

데이터마이닝 기법을 이용한 제조 공정내의 불량항목별 예측방법

변성규* · 강창욱** · 심성보***

*삼성전자로지텍(주) 국관물류팀

**한양대학교 정보경영공학과

***한양대학교 산업공학과

Defect Type Prediction Method in Manufacturing Process Using Data Mining Technique

Sung-Kyu Byeon* · Chang-Wook Kang** · Seong-Bo Sim***

*Domestic Logistics Team, Samsung Electronics Logitech Co. LTD

**Dept. of Information & Industrial Engineering, Hanyang University

***Dept. of Industrial Engineering, Hanyang University

Data mining technique is the exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules. This paper uses a data mining technique for the prediction of defect types in manufacturing process. The purpose of this paper is to model the recognition of defect type patterns and prediction of each defect type before it occurs in manufacturing process. The proposed model consists of data handling, defect type analysis, and defect type prediction stages. The performance measurement shows that it is higher in prediction accuracy than logistic regression model.

Keywords : Data mining technique, Defect type, Process data

1. 서론

최근 우리나라에서는 과거 어느 때보다도 제조공정에서의 품질관리 및 공정관리의 중요성이 대두되고 있는 실정이다. 대부분의 제조 기업에서는 이런 관리를 위해 주로 공정단계에서 원인분석을 통해 불량발생 예방활동을 중심으로 하는 통계적 공정관리(SPC : statistical process control)를 이용하고 있다. 통계적 공정관리의 대표적인 기법인 관리도(control chart)를 이용하여 공정을 모니터링 하고 있다. 제조공정에서 수집된 공정데이터를 이용하여 $\bar{x} - R$ 등의 관리도를 통해서 공정에 이상원인이 발생하는 경우 이를 탐지하여 수정·조치함으로써 불량률의 발생을 사전에 억제하고 있다. 또한 상관분석,

회귀분석, 실험계획법 등 다양한 통계적 기법이 공정관리에 사용되고 있다[4]. 이와 같은 기법들은 현 시점에서 발생하는 불량원인들을 분석하여 주로 예방활동에 치중하며, 불량이 발생하기 이전에 불량에 대한 예측은 이루어지지 않고 있다. 최종검사단계에서 불량이라고 판명이 난 경우는 이미 불량이 만들어지고 난 이후가 된다. 그러므로 최종검사단계가 아닌 공정단계에서 불량을 미리 발견할 수 있다면 공정이상원인을 해결하는 과정에 소요되는 시간을 단축시킬 수 있을 뿐만 아니라, 불량률을 사전에 감지함으로써 불량비용과 원자재비용 등을 감소할 수 있을 것이다.

최근에는 자동화 시스템의 확산과 더불어 컴퓨터나 센서를 이용하여 실시간으로 공정데이터를 수집하고 있

다. 이와 같이 수집된 데이터들은 수많은 공정변수와 방대한 양의 데이터로 인해 공정개선에 관련된 유용한 정보를 찾아내는 어려움과 분석의 어려움이 따른다. 이와 같은 문제를 해결하기 위해 전략적인 의사결정에 필요한 정보를 추출하는 방법론으로서 데이터마이닝(Data Mining)기법이 활용되고 있다[3]. 공정데이터를 이용하여 품질개선에 활용한 연구로는 Hancock et al.(1998)에 의한 자동차 밸브 주조 공정에서 공정변수들과 품질 특성치간의 관계를 규명하여 불량을 감소시키기 위한 공정조건을 설정한 사례[9]와 R.-S. Guh and Y.C. Hsieh. (1999)에 의한 신경망 모델을 이용한 관리도의 비정상패턴(abnormal pattern : shift, trend, cycle)의 모수를 추정한 연구[7]가 있다.

본 연구에서는 데이터마이닝 기법을 이용하여 최종검사단계가 아닌 공정단계에서 불량에 대한 패턴과 분석유형을 파악하고, 분석된 결과를 이용하여 불량을 사전에 예측하는 방법을 제시하고자 한다. 이를 위해 데이터마이닝 기법 및 분석유형에 대하여 알아보고, 분석유형에 적합한 모형을 도출한다. 도출된 모형을 공정데이터에 적용하여 분석한다. 분석된 결과를 예측모형에 유용하게 사용되는 로지스틱 회귀모형과 수행도를 비교한다.

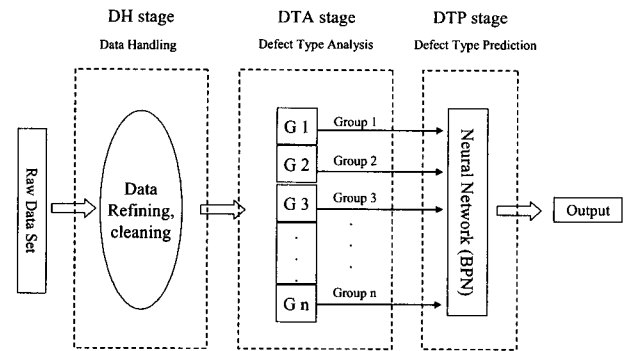
2. 불량항목별 예측방법모형

2.1 수행절차

공정데이터에 대한 분석에 앞서 원 데이터에 대해서 충분한 사전처리단계를 거치게 된다. 제조현장에서 공정관리 및 개선은 측정된 데이터에 근거하여 이루어지므로 정확한 데이터를 수집하여 분석하는 것이 중요하다. 또한, 효과적인 공정관리를 위해서는 품질 특성치에 영향을 주는 요인들과의 관계를 체계적으로 정리해야한다. 기존에는 이러한 정보를 얻기 위해서 통계분석을 사용하여 왔다. 그러나 공정데이터의 특성상 미처 파악하지 못하는 요인이 다수 존재하게 된다[4]. 특히, 수많은 공정변수와 방대한 양의 데이터를 인해 모든 변수를 분석하는데 어려움이 있다. 이와 같은 문제를 해결할 수 있는 기법의 하나인 데이터마이닝 기법을 연구에 이용하였다.

본 연구에서 제안하는 모델은 세가지단계로 구성된다. 첫 번째 단계는 원 데이터에 대하여 결측치의 처리, 이상치 제거 등 분석에 용이하고, 정확한 데이터로 재구성하기 위한 사전처리단계(DH : Data Handling)이다. 두 번째 단계로는 불량을 항목별로 분류하고 분류된 불량항목에 대하여 공정패턴을 분석하는 불량항목별 분석

단계(DTA : Defect Type Analysis)이며, 마지막 세 번째 단계는 최종적으로 불량항목별 예측단계(DTP : Defect Type Prediction)에서 불량을 항목별로 예측하게 된다. 각 단계별 수행 절차 모형은 다음<그림 1>과 같다.



<그림 1> 불량항목별 예측에 대한 수행 절차 모형

2.2 사전처리 단계 (DH : Data Handling)

데이터 사전처리 단계는 데이터마이닝 기법을 적용하기 전에 분석에 용이한 데이터로 재구성하기 위해 데이터 정제 및 변환하는 단계이다. 데이터마이닝의 대상이 되는 데이터는 많은 불확실성을 내재하고 있으므로 사전 처리 작업이 필요하다[3].

결측치 처리는 결측치가 포함된 레코드 삭제, 결측치가 포함된 필드(변수) 삭제, 사전에 정해놓은 고정값을 부여, 모형으로부터 결측치값을 추론하여 부여하는 방법 등이 있다[5]. 본 연구에서는 현장데이터를 직접 이용하여 실험하였으므로, 고정값이나 결측치값을 추론하여 부여한다는 것은 데이터를 왜곡하여 나타낼 수 있는 위험이 있기 때문에 결측치가 포함된 레코드나 필드를 삭제시켜서 데이터를 재구성하였다.

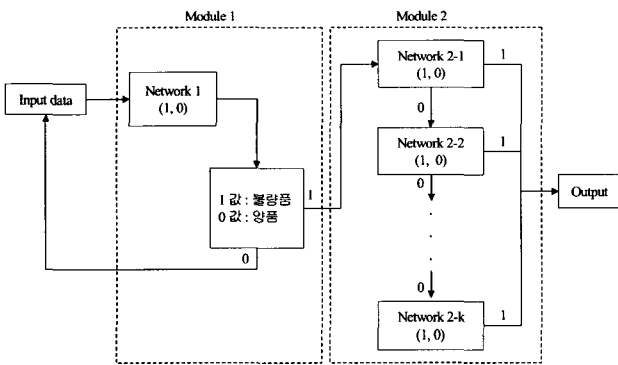
2.3 불량항목별 분석단계 (DTA : Defect Type Analysis)

불량항목별 분석단계에서는 사전처리단계가 끝난 공정데이터에 대하여 의사결정나무(decision tree)기법을 사용하여 불량을 항목별로 분류하고, 분류된 각 불량항목에 대해서 공정패턴을 분석하는 단계이다. 의사결정나무 기법은 분류와 공정패턴을 명확하게 나타내는데 매우 효과적인 기법이다[6]. 목표변수가 범주형인 경우 사용되는 알고리즘으로 CHAID(Chi-squared Automatic Interaction Detection)알고리즘이 있다. CHAID알고리즘의 분할단계는 첫 단계에서 반응변수 중에서 통계적으로 유의한 차이를 주지 않는 모든 입력변수들이 함께 묶여진다. 둘째

단계에서 셋 이상의 변수로 된 각 그룹은 모든 가능한 이원분리에 의해 재분할된다[6]. 제조공정에서는 최종제품이 생산되기까지 여러 단계의 공정을 거치게 되므로 최종제품의 품질특성은 공정단계의 여러 공정변수에 의해 영향을 받게 된다[7]. 따라서 최종제품의 품질향상을 위해서는 최종단계뿐만 아니라, 이전 단계의 공정변수들을 파악해야 할 것이다. 분류이후에 종속변수인 불량에 대하여 영향을 미치는 독립변수를 파악하여 각 독립변수들의 조합으로 이루어진 공정패턴을 인식하게 된다.

2.4 불량항목별 예측단계 (DTP : Defect Type Prediction)

불량항목별 예측단계에서는 분석단계에서 구해진 결과를 신경망(neural network)기법을 이용하여 새로운 공정데이터에 대해 불량을 항목별로 예측한다. 신경망은 다른 기법과는 달리 결과값을 입력변수들의 비선형조합(non-linear combination)을 통해 제공하기 때문에 예측력이 우수하다[8]. 신경망은 연결강도에 의하여 상호 연결되어진 노드들의 집합이다. 노드들이 서로 독립적으로 작동되므로 병렬 및 분산처리가 가능한 구조를 갖고 있으며, 이에 따라 빠른 정보처리 및 외부잡음에 강건한 특성을 가지고 있다[5]. 신경망 기법을 이용하여 불량항목별 예측단계에 대한 모형은 다음<그림 2>와 같다.



<그림 2> 불량항목별 예측모형

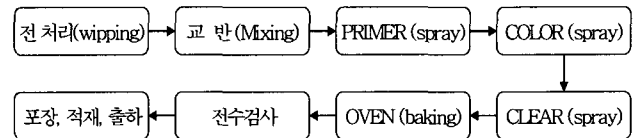
데이터가 입력되면 module 1단계에서 network 1을 통해 입력데이터가 양품인지 불량품인지에 대하여 예측하게 된다. network 1에서 양품(0)인 경우 다시 입력으로, 불량(1)인 경우에는 module 2단계에서 각 불량항목으로 이루어진 network를 통해 예측이 이루어진다. module 2단계의 network 2-1에서 0값인 경우에는 network 2-2로 진행되어지고, 1값인 경우에는 network 2-1에 해당하는 불량으로 결정된다.

3. 사례연구

본 사례연구에 사용된 데이터는 자동차 내·외장품 생산업체인 국내 D기업의 범퍼공정에서 학습용 데이터(training data)로 700개, 평가용 데이터(test data)로 300개를 수집하였으며, 분석도구로 SAS와 SAS E-Miner를 사용한다.

3.1 공정소개

범퍼는 자동차 차체를 보호하기 위해 달려있는 우레탄계열 플라스틱으로 사출(고열, 고압에서 원료를 녹여 제품을 금형에서 찍어내는 일)하여 사출물에 도장작업을 거치게 된다. 범퍼는 우레탄계열의 제품이어서 열 반응에 쉽게 치수가 줄어들거나 늘어날 수 있다. 이 공장의 범퍼 생산 공정은 다음<그림 3>과 같다.



<그림 3> 범퍼 생산 공정

범퍼 공정에서는 전 처리단계에서 CLEAR단계까지를 주요 공정으로 관리하고 있다. 제품이 플라스틱을 사출하여 생산되므로 주요 공정에서 온도와 습도에 매우 민감한 반응을 보인다. 대부분의 불량이 적정온도(28±3°C)와 적정습도(70±10%)를 벗어나서 이색, 이물질, 분화구 등의 불량이 발생한다. 그리고 불량의 원인으로서는 작업조건의 이상으로 인한 불량, 작업자의 작업 미숙에 의한 불량, 설비에 의한 불량 등이 존재한다. 여기서 작업자의 작업 미숙과 설비의 이상에 의한 불량은 매우 유동적이고 데이터 수집이 용이하지

<표 1> 변수 정의

변 수	변 수 명
X1	교반시간
X2	전처리(wipping)온도
X3	전처리(wipping)습도
X4	PRIMER 온도
X5	PRIMER 습도
X6	COLOR 온도
X7	COLOR 습도
X8	CLEAR 온도
X9	CLEAR 습도
Y	불 량

않으므로 본 연구에서는 여러 불량 원인 중 실질적으로 제품에 많은 영향을 미치는 작업조건의 이상으로 인한 불량을 고려하였다. 본 사례의 범퍼공정에서는 전처리공정, primer 공정, color공정, clear공정, 교반공정을 주요공정으로 관리하고 있다. 그래서 실험에 사용된 공정변수로 주요공정과 동일하게 선택하였으며, 교반공정을 제외한 주요공정의 관리항목인 온도와 습도, 교반공정에서는 간을 선택하였다.

실험에 사용된 변수는 <표 1>에 정의하였다.

3.2 사전처리 단계

학습용 데이터 700개에서 변수 X3의 453번째 레코드와 X8의 324번째의 레코드에서 결측치를 확인하여 삭제시켰다. 데이터의 재구성을 위한 데이터의 정제 및 변환을 실시하였다. 학습용 데이터 698개를 이용하여 변수들 간의 관련성을 파악하고 현상분석을 위해 기초 통계량 [표 2]과 상관관계분석을 실시하였으며, 변수들 간의 관련성의 척도로 피어슨(Pearson) R 계수를 이용하였다. 구한 결과는 <표 3>과 같다.

<표 2> 주요공정변수별 기초통계량

변수	N	평균값	표준편차	최소값	최대값
X1	698	30.38	1.61	22	34
X2	698	25.78	1.83	20	29
X3	698	63.34	4.93	50	79
X4	698	26.19	1.55	21	29
X5	698	65.06	5.48	50	79
X6	698	26.72	2.04	21	33
X7	698	64.64	6.11	42	87
X8	698	25.76	1.67	20	29
X9	698	62.52	5.63	45	81

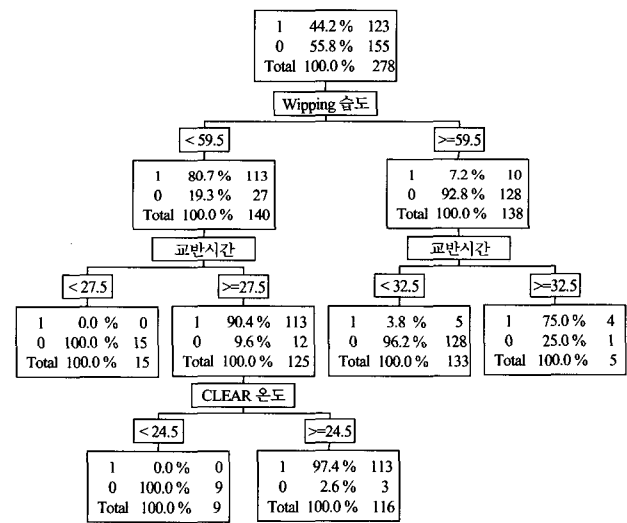
[표 3] 주요공정변수 간 상관행렬

	X1	X2	X3	X4	X5	X6	X7	X8	X9
X2	-0.060								
X3	-0.037	0.396							
X4	0.133	0.094	0.034						
X5	-0.166	0.265	0.327	0.169					
X6	-0.098	0.095	0.097	0.075	0.116				
X7	-0.106	0.108	0.156	-0.138	0.214	0.089			
X8	0.008	-0.003	-0.126	0.308	0.019	0.164	-0.185		
X9	-0.152	0.368	0.299	0.191	0.412	0.253	0.237	0.322	
Y	-0.092	-0.457	-0.451	-0.322	-0.429	-0.343	-0.174	-0.296	-0.675
	0.015	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

<표 3>의 결과 종속변수인 불량에 모든 독립변수가 관계가 있으며, p-value 값이 모두 <0.05으로 모든 독립변수가 유의한 결과를 보였다. 이는 주요공정에서의 작업조건인 온도와 습도가 불량에 많은 영향을 미침을 알 수 있다.

3.3 불량항목별 분석단계

사전처리단계가 끝난 후 재구성된 데이터에 대하여 의사결정나무 기법을 이용하여 불량항목별 공정패턴을 분석한다. 범퍼공정에서는 공정특성상 불량크게 4가지(이물질, 칠흐름, 분화구, 이색)로 분류되어진다. 이물질 불량크는 경우 대부분 작업실(Booth)내의 온·습도가 맞지 않을 때, 특히 습도가 맞지 않을 경우 대기의 부유 중된 먼지들이 묻어서 발생한다. 다음 <그림 4>은 이물질 불량크의 패턴을 찾기 위해서 의사결정나무 분석을 한 결과이다.



<그림 4> 이물질 불량크의 공정패턴결과

<그림 4>의 결과에서 이물질 불량크를 1값으로, 나머지에 대해서는 0값으로 변환하였다. 전체 불량 수(278개)에서 이물질 불량크는 123개로 전체 44.2%를 차지하고 있다. 마지막 끝 마디에서 순수도가 97.4%로 나타났다. 첫 번째 노드로 전처리 습도에 따라 이물질 불량크의 범주를 나눌 수 있다. 즉, 전처리 습도가 59.5%이하일 때 순수도 80.7%로 분리되어짐을 알 수 있으며, 이는 전처리 습도변수가 이물질 불량크에 많은 영향을 미치는 변수임을 알 수 있다. 다음으로 교반시간이 27.5분 이상이고, CLEAR 온도가 24.5°C이상일 때 이물질 불량크가 발생함을 알 수 있다. 위 패턴 결과를 통해서 볼 때 대부분 소재 표면에 묻

어있는 미세 먼지 및 이물질을 휘발성이 강한 약품으로 닦아 주는 작업인 전처리 과정에서 이물질 불량이 일어남을 알 수 있다.

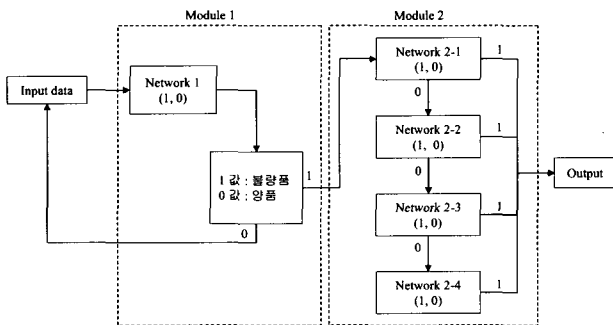
각 불량 항목에 대한 공정패턴 결과는 다음<표 5>과 같다.

<표 5> 불량항목별 공정패턴 결과

불량항목	패턴
이물질	· 전처리 습도(<59.51%) → 교반시간 (≥27.5s) → CLEAR 온도(≥24.5°C)
칠흐름	· CLEAR 온도(<24.5°C) → COLOR 온도(<24.5°C) → 교반시간(≥28s)
분화구	· COLOR 온도(<24.5°C) → CLEAR 온도(<24.5°C) → COLOR 습도(<60%)
이색	· 교반시간(<28.5s) → 교반시간(<27.5s) → PRIMER 습도(≥53%)

3.4 불량항목별 예측단계

신경망 모형은 모든 독립변수와 종속변수가 0에서 1 사이 범위의 값을 취할 때 최적의 성능을 제공한다[2]. 따라서 변수의 성격에 상관없이 모든 변수의 값을 0과 1사이의 값으로 변환시키는 것이 바람직하다. 변환값(massaged value)은 (실제값-최소값)/(최대값-최소값)으로 값을 변환하였다. 여기에서 교반시간 30은 (30-22) / (34-22) = 0.67값으로 변환된다. 종속변수인 불량은 이항형(binary) 변수이므로 알고자하는 불량을 '1'로, 양품을 '0'으로 변환하여 예측을 실시한다. 범퍼 공정의 예측 모형의 module 2단계의 각 network에는 4가지 불량(network2-1:이물질, network2-2:칠흐름, network2-3:분화구, network2-4:이색)이 대응된다. 예측단계에 대한 모형은 다음<그림 5>와 같다.



<그림 5> 불량항목별 예측모형(범퍼공정)

입력된 데이터는 module 1단계의 network 1을 통해 양

품인지 불량품인지를 예측하고, module 2단계에서의 각 network를 통해 불량 항목별 예측이 이루어진다. 출력값은 다음 <표 6>과 같다.

<표 6> 불량항목별 출력값

불량항목	Output			
	network 2-1	network 2-2	network 2-3	network2-4
이물질	1	0	0	0
칠흐름	0	1	0	0
분화구	0	0	1	0
이색	0	0	0	1

예측결과 module 1에서의 network 1의 양·불량예측은 평가데이터 300개에서 284개(94.6%)의 정확도를 보였다. 양·불량 예측결과에서 1값은 112개, 0값은 172개로 나타났다으며, 1값을 가지는 112개의 데이터에는 양품을 불량품이라고 예측한 잘못된 결과값이 9개가 존재한다. 이 값은 module 2단계에서도 잘못된 예측결과를 가져오게 된다. 이 값을 고려한 불량항목별로 예측결과는 다음<표 7>과 같다.

<표 7> 불량항목별 예측결과

불량항목	No. of correct	No. of wrong	% of correct	% of wrong
이물질	107	5	95.5	4.5
칠흐름	100	12	89.3	10.7
분화구	106	6	94.6	5.4
이색	110	2	98.2	1.8

3.5 수행도 비교

본 연구에서 제시한 예측모형을 예측에 유용하게 사용되는 통계적 기법 중에서 로지스틱 회귀모형과 수행도를 비교한다. 로지스틱 회귀모형은 종속변수가 이항형(binary-type) 또는 순서형(ordinal-type)으로 나타나는 경우에 선택회귀모형의 단점을 극복하기 위해 주로 사용되는 분석방법이다[2]. 그러므로 제안모형과 로지스틱 회귀모형에 대하여 주요 통계량인 오분류율(Misclassification rate), 평균제곱오차(Average Squared Error), 제곱합오차(Sum of Squared error)를 이용하여 모형에 대한 평가를 실시하였다. module 1의 양·불량에 대한 모형평가결과는 다음<표 8>과 같다.

모형평가결과 모든 항목에서 제안모형이 로지스틱 회

귀모형보다 좋은 결과를 나타내고 있다.

<표 8> 양·불량에 대한 모형평가결과

항 목	제안모형	Logistic
오 분 류 율	0	0.04286
평균 제 품 오 차	0.00074	0.04287
제 곱 합 오 차	0.61769	30.02238

특히, 제안 모형의 오분류율이 0값으로 충분한 훈련이 되었음을 알 수 있다. 평가 데이터를 설정된 모형에 적용한 예측결과는 다음 <표 9>와 같다.

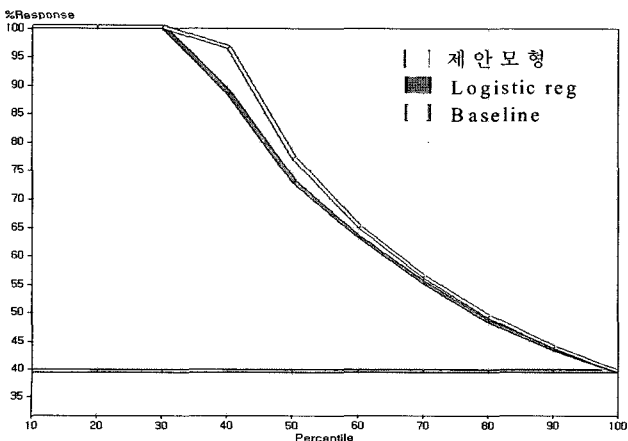
<표 9> 양·불량에 대한 예측결과

항 목	제안모형	Logistic
정 확 도	94.6%	92.7%

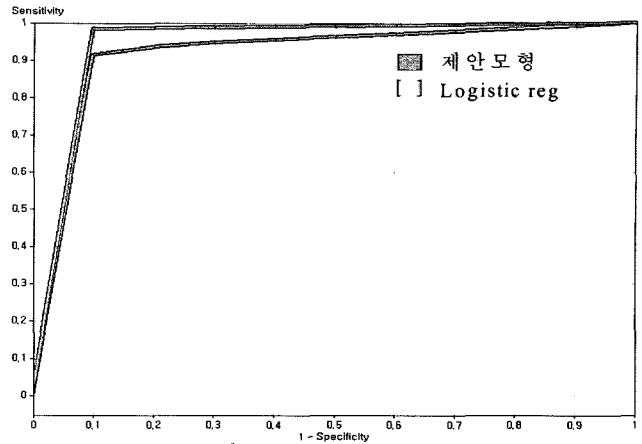
예측결과 정확도에서 제안모형의 정확도가 94.6%로 로지스틱 회귀모형에 비해 1.9%p 더 좋은 결과를 나타내고 있다.

다음 <그림 6>는 예측의 정확도를 나타내는 Lift Chart와 <그림 7>은 이진형 종속변수를 가지는 모형들의 성능을 비교, 평가하는데 매우 유용한 도표인 Roc Curve를 이용하여 평가한 결과이다.

Lift Chart는 전체 집단에서 해당집단 X축(percentile)에 대하여 Y축(%response)에 종속변수에 대한 특정범주의 빈도를 나타내는 도표이고, Roc Curve는 민감도(Sensitivity)와 특이도(Specificity)를 이용하여 X축에는 1-특이도를, Y축에는 민감도를 각 분류 기준값에 대해 나타낸 도표이다[6]. 두 그래프 모두 도표의 왼쪽 상단에 가까울수록 성능 면에서 우수한 모형으로 판단된다.



<그림 6> Lift chart



<그림 7> Roc curve

민감도와 특이도는 다음과 같이 정의 된다.

$$\text{민감도} = \frac{(\text{실제1, 예측1})\text{인관찰치의빈도}}{\text{실제1인관찰치의빈도}} \dots\dots\dots (1)$$

$$\text{특이도} = \frac{(\text{실제0, 예측0})\text{인관찰치의빈도}}{\text{실제0인관찰치의빈도}} \dots\dots\dots (2)$$

<그림 6>의 Lift chart는 해당집단에서 종속변수의 특정범주의 빈도를 나타내는 Y축의 %Response를 보면 각각 모형을 적용 후 불량률의 가능성이 높은 상위 40%의 데이터에 대한 반응의 정도(정확도)가 제안모형은 96%, 로지스틱 회귀모형은 89%로 제안모형의 반응의 정도가 7% 더 좋은 결과를 나타낸다. <그림 7>의 Roc curve를 통해 비교하면 1-특이도(실제값은 0값인데 1값으로 잘못 예측할 확률)에 대한 민감도(실제값 1값을 1값으로 정확히 예측할 확률)의 정도가 제안모형이 로지스틱 회귀모형보다 0.1정도 높게 나타났으며, 1-특이도값이 증가하더라도 제안모형은 크게 영향을 받지 않으므로 로지스틱 회귀모형에 비해 안정적이다. 두 도표의 결과 제안모형이 로지스틱다음<표 10>는 불량항목별에 대한 모형평가와 예측의 정확도에 관한 수행도를 비교한 결과이다.

<표 10> 이물질, 칠흐름 수행도 비교 결과

항 목	이 물 질		칠 흐 름	
	제안모형	Logistic	제안모형	Logistic
정 확 도	95.5	92.9	89.3	85.7
오 분 류 율	0	0.0207	0.0479	0.1802
평균 제 곱 오 차	0.0002	0.0277	0.0486	0.1555
제 곱 합 오 차	0.0685	6.1456	16.2427	34.5309

<표 11> 분화구, 이색 수행도 비교 결과

항 목	분 화 구		이 색	
	제안모형	Logistic	제안모형	Logistic
정 확 도	94.6	93.8	98.2	91.1
오 분 류 율	0.0299	0.0135	0.0006	0.0450
평균제곱오차	0.0913	0.1083	0.0050	0.0323
제곱합오차	16.2427	24.035	1.6780	7.1701

각 불량항목별 수행도 비교 결과, 모든 모형평가항목에서 제안모형이 로지스틱 회귀모형에 비해 성능이 우수함을 알 수 있으며, 예측의 정확도에서도 제안모형이 더 좋은 결과를 나타내고 있다. 특히, 이색 불량에 대해서는 제안모형의 정확도가 98.2%로 로지스틱 회귀모형보다 7.1%p 더 좋은 예측결과를 나타내고 있다.

4. 결 론

본 연구에서는 제조공정에서 발생하는 대용량 공정데이터를 이용하여 불량항목별 예측을 위한 분석절차와 모형을 제시하였다. 실시간으로 수집되는 대용량의 데이터베이스로부터 분석적인 데이터나 유용한 정보를 찾아내는 어려움을 해결하기 위해 데이터베이스 마케팅 분야를 중심으로 의사결정에 필요한 정보를 제공해주는 실용적인 방법론인 데이터마이닝 기법을 이용하였다.

대부분 기업에서는 검사단계에서 불량검사가 이루어지며 검사단계에서 불량이라고 판명이 난 경우는 이미 불량이 만들어지고 난 이후가 된다. 본 연구에서 제시하는 분석절차와 모형을 통해 불량이 만들어지기 전에 공정단계에서 불량을 미리 예측하여 공정이상원인을 해결하는 과정에 소요되는 시간을 단축시킬 수 있을 것이다. 또한, 불량을 사전에 감소시킴으로서 불량발생에 의해 수반되는 비용과 원자재비용 등을 감소할 수 있을 것이다.

참고문헌

[1] 장남식, 홍성완, 장재호 “데이터마이닝”대청, 1999
 [2] 최국렬 외 9명, “데이터마이닝 이론과 실습”청구문화사, 2001
 [3] Banks, D.L, Parmigiani, “Pre-Analysis of Superlarge Industrial Data Sets”, Journal of Quality Technology, Vol.24, pp.115-129, 1992
 [4] Douglas C.Montgomery, “Introduction to statistical quality control” John Wiley & Sons, 1996

[5] Feng Yu Lin, Sally McClean, “A data mining approach to the prediction of corporate failure”, Knowledge-Based systems, Vol 14. pp.189-195, 2001
 [6] Michael J.A.Berry, “Data Mining Techniques”, John Wiley & Sons, 1997
 [7] R.-S. Guh, Y.C. Hsieh, “A neural network based model for abnormal pattern recognition of control charts”, Computers & Industrial Engineering Vol. 36, pp. 97-108, 1999.
 [8] Velasco T, Rowe MR. “Back propagation artificial neural networks for the analysis of quality control charts”. , Computers & Industrial Engineering Vol. 25, pp. 397-400, 1993.
 [9] W.M.Hancock, J.W.Yoon & R.Plot, “Use of Ridge Regression in the Improved Control of Casting Process”, Quality Engineering, Vol.8(3), pp.395-403, 1998.