

이산 제조 공정에서의 수율 향상을 위한 분석 프레임워크의 개발에 관한 연구

송치욱* · 노금중** · 박동진***

〈 목 차 〉	
I. 서론	III. 연구 대상 및 방법
II. 관련 연구	3.1 자료 수집 및 모델 수립 방법
2.1 시뮬레이션 기반의 사이버물리제조시스템	3.2 데이터 전처리
2.2 기존 체계의 문제점	3.3 하이브리드 분석 프레임워크
2.3 연구의 방법	IV. 실험 결과
2.4 로지스틱 회귀 분석	V. 결 론
2.5 연관 규칙 탐색	참고문헌
	<Abstract>

I. 서론

현대의 제조 기업은 빠른 제품 개발 기간, 소비자의 제품 품질에 높은 요구, 시장의 다변화와 같은 문제에 직면하고 있다. 이에 기업에서는 생산성 향상, 제품주기 단축, 수율향상 등의 경쟁력 확보를 위한 역량을 집중하고 있다. 이에 대해서 제조 기업에서는 공정변동, 공정재고, 생산주기, 장비효율성, 수율 등의 공정 및 품질에 대한 다양한 지표를 관리하고 있다. 이런 제조업 생산성 향상 지표 중 수율은 특정기

간 동안 공정에서 총 생산된 제품에서 제작업이나 폐기가 없이 양품으로 생산된 비율을 의미한다. 수율은 공정의 품질을 측정하는 주요한 측정도구로 사용된다. 제품개발 단계에서 시제품이 인도되고 양산라인에 투입된 후 양산수율을 높이기 위해서 관계된 부서가 협업을 하게 되는데, 수율을 부서, 작업자, 엔지니어 간의 표준화된 지표로 사용 한다. 따라서 수율은 제조 기업 경쟁력의 가장 중요한 요소이며, 일정 수준 이상의 수율을 확보하지 못 한다는 것은 시장에서의 실패를 가져 올 수 있는 주요한 요인 이므로 제조 기업에서는 양품 수율 향상을 위

* SK 주식회사 C&C SF사업 1팀, chiwook.song@gmail.com(주저자)
** 삼성디스플레이 OLED 제조혁신팀, rulerrain@gmail.com
*** 공주대학교, mispdj@kongju.ac.kr(교신저자)

해서 다양한 도전을 하고 있다.

수율관리를 포함하는 공정품질 및 제품품질 분야도 최근 이슈가 되고 있는 Industry4.0 및 스마트팩토리 구현의 주요 대상 분야이다. 특히 기존 기술에 의한 정보화를 넘어서 인공지능, 빅데이터, 그리고 데이터분석(Data Analytics) 등의 첨단 분석 및 ICT 기술이 응용됨에 따라 획기적으로 품질 향상의 목표를 달성하여 제조 경쟁력을 확보할 수 있게 한다. 최근 제조기업에서 이를 구현하는 기술로는 사이버물리시스템(CPS: Cyber-Physical Systems)으로 대두되고 있다. 이런 관점에서 수율관리 분야의 CPS 적용은 시스템 통합을 위한 ICT의 혁신적인 접근으로 볼 수 있다. 기존의 통계적 기법에 의한 접근(Duane et al., 1999)과 데이터 마이닝 기법에 의한 접근(Susana et al., 2011)은 데이터의 수집 후에 오프라인 상에서 하나의 독립적인 업무로 진행되고 업무에 적용되는 형식이었다. 본 연구에서는 ICT 기술을 적용하여 데이터의 수집, 저장, 분석 및 결과의 적용 및 피드백에 이르기까지 온라인상에서 체계적으로 진행되는 프레임워크를 제시하고, 프로토타입을 개발하여 적용한 사례를 보여준다. 본 프레임워크는 반도체, 디스플레이, 자동차, 배터리 등과 같은 제조형태인 이산공정에 적합하도록 설계되었다

본 연구에서 개발한 HLRAM(Hybrid Logistic Regression and Association Rule Mining) 프레임워크는 첫째, 제품 품질에 영향을 주는 불량률 조기에 찾을 수 있게 한다. 특히 자동화된 시스템 환경은 품질 분석가, 제조공정 엔지니어와 작업자들이 기존의 복잡하고 분석에 오랜 시간이 걸리는 문제를 극복 할 수 가

있게 한다. 둘째, CIM 및 MES에서 생산되는 대용량의 제조 빅데이터를 특정 분석을 위한 데이터셋으로 재구성하는데 있어서 사용자 친화적으로 설계하였다, 마지막으로 통계기법인 로지스틱 회귀분석과 데이터마이닝기법인 연관규칙을 통합하여 현장의 분석 비전문가도 쉽게 분석하고 결과를 적용하도록 하였다. 본 연구는 제1장 서론과 아울러, 제2장은 연구의 배경 및 분석기법에 대한 관련연구를 소개하고, 제3장에서는 프로토타입의 개발내용을 설명하고, 제4장에서는 프로토타입을 통한 실험의 결과를 분석하고, 마지막으로 5장에서는 결론을 맺는다.

II. 관련 연구

2.1 시뮬레이션 기반의 사이버물리제조시스템

최근 선진 제조기업에서는 사이버 물리 시스템을 적용하여 차세대의 제조형태로의 발전을 도모하고 있다. 특별히 제조업에 적용하는 CPS를 사이버물리제조시스템(CPPS: Cyber-Physical Production Systems)이라 부르며(L. Monostori et al., 2016), 아직은 초기단계이지만 스마트 팩토리 및 Industry 4.0의 구체적인 실현의 방안으로 많은 관심을 받고 있다. CPPS는 2개의 주요 기능이 있는데, 첫째는 물리적 세계로부터 실시간 데이터를 획득하고 사이버 상으로 부터 정보를 피드백하는 높은 수준의 연결기능(advanced connectivity)이다. 둘째는 사이버 상에서 이루어지는 지능적인 데이터 관

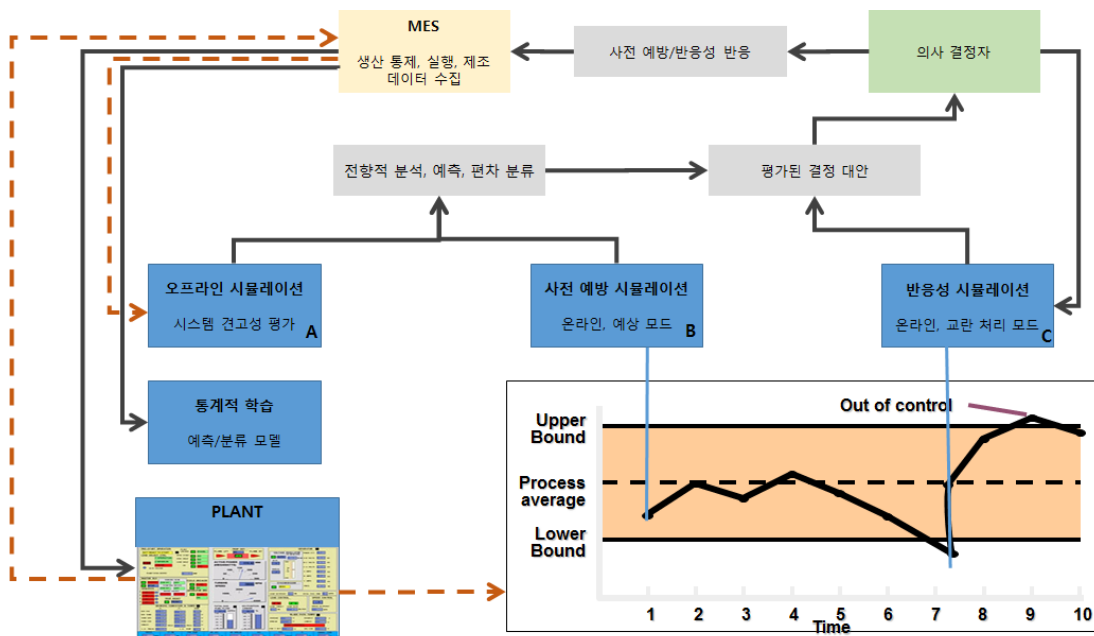
리, 분석, 그리고 계산능력 등이다(J. Lee et al., 2015).

최근 수요의 변화에 따라 탄력적인 제조 공정을 운영하는 것이 더욱 중요해지며, 또한 유연 생산체제로부터 피드백을 받아 공정에 다시 반영 하는 것이 필수적이다. 이를 위해서는 CPPS에서 강조하는 디지털과 물리적 공장의 체계적인 결합이 필요하다. <그림 1>은 이산제조 환경에서 이산사건시물레이션(DES: Discrete Event Simulation)을 기반으로 공정을 제어하는 CPPS의 구현 예이다(A. Pfeiffer et al.,) <그림 1>의 시물레이션 A는 오프라인 시물레이션으로 제조 실행 전에 불확실성을 대비하여 생산일정에 대하여 민감도분석을 실시하는 것이며, 시물레이션 B는 전향적(proactive) 시물레이션으로 계획된 일정에 따라 실시할 단기 행동(short-term actions)에 대한 결과 상황을 예

견하기 위하여 시물레이션을 실시하는 것이다. 시물레이션 C는 사후 시물레이션으로 이미 실행된 제조 결과에 대한 분석 시물레이션으로 문제의 원인을 찾아 발생 가능한 오류와 불량을 최소화하기 위한 노력이다. 이상과 같은 체제는 MES로부터 제조정보를 온라인 혹은 오프라인으로 박아서 시물레이션을 통하여 의사결정을 지원받거나 자동으로 조치하는 CPPS의 구조이다. 본 연구를 통하여 제안하는 프레임워크는 이상과 같은 기능을 포함한다.

2.2 기존 체계의 문제점

이산공정에서 수율향상이 어려운 이유는 첫째, 공정이 미세화, 소형화됨에 따라 관리해야 할 공정변수가 기하급수적으로 증가해서 수율산포에 대한 원인을 신속, 정확하게 파악하기



<그림 1> 이산제조 환경의 이산사건 시물레이션 기반 CPPS구조

어려움에 있다. 둘째, 수율이 낮은 저수율이 발생되고 있는 공정이 자(自)공정에서만 기인해서 발생되지 않고 다수의 공정에서의 공정변수 간에 상호 영향을 받으면서 발생하고 있다. 이와 같이 개별 공정에서 발생되고 있는 공정변수를 통해서 수율을 관리하게 되면 이후 공정에서 공정변수의 산포가 누적되어서 복잡한 수율관리가 필요하게 되며, 전체 공정의 공정 변수들 간의 선형, 비선형의 복잡한 관계를 파악하기가 어렵게 된다. 이와 같은 문제를 해결하기 위해서 공정에서 발생한 데이터를 수집한 후 통계적 기법과 데이터 마이닝 기법으로 분석하고 해결책을 제시하였다(Duane S. Boning et al., 1999; Susana Ferreiro et al., 2011). 이중 통계적 기법으로는 다양한 계측 데이터를 이용하여 x 관리도, \bar{x} -R 관리도 등을 통한 생산 공정의 품질을 관리하거나 장비 이상발생 이력을 관리하고 있다. 상관분석, 회귀분석, 유의차 분석 등의 기법이 사용되고 있으나, 이와 같은 분석 기법으로는 수율저하의 원인을 파악하기가 어렵다. 전술한 바와 같이 통계적 기법은 수많은 공정 변수의 복잡한 관계를 파악하기가 어려울 뿐만 아니라 현재와 같이 대용량의 데이터가 발생되고 있는 전체 공정을 고려한 종합적 수율 개선을 기대하기가 어렵다.

데이터마이닝 기법은 대용량의 데이터로부터 이들 내에 존재하는 관계, 패턴, 규칙 등을 탐색하고 모형화 할 수 있으며, 수율향상을 위해서 공정변수 내에 숨겨져 있는 지식을 발견할 수 있는 유용한 방법이다(정혜진 외, 2007). 본 연구에서는 데이터에 대한 전처리후에 분류를 위해서 로지스틱 회귀분석을 사용하고, 이에 대한 저수율의 규칙성을 찾기 위해서 연관관계

탐색을 사용 하였다.

생산 중 수집되고 있는 데이터를 효과적으로 사용하지 못하게 되면 공정 제어의 불안정 및 균일하지 못한 수율 분포를 가져온다(Yu, Peng et al., 2012). 일반적인 제조라인에서 발생하는 대부분의 데이터는 인라인(in-line) 과 라인종단(end of line)에서 수집된다. 인라인 데이터는 생산 중인 반제품과 자재가 가공이 될 때 생산 장비, 계측기를 통해서 PLC, HMI, 그리고 PC 등에서 데이터가 집계된다. 이렇게 집계된 데이터는 최종적으로 제조실행시스템(MES: Manufacturing Execution Systems)의 데이터베이스, 파일의 형태로 저장이 된다. 이러한 생산 데이터의 종류로는 강도, 두께, 입자 개수, 면적, 저항률 등과 같은 반제품의 물리적 화학적인 특성을 가진 값들도 포함한다. 또한 일부 시계열성인 정보도 장비의 중단에서 기록이 되고 있다. 이외에도 라인에서의 장비 환경 데이터, 공기구 정보, 공장 내의 ESH(Environment Safety Health) 정보들이 공장의 안정적인 모니터링과 통제에 이용되고 있다.

집계된 정보에 대해서 통계적으로 정확한 해석이 이루어지지 않는다면 수많은 데이터가 집계되어도 활용할 수가 없다. SPC(Statistical Process Control)와 같은 공정통계 기법은 공정을 안정적으로 제어하는데 기여를 하고 있지만, 수율에 영향을 주고 있는 인자를 찾는 과정은 공정정보, 계측정보의 비선형성, 복잡성의 성질을 지니고 있는 문제로 특정한 분석기법을 사용해야 한다. 기존에 수율에 영향을 주는 인자를 찾는 방법에 대해서는 여러 연구가 진행이 되었지만 통계적인 해석의 어려움과 분석 툴의 사용을 위한 과도한 양의 데이터 집계로 인한

업무 가중의 부담은 엔지니어나 분석가들에게 분석에 많은 어려움을 주었다(Wei-Chou Chen et al., 2005).

2.3 연구의 방법

본 연구에서는 일반적인 이산제조공정에서 사용되고 있는 설비, 장비에서 발생하는 데이터를 집계하는 데이터 수집서버에서, 생산 프로세스 정보, 장비 알람정보, 장비 측정정보로의 데이터 마트 구성을 위해서 데이터웨어하우징을 실시한다. 분석가가 지정한 수율이 좋음, 나쁨을 이진형으로 분류하기 위해서 로지스틱 회귀 분석을 실시한다. 로지스틱 회귀 분석은 금융, 의료, 제조 등의 분야에서 단변량 보다는 다변량 분석을 요하는 분석에서 많이 사용되었다. 물론 데이터 마이닝을 이용해서 고도화된 분류 작업을 진행 할 수 있으나, 학습데이터의 양이 확보하기가 어려운 상황이거나 본연구와 같이 제조현장의 복잡도가 높지 않은 실험에서는 의사결정나무나 인공신경망 같은 분류 방법은 지양 하였다. 수율을 높이기 위해서 주성분 분석을 이용한 다변량 분석 기법을 반도체 계측에 적용한 연구가 있었다(G. A. Cherry et al., 2012). (Brian E. Goodlin et al., 2003)은 여러 개의 관리도 중에서 불량률의 유형별로 이상원인을 식별 할 수 있는 통계적인 관리도 기법을 제안하였다. (Ming-Da Ma et al., 2010)은 반도체 공정 중 하나인 CVD(Chemical Vapor Deposition)공정에 품질에 영향을 미치는 주요 파라미터를 관리하고 수율을 높이는 방법을 Wilcoxon rank sum test, clustering regression 을 사용하여서 연구를 하였다. 이와 같이 기존

의 연구에서는 수율에 영향을 주는 파라미터에 대해서 분석을 하고 관리를 할 수 있도록 하는데 중점을 두고 연구를 하였지만 본 연구에서는 로지스틱 회귀를 통해서 분류된 데이터 그룹에서 저수율 규칙을 찾기 위해서 연관 규칙 탐색 기법을 사용 한다. 연관관계 규칙을 사용하게 됨으로서 단위공정 간에 흩어져서 저수율의 규칙을 찾지 못 하였던 숨겨진 장비 파라미터간의 규칙성을 찾아 낼 수 있다. 그리고 제품 투입부터 완료 까지 공정 간의 저수율의 규칙을 회피하도록 장비 파라미터의 조정을 할 수 있다.

2.4 로지스틱 회귀 분석

생산관리 관점에서의 수율은 전체 생산량 대비 양품의 비율로서 수치형으로 표현이 되며, 생산관리자는 현재 생산 수율을 기준치 이상, 이하로 구분하여 이하일 경우 생산 현장에 여러 가지 조치를 취하게 된다. 이를 위해서 수율에 대해서 기준치 이상, 이하의 범주를 통해 분류를 해야 되며, 로지스틱회귀모델은 이처럼 범주형 종속변수 대비 각 범주의 특성을 파악하기 위한 해석이 용이하며, 결과를 이용한 사후 확률 판별이 용이하다. 출력변수가 이항 범주이므로 아래와 같이 다항 로지스틱 회귀 모형은 아래와 같다.

$$P(Y=1|x) = p(B_0 + B_1 x_1 + \dots + B_p x_p)$$

여기에서 $p(x) = \frac{\exp(x)}{1+\exp(x)}$ 이다. $p/(1-p)$ 에서 성공할 확률과 실패할 확률의 비를 승산비(odds ratio)라고 하며, 로지스틱 회귀

모형에 대한 종속, 독립변수간의 관계를 설명할 수 있다.

로지스틱 회귀는 주어진 x 값에 대한 출력변수 y 가 1이 될 확률 $p(y=1|x)$ 을 추정하는데, 0과 1사이의 적당한 수를 절단값 c 로 하여서 $y=1$, $y=0$ 인 클래스를 분류를 할 수가 있다. 분류경계가 선형일 경우에는 로지스틱 분류를 사용할 수 있다. 분류경계가 선형일 경우에는 로지스틱 분류를 사용할 수 있고 비선형일 경우에는 의사결정 나무나 신경망 등의 비선형 분류 경계를 찾는 방법을 고려해야 한다. 로지스틱 회귀분석은 국내 자동차 부품기업의 재무 효율성 분석과 재무전략을 수립하기 위해서 기업의 재무비율 항목 중 비효율적인 부분을 분류하기 위해서 판단의 지표로도 사용이 되었다.(신정훈 외, 2016) 비효율적인 항목 P 의 확률이 일정 수준(0.5)을 넘으면 비효율적인 재무 비율 항목으로 간주 하는 모델을 사용 하였다.(Wei-Chou Chen et al, 2005)

2.5 연관 규칙 탐색

연관 규칙 탐사는 데이터 마이닝 안에 존재하는 항목들 간의 빈발성 규칙(frequent pattern)을 발견하는 과정을 말하며, 비지도 학습법의 일종이다(Rakesh Agrawal et al., 1993). 연관규칙은 유용한 규칙, 자명한 규칙, 설명이 불가능한 규칙으로 구분 할 수 있다. 유용한 규칙에 대한 필요조건인 지지도(support), 신뢰도(confidence), 향상도(lift)에 대해서 만족을 하는 if-then 거래 집합(transaction set)을 찾는 것이 연관 규칙 분석이다(Bernard Kamsu-Foguem et al., 2013). 거래 집합에서 분석가가

특정한 최소지지도를 만족하는 빈발집합을 찾는 과정은 시간이 많이 걸리는 작업이다. 이와 같은 작업에 대한 시간 및 공간을 줄이는 방법이 많이 제안되고 있으며, 단계별 후보군을 통한 빈발한 데이터 집합을 찾는 Apriori 알고리즘이 많이 사용되고 있다.

연관 관계 분석을 이용한 조선 생산 실행 시스템 구축 연구 사례는 통합된 제조 실행 시스템의 구성을 위해서 기존 레가시 시스템인 생산계획, Shipyard 관리 시스템, 인력관리 시스템, 자재관리 시스템에서의 프로그램에 대한 사용빈도, 모듈 별 연관성을 이용해서 연관 관계 규칙의 지지도, 신뢰도가 높은 모듈을 실제 통합 생산실행시스템에서 사용해서 응용을 하였다(이필립 외, 2013).

POS 데이터를 연관 관계 분석 알고리즘을 이용해서 기존의 상품 추천 서비스 이외에 상품 판매의 상승/하향 규칙을 추가 하여서 경영자에게 유용한 판매 분석 정보를 제공해주었다(안경찬 외, 2012).

지지도, $s(X \rightarrow Y) =$

$$\frac{\sigma(X \cup Y)}{N} \left(\frac{X \text{와 } Y \text{가 동시에 포함된 거래수}}{\text{전체 거래수}} \right)$$

신뢰도, $c(X \rightarrow Y) =$

$$\frac{\sigma(X \cup Y)}{\sigma(X)} \left(\frac{X \text{와 } Y \text{가 동시에 포함된 거래수}}{X \text{를 포함하는 거래수}} \right)$$

Ⅲ. 프로토타입의 개발

3.1 자료 수집 및 모델 수립 방법

본 연구에서는 제조업체 중에서 이산 제조

공정(조립공정)을 가지고 있는 제조업체 인라인 공정에서 발생하는 지정된 장비 그룹에서의 공정 데이터, 알람 데이터 그리고 장비에서 측정된 사이클 타임, 피치(pitch time)을 분석에 사용 하였다.

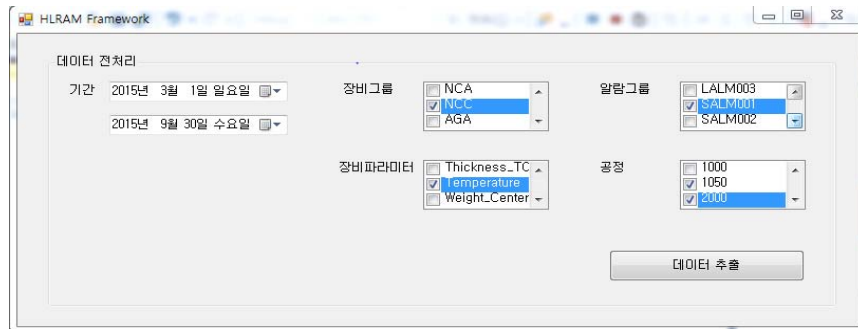
데이터 분석을 하고 모델을 수립하기 위한 절차는 첫째, 최종적인 연관규칙을 찾기 위해서는 공정 중 발생하는 데이터에 대한 다차원의 메타 데이터 적재가 필요하다. 분석용 데이터 마트 구성을 위해서 공정 데이터, 알람, 측정(사이클 및 피치 타임) 그리고 수율에 대해서 데이터를 적재한다. 둘째, 데이터를 분석하기 위해서 기간, 공정, 장비그룹, 알람그룹, 장비 파라미터를 선택해서 대용량 데이터 처리에 사용하는 데이터웨어하우스 작업을 진행한다. 셋째, 로지스틱 회귀 분류 모델을 수립하기 위해서 통계 패키지 프로그램인 R을 분석엔진으로 사용 하였다. 데이터웨어하우스 작업이 완료된 데이터마트를 R프로그램에서 해석 할 수 있도록 JAVA 언어에서 전처리를 하였고, R에서 제공하는 API를 통해서 전처리가 완료된 데이터를 입력값으로 처리 할 수 있도록 프로그램을 작성 하였다. 입력값을 입력하기 전에 데이터 포맷은 유효한 p-value를 찾기 위해서 수율을 종속변수로 하여서 각 독립변수 중 유효한 값을 찾을 수 있도록 구성 하였다. 분류 모델은 수율 목표치를 합격하는 데이터 군과 그렇지 못한 군으로 분류를 할 수 있도록 한다. 연구에서 목표로 하고 있는 수율 목표 불합격 데이터 집합은 $P(Y=1|x)$ 가 절단값보다 낮은 집합군으로 분류가 되게 R프로그램에서 설정을 하였다. 넷째 로지스틱 회귀 분류모델을 통해서 분류된 저수율 클래스 데이터 집합 중 수율이 목표치

보다 낮게 나올 수 있는 규칙을 찾기 위해서 연관규칙 탐사 기법을 사용한다. 연관규칙 탐사는 분류된 데이터 집합에서 규칙성을 갖는 데이터 집합을 찾기 위해서 신뢰도, 지지도, 향상도의 세가지 척도를 통해서 해당하는 규칙을 추출한다.

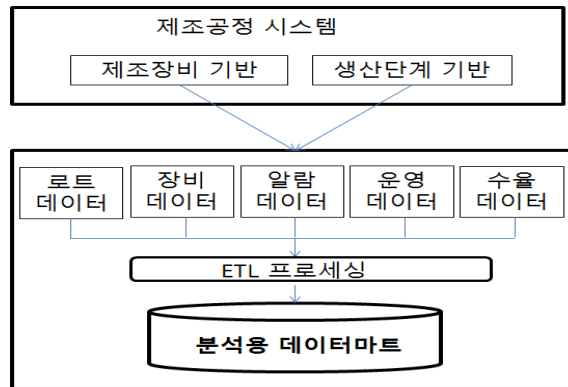
3.2 데이터 전처리

분석을 위해서 6개월 동안 생산이 완료된 2차 전지 제품의 생산 이력정보를 사용 한다. 각 생산 오더에는 로트 기준의 생산지시 정보가 담겨져 있으며, 생산지시 정보에는 N개의 로트정보가 있다. 생산이 진행되었던 로트는 생산 설비, 장비에서 진행이 되었던 각종 공정, 알람, 측정 정보가 기록되어 있으며, 생산완료 시점의 수율 정보가 담겨 있다. 데이터 전처리는 공정, 알람, 측정, 수율 분류로 데이터를 집계한다. <그림 2>와 같은 데이터 전처리 세팅화면에서 기간, 장비그룹, 알람그룹, 장비 파라미터, 공정으로 데이터를 수집 할 수 있는 기준을 마련하였다. 장비그룹을 선택하게 되면 제품을 생산하기 위해 필요한 자재, 반제품을 가공 할 때 발생하는 선택된 장비의 상태정보가 수집이 되며, 알람그룹을 선택하면 제품의 생산 시 장비, 설비에서 발생되고 있는 각종 경고, 이상 표시 등이 수집된다. 장비 파라미터를 선택하면 생산, 측정 장비에서 엔지니어가 분석 하고자 하는 각종 정보에 대한 설정에 대한 처리 값을 수집한다.

이렇게 각종 선택을 완료 하면 데이터 추출을 진행하여서 미리 구성된 데이터마트로 ETL(Extracting Transformation Loading) 프로



<그림 2> 데이터 전처리 설정 화면



<그림 3> 데이터 처리 흐름

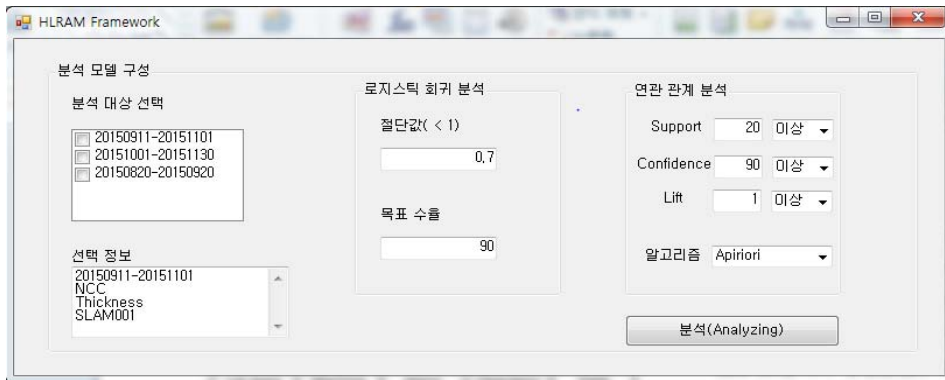
세스를 진행이 되도록 구성을 하였다. 로트 별 수율은 생산오더가 완료되어야 전체 생산 량 대비 양품으로 계산이 가능하도록 제한을 하여야 하며, 사이클 타임은 장비에서의 생산 시작, 종료의 평균값을 가져 올 수 있도록 로직을 구성 하였다. ETL이 완료된 데이터 집합은 추출이 편리 하도록 관계형 데이터베이스에 저장을 하도록 한다.

데이터의 흐름은 <그림 3>에서 확인 할 수 있다. 제조 현장(Shop floor)에서 발생 되는 데이터를 제조장비 기반과 생산단계 기반으로 구분하였다. 데이터 프로세스 단계에서는 데이터 베이스에서의 데이터를 ETL 프로세싱을 거쳐서 로트기준정보 마트, 장비 마트, 알람마트, 공

정정보 마트로 구분이 되게 적재가 되도록 한다.

3.3 하이브리드 분석 프레임워크

로지스틱 회귀분석과 연관관계 분석을 위한 통합화면은 <그림 4>와 같다. 데이터 전처리에서 분석을 위해 선택된 정보는 기간을 키값으로 하단에 선택된 장비파라미터, 알람 그룹 정보를 알 수 가 있다. 로지스틱 분석을 위해서 종속 변수인 수율의 목표 값을 입력한다. 목표 수율보다 높은 수율은 분석에서 제외되며, 낮게 분포되어 있는 수율정보를 가지는 데이터가 분류가 되게 된다. 목표로 하는 수율에 영향을 주



<그림 4> 데이터 분석 설정 화면

는 인자를 찾기 위한 로지스틱 회귀 모델에서 $x^2 - test$ 결과 p-value가 유의한 독립변수를 찾는다. 공정 엔지니어는 이번 단계에서 장비, 알람, 공정 분류 중에서 종속변수인 수율에 대한 오즈비(Odd ratio)를 통한 해석을 할 수 있다.

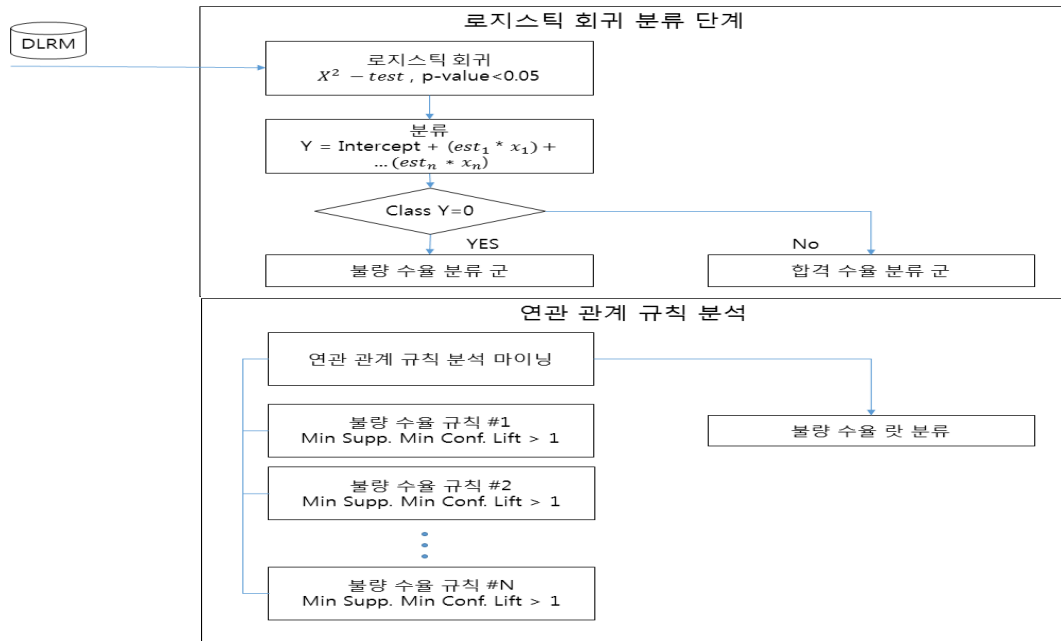
로지스틱 회귀모델을 이용한 분류를 하기 위해서 절단값 중 적당한 값(1 이하)을 입력하게 되면 산출된 유의한 p-value를 구하는 회귀식을 통해서 수율이 좋지 않은 데이터 셋을 분류 할 수 있게 된다. 이렇게 분류가 된 데이터 집합은 별도의 분석을 위한 데이터마트로 구성 된다.

이어서 연관 관계 분석을 위해서는 <그림 4>에서 사용자는 지지도, 신뢰도, 향상도를 입력한다. 지지도와 신뢰도를 다르게 입력함으로써 유용하지 않은 규칙을 제거 할 수 있으며, 신뢰도가 높다고 하더라도 해당 규칙에 대한 거래가 일어날 가능성이 낮으면 해당 연관규칙에 대한 실제 의미 부여가 힘들 수도 유용한 규칙이 아닐 수도 있기 때문에 지지도, 신뢰도, 향상도의 입력값을 다르게 해서 최적의 해를 찾을 수 있다. 알고리즘으로는 Apriori, FP-Growth 두 가지를 제시하고 있는데 Apriori는

Breadth-first 탐색 기법으로 반복되는 패턴에 대한 쌍에 대해서 탐색을 진행하고 FP-Growth는 Depth-first 탐색 기법으로 반복되는 패턴에 대해서 수직으로 반복적인 연관 집합을 찾게 한다.

로지스틱 회귀 모델에서 Y=0 인 클래스로 분류된 트랜잭션 집합 중에서 X→Y를 만족하는 트랜잭션 집합을 지지도, 신뢰도를 통해서 유의한 규칙을 판단하고 2가지 척도를 통해서 분류된 데이터 셋 중에서 향상도를 통해서 우수한 상관관계가 있는 규칙을 추출한다.

<그림 5>는 <그림 4>화면과 연결되는 내부 분석 프로세스를 보여준다. 즉 데이터 마트가 준비되면 <그림 4>에서 입력한 데이터를 이용하여 로지스틱 회귀분석과 연관관계 규칙 마이닝 모델 분석을 하게 한다. 데이터 마트가 구성이 되면 로지스틱 회귀 분류에 따라서 불량 수율 분류군(Class Y=0)으로 분류된 정보는 연관 관계 규칙 분석을 위한 데이터 셋으로 분류되게 된다. 연관관계 규칙 분석 마이닝을 통해서 해당 데이터 집합의 불량 수율을 보여 주는 규칙을 신뢰도, 지지도, 향상도를 통해서 찾게 되면 이렇게 파악된 패턴을 최종적인 불량 수율



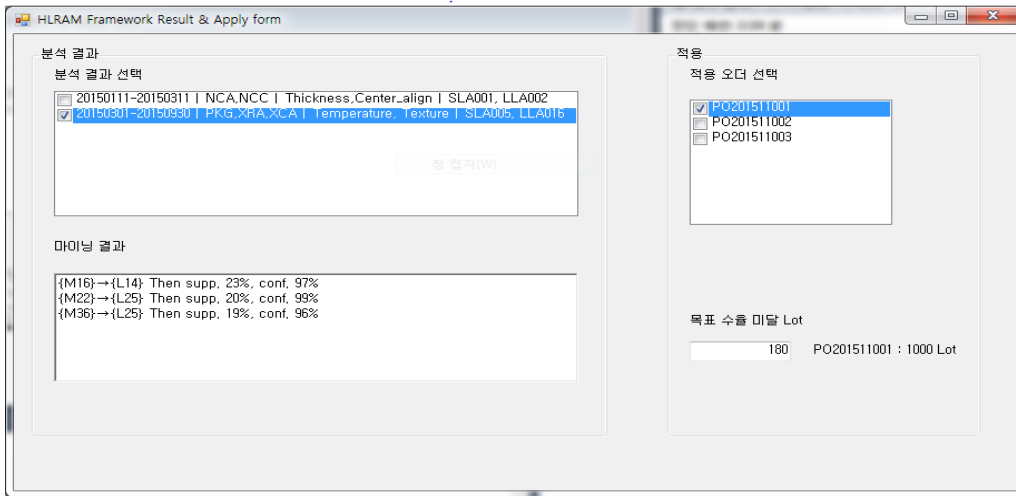
<그림 5> 로지스틱 회귀와 데이터 마이닝 모델

을 판단 할 수 있는 결과를 사용 되게 된다.

<그림 4> 에서의 분석 예를 보게 되면 분석 대상 선택에서 분석 대상 기간을 선택 하고, 선택정보에서 로지스틱 회귀분석에서 사용 할 장비그룹, 장비 파라미터, 알람그룹, 공정에 대한 전처리된 데이터를 선택하게 된다. 목표 수율을 90%로 설정하고 절단값을 0.7로 주어 졌을 때 목표 수율보다 낮은 데이터 집합이 필터링 되고 이에 대해서 설정된 지지도, 신뢰도, 향상도에 따라서 그중 낮은 수율을 보이는 패턴을 가지는 규칙을 알 수 있다.

<그림 5>는 HLAM 프레임워크를 통해서 찾은 목표에 도달하지 못하는 공정정보와 장비상태에 대한 패턴을 추출함으로써 현재 생산 진행 중인 생산 오더의 로트 중에서 몇 개의 로트가 목표 수율 이하의 로트가 될지 예상을 할 수 있다. 분석결과 화면에서 공정엔지니어, 생산관

리는 생산기간, 생산 구간, 공정 파라미터 등을 과거의 데이터를 근간으로 선택을 하게 된다. 여기에서의 제한 사항은 유사한 생산제품, 장비를 선택 했을 때에 정확한 마이닝 결과를 얻을 수 있다. 적용 오더 선택 화면은 분석 결과를 선택 하였을 때 현재 생산 중인 생산 오더 중에서 분석 결과와 유사한 생산 제품, 장비 구성 환경을 가지고 있는 정보를 보여 준다. 이렇게 선택된 생산 오더는 목표 수율 이하의 연관 관계 규칙을 가지고 있는 패턴이 적용이 되어서 현재 공정, 장비 정보 중에서 사용자가 추출하기 원하는 목표 수율 이하의 로트 개수를 전체 생산 로트에 대비 하여서 보여준다. 이 결과를 통해서 생산관리자, 엔지니어는 현재 생산 중인 재공에 대한 장비, 설비 설정이 정상적으로 되어 있는지를 판단 할 수 가 있으며, 최종 완성품에 대한 수율을 예측을 할 수 가 있다.



<그림 6> 분석 결과와 적용 화면

<그림 6>의 마이닝 결과 창을 보게 되면 생산 중 저수율이 발생 할 가능성이 높은 장비, 설비 파라미터를 알 수 가 있다. 생산관리자는 아래의 결과를 통해서 장비엔지니어에게 장비의 파라미터 값 조정을 요청하게 되며, 장비엔지니어는 해당하는 파라미터 셋팅을 변경 한다. 이후 생산관리자는 마이닝 결과에 대해서 저수율 로트가 발생하는지 모니터링 활동을 한다.

IV. 실험 결과

실험을 위해서 이산 제조 공정을 가지고 있는 2차 전지 제조 공장의 1년 중 발생된 생산 정보를 MES 데이터베이스에서 수집을 하였다. 분석용 데이터는 03월부터 09월까지의 6개월분의 생산 데이터를 기초로 ETL 프로세스를 진행 하여서 분석에 필요한 형태의 데이터마트 데이터베이스를 구성 하였다. 로지스틱 회귀분석을 위해서 범용적으로 많이 사용되고 있는

통계 패키지인 R을 엔진으로 사용자 UI에서 분석을 위해서 입력을 받았고 분석에 필요한 회귀식, 연관관계 탐사를 HLAM 프레임워크 기반으로 진행을 하였다.

로지스틱 회귀 결과에서의 M_1, \dots, M_{12} 는 장비에서 수집된 온도, 습도, 굴곡도, 편평도, 공극률 등의 가공, 측정 변수를 의미하며 L_1, \dots, L_5 와 같은 알람 정보는 서보(Servo) 회전 이상, 자재 이송 불량, 진공센서 이상, HMI통신 이상 등의 장비에서의 이상정보를 표시한다. P_1, \dots, P_2 는 각 장비에서의 가동률이 최고로 지속되고 있는 시간인 피치(Pitch) 시간을 의미한다.

<표 1> 에서 x^2 -test 결과 p-value 가 장비변수 M_1, M_2, M_{12} , 알람변수 L_3, L_4 가 유의한 결과를 얻을 수 있었다. 회귀결과를 해석하기 위해서 95% 신뢰구간내의 오즈비를 구해보면 <표 2> 에서와 같이 장비에서 수집된 M_1, M_2, M_{12} 변수가 각 종속변수인 수율을 각 $4.496172, 2.670069e^{17}, 1.883956e^{-3}$ 배 높임

<표 1> 로지스틱 회귀 결과

입력 변수	회귀 계수	표준 오차	X^2	유의 확률
(Intercept)	-32.187774	12.389786	-2.598	0.00938 **
M_1	1.503226	0.538276	2.793	0.00523 **
M_3	-0.026305	0.065555	-0.401	0.68823
M_6	40.126051	14.501754	2.767	0.00566 **
M_{11}	2.886185	1.948931	1.481	0.13863
M_{12}	-6.274381	2.274672	-2.758	0.00581 **
P_1	0.009109	0.012813	0.711	0.47711
P_2	0.023933	0.027412	0.873	0.38262
L_1	0.169446	0.401911	0.422	0.67332
L_2	0.045487	0.402134	0.113	0.90994
L_3	-0.555068	0.246873	-2.248	0.02455 *
L_4	-0.928060	0.358333	-2.590	0.00960 **
L_5	-0.452689	0.400645	-1.130	0.25852

을 알 수 있으며, 알람 빈도가 가장 높은 순으로 수집한 $L_1 \sim L_6$ 중에서 L_3, L_4 가 $5.740334 e^{-1}$, $3.953199 e^{-1}$ 배 증가, 즉 1.7421, 2.5296 배 감소되는 것으로 분석 되었다. 공정에 관련된 변수와 알람정보는 수율에 영향을 주는 것으로 알 수 있으나 Cycle Time, Pitch Time은 수율에 연관성이 없는 것으로 판별되었다.

<표 1>에 있는 회귀결과를 기반으로 분류모델을 만들기 위해서 아래 식을 풀은 후에 $P(Y = 1|x)$ 가 절단값 보다 작은 클래스를 수율이 목표치에 미달 할 확률이 높은 클래스로 분류를 한다.

$$\ln \frac{p}{1-p} = -32.187774 + 1.503226 X_1 + 40.126051 X_2 - 6.274381 X_3 - 0.555068 X_4 - 0.928060 X_5$$

추정된 회귀식을 이용해서 $Y = 0$ 인 클래스를 분류해서 382개의 결과를 얻을 수 있었다. 이렇게 수집된 정보는 연관규칙 분석을 위해서

<표 2> 회귀 결과 Odds 비

입력 변수	Odds ratio
M_1	4.496172
M_6	2.670069,
M_{12}	1.883956
L_3	5.740334,
L_4	3.953199

데이터 마트를 구성 하였고 빈발품목 집합(frequent itemset)을 찾기 위해서 Apriori 알고리즘을 사용 하였다. 분석대상의 Data set 집합을 분석하기 위해서 범주형 Data로 변환작업을 진행하였는데, 사이클, 피치 타임은 수율에 연관성이 적은 것으로 판단되어서 제외 하였고, 연관관계탐색 대상을 장비변수 3개, 알람변수 2개를 선택해서 범주형 변수로 변환을 하였다. 최소지지도(Minimum support)를 0.05, 최소 신뢰도(Minimum confidence)를 0.8로 지정 했을 때 발견된 규칙은 222개가 발견 되었으며 지지도 기준으로 빈발하는 집합으로 20개의 규칙을 <표 3>에서 발견 할 수 있다.

<표 3> 연관관계 규칙 분석 결과

Rule	Support	Confidence	Lift
{L14}→{L25}	0.743	0.900	0.993
{L25}→{L14}	0.743	0.820	0.993
{M35}→{L25}	0.327	0.937	1.034
{M35,L14}→{L25}	0.287	0.937	1.034
{M35,L25}→{L14}	0.287	0.881	1.067
{M16}→{L14}	0.232	0.967	1.169
{M16}→{L25}	0.229	0.954	1.053
{M31}→{L25}	0.227	0.921	1.017
{M16,L14}→{L25}	0.221	0.952	1.051
{M16,L25}→{L14}	0.221	0.964	1.167
{M23}→{L14}	0.212	0.951	1.151
{M23}→{L25}	0.213	0.951	1.049
{M31}→{L14}	0.202	0.948	1.046
{M23,L25}→{L14}	0.202	0.948	1.149
{M22}→{L25}	0.196	0.986	1.088
{M36}→{L25}	0.193	0.959	1.058
{M36}→{L14}	0.188	0.931	1.127

연관관계 규칙 분석 결과 특이한 사항으로는 후항이 L_i , 즉 알람에 관련된 규칙이 지지도, 신뢰도, 향상도 척도를 만족하는 결과를 보여주고 있다. 이는 제품 생산 중 장비에서 관리 되고 있는 파라미터들이 알람에 높은 영향을 주고 있음을 시사한다. 장비에서 특정 파라미터의 가 공값(Process Value)들이 발생되고 있으면 수 율에 좋지 않은 영향을 주고 있는 알람들을 발 생시키고 있음을 알 수 있고, 이와 같은 규칙은 높은 신뢰도를 보이는 규칙에서는 향상도 척도 1.0 이상의 값을 가지므로 낮은 수율을 가져 올 확률이 높은 공정변수를 판별 할 수 있는 규칙 이라 할 수 있겠다.

IF {M16}→{L14} Then supp. 23%, conf. 97% (1)

IF {M22}→{L25} Then supp. 20%, conf. 99% (2)

IF {M36}→{L25} Then supp. 19%, conf. 96% (3)

IF {M16, L25}→{L14} Then supp. 21%, conf. 96% (4)

(1), (2), (3), (4) 의 규칙을 보게 되면 특정 공정 변수가 빈번하게 등장하게 되면 특정한 종류의 알람을 발생하게 하고 최종적으로 좋지 않은 수율이 발생하게 되는 확률이 상승하게 된다. 상기의 규칙을 수율 판별 프레임워크 구성에 이용하게 되면 신규 로트가 공정을 진행 하고 있을 때 엔지니어나 관리자가 빠른 시간 내에 문제를 조기에 인지하게 되고 공정을 수 행하는데 집중을 할 수 있다. 상기 규칙을 적용 해서 한 달 동안의 결과를 분석 해보면 전체 생 산된 1352개의 로트에서 수율이 미달되는 50개 중에서 48개를 HLRAM 프레임워크를 통해서 필터링을 할 수 가 있었다. HLRAM 프레임워 크에서 필터링 하지 못한 2개의 로트는 해당 실 험에서 사용한 연관규칙을 통해서 찾아내지 못 한 저수율 로트로서 연관규칙 대상에 대해서 범위를 확대하거나 다른 마이닝 방법을 사용해서 정밀도를 높여야 한다.

V. 결론

본 연구에서는 공정상에서 수율에 영향을 미 치는 요소들을 찾기 위해서 통계적 기법과 데 이터마이닝 기법을 결합한 방법론(HLRAM)을 제안하고 이것의 프로토타입을 개발하였다. 그 리고 실제 데이터를 이용하여 제조 기업에 직 접 적용해 보았다. 구체적으로 이산공정(조립

공정)을 가지는 제조업에서의 수율에 영향을 주는 공정, 장비에서의 변수를 로지스틱 회귀분석을 통해서 1차적으로 분류 한 뒤에 분류된 데이터 셋 중에서 지정된 수율에 미달하는 규칙을 연관관계 탐색을 이용해서 찾아낼 수가 있었다. 연관관계 탐색을 통해서 파악된 수율에 안 좋은 영향을 주고 있는 규칙을 생산관리 시스템에 필터 설정을 하게 되면 생산 중 해당 값들이 장비에서 탐지가 될 때 관리자에게 조기에 인지시킬 수 있어 생산 장비 설정 값 변경, 자재교체, 생산 라인 중단 등의 후속적인 활동을 할 수 있게 된다. 수율에 영향을 줄 수 있는 공정의 변수를 생산 장비 프로세싱 정보, 알람 정보, Cycle/Pitch time의 영역으로 구분해서 어떠한 인자들이 영향을 줄 수 있는지 파악을 하였는데, 특정장비 그룹과 알람에서 최종적인 수율에 영향을 주고 있음을 파악 할 수 있었지만, Cycle/Pitch time은 수율에 영향을 주지는 않는 것으로 본 연구를 통해서 알 수 있다.

HLRAM 프레임워크를 사용 하였을 때 제조 공정 엔지니어, 작업자들은 기존의 복잡하고 분석에 오랜 시간이 걸리는 문제를 극복 할 수가 있다. 또한 분석 기법 중의 데이터 마이닝, 덤핑 등의 방법을 사용 했을 때의 학습, 검증데이터의 확보와 같은 패턴학습에 필요한 추가적인 작업이 없이 분석이 가능하다는 이점을 제공할 수 있었다. 그리고 분석 시 어떠한 공정장비, 알람이 수율에 좋지 않은 영향을 주는지 쉽게 인지 할 수가 있다. 생산 중에도 불량률 조기에 감지 할 수 있는 화면을 통해서 제조공정 담당 엔지니어, 작업자 등의 생산에 관련된 파트에서는 되어 불량률 발생 할 확률이 높은 로트를 생산 초기에 감지 할 수가 있다. 이를 통해서 제

조 원가의 절감, 불량 선별의 자동화 등의 이점을 가져 올 수가 있다. 또한 본 연구는 최근 국내외에서 많은 관심이 집중되고 있는 스마트팩토리 구현을 위한 초창기 연구로 보아질 수 있다. 특히 본 연구에서 구현한 데이터획득, 분석, 모델링, 그리고 시뮬레이션 등을 포함하는 프레임워크는 CPPS 구현을 위한 기반구조로서의 역할을 할 수 있다.

그러나 제안된 프레임워크에서는 엔지니어가 기민하게(Agile)하게 사용 할 수 있도록 비교적 가벼운 프레임워크를 사용 하였지만 향후 연구에서는 보다 좋은 성능과 정확한 결과를 가져 올 수 있는 빅데이터처리 및 인공지능기법 들을 포함하는 연구가 보강이 되어야 할 것이고, 적용 영역도 공정인자 뿐만 아니라 품질 관련 인자에 대한 연구의 확장도 필요하다.

참고문헌

- 신정훈, 황승준, “DEA와 로지스틱 회귀분석을 이용한 자동차부품기업의 효율성 분석 및 재무전략”, 한국경영과학 학회지, 41(1), 2016, pp. 127-143.
- 안경찬, 문창배, 김병만, 이종열, 장득현, 김지윤, 성혜정 “연관규칙 마이닝 알고리즘을 이용한 POS 데이터 분석 시스템”, 대한 산업공학회 춘계 공동 학술 대회 논문집, 2012, pp. 2542-2547.
- 이필립, 박영미, 최양연 “조선 생산 실행 시스템 업무 분석을 위한 연관성 규칙 방법 적용”, 한국 CAD/CAM 학회 하계 학술대회 논문집, 2013, pp. 117-122.

- 정혜진, 구본철, “데이터 마이닝을 이용한 로버스트 설계 모형의 최적화”, 산업 경영 시스템 학회지, 30(2), 2007, pp. 99-105.
- Brian E. Goodlin, Duane S. Boning, and Herbert H. Sawin, “Simultaneous Fault Detection and Classification for Semiconductor Manufacturing Tools.” *Journal of the Electrochemical Society*, Vol.150, No.12, 2003, pp.778-784.
- Bernard Kamsu-Foguem, Fabien Rigal, Félix Mauget “Mining association rules for the quality improvement of the production process”, *Expert Systems with Applications*, 40(4), 2013, pp. 1034-1045.
- Duane S. Boning , Jerry Stefani, Stephanie W. Butler., “Statistical Methods for Semiconductor Manufacturing”, *Encyclopedia of Electrical Engineering*, 1999, pp. 1-22.
- G. A. Cherry, S. J. Qin, “Multiblock Principal Component Analysis Based on a Combined Index for Semiconductor Fault Detection and Diagnosis”, *IEEE Transactions on Semiconductor Manufacturing*, Vol.19, No.2, 2006, pp.159-172.
- L. Monostori, B. Kadar, T. Bauernhansl, S. Kondoh, S. Kumara, G. Reihart, O. Sauer, G. Schuh, W. Shin, K. Ueda, “Cyber-physical system in manufacturing,” *CIRP Annals, Manufacturing Technology* 65, 2016, pp.621-641.
- Lee, J, Bagheri, B and Kao, H, A Cyber-Physical Systems architecture for Industry 4.0 based manufacturing systems. *Manufacturing Letters* 3 (2015), pp18-2.
- Ming-Da Ma. David Shan-Hill Wong, Shi-Shang Jang, Sheng-Tsaing Tseng, “Fault detection based on statistical multivariate analysis and microarray visualization,” *Industrial Informatics, IEEE Transactions on*, Vol.6, No.1, 2010, pp.18-24.
- Rakesh Agrawal, Tomasz Imielinski & Arun Swami, “Mining association rules between sets of items in large database”, *ACM SIGMOD Conference*, 1993, pp. 207-216.
- Susana Ferreira, Basilio Sierra, Itziar Irigoien, Eneko Gorritxategi, “Data Mining for quality control : Burr detection in the drilling process”, *Computer & Industrial Engineering*, 60(4), 2011, pp. 801-810.
- Wei-Chou Chen, Shian-Shyong Tseng. Ching-Yao Wang, “A novel manufacturing defect detection method using association rule mining techniques”, *Expert System with Application* 29, 2005, pp.807-815.
- Yu, Peng. Lee. Jong-Nam, Lee, Jang-Hee “A Quality management using Data mining Techniques for Small and Medium Manufacturing Companies”, *대한 경영 학회지*, 25(9), 2012, pp. 3579-3599.

송 치 욱 (Song, Chi-Wook)



충남대학교 해양학과와 대전대학교 컴퓨터 공학 석사를 취득하였다. 공주대학교 산업공학과 박사과정을 수료 하였다. 현재 SK(주) HiTech자동차 개발팀에 재직하고 있으며, 주요 관심분야는 생산 자동화, 공정 최적화, Process Innovation 등이다.

노 금 종 (Roh, Geum-Jong)



부산대학교 물리학과를 졸업하고 LG디스플레이에 근무하였다. 이후 성균관대학교 경영대학원 경영학과를 졸업하고, 현재 공주대학교 일반대학원 산업시스템공학과 박사과정에 재학중이다. 관심분야는 Industry4.0, CPS, CPPS, Python을 이용한 업무 자동화 분야이다.

박 동 진 (Park, Dong-Jin)



아주대학교 산업공학과를 졸업하고 한국외국어대학교에서 석사학위를, 그리고 아주대학교에서 경영정보학 전공으로 박사학위를 취득하였다. 현재 공주대학교 공과대학 산업시스템공학과에서 교수로 재직 중이다. 주요 관심분야는 기업정보화, 데이터과학, CPS, 머신러닝 등이다

<Abstract>

A Study on analysis framework development for yield improvement in discrete manufacturing

Song, Chi-Wook · Roh, Geum-Jong · Park, Dong-Jin

Purpose

It is a major goal to improve the product yields during production operations in the manufacturing industry. Therefore, factory is trying to keep the good quality materials and proper production resources, also find the proper condition of facilities and manufacturing environment for yields improvement.

Design/methodology/approach

We propose the hybrid framework to analyze to dataset extracted from MES. Those data is about the alarm information generated from equipment, both measurement and equipment process value from production and cycle/pitch time measured from production data these covered products during production. We adapt a data warehousing techniques for organizing dataset, a logistic regression for finding out the significant factors, and a association analysis for drawing the rules which affect the product yields. And then we validate the framework by applying the real data generated from the discrete process in secondary cell battery manufacturing.

Findings

This paper deals with challenges to apply the full potential of modeling and simulation within CPPS(Cyber-Physical Production System) and Smart Factory implementation. The framework is being applied in one of the most advanced and complex industrial sectors like semiconductor, display, and automotive industry.

Keyword: Discrete Manufacturing, Yield Improvements, Data Mining

* 이 논문은 2017년 2월 7일 접수, 2017년 2월 24일 1차 심사, 2017년 5월 30일 게재 확정되었습니다.