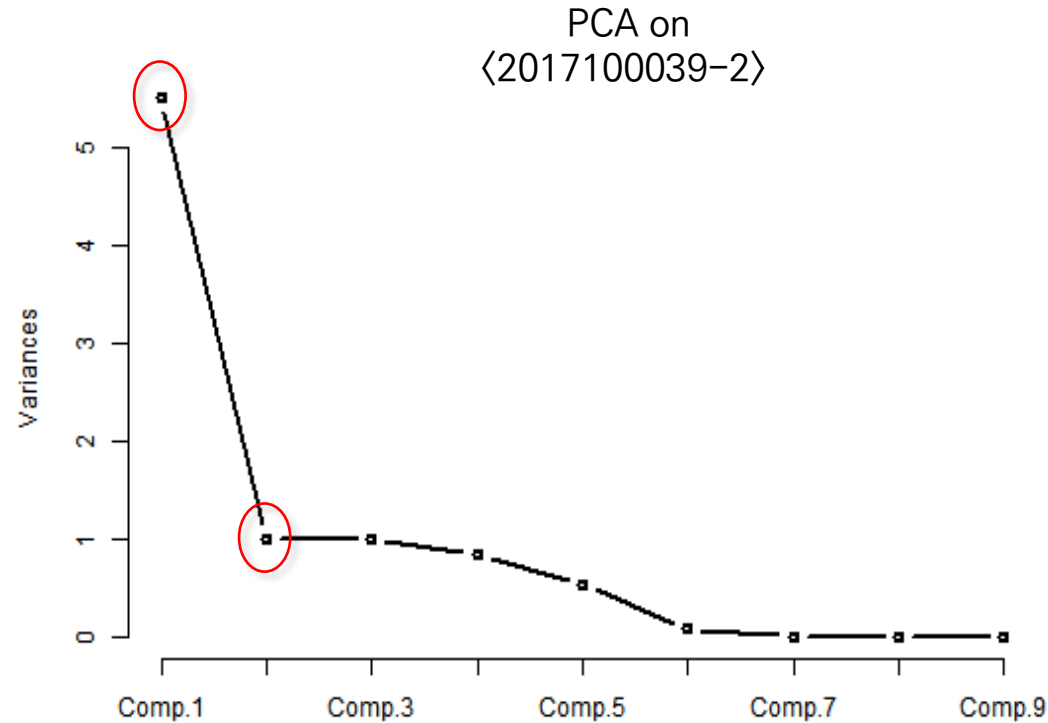


〈Principal Component Analysis〉

- 고차원 데이터에서 상관관계가 있는 변수들을 활용하여, 새로운 변수(PC1, PC2, ..., PCn)를 만드는 것
- PC1은 데이터 분포를 가장 잘 나타내는 변수이며, PC2, PC3, ..., PCn로 갈수록 분포 설명에 불필요한 정보들로 구성된 변수 (이상치 분석에 활용 가능)
- PC1, PC2 변수가 데이터들의 분포를 최대한 설명하고 있지만, 결국 원 데이터와의 차이가 존재 (Reconstruction error)
- Reconstruction error가 클수록, 변수간 상관관계를 통한 해석이 매우 어렵기 때문에, 이상치로 간주할 여지가 생김

# PCA 기반 이상감지 및 해석

- PC변수 개수 선정 기준

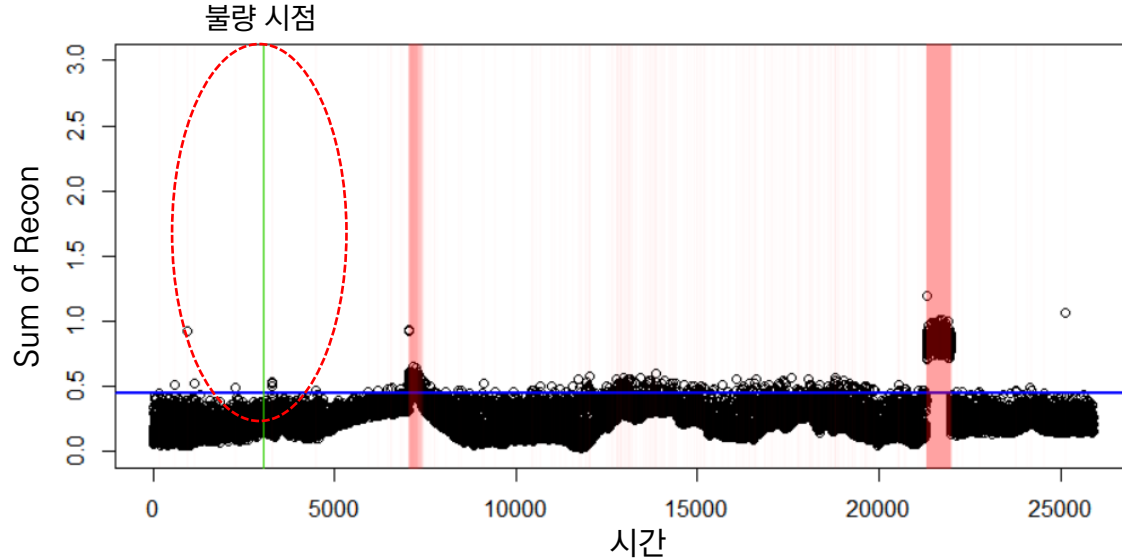


- Row 수가 가장 많은 작업지시번호 〈2017100039-2〉로 PCA 실행
- 추출된 변수들 중 분산이 급락하는 부분인 Elbow Point에 해당하는 지점으로 주성분 수를 선택 (2개 : PC1, PC2 사용)
- 또는, 분산 설명력의 누적값이 전체 정보량의 80%가 넘는 수준의 PC를 사용

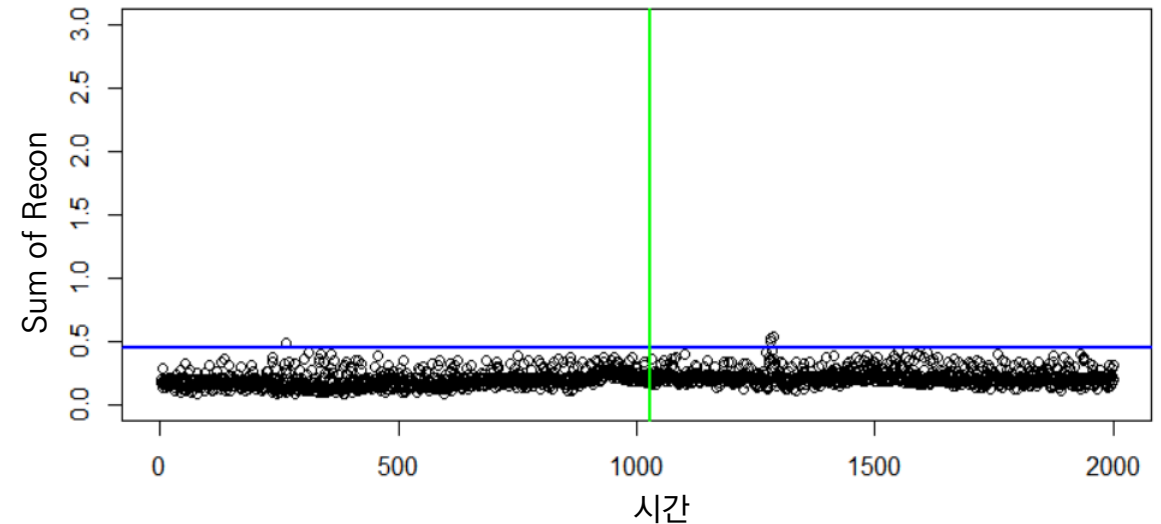
# PCA 기반 이상감지 및 해석

## PCA 기반 이상감지

〈PCA 기반 이상감지〉



〈실제 불량 발생 시점〉



- PCA의 Reconstruction error를 활용하여 이상감지 실시
- Bootstrap을 활용하여 임계값 추정 후, 유의수준 0.05에서 이상감지 실시
- 실제 제공받은 9호기 불량이력 데이터를 대조 분석하였을 때, 정상분포에 속해있기 때문에, 불량 이력데이터를 통한 유의미한 결과 도출이 어려움

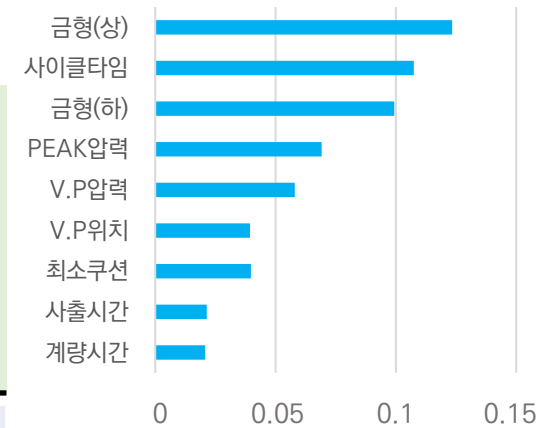
# PCA 기반 이상감지 및 해석

## PCA 기반 이상해석

〈각 공정 변수가 이상치 발생에 미치는 영향 파악〉 / PC3 ~ PC9

Component variance( $\lambda$ )		the weight of an attribute								
		V.P압력	사출시간	계량시간	사이클타임	금형(상)	금형(하)	PEAK압력	V.P위치	최소쿠션
0.612	PC1	0.382155	0.411553	0.420464	0.0211939	0.200045	0.119393	0.368657	0.398574	0.397869
0.112	PC2	0.167307	0.006926	0.043646	0.430739	0.426317	0.705319	0.241998	0.152782	0.152947
0.111	PC3	0.081773	0.003057	0.023306	0.899028	0.272079	0.285123	0.125659	0.081804	0.081888
0.094	PC4	0.17489	0.091377	0.065519	0.0752665	0.692556	0.617579	0.296171	0.034339	0.034224
0.060	PC5	0.480702	0.06842	0.163354	0.00961688	0.460419	0.159863	0.426846	0.397373	0.399778
0.010	PC6	0.294281	0.807512	0.180703	0.00143828	0.0305356	0.00338592	0.113007	0.308793	0.345833
0.001	PC7	0.648382	0.259562	0.327281	0.0019374	0.102416	0.00184454	0.624498	0.064836	0.020451
0.000	PC8	0.224943	0.310284	0.805257	0.000784978	0.0341677	0.000447362	0.342059	0.23573	0.175912
0.000	PC9	0.004152	0.042575	0.039749	0.0000289383	0.0000961519	0.0000511222	0.007182	0.704908	0.706855
$\Sigma \lambda W$		0.123334	0.10746	0.099329	0.069153551	0.057949797	0.039303193	0.039772	0.021369	0.020681

PC3 ~ PC9의  $\Sigma \lambda W$



- PCA를 기반으로 각 공정 변수가 이상치 발생에 얼마나 영향을 미치는지 전체적으로 파악 가능
- 해당 시점은 금형(상) 변수가 이상치에 발생에 가장 영향을 많이 끼칠 확률이 높다는 의미로 해석됨
- $T^2$  decomposition은 각 데이터별로 이상치 해석을 한다면, PCA는 전체 데이터 관점에서 이상 해석을 실시함 → 중요 변수가 다르게 나올 가능성 有

---

# $T^2$ 통계량 기반 이상감지 및 해석

---

# $T^2$ 통계량 기반 이상감지 및 해석

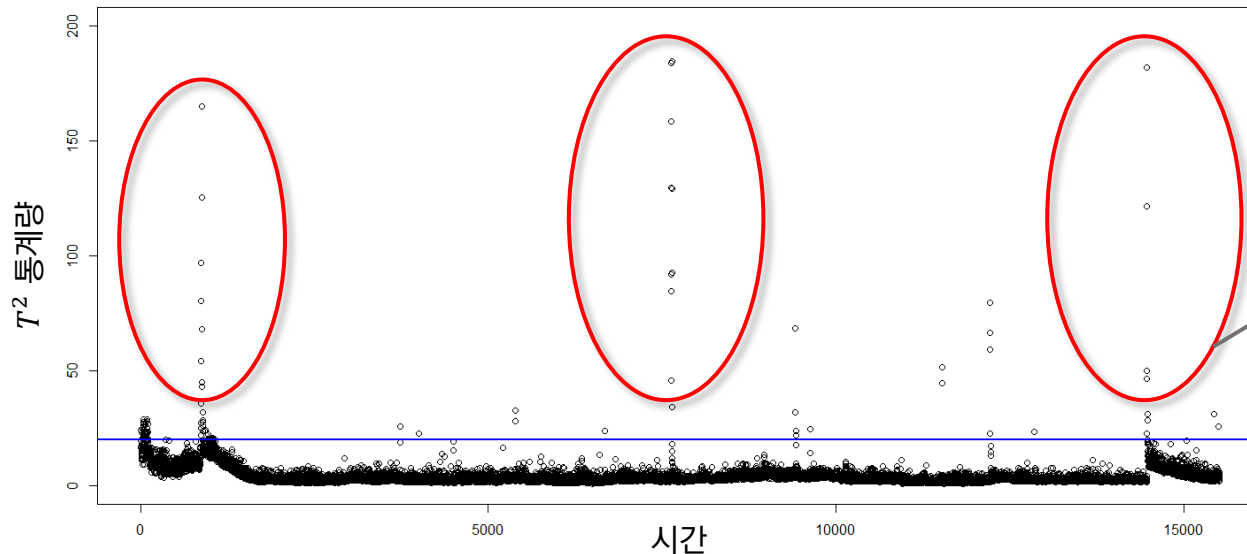
## ■ $T^2$ 통계량 기반 이상감지 : 주요 개념 및 이슈

- 모니터링 통계량

$$T^2 = (X - \bar{X})^T S^{-1} (X - \bar{X})$$

- Bootstrap based  $T^2$  Control limit 산정 후 관리도 평가
- 다변량  $T^2$  관리도의 이상구간 발생시  $T^2$  값은 모든 변수의 정보가 요약되어 있어 어떤 변수가 이상치의 원인인지 파악할 수 없음

〈작업지시번호 2018020002-11의  $T^2$  관리도〉



모든 변수의 정보가 요약되어 있음  
→ 어떤 변수가 이상을 일으켰는지 이상원인 분석이 필요함



# $T^2$ 통계량 기반 이상감지 및 해석

## ■ $T^2$ Decomposition 기반 이상원인 분석

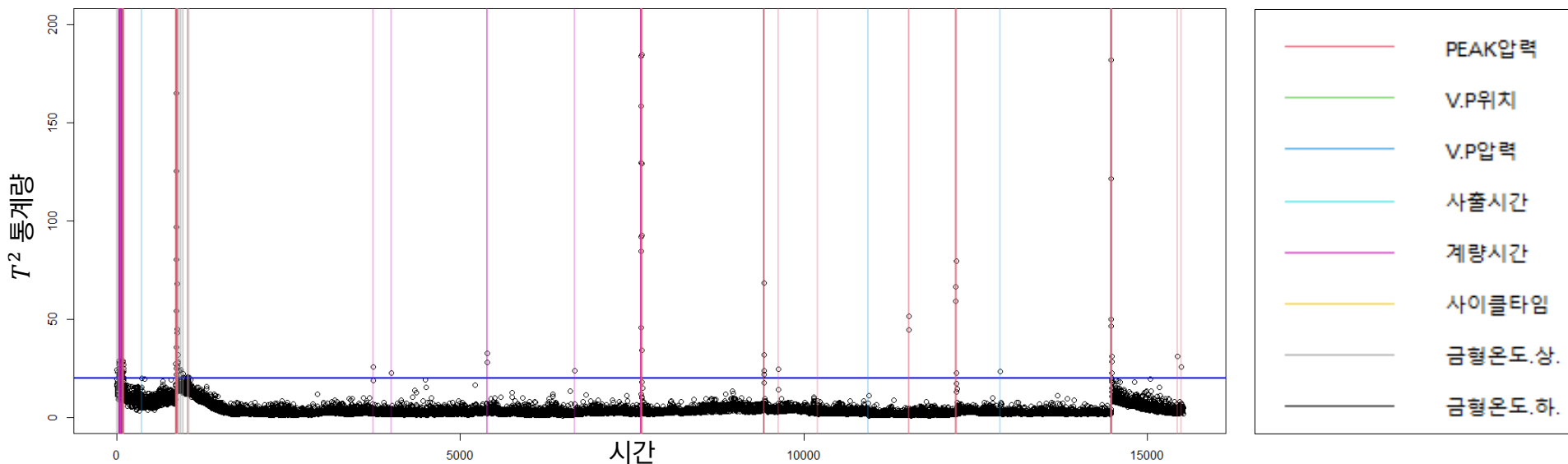
- 이상 원인 분석  $T^2$  Decomposition

$$d_i = T^2 - T^2_i$$

전체 변수를 통해 구한  $T^2$  값과 각 변수를 제외하여 구한  $T^2$  값의 차이를 활용하여 이상치에 대한 원인 변수 파악

$d_i$  값이 클수록  $i$ 번째 변수가 이상치에 대한 원인일 것

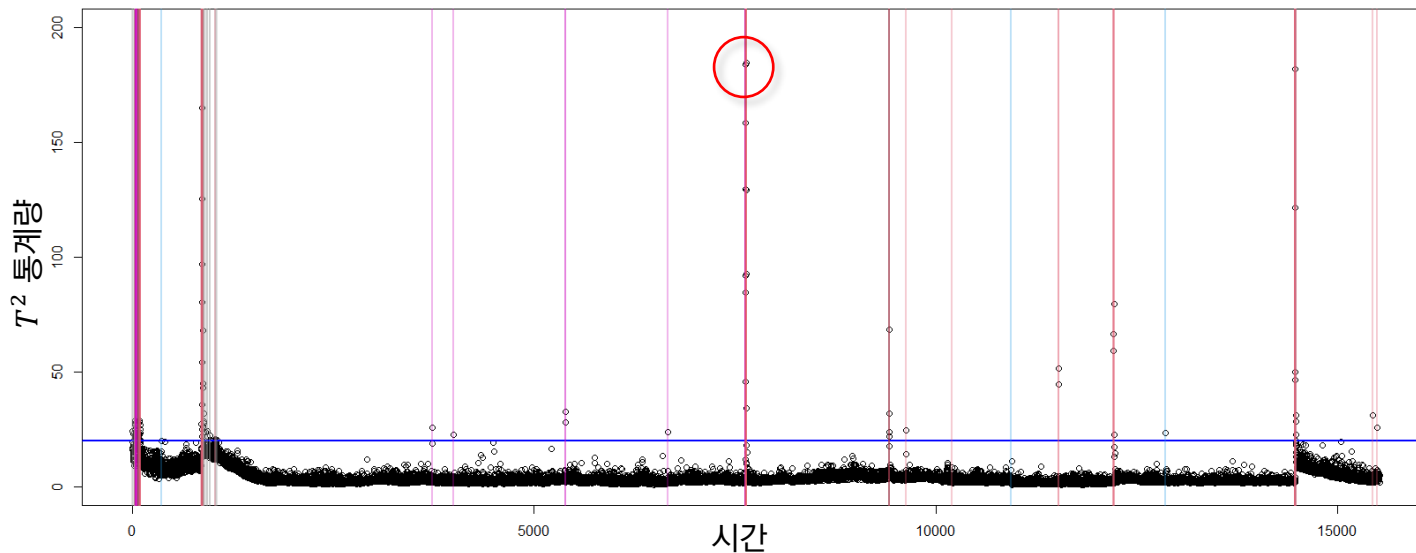
〈작업지시번호 2018020002-11의  $T^2$  Decomposition〉



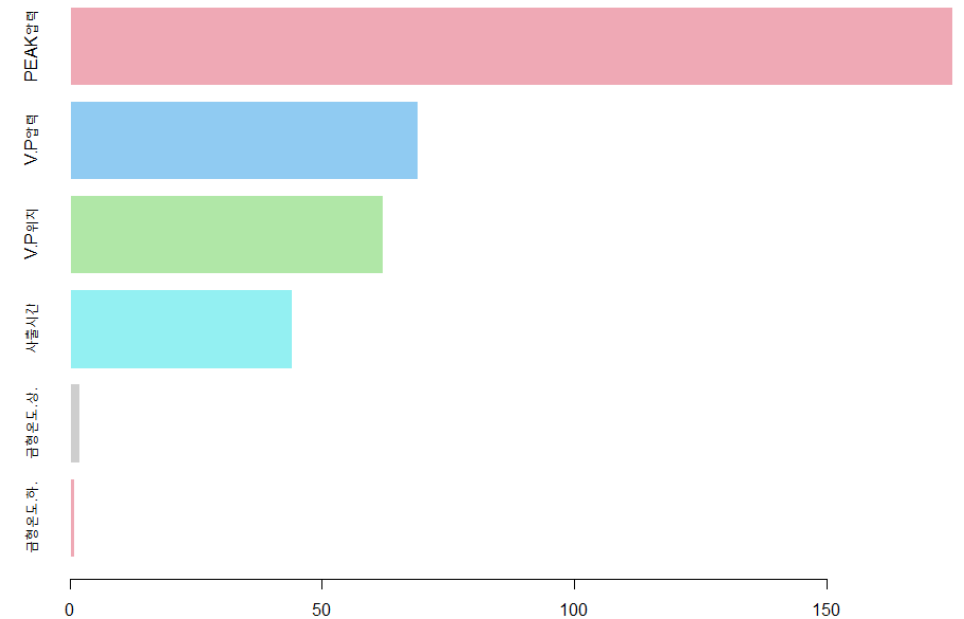
# $T^2$ 통계량 기반 이상감지 및 해석

## ■ $T^2$ Decomposition 기반 이상원인 분석

〈작업지시번호 2018020002-11의  $T^2$  Decomposition〉



〈해당 관측치에 영향을 끼친 변수〉

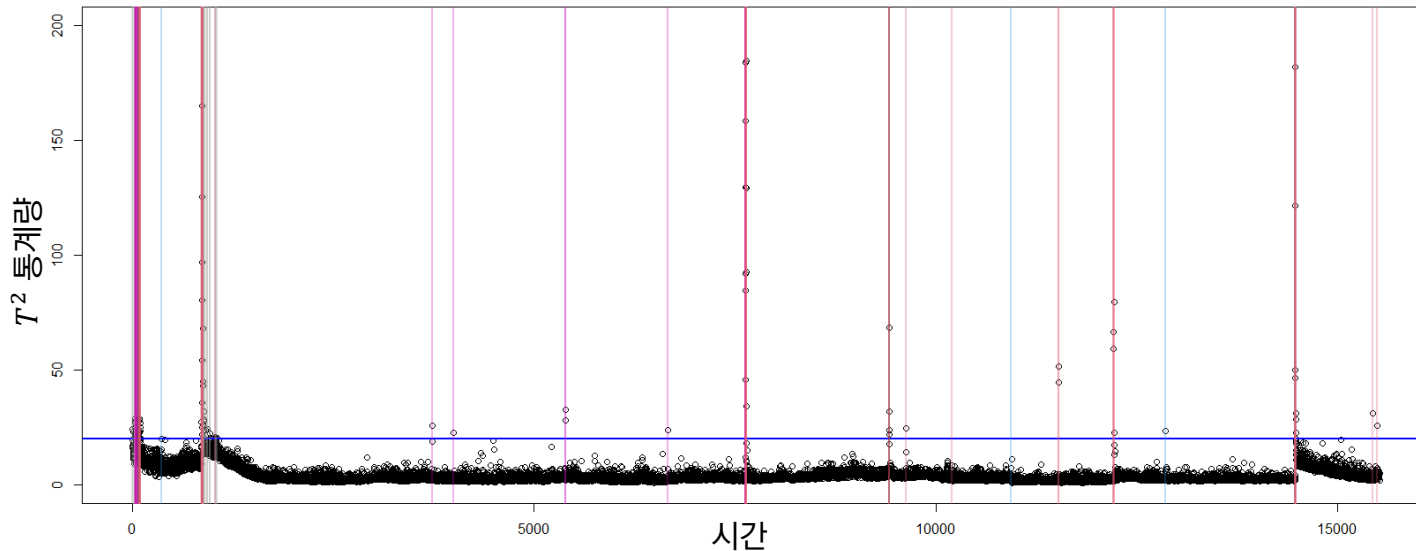


- 관리 한계선을 벗어난 해당 관측치에 영향을 끼친 변수
- PEAK압력, VP압력, VP위치, 사출시간, 금형온도(상), 금형온도(하)로 해석이 가능함

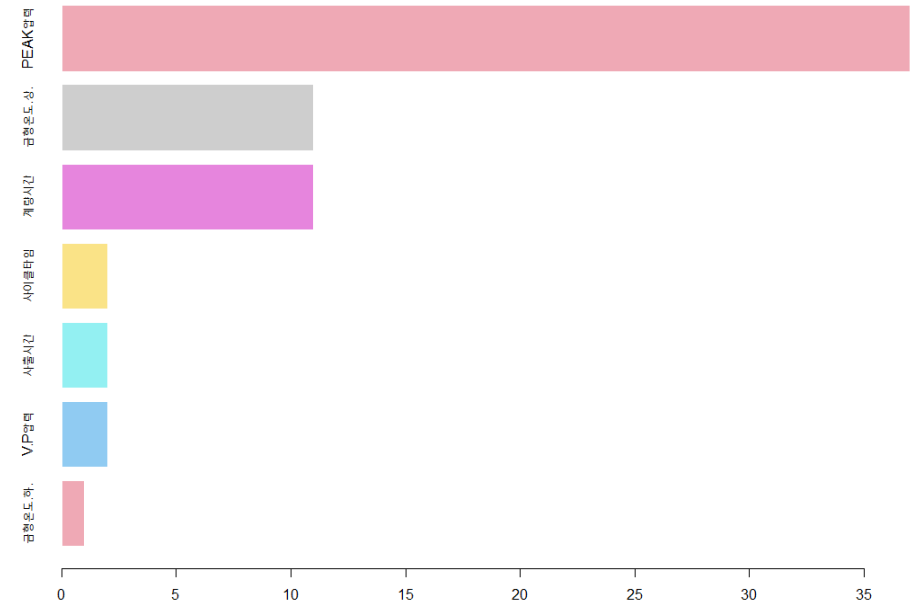
# $T^2$ 통계량 기반 이상감지 및 해석

## ■ $T^2$ Decomposition 기반 이상원인 분석

〈작업지시번호 2018020002-11의  $T^2$  Decomposition〉



〈이상치에 가장 큰 원인이 되는 변수의 빈도 수〉



- 관리 한계선을 벗어난 관측치에 가장 큰 원인이 되는 변수의 빈도 수
- PEAK압력, VP위치, 금형온도(상), 사출시간, 사이클타임, VP압력 순으로 이상 관측치가 많이 나타남
- 사출 공정 품질 향상을 위해서는 작업지시번호별 주요 변수의 Set Value값을 최적화 해야함

---

# BURR 불량 이력 분석

---

# BURR 불량 이력 분석

## ▪ BURR 불량 이력 데이터셋 (작업지시번호 2018020002-11)

일시	PEAK압력	최소쿠션	V-P위치	V-P압력	사출시간	계량시간	사이클타임	금형(상)	금형(하)
18-02-27 20:12:20	1989.2	9.6	9.6	1900.3	2.5	1.55	18.66	19.7	20.6
18-02-27 20:33:55	1654.4	1.3	2	1493.3	0.19	3.86	18.66	19.9	20.9
18-02-27 20:34:27	1937.7	11.3	11.3	1900.1	2.5	1.15	18.66	19.9	20.8
18-02-27 20:34:51	1939.8	10.9	10.9	1899.8	2.5	1.24	17.8	19.8	20.7
18-02-27 20:36:06	1949.3	11	11	1899.8	2.5	1.23	17.8	19.9	20.8
18-02-27 20:36:40	1945.4	10.9	10.9	1900.6	2.5	1.24	17.86	19.8	20.6
18-02-27 20:41:35	1955.5	11	11	1899.8	2.5	1.22	17.8	19.7	20.7
18-02-27 20:41:54	1949.2	10.9	10.9	1899.5	2.5	1.21	18.66	19.8	20.8
18-02-27 20:42:13	1964.4	11	11	1900.1	2.5	1.18	18.64	20	20.7
18-02-27 20:42:31	1967.1	11.1	11.1	1899.5	2.5	1.17	18.66	20	20.8
18-02-27 20:47:41	1656.3	1.4	2	1471.4	0.19	3.85	18.66	19.7	20.5
18-02-27 20:48:03	1937.4	11.1	11.1	1899.9	2.5	1.2	18.66	19.7	20.6
18-02-27 20:48:21	1945	10.9	10.9	1899.9	2.5	1.24	18.66	19.6	20.5
18-02-27 20:48:40	1950	10.9	10.9	1900.3	2.5	1.24	18.68	19.5	20.6
18-02-27 20:48:59	1958.5	10.9	10.9	1899.8	2.5	1.23	18.64	19.5	20.5
18-02-27 20:49:17	1964.4	11	11	1899.9	2.5	1.23	18.66	19.5	20.5

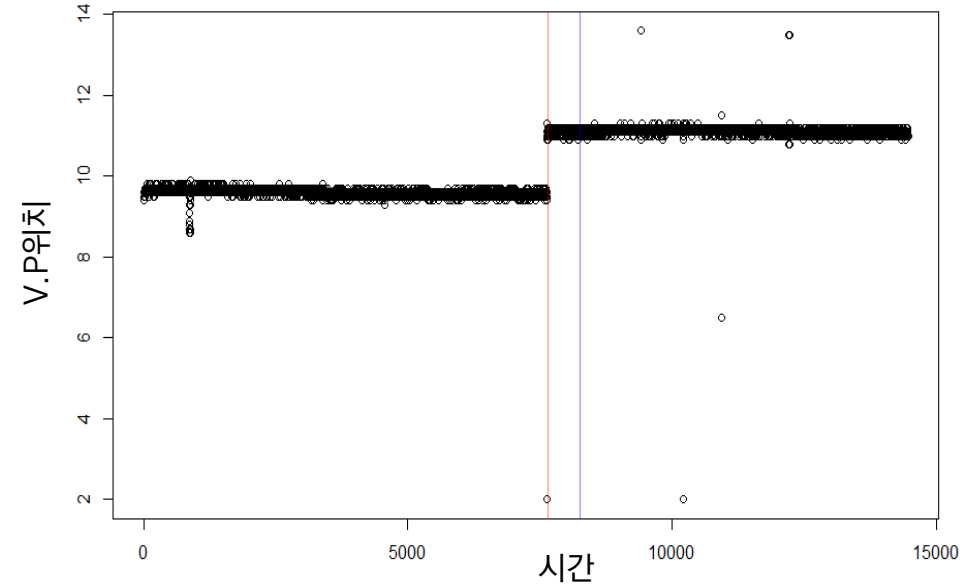
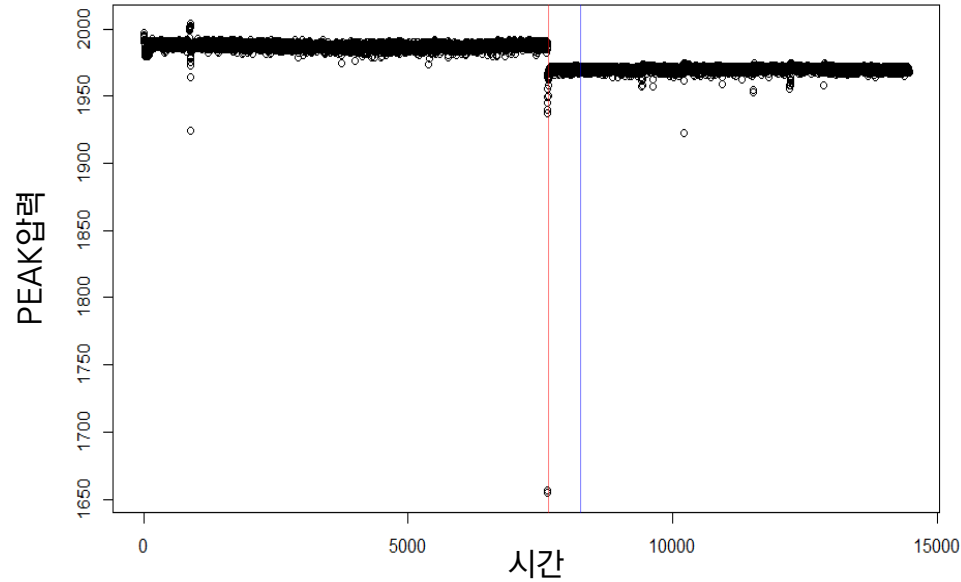


BURR불량 발생 의심 row : 14개

- 2018-02-24 ~ 2018-03-01 구간의 데이터셋 14,4470개 中 BURR 불량 발생 의심 데이터 14개
- 빨간색 row는 매우 값이 튀는 이상치
- 해당 구간의 변화가 BURR 불량을 야기하였는지에 대한 해석 필요

# BURR 불량 이력 분석

## EDA



- 빨간선 - 불량 발생 의심구간 / 파란선 - 출하검사일
- 공정변수 EDA 결과, 분포에 변화가 있던 변수들은 점진적으로 바뀌지 않고, 이산적인 형태의 변화를 보임
- 따라서, 해당 구간에서 이상 발생에 따른 변화보다는 레시피 변동이 있었을 것으로 판단됨

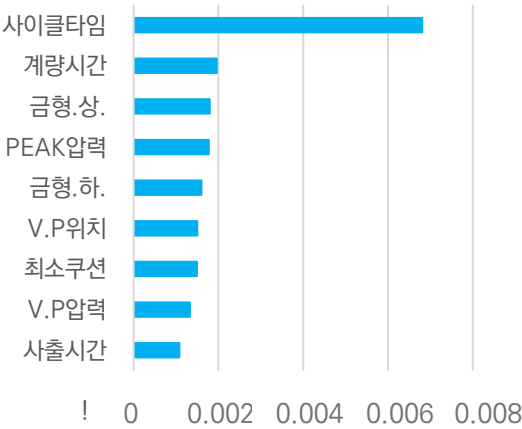
# BURR 불량 이력 분석

## PCA 기반 이상 해석

〈각 공정 변수가 이상치 발생에 미치는 영향 파악〉 / PC4 ~ PC9

Component variance( $\lambda$ )		the weight of an attribute								
		V-P위치	V-P압력	금형(상)	금형(하)	PEAK압력	최소쿠션	사출시간	계량시간	사이클타임
0.468	PC1	0.46961	0.018419	0.24238	0.254947	0.422558	0.469101	0.029175	0.461597	0.206996
0.236	PC2	0.095203	0.654384	0.132166	0.130572	0.280851	0.102057	0.65167	0.10966	0.047011
0.198	PC3	0.123578	0.138576	0.616159	0.607927	0.146603	0.121959	0.128956	0.115458	0.387262
0.076	PC4	0.162232	0.033257	0.226564	0.200263	0.140889	0.161716	0.005751	0.175104	0.896907
0.013	PC5	0.045807	0.645566	0.005903	0.004161	0.154792	0.046152	0.74108	0.075294	0.011632
0.005	PC6	0.113951	0.274838	0.021978	0.01842	0.577881	0.115252	0.05721	0.748288	0.014815
0.003	PC7	0.460086	0.241909	0.046276	0.083313	0.584089	0.458236	0.072071	0.405436	0.002543
0.001	PC8	0.051162	0.017482	0.700101	0.707796	0.046465	0.050691	0.011649	0.03048	0.012945
0	PC9	0.706705	0.004068	0.000134	0.000156	0.00046	0.707489	0.003267	0.0000398073	0.000028
		사이클타임	계량시간	금형(상)	PEAK압력	금형(하)	V.P위치	최소쿠션	V.P압력	사출시간
$\Sigma \lambda W$		0.068038	0.019587	0.017935	0.017711	0.016029	0.014986	0.014953	0.133064	0.0176631

PC4 ~ PC9의  $\Sigma \lambda W$

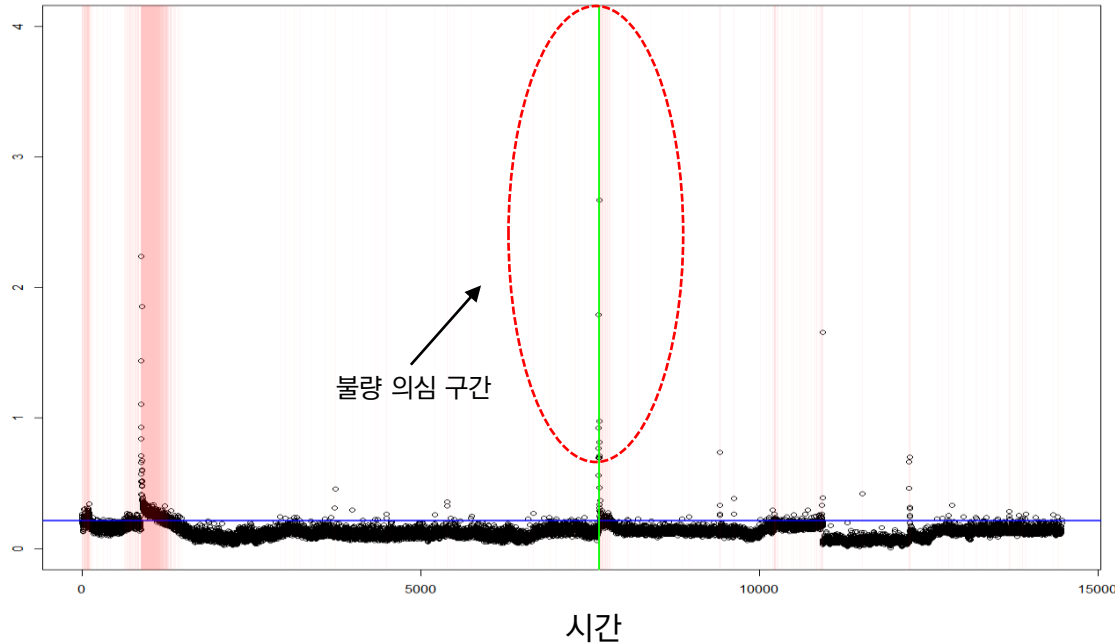


- BURR 불량 발생 의심구간에 대한 PCA 기반 이상 해석 실시
- 해당 구간은 사이클타임 변수가 이상치에 발생에 가장 영향을 많이 끼칠 확률이 높다는 의미로 해석됨

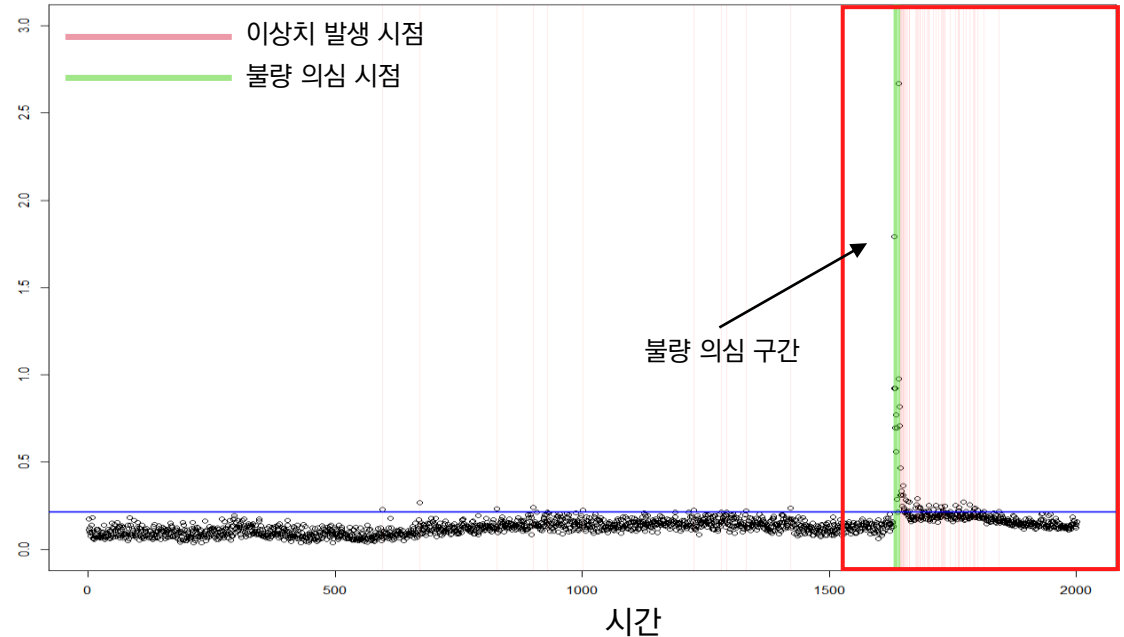
# BURR 불량 이력 분석

## PCA 기반 이상감지

〈PCDD 기반 이상감지〉



〈불량 의심 구간 확대 시각화〉



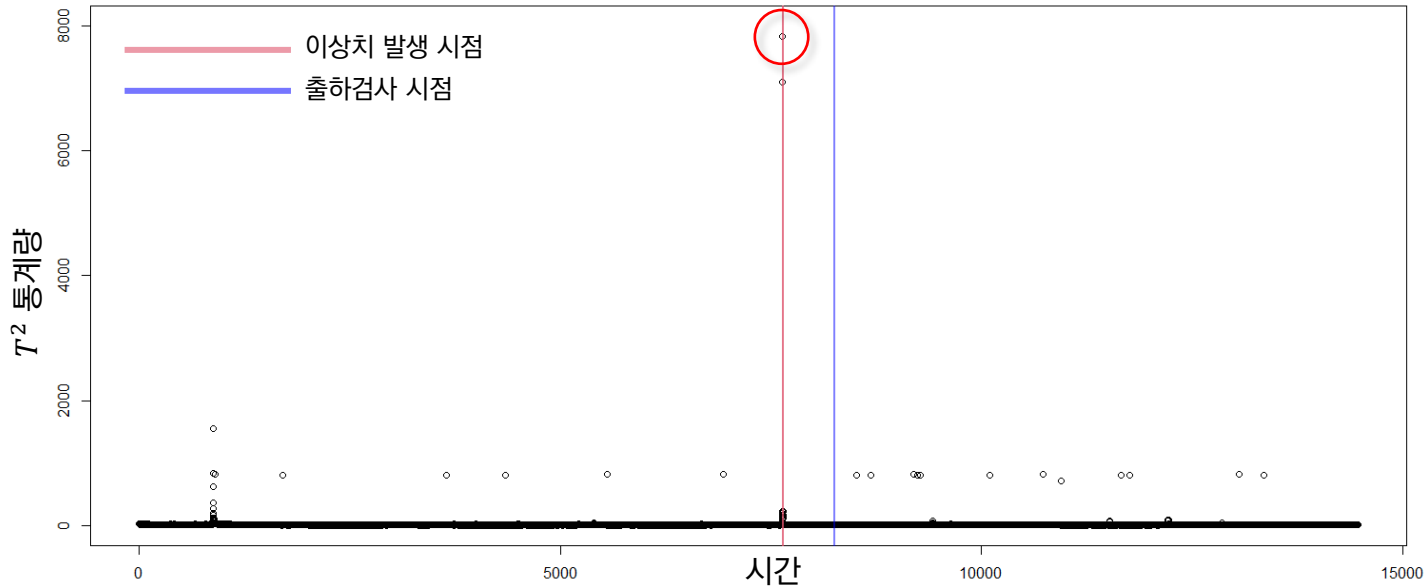
- Bootstrap을 활용하여 임계값 추정 후, 유의수준 0.05에서 이상감지 실시
- BURR 불량 의심 구간 이후로, 이상치 발생 빈도가 다소 높아지는 모습을 보임
- 작업자가 레시피를 조정하는 단계에서 발생한 이상치일 것이라고 예상됨



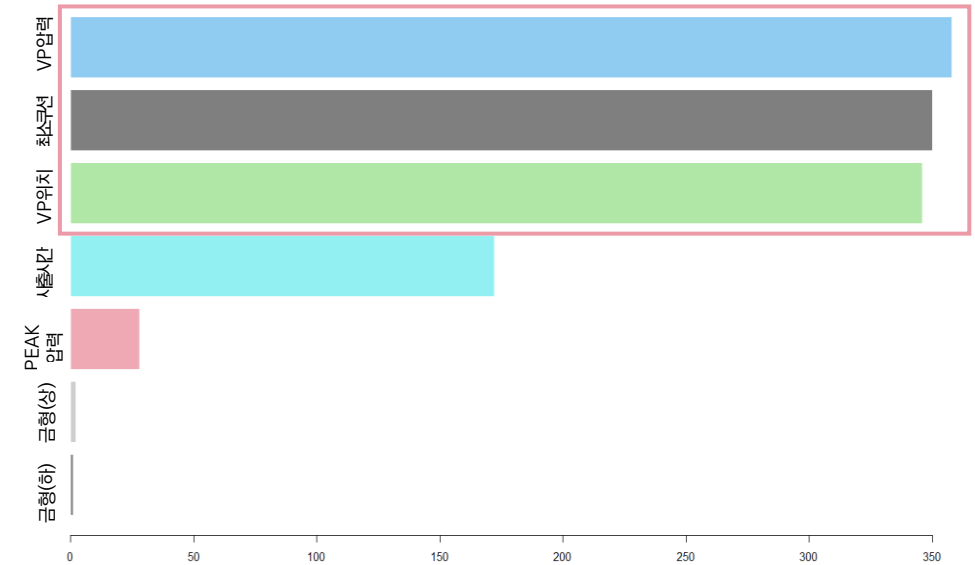
# BURR 불량 이력 분석

## ▪ $T^2$ Decomposition 기반 이상원인 분석

〈작업지시번호 2018020002-11의  $T^2$  통계량〉



〈해당 관측치에 영향을 끼친 변수〉



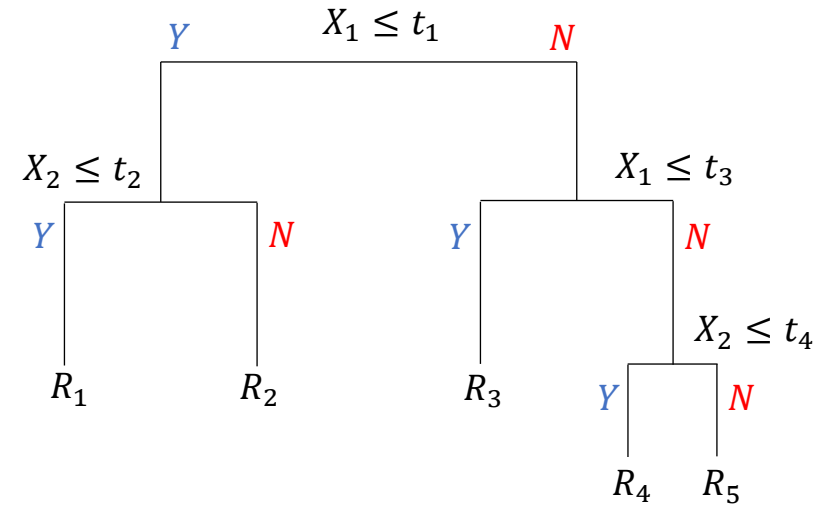
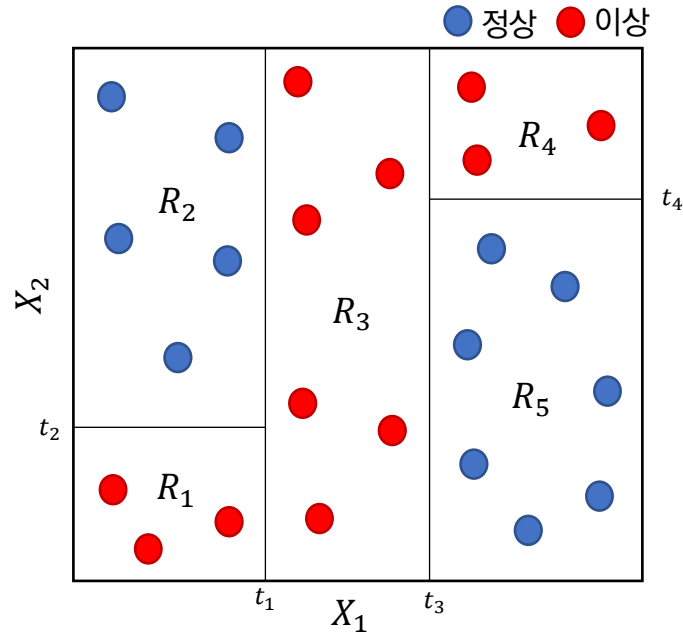
일시	PEAK압력	최소쿠션	V-P위치	V-P압력	사출시간	계량시간	사이클타임	금형(상)	금형(하)
2018-02-27 20:12:01	1987.1	9.4	9.4	1899.5	2.5	1.58	18.64	19.7	20.6
2018-02-27 20:12:20	1989.2	9.6	9.6	1900.3	2.5	1.55	18.66	19.7	20.6
2018-02-27 20:33:55	1654.4	1.3	2	1493.3	0.19	3.86	18.66	19.9	20.9
2018-02-27 20:34:27	1937.7	11.3	11.3	1900.1	2.5	1.15	18.66	19.9	20.8
2018-02-27 20:34:51	1939.8	10.9	10.9	1899.8	2.5	1.24	17.8	19.8	20.7

: 해당 관측치가 BURR불량 시점이 맞다면  
VP압력, 최소쿠션, VP위치를 제어해야함

- $T^2$  Decomposition을 활용하여 outlier값의 원인 분석이 가능함
- 단변량인자에서 확인 가능함

# BURR 불량 이력 분석

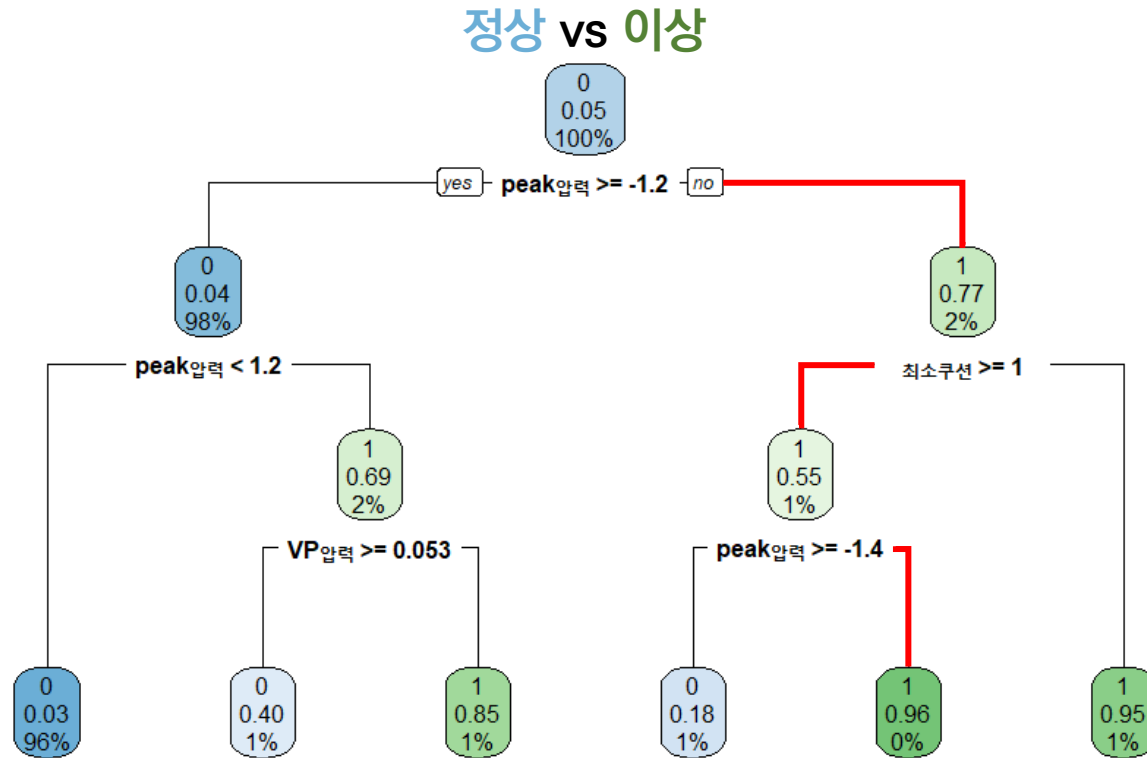
## Decision Tree란?



- 의사결정나무는 데이터마이닝의 대표적인 기법으로 두 개(혹은 그 이상)의 집단이 있을 때, 그것들을 구분하는 Rule을 제공하는 알고리즘
- 정상/이상 집단에 대한 판정 Rule들에 대한 직관적 확인 가능
- 제조/의료/서비스 등 다양한 분야에서 활용되고 있음

# BURR 불량 이력 분석

## Decision Tree 기반 이상해석

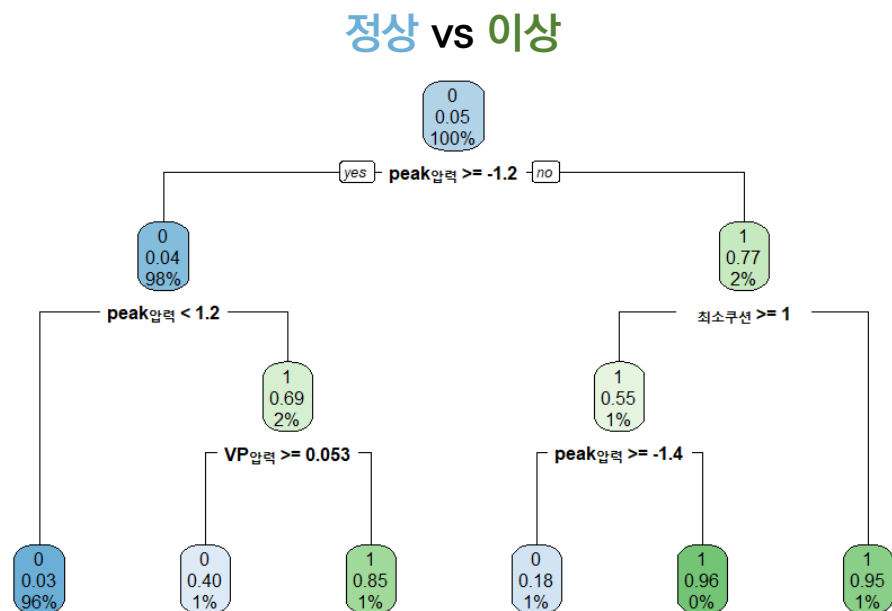


If  
peak압력 < -1.2  
&&  
If  
최소쿠션 ≥ 1  
&&  
Peak압력 < -1.4  
Then  
이상발생

- 데이터마이닝의 대표적인 기법으로 다변량 공정 모니터링 시 이상구간에 대한 Rule 제공
- If – Then Rule 기반으로 이상경보에 대한 해석 가능

# BURR 불량 이력 분석

## Decision Tree 기반 이상해석



이상 발생 해석				
PEAK압력 < -1.2	→	최소쿠션 ≥ 1	→	PEAK압력 < -1.4 → 이상발생
PEAK압력 < -1.2	→	최소쿠션 < 1	→	이상발생
PEAK압력 ≥ -1.2	→	PEAK압력 ≥ 1.2	→	VP압력 < 0.053 → 이상발생

- 관리한계선을 벗어난  $T^2$  통계량의 원인 인자로는 PEAK압력, 최소쿠션, VP압력의 변동이 주요 원인
- Decision Tree의 다변량 조건으로 관리시 사출공정의 이상을 제어할 수 있음

# 이상 원인 항목 분석 프로세스

## Monitoring & Detection



- 다변량 이상감지(PCDD)
- $T^2$  Control Chart
- 비모수 추정법 기반 임계치 산정

## Identification

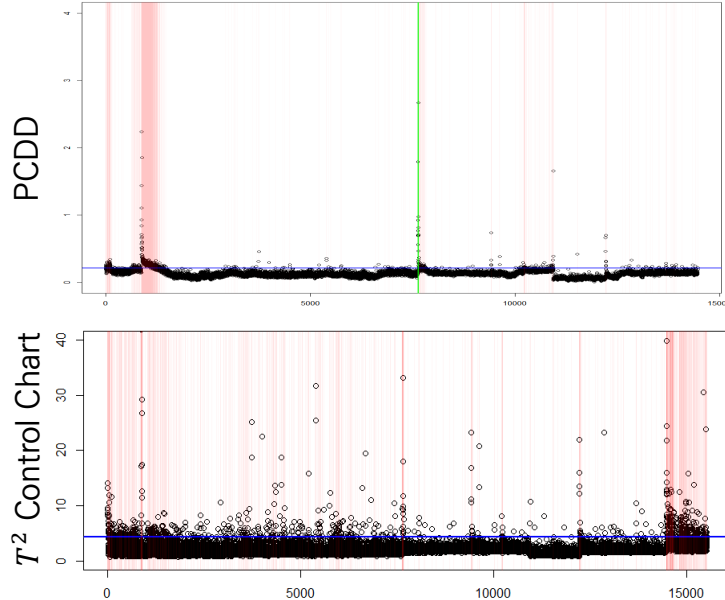


- $T^2$  Decomposition

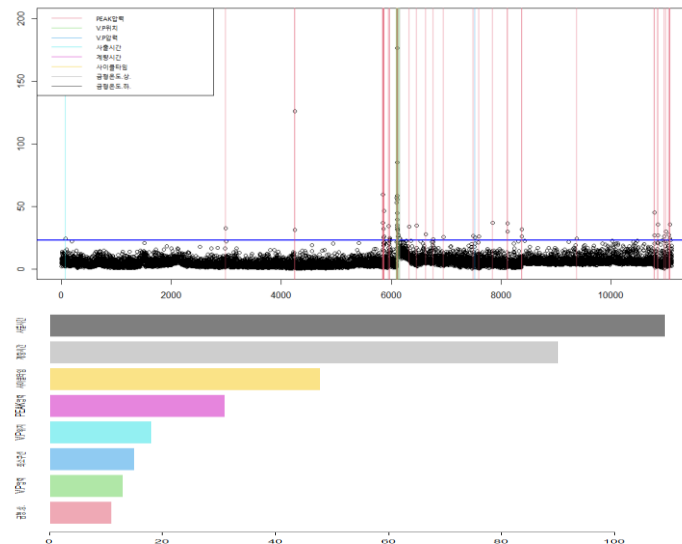
## Diagnosis & Interpretation



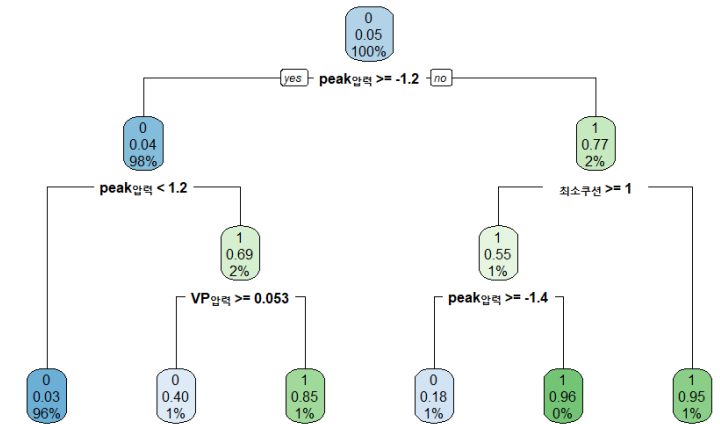
- Decision Tree



## $T^2$ Decomposition



## Decision Tree



- 사출 공정의 이상을 조기 감지하여 대응 조치 취할 수 있도록 하는 모니터링 시스템 방법론을 제안함
- PCA로 모니터링 모델을 구성하여 다변량 이상 감지
- 이상 여부를 확인하기 위해  $T^2$  통계량 비교 및 Decomposition으로 이상 해석
- 이상 조기 감지하여 이상의 원인을 진단하고 이상 원인 변수 제어를 통한 공정최적화 실현

---

감사합니다

---