

# 데이터 마이닝을 이용한 생산공정 데이터 분석 시나리오

이 형 옥<sup>\*1)</sup> · 배 성 민<sup>2)</sup>

한국교통대학교 에너지시스템공학과<sup>1)</sup> · 한밭대학교 산업경영공학과<sup>2)</sup>

## Scenarios for Manufacturing Process Data Analysis using Data Mining

Hyoung-wook Lee<sup>\*1)</sup> · Sung-min Bae<sup>2)</sup>

<sup>1)</sup>Department of Energy System Engineering, Korea National University of Transportation, 50 Daehak-ro Chungju-si, Chung-buk, 380-702, Korea

<sup>2)</sup>Department of Industrial & Management Engineering, Hanbat National University, 125 Dongseo-Daero, Yooseong-Gu, Daejeon, 305-719, Korea

(Received 2013. 03. 05 / Accepted 2013. 04. 29)

**Abstract** : Process and manufacturing data are numerous accumulated to the enterprise database in industries but little of those data are utilized. Data mining can support a decision to manager in process from the data. However, it is not easy to field managers because a proper adoption of various schemes is very difficult. In this paper, six scenarios are conducted using data mining schemes for the various situations of field claims such as yield problem, trend analysis and prediction of yield according to changes of operating conditions, etc. Scenarios, like templates, of various analysis situations are helpful to users.

**Key words** : Manufacturing process data, Data mining, Neural network, Decision tree, Clustering, Scenario

### 1. 서 론

제조업체의 생산공정 관리의 현실은 양산라인의 생산공정에서 나오는 POP 데이터 등의 관리는 SPC를 통하여 ERP 등에 저장되고 있으나, 통계적 관리와 관리도적 측면에서만 사용되고 있다. 저장된 데이터에서 많은 지식을 추출할 수 있지만 공정개선에의 활용성이 떨어지고 있는 현실이며, 너무 방대한 데이터 량으로 인하여, 클레임 발생 시 분석할 대상조차 찾지 못하고 있다.<sup>1,2)</sup>

만일 생산공정 및 제품설계에 대한 다방면의 전문가가 있는 경우에는 공정개선 시 많은 데이터 중에서 직관에 의한 필요한 데이터의 선별이나 QFD 전개를 통하여 미리 필요한 데이터의 관리를 수행하고, 변수를 줄일 수 있지만, 현실적으로 이러한 전문가를 보유

하기는 쉽지 않다. 또한, 공정개선과 불량에 대한 대비 측면에서 공정 FMEA를 기업에서 TFT를 구성하여 작성하지만 지속적인 개선이나 업데이트가 이루어지지 힘든 현실이며, 개선의 방향을 찾는 것 또한 쉽지 않다.

그러나 생산 공정 데이터 안에 들어있는 지식은 이러한 공정 FMEA의 중요한 부분을 함축하고 있어 생산 공정 데이터 안의 지식을 추출할 수 있다면 공정 FMEA의 개선과 함께, 양산라인의 혁신을 이룰 수 있다.

데이터마이닝에 대한 알고리즘이나 프로그램들은 많이 개발되어 있지만, 현장의 작업자가 실제로 데이터를 분석하고자 할 때는 각 기법들을 어떻게 조합해서 결과를 얻어야 하는지 모르는 경우가 대부분이다. 본 논문에서는 제조업 생산현장에서 주요한 문제가 생길 수 있는 상황을 제시하고, 해당 상황에서 주요 데이터 마이닝 기법을 이용하여 생산공정 데이터를 분석하여 의사결정에 도움을 받을 수 있도록 하는 시나리오를 제시하고자 한다.

\*Corresponding author, E-mail: hwlee@ut.ac.kr.

## 2. 생산공정 데이터마이닝 시스템

현장의 SPC, PDM, ERP 등의 데이터로부터 구축된 생산 데이터웨어하우스에서 데이터 추출모듈을 이용하여 분석할 데이터마트를 구성하고, 신경망회로, 군집분석, 의사결정나무 등의 분석모듈을 이용하여, 생산 공정의 지식, 공정 조건변화에 대한 예측을 수행하게 된다.

Fig. 1에서와 같이 각 공정별로 구성된 공정 FMEA를 개선하기 위하여 또는, 불량이 발생한 경우, 공정을 선택하여 (P#11) 공정정형화 모듈에 들어있는 공정 DB 정보를 불러온다. 이 공정 DB 정보로부터 생산 Data Warehouse에서 필요한 정보를 추출하여 Data Mart에 저장하고, 데이터마이닝 분석모듈을 가동하여 생산규칙을 도출한다. 도출된 규칙은 Rule DB에 저장되고 공정 FMEA를 수정하기 위하여 수정화면을 불러오면, 해당 공정에서 추출된 규칙을 보여준다.<sup>1)</sup>

Fig. 2에는 데이터마이닝 시스템의 개념도를 도시하였다. 생산 공정 중에서 저장되는 데이터는 장비의

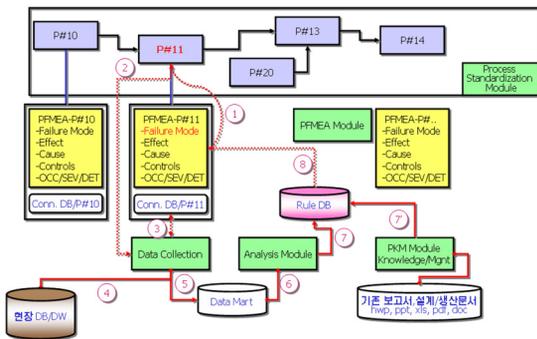


Fig. 1 Data mining process

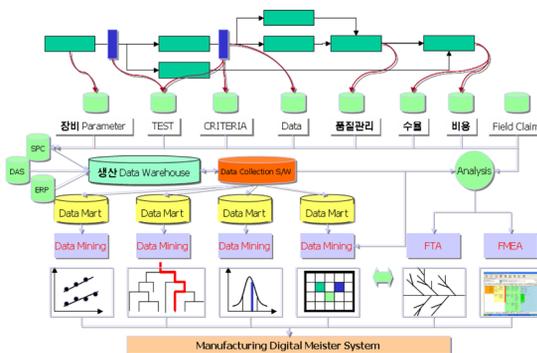


Fig. 2 Schematic diagram of data mining system

파라미터, Test 값, 결과, 판별치, 품질관리 데이터, 수율 데이터, 비용데이터, 필드 클레임 데이터 등이며, 이러한 데이터는 SPC와 ERP를 통하여 Data Warehouse에 저장된다. 데이터마이닝을 구동하기 위해 필요한 Data Mart는 Data warehouse에서 데이터 추출모듈을 통하여 이루어지며, 얻어진 데이터로부터 데이터 마이닝 모듈을 구동하여 군집분석, 의사결정나무, 분류분석, 신경망회로를 통한 예측 공정을 통하여 FMEA와 연계되도록 한다.

생산공정 데이터를 분석하기 위한 가장 기본적인 조건은 데이터의 Traceability로 동일 아이템에 대하여 일렬로 데이터가 정리되어야 분석이 가능하다. 이러한 데이터는 현장에서 데이터를 수집하는 순간부터 저장될 때까지 모두 데이터 관리가 되어야 한다. Fig. 3에는 이러한 데이터의 양식을 예로 보여준다.

## 3. 분석 시나리오

Fig. 4에는 양산라인 또는 생산공정라인에서 보유하고 있는 공정데이터의 종류에 따른 지식 추출방안과 데이터의 활용방안을 요약하였다. 크게 데이터의 경향파악과 달라진 경향에 대한 이유 분석, 판별치에

공정장비의 셋업 조건도 사용가능

Fig. 3 Example of data mart format

No	장비조건 공정측정 DB	성능측정 DB	최종 판별 값	분석 모듈	활용방안
1	○	○	○	Clusterin g	• 시가별 데이터 군집화 비교를 통한 공정의 안정화, 경향도 파악 • 시가별 데이터에서 지어진 경우 1-에서 영향 변수 파악
1-1	○	○	○	Classifica tion	• 군집화 분석에서 지어진 경우 기준 중의 영향 변수 지어올리는 주요 영향 변수의 파악, 특성분석 → 관리대상 설정(FMEA)
2	○	○	○	Classifica tion	• 판별값에 영향을 주는 성능측정 변수의 ranking 파악 • 판별을 위한 rule 도출
3	○	○	○	Classifica tion	• 판별값에 영향을 주는 공정측정 변수(장비조건)의 ranking 파악 • 판별을 위한 rule 도출 → 많은 공정측정 변수(장비조건) 중 영향력 있 는 변수만 따로 추출, 관리대상 설정(FMEA) 개선
3-1	(주변수)	○	○	Neural Network	• 각 공정에서의 공정측정 변수 최종 판별값 예측 • >1에서 공정측정 변수(장비조건)와 최종 판별값의 관계 설정으로 미려한 공정측정 변수(장비조건)에 대한 수율 예측 및 판단 예측
4	○	○	○	Classifica tion	• 최종 판별값에 영향을 주는 공정측정(장비조건), 성능측정 변수의 ranking 추출 • 주 영향 변수에 대한 관리대상 설정 및 FMEA 개선 (4-1)
4-1	○	○	○	Classifica tion	• 최종 판별값 "NG" 또는 "OK"에 대한 Rule로부터 미리 선정된 주요 공 정 변수에 대한 기준을 만족시키는 신규 기준 설정 및 고정된 수율
5	○	○	○	Classifica tion	• 장비조건과 공정측정 결과의 경향도 파악, 주변수 파악으로 변수 중립

Fig. 4 Usages of data mining analysis algorithm according to manufacturing data base

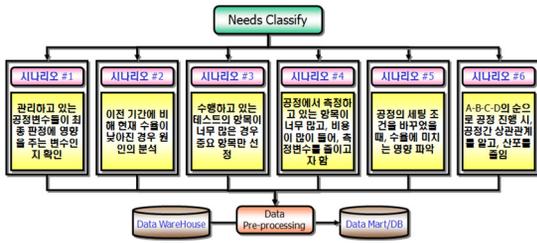


Fig. 5 Scenarios of data mining analysis for manufacturing process data

영향을 주는 인자를 구해내는 의사결정나무, 그리고, 변화된 조건에 따른 수율 예측을 위한 신경망회로로 구성되어 있으며, 각 상황과 데이터의 보유 공정에 따라 순차적 또는 개별적인 적용을 통하여 지식을 추출할 수 있다.

생산공정데이터를 이용한 제조업체의 문제상황들에 대한 분석시나리오를 Fig. 5와 같이 정리할 수 있다. 각각의 경우에 대해서 분석방법을 정리하면 다음과 같다.

1) 현재 우리가 관리하고 있는 공정변수들이 우리 제품의 최종 판정에 영향을 주고 있는 변수들인지 확인하고 싶은 경우, 우선 공정변수들의 측정값과 최종 판정치를 이용하여 Classification 작업(3번)을 수행 후, 의사결정나무에서 Root Node에 있는 변수들로부터 아래로 변수 리스트 도출하고, 의사결정나무에 나타난 Criteria를 기존 관리기준과 비교하여 공정 FMEA를 수정하도록 한다.

2) 이전 달 (또는 직전분기)과 비교해서 이번 달 (또는 이번 분기)의 수율이 낮아진 경우 원인을 분석하고 싶을 때는 전 달과 이번 달의 공정 측정 데이터 (특정 공정 또는 모든 공정)의 트렌드를 비교하여 달라진 점이 있는 지 확인(No 1)하고, 달라진 점이 있다면 두 그룹이 차이가 나게 하는 변수의 특성을 파악하여 어떤 변수에서 차이가 나고 있는지를 확인(No 1-1)한다. 이때 변수의 특성이 기존 관리 조건과 다르다면 공정 FMEA 수정이 가능하고, 달라진 점이 없다면 장비의 세팅 조건 및 장비 노후화를 의심해야 한다.

3) 지금 우리가 수행하고 있는 성능테스트 (또는 최종 제품 테스트) 항목이 너무 많아 줄이고자 할 때는 성능테스트 데이터와 최종 판정 데이터를 조합하여, Classification (No 2)을 수행한다. 의사결정나무의 root node로부터 도출되는 순서가 최종 수율에 영향을 많

이 주는 순서가 된다. 공정에 영향이 없다면, 성능테스트 순서를 중요도 순서대로 바꾸면 좀 더 시간을 절약할 수 있다.

4) A 공정에서 측정하는 변수가 너무 많고, 측정하는데 드는 비용이 만만치가 않다. 그래서, 측정변수의 수를 줄이고 싶을 때는 No 5의 방식으로 A공정 측정값과 A공정 측정치의 판정값을 이용해 Classification을 수행하여, 변수의 중요도 순서대로 변수의 수를 줄인다. 측정변수의 수를 줄이고 난 뒤, 수율에의 영향력을 파악하기 위해서는 No 3-1을 이용하여 변수의 수를 줄이고 난 뒤의 측정값과 최종 판정치를 가지고 검증해야 한다.

5) A 공정의 세팅 조건을 바꾸었을 때 수율에 얼마만큼의 영향을 미치는지 궁금할 경우, 기존 A공정의 측정값과 판정치를 가지고 No3-1과 같이 신경망을 이용하여 weight 파일을 생성하고, 바뀐 세팅 조건에서 생성된 측정값을 기존 weight 파일을 이용해서 예측할 수 있다. 이를 이용하여, 기존 공정과 바뀐 공정의 수율의 차이를 계산할 수 있다.

6) A → B → C → D 공정인 경우, A공정과 C공정간의 상관관계를 알고 싶은 경우, A공정 측정데이터 중 OK 판정 받은 것들 (즉, 다음 공정으로 넘어간 제품의 측정데이터) + C공정 판정데이터를 대상으로 Classification을 이용하여 의사결정나무를 만들고, root node 부터 아래쪽으로 관련성이 높은 것들을 나타낼 수 있다. 이것은 공정간의 데이터 traceability가 보장된다는 가정 하에 수행 가능하다.

#### 4. 결론

제조업체의 생산공정 데이터는 지속적으로 기업 데이터베이스에 쌓이지만 제대로 활용하지 못하고 있다. 데이터 마이닝 기법은 이러한 데이터로부터 의사결정의 위한 지식을 제공할 수 있다. 하지만 다양한 기법을 적용하는 방법이 어려워 현장에서는 쉽게 활용하지 못한다. 데이터 분석 시나리오는 다양한 기법의 데이터 마이닝 방법을 활용하기 쉽게 템플릿화 한 것으로 현장에서 활용하기 쉽게 해준다. 현장에서는 수율문제, 경향분석, 기계의 조건의 변화에 따른 수율 예측 등 다양한 문제가 발생할 수 있고, 본 논문에서는 현장에서 발생할 수 있는 이러한 다양한 문제에 대하여 6가지의 활용시나리오를 제시하였다.

## References

- 1) 이형욱, 이근안, 최석우, 배성민, 박홍균, “생산공정데이터를 활용한 생산디지털마이스터 프레임워크 개발”, 한국정밀공학회 춘계학술대회 논문집, 2007.
- 2) 박홍균, 이근안, 최석우, 이형욱, 배성민, “데이터 마이닝을 위한 생산공정 데이터 추출”, 한국정밀공학회 춘계학술대회, 2005.